Challenges and Advances
in Computational Chemistry and Physics 28
Series Editor: Jerzy Leszczynski

Nir Goldman *Editor*

Computational Approaches for Chemistry Under Extreme Conditions

Springer

Challenges and Advances in Computational Chemistry and Physics

Volume 28

Series editor

Jerzy Leszczynski
Department of Chemistry and Biochemistry
Jackson State University, Jackson, MS, USA

This book series provides reviews on the most recent developments in computational chemistry and physics. It covers both the method developments and their applications. Each volume consists of chapters devoted to the one research area. The series highlights the most notable advances in applications of the computational methods. The volumes include nanotechnology, material sciences, molecular biology, structures and bonding in molecular complexes, and atmospheric chemistry. The authors are recruited from among the most prominent researchers in their research areas. As computational chemistry and physics is one of the most rapidly advancing scientific areas such timely overviews are desired by chemists, physicists, molecular biologists and material scientists. The books are intended for graduate students and researchers.

All contributions to edited volumes should undergo standard peer review to ensure high scientific quality, while monographs should be reviewed by at least two experts in the field. Submitted manuscripts will be reviewed and decided by the series editor, Prof. Jerzy Leszczynski.

More information about this series at http://www.springer.com/series/6918

Nir Goldman
Editor

Computational Approaches for Chemistry Under Extreme Conditions

Springer

Editor
Nir Goldman
Materials Science Division
Lawrence Livermore National Laboratory
Livermore, CA, USA

ISSN 2542-4491 ISSN 2542-4483 (electronic)
Challenges and Advances in Computational Chemistry and Physics
ISBN 978-3-030-05599-8 ISBN 978-3-030-05600-1 (eBook)
https://doi.org/10.1007/978-3-030-05600-1

Library of Congress Control Number: 2018964691

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book presents recently developed computational approaches for the study of reactive materials under extreme conditions, with an emphasis on atomistic methods or those derived from atomistic calculations. Our intention is to include state-of-the-art developments in a single source, spanning high pressures (e.g., 10s to 100s of GPa), high temperatures (up to 1000s of degrees Kelvin), and even strong electrical fields. The methods presented here include *ab initio* approaches such as Density Functional Theory (DFT) and semi-empirical quantum simulation methods (spanning nanometer length scales and picosecond timescales) as well as reactive force field and coarse-grained approaches (spanning microns, nanoseconds, and beyond). These approaches are readily applied across a broad range of fields, including prebiotic chemistry in impacting comets, studies of planetary interiors, high-pressure synthesis of new compounds, and detonations of energetic materials. We have aimed to emphasize the strong connection with experiments throughout the book. Our hope is that this will prove to be a useful reference for both computational scientists wishing to learn more about specific subfields in this area as well as experimental scientists wishing to gain familiarity with well-known computational approaches for simulation of their experiments.

The book can be subdivided into a number of ways that will hopefully prove instructive to both experts and the uninitiated. Broadly speaking, Chaps. 1–5 have focused on some sort of quantum simulation approach, mainly with Kohn–Sham DFT as the starting point. Chapters 1, 4, 6, 7, and 8 discuss force field simulation and development for classical molecular dynamics calculations. Chapters 8, 9, and 10 discuss coarse-graining approaches to extend prediction of physical and chemical properties to closer to experimental time and length scales. In more detail, Chap. 2 focuses on crystal composition and structure prediction through evolutionary algorithms. Chapters 3 and 4 discuss semi-empirical quantum method developments, which can yield significantly longer timescales for molecular dynamics simulations while still providing information about electronic states. Chapter 4 places an emphasis on refining both semi-empirical and classical molecular dynamics simulations through force matching to DFT-MD trajectories,

including approaches for speeding up the fitting process through linear least-squares fitting. Chapters 4, 5, and 6 discuss extending quantum simulations to long time-scales through free energy calculations and accelerated simulation approaches. Chapter 8 has some focus on novel reactive coarse-graining approaches for shock-induced volume collapsing reactions. Chapters 9 and 10 place additional emphasis on use of machine-learning methods to extend physical–chemical simulation approaches to closer experimental time and length scales.

The chapters in this book can also be neatly subdivided by application area. Chapter 1 discusses hydrocarbon polymers under extremely high pressures and temperatures. Chapters 2, 8, and 10 explore energetic materials, with Chaps. 8 and 10 exploring reactive conditions. Chapters 4, 5, and 6 explore prebiotic chemistry, and the role extreme conditions can play in the synthesis of simple, life-building precursors. Chapters 3 and 8 focus on the chemistry of simple organics under shock compression conditions. Chapter 9 investigates the reactivity of jet fuels under combustion (high temperature) conditions.

Ultimately, we wish for this volume to be a useful pedagogical tool for a wide variety of researchers in extreme physics and chemistry. Though our specific emphasis has been on elevated conditions, the work presented here can be generalized very easily to other materials, pressures, and temperatures. Our collective efforts can be broadly applied to any number of scientific efforts spanning many different types of compounds and reactive conditions. Thus, the aim for our book is to be impactful for any research area that relies on atomistic simulation approaches to guide and elucidate experimental studies and materials discovery.

Livermore, USA Nir Goldman
October 2018

Contents

Chapter 1
Simulations of Hydrocarbon Polymers Related to Compression Experiments on Sandia's Z Machine

Thomas R. Mattsson, Kyle R. Cochrane, J. Matthew D. Lane and Seth Root

Abstract High-fidelity modeling of hydrocarbon polymers is important for gaining fundamental understanding of the underlying material behavior as well as for designing high energy density (HED) experiments. In this chapter, we describe multi-scale modeling/simulation of hydrocarbon polymers done at Sandia and corresponding experiments on Sandia's Z machine. For polymers, a combination of first-principles simulations using density functional theory (DFT) and atomistic simulations using classical molecular dynamics has proven to effectively model different aspects of the system and we will present both. Throughout, we find that the simulations are in qualitative and quantitative agreement with experiments, suggesting that the hierarchy of simulations can be used to increase our understanding of polymers under dynamic loading conditions.

1.1 Introduction

Hydrocarbon polymers are versatile and therefore ubiquitous materials in experimental loads on high energy density (HED) facilities like the National Ignition Facility (NIF), the Omega laser facility at Laboratory for Laser Energetics (LLE), and the Z machine at Sandia National Laboratories. The physics of polymers span large length- and timescales and the way of modeling polymers has to change to match the different physics. For example, under successively stronger shock compression, polymers melt and dissociate into a hydrogen and carbon fluid/dense plasma. For strong shocks, an accurate modeling of atomic dissociation is important, suggesting that atomistic simulations based on quantum mechanics are the appropriate approach. Polymers are also the building blocks of different types of foams, which have porosity and other internal structure. Foams are extensively used to tailor the behavior in high-pressure experiments not only on a pulsed power driver like the Z machine [41] but also on laser drivers [45]. For foams, an additional scale enters, the scale of pores.

T. R. Mattsson (✉) · K. R. Cochrane · J. M. D. Lane · S. Root
Sandia National Laboratories, Albuquerque, NM 87185, USA
e-mail: trmatts@sandia.gov

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_1

For foams, the use of classical molecular dynamics (MD) simulations using reactive force fields has given new insights into formation of hotspots and other aspects of shock dynamics in heterogeneous materials.

The outline of the chapter is as follows. The first section is a discussion on first-principles simulations using density functional theory (DFT), including post-processing to analyze the chemical composition of the material. The second section presents classical MD simulations of hydrocarbon polymers and—foams, we show how the shock propagates in the foam and how local hotspots are formed. In the third section, we discuss Z experimental results on PMP plastic and demonstrate how Z and simulations give a complete picture of the PMP shock response. The chapter concludes with a summary and conclusions.

1.2 First-Principles Simulations of Shocked Polymers

First-principles simulations using Density Functional Theory have brought important insights into how matter behaves under extreme conditions and, importantly, key predictions of thermodynamic properties from DFT simulations have been confirmed experimentally. For example, Sandia's Z machine has been employed to experimentally determine the shock Hugoniot of a number of elements and compounds, including carbon/diamond [25], xenon [47], water [28], carbon dioxide [48], hydrogen [23, 26], krypton [39], and quartz [27]. For these elements and compounds, DFT simulations have been shown to predict the behavior with high fidelity. The DFT modeling of shock compression of hydrogen, carbon, water, carbon dioxide strongly suggests that it is possible to use DFT to successfully simulate hydrocarbon polymers as well. That said, polymers add complexity and the excellent agreement between DFT simulations and experiments demonstrated in 2015 [50] was far from a foregone conclusion.

Density functional theory is based on a theoretical discovery by Walter Kohn [18, 29] that the energy of the electron gas of N electrons can be formulated as a functional of the electron density in real space (3 spatial coordinates) without an explicit use of the wave functions (3 N spatial coordinates, one for each electron). The resulting simplification is tremendous since the problem is reduced from 3 N dimension to 3. The discovery soon resulted in practical quantum simulations of electronic structures of solids, atoms, and molecules. The field has expanded, and DFT simulations are now commonplace across physics, chemistry, biology, and materials science. Although a review of applications of DFT is well outside the scope of this chapter, we will discuss the most important factors of applying DFT to high energy density conditions in general and post-processing for polymers in particular. For a more thorough review on designing DFT calculations for materials science, including HED, please see [37].

1.2.1 First-Principles Thermodynamics of Shock Compression

DFT is applied to problems in physics, material science, and chemistry to calculate the energy in atomic systems. DFT calculations are used to compare energy differences between structures and determine energy barriers between transitions. One of the most common DFT calculations is ionic position optimizations to minimize the system energy and to determine transition paths in the energy landscape. The electronic structure is mostly sought for the lowest, zero Kelvin, energy state, where the Fermi distribution is a step function. The discontinuous Fermi distribution requires extensive k-point sampling of reciprocal space in order to capture the behavior well. Shock compression, on the other hand, involves a discontinuous jump from an initial state to a final state that often is at very high temperature (Fig. 1.1).

Shock physics is driven by the Rankine–Hugoniot (RH) jump conditions, which link the initial and final states for a system with a propagating shock wave of speed u_s and particle velocity after the shock u_p. The RH equations are derived from the conservation of mass, momentum, and energy across the front discontinuity:

$$(P_2 - P_1) = \rho_1(u_s - u_1)(u_2 - u_1) \tag{1.1}$$

$$\rho_{2/}\rho_1 = (u_s - u_1)/(u_s - u_2) \tag{1.2}$$

$$(e_2 - e_1) = \frac{1}{2}(P_2 + P_1)(v_1 - v_2) \tag{1.3}$$

where e is the specific internal energy, P is pressure, and ρ is the density. The subscripts 1 and 2 designate the initial and final states, respectively. The result is a thermodynamics relationship between pressure, specific volume ($v = 1/\rho$), and internal energy of the final and initial states, the so-called Rankine–Hugoniot relationship (1.3) [6].

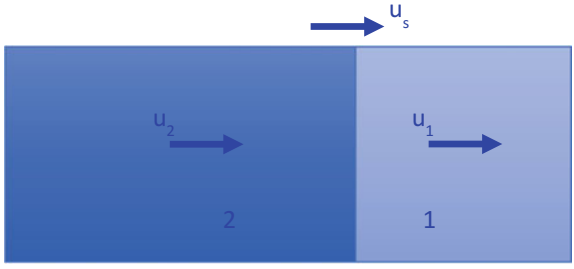


Fig. 1.1 Shock wave in a material, the shock is supersonic and causes a discontinuous transition from thermodynamic state 1 to state 2. The shock velocity is u_s and the net particle velocities before and after the shock are u_1 and u_2. The initial and final states are linked by conservation laws. The analysis can, for example, be made as a thermodynamic control volume

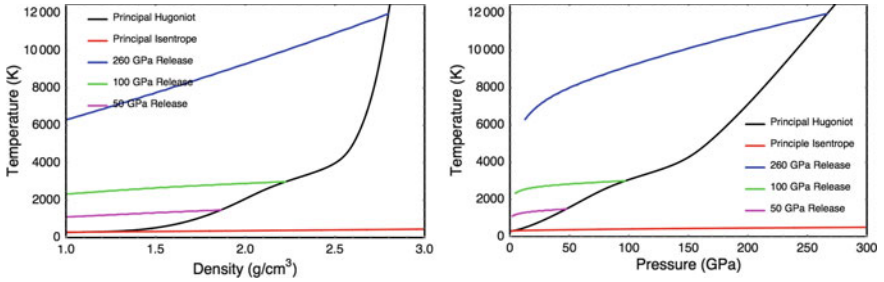


Fig. 1.2 The end states of shock compression of polyethylene [9] (black) and three release paths from different initial shock pressures. Note the shoulder in the Hugoniot between 100 and 150 GPa caused by dissociation of the polymer. Above 150 GPa, there are no longer any molecular fragments, see Fig. 1.7

In thermodynamic space, the loci of end states generated from successively stronger shocks starting from a common initial state are called the Hugoniot. Figure 1.2 shows the polyethylene Hugoniot in density–temperature space (ρ–T) and pressure–temperature (P–T). Figure 1.2 also shows release isentropes: the path the initial shock state relaxes from when the driving pressure is released. Note that the RH equations do not depend on the temperature except via the internal energy. Temperature is extremely difficult to measure experimentally, so there is more uncertainty involving temperature than pressure and density.

The RH energy relationship points toward the need for high accuracy calculations of internal energy and pressure in simulations of shock compression. It is evident that temperature plays a key role, making molecular dynamics (MD) simulations appropriate for sampling the thermodynamics ensemble [1]. At the high temperatures of shock compression, MD is a highly efficient method of sampling the thermodynamic phase space [1].

Importantly, temperature also affects the electronic structure making it necessary to use the Mermin finite-temperature formulation of DFT [42] which involves calculating the full Fermi distribution of electronic states. At high temperature, the full Fermi distribution includes many partially occupied continuum states, requiring significantly more electronic bands than a normal "cold" DFT calculation does. *The consistent use of finite-temperature DFT is one of the most important aspects of performing high-fidelity simulation in the HED regime.*

While the thermal–structural disorder at high temperature relaxes the demands for using a dense k-point sampling compared to that required for calculations of ordered solids at zero Kelvin, it is necessary to apply a high plane-wave cutoff energy in order to converge the electronic pressure [37]. The pressure is calculated with derivatives of the electronic structure and hence converges more slowly than the internal energy. The use of finite-temperature and systematic convergence were two main reasons behind the first high-precision calculations of the deuterium Hugoniot [11]. To summarize, DFT-based molecular dynamics calculations, often referred to as QMD—quantum molecular dynamics—of shock compression requires high

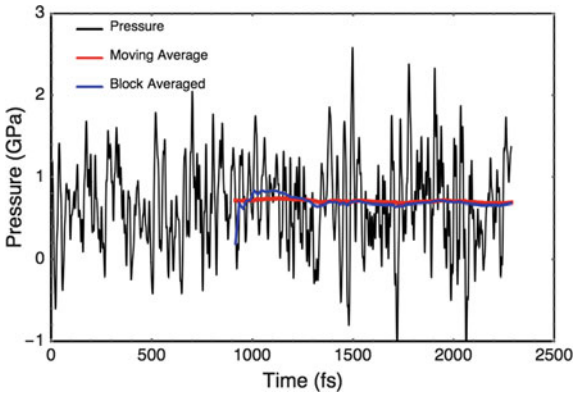


Fig. 1.3 Obtaining thermodynamic information from a long MD simulation. The moving average is a useful measure for when the periodic oscillations in the simulation no longer influence the result. A drift toward lower pressure during the simulation, for example, would indicate relaxation in the system and a longer simulation time is needed

plane-wave cutoff energies, finite-temperature DFT using many partially occupied bands, and the use of pseudo-potentials that are of high quality for normal conditions as well as high-pressure and high-temperature conditions.

Figure 1.3 shows the electronic pressure for a polystyrene simulation at 1.05 g/cc and 300 K. The time step is 0.5 fs. The red line is the moving average of the pressure and is strongly influenced by the correlation between atom positions. In this plot, the initial time for starting the moving average was picked to be close to the estimated mean pressure, so, while the mean of the electronic pressure is 0.67 ± 0.015 GPa, this is misleading as picking a different starting position will yield the same standard deviation but a different mean until many more time steps. Block averaging is a powerful way of extracting uncertainties from correlated data [1]. In this case, the block size used was 37 but made little difference between 20 and 45. The block averaged mean is 0.70 ± 0.08 GPa.

1.2.2 Analysis of Chemical Composition—Tracking Bonds

Of particular importance to simulation of polymers compared to elements is the presence of covalent bonds and the energy associated with bond breaking. The breaking of covalent bonds takes energy, energy that otherwise would be available for increasing the temperature of the shocked material. The result on the Hugoniot curve is the shoulder in temperature evident in Fig. 1.2 at 2.25 g/cm^3 / 120 GPa. In this section, we present a way of identifying molecular structure and apply it to conditions commonly created by strong shock waves in polymer systems.

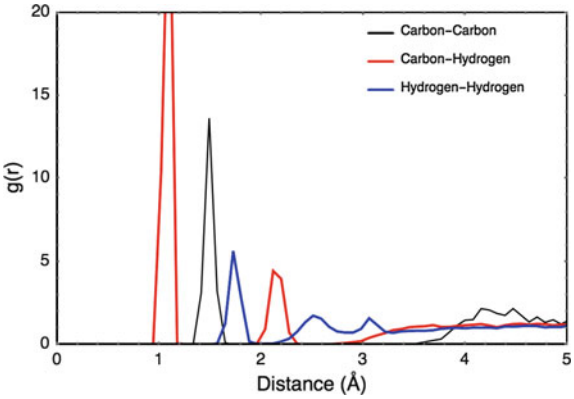


Fig. 1.4 Species-dependent pair correlation function for ethane at 0.571 g/cm^3 and 300 K. The pair correlation plateaus at 1.00 since the pair correlation function is normalized to the random probability of finding an atom in space, hence without any correlation the value is 1.00

Tracking bonds in a simulation is a three-step process. We determine bond lengths, determine bond times, and then, based on those lengths and times, identify and track the bonded atoms. Atoms are considered bonded if they stay within a given distance of each other for a defined amount of time. If, at any time, one of the atoms drifts outside of the prescribed distance and then drifts back in, the time is reset to zero and must again be satisfied.

To determine bond length, we use the species-dependent pair correlation function g(r) to find the distribution of distances between atoms in the reference system, usually at ambient density and temperature. Figure 1.4 shows the pair correlation function of carbon to carbon, carbon to hydrogen, and hydrogen to hydrogen for ethane at 0.571 g/cm^3 and 300 K. These simulations are several picoseconds long, so we use multiple averaging times to obtain different samplings which allow us to check that we do not have a large variation between each averaging time. We then use these values in the bond tracking method on the reference system to confirm that it yields the species of the initial state. Because we use a finite time step and have a small number of atoms, we find that first minimum of the pair correlation for each species is the best choice for bond lengths.

For the following example, we incremented the density of ethane from 0.6 to 4.0 g/cm^3 by 0.1 g/cm^3 and equilibrated each simulation at that density and constant temperature in order to map out the Hugoniot locus points. When the Hugoniot developed a slope inflection, which is often indicative of the onset of dissociation, we calculated the pair correlation on the simulation just prior to the inflection density to confirm that the molecules were still intact. This can be seen in Fig. 1.5. The carbon–carbon bond length is the same. However, the carbon–hydrogen bond is slightly longer, changing from 1.19 to 1.28 Å. Because these values are taken from simulations closer to the dissociation regime in question, these are the values that would be used for the bond tracking.

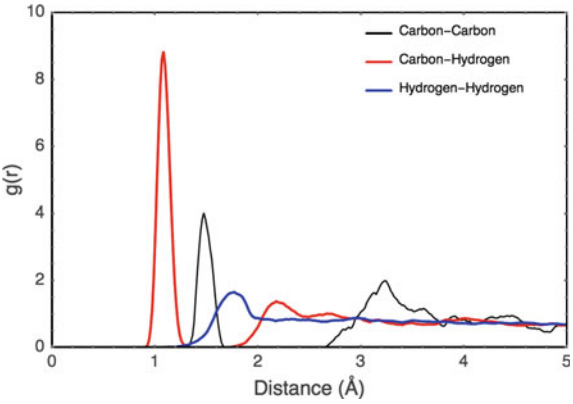


Fig. 1.5 Species-dependent pair correlation function for ethane at 1.2 g/cm^3 and 1050 K. The probability approaches the random value 1.00 at a shorter distance—reflecting less correlation in the positions of atoms than at lower temperature

The next step is to determine the amount of time atoms must be in close proximity to each other to be considered bonded. To start, we run the tracker on the simulation in question and vary the C–C bond distance, C–H bond distance, H–H bond distance, and time requirement.

At 1.5 g/cm^3, 3100 K, we varied the required bond lengths and times in order to determine that there is no sensitive dependence on a particular cutoff parameter. The results from this sensitivity study are shown in Fig. 1.6. The hydrogen–hydrogen is not shown as it made no difference. Each plot uses different carbon–carbon bond lengths, varied between 1.60 and 1.72 Å. The abscissa is the carbon–hydrogen bond length and is the same for all figures. Finally, the different colored lines denote different proximity time requirements from 60 to 120 fs. In a transient system such as this, as time goes to infinity, we expect the free carbon and free hydrogen to approach their atomic ratios (carbon to 33% and hydrogen to 66%). It can also be seen that there appear to be no clear breakpoints for picking one bond distance over another, hence using the pair correlation functions as a guide. If there were distinct features in Fig. 1.6, we would need to investigate the cause for them before continuing.

We use atom velocity as a guide to select a suitable duration. For this example, we used the kinetic theory of gasses and ideal gas law as a crude approximation to the velocity.

$$\frac{1}{2}mv^2 = \frac{3}{2}k_{\mathrm{B}}T$$

where k_{B} is the Boltzmann's constant, T is the absolute temperature, m is the mass (in our case of a single atom), and v is the average Maxwellian velocity. With $T =$ 3100 K, for carbon, we get a velocity of 2538 m/s or 0.02538 Å/fs. Therefore, over 100 fs, we can expect a single atom to move about 2.5 Å. For hydrogen, this would be about 25 Å. We assume that for C–H, the atoms must be bonded if the hydrogen does not rapidly move away from the carbon. Similarly, unless both carbons are moving parallel, they would still drift apart quickly enough to fail the time/distance

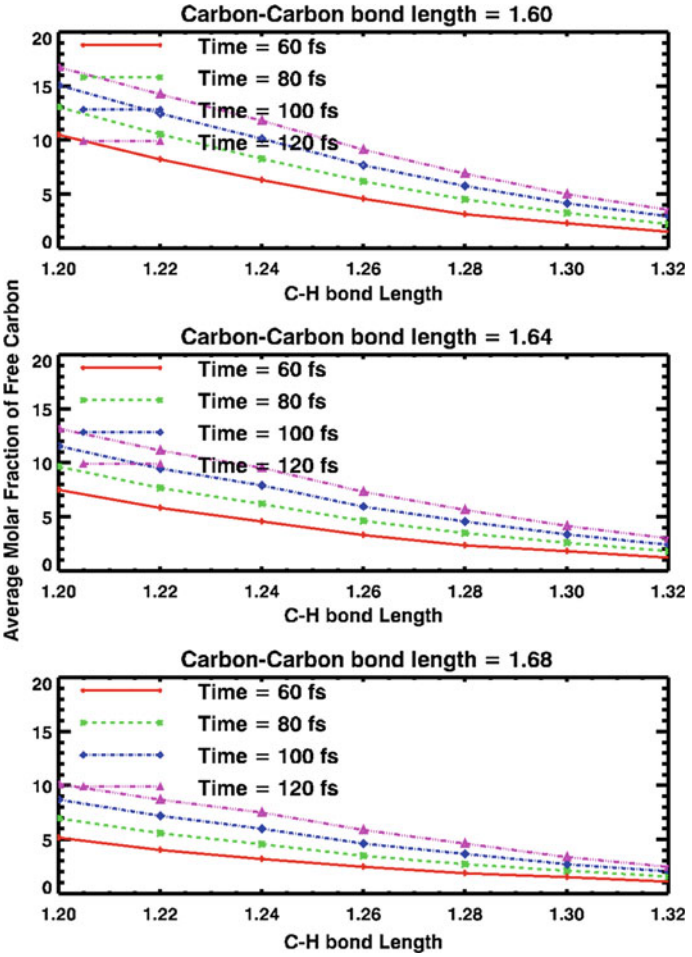


Fig. 1.6 Change in perceived composition as a function of bond distances for ethane at 1.5 g/cm^3 and 3100 K. The changes are smooth, so the selection of a particular bond cutoff does not change qualitative behavior

constraint. We use the vibrational period as a reference for our bond time calculation. A carbon–carbon triple bond has a full period of about 18 fs. For a double bond such as ethylene, the period is about 21 fs, and the period for a single bond in ethane is about 33 fs. The 100 fs time cutoff was chosen to be significantly longer than multiple vibration periods. The tracking method has a hard reset to a given molecule interaction meaning if two atoms move apart farther than the distance cutoff, they must satisfy the distance constraint for the entire time constraint again before they are considered bonded. This allows us to better distinguish between a scattering event and a bond.

After identifying all of the bonded atoms, we can begin to identify molecules. Our method uses a recursive algorithm capable of identifying branched molecules beyond a nearest-neighbor picture. One way to conceptualize this information is by building a three-dimensional construct where the first two dimensions are a matrix of atom-to-atom bonds (designated by either a 0 or 1 but more information can be stored) and the third dimension is time. The algorithm walks through the time-dependent upper diagonal, following all bonded pairs until no more bonds are found in that chain (at which point it starts with the next atom not in that chain) or until all atoms are accounted for.

The recursive nature of the algorithm makes it fast and reliable. The ability of looking beyond nearest neighbor is particularly important for the purpose of changing chemistry. In the case of polyethylene, the polymer structure is never recovered if cooled, but other species may be formed which helps us to understand how the system evolves from a metastable polymer to thermodynamic equilibrium along a given trajectory in a phase diagram.

Figure 1.7 shows the stoichiometry for four strands of polyethylene along the principle Hugoniot calculated according to the method described above. This series of simulations was performed with four strands of CH_2, 16 carbon atoms long and with an extra hydrogen atom at each end to cap the chain and prevent cross bonding (for a total of 200 atoms). For densities of 1.1 to 1.7 g/cm^3, all four strands are intact. At 2.6 g/cm^3, the system is almost entirely an atomic gas. Between 1.7 and 2.7 g/cm^3 is the dissociation regime responsible for the inflection in the Hugoniot. Outside of this regime, the stoichiometry is easy to calculate, but inside the variety and species of molecules is much harder to understand and additional constraints are often required, not on the bond tracking program itself but on the results. For example, one might

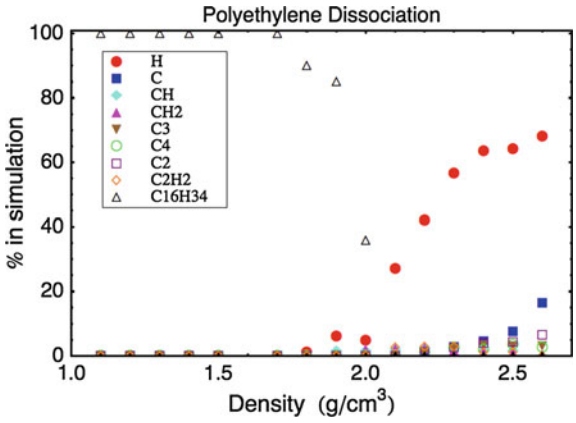


Fig. 1.7 Polyethylene stoichiometry along the principal Hugoniot. These simulations used 4 strands of 16 carbons and 32 hydrogens along the chain plus 2 hydrogens (one at each end) to cap the strand and prevent cross bonding. 1.1 g/cm^3 through 1.7 g/cm^3 show only $C_{16}H_{34}$ which indicates that all four strands are intact. At 2.6 g/cm^3 the system is almost entirely an atomic gas

see a C_1 and C_7H_5 for a few time steps and then see a C_2 and C_6H_5 for a few time steps. These are transients that do exist but are difficult to quantify in a meaningful way given the very small number of atoms present in a DFT simulation. Figure 1.7 shows only the most persistent of the species.

With this tool, we are able to carefully follow the dynamic chemistry of polymer dissociation as a function of shock strength. Accurate accounting of species is important for understanding the role of bond breaking in release of internal energy, for comparisons with classical molecular dynamics simulations using reactive force fields, and for comparison with chemical reactivity models.

1.3 Classical Molecular Dynamics Simulations of Polymers Under Shock Compression

DFT allows extremely high-fidelity descriptions of chemistry over a huge range of compressions; however, its system sizes are computationally constrained to just a few hundred atoms. This can be a significant obstacle, particularly in polymeric systems, where even single chains can easily exceed this number of atoms. Furthermore, in the case of structured polymers and foams, nanoscale features cannot be captured on length scales available to DFT. Thus, for the study of void collapse and hotspot heating, another simulation tool is required.

Classical molecular dynamics (MD) can extend the accessible simulation length scales to tens or hundreds of nanometers (and many millions to billions of atoms), by eliminating the electronic degrees of freedom and reducing quantum mechanical interactions into a classical ion–ion interatomic potential. These potentials are generally quantitatively accurate over smaller compression ranges than DFT, but can be highly accurate within these ranges.

Our classical MD simulations are run with Sandia's LAMMPS parallel molecular dynamics code [44]. LAMMPS allows highly efficient large-scale calculation and offers a broad library of established interaction potentials. For hydrocarbons, we have taken advantage of this library to compare the shock response of several published potentials under identical loading conditions.

In Mattsson et al. [38], we compared the results of two reactive and two bonded interatomic potentials. The reactive force fields, ReaxFF [7] and AIREBO [53] allow for chemical reactions and dissociation. The bonded force fields, OPLS-AA [21] and the Borodin et al. exponential-6 [4, 5] are computationally much more efficient but cannot capture dissociation reactions.

It had been expected that the reactive potentials would do better than the less sophisticated bonded potentials. However, we found that core stiffness was a stronger predictor of a potential's accuracy than its reactivity. Specifically, we found that the AIREBO and OPLS potentials were both too stiff in compression, and diverged from DFT and experiments well before the onset of dissociation chemistry. Since our 2010 study, however, both the AIREBO and OPLS potentials have been reworked to sig-

nificantly address the observed shortcomings in their high compression response. AIREBO-M [43] was refit to high-pressure graphene and polyethylene data, and now reproduces DFT results for shocked PE at least up to 40 GPa. Similarly, Sewell, Frohlich, and Thompson [13] showed how OPLS-AA can be refit to use exp-6 Buckingham-style interactions to replace the stiffer 12–6 Lennard-Jones cores. Given these improvements since our last comparison, we focus our discussion on the ReaxFF and Borodin potentials.

Shock compression states were produced using two MD methodologies. In the case of homogeneous full-dense polymer studies, the Hugoniostat [46] method was used to homogeneously bring the material to the final shock state. This equilibrium approach uses modified equations of motion to correlate the system compression to the internal energy thermostat, so as to enforce the RH relation, (1.3). For shocked foams, we used a full nonequilibrium (NEMD) propagating wave approach [20, 22] in order to capture the nonuniform dynamical collapse around the void spaces. NEMD uses a piston driver on one end of the system, to provide a sustained drive into the sample, producing a shock front, which then propagates across the system. Three representative ambient-density polymer systems are shown in Fig. 1.8. For these studies, we built polymeric systems from either polyethylene (PE) or poly(4-methyl 1-pentene) (PMP). While both materials have the same monomer C:H ratio of 1:2, and nearly identical local structure, they differ in extended structure. PE has a linear backbone, while PMP has a branched structure. Full-density systems for PE (0.930 g/cm^3, Fig. 1.8a), and PMP (0.801 g/cm^3, Fig. 1.8b) were built and equilibrated. In addition, the amorphous PMP was conducive to building distended foam systems (0.30 g/cm^3, Fig. 1.8c) as well. Details of the building and equilibration have been published [38, 49].

Figure 1.9 shows two plots of polymer Hugoniot response from Z experiments, DFT, and MD simulations. The MD simulations shown are for the ReaxFF and Borodin et al. potentials. We note that DFT best reproduces experiments over the full range of compression. The MD does reasonably well at lower compression but is generally stiffer than experimental measurements as pressure increases. In both the PE and PMP, the ReaxFF potential gives the better agreement with experiment between the two MD potentials shown. To within experimental uncertainty, it is accurate from ambient density up to approximately 2× ambient density. The Borodin potential is good from ambient up to ~1.6 g/cm^3. Moreover, the ReaxFF potential shows good agreement with DFT dissociation onset, as shown in Fig. 1.10, which begins at just above 1.8 g/cm^3 (vertical dashed line) in polyethylene. Figure 1.10 plots the percentage of carbon atoms bonded to two other carbons as a function of compression. This simple measure of polymer chemistry shows agreement between DFT and ReaxFF MD in both onset density and density trend to complete vaporization.

Based on Hugoniot response and dissociation, we conclude that MD with the ReaxFF potential gives comparable quantitatively accurate hydrocarbon simulations to DFT under compression up to ~1.8 g/cm^3. This validation allows us to extend atomistic simulations to length scales which are amenable to study of nanoscale foams.

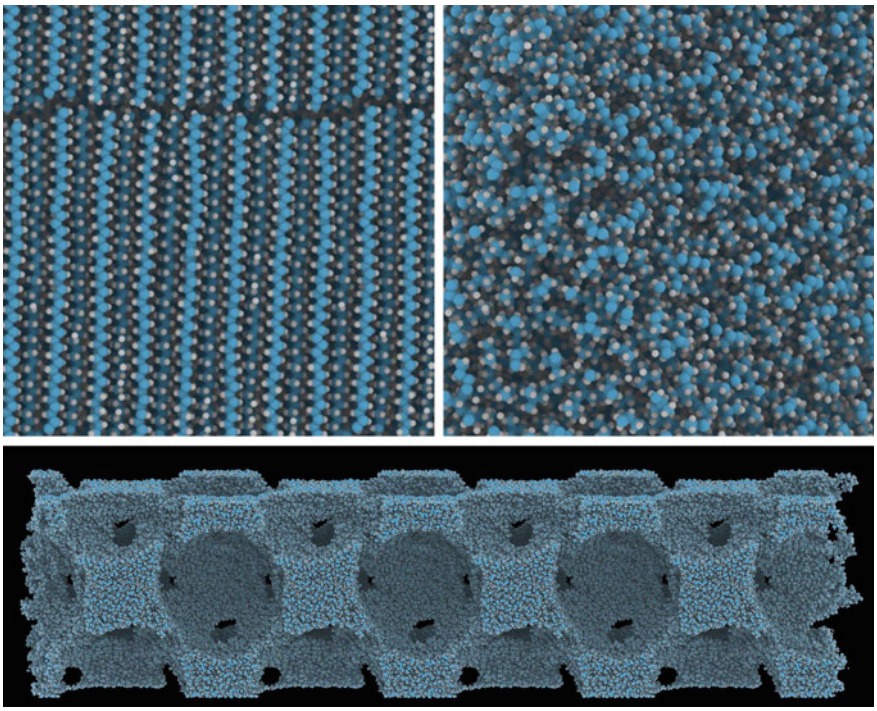


Fig. 1.8 Snapshots from initial configurations of molecular dynamics simulations for semicrystalline polyethylene (top left), poly(4-methyl 1-pentene), i.e., PMP (top right), and a model PMP foam with density of 0.30 g/cm^3 (bottom)

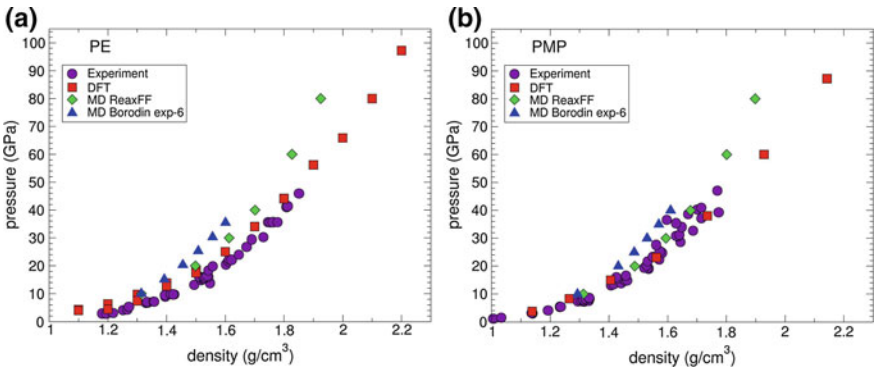


Fig. 1.9 Classical molecular dynamics Hugoniot response compared to experiments [36] and DFT data for full-density polyethylene (left) and Poly(4-methyl 1-pentene) (right) [38]

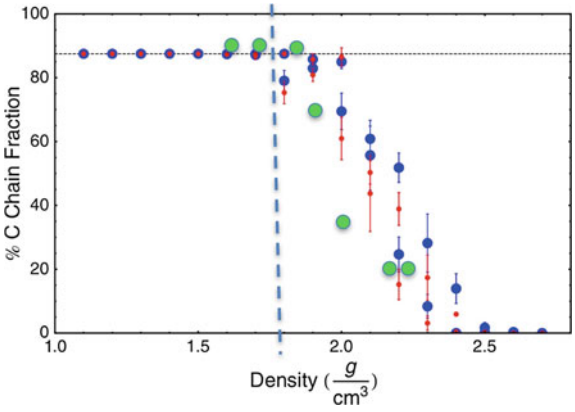


Fig. 1.10 Dissociation, as percent carbon chain, for DFT (20 and 50 fs bond criteria are blue and red, respectively) and ReaxFF classical MD (green) [38]. Maximum percent fraction is 87.5% for the C16 polyethylene chains, and 95% for the C44 chains, shown

While foams and porous materials are frequently useful and necessary in shaping high-pressure waves, their response is notoriously difficult to characterize. Experimental error bars are usually large due to sample-to-sample foam variation, as well as to inhomogeneities within individual samples. Issues of sample control combined with a lack of direct experimental measures of temperature and chemistry make analysis difficult. Root et al. conducted Z machine shock experiments on PMP foams with average density of 0.3 g/cm^3 [49]. The experimental results were published with extensive atomistic, and continuum simulations exploring the effects of porosity on average temperature, vaporization/ejecta, and wave propagation profiles. Subsequent work [30, 31] explored the temporal evolution of void collapse and local temperature hotspot formation.

A foam Hugoniot comparison between Z machine experiments, MD simulation, and Equation of State (EOS) calculations are shown in Fig. 1.11a. These results show that the final densities of these compressed systems do not reach nearly the compaction that the fully dense polymer samples reach these pressures. This is because the void collapse in foams leads to significant heating, which causes expansion in the final states. Because of this, we remain well within the density range for which ReaxFF has been shown to be reliable.

The Hugoniot curves for PMP foam have two distinct regions. At low pressures, the foam compacts to approximately four times its initial density, at which point the curve turns steeply vertical, with little increased density for large increases in pressure. In both the EOS calculations and the MD simulations, the data is tightly aligned along near-vertical lines. The experimental data, colored to match the same initial densities of MD and EOS data, have significant horizontal variation. The simulations help to build confidence for this being due to variation in foam density across different samples.

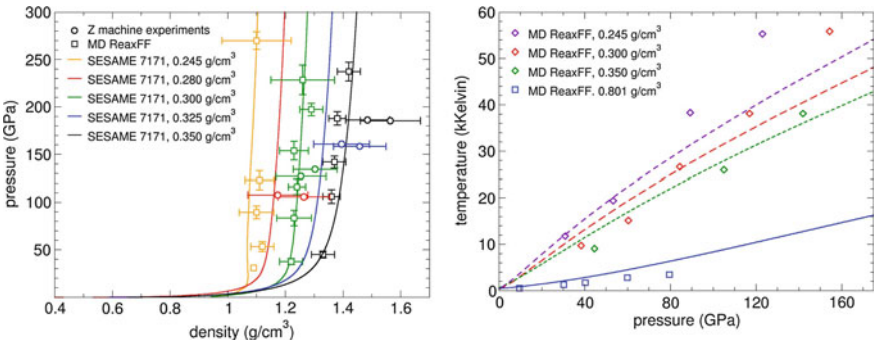


Fig. 1.11 Comparison of the shock Hugoniot from classical MD and experiments on Z [48, 49]. Equation of state results are also shown for SESAME 7171 [12, 19]

Figure 1.11b shows the final average temperatures for the same shot data. Experimental temperatures are not available. The data shows the strong dependence of final temperature on the initial density of the foam. The fully dense polymer temperatures are almost an order of magnitude lower than the foams. And the slope of increased temperature with pressure increases with decreasing initial density. This is consistent with the conclusion that increased void size/number leads to increasing hotspots and higher temperatures. One also sees in the plots that the MD temperatures show the same trends as the EOS temperatures and agree at lower pressures. However, at higher pressures, the SESAME 7171 values are consistently lower than MD temperatures, even when the pressure vs. density values agree. This is likely a result of the fact that the EOS model does not account for the polymer dissociation, which would indicate that the MD temperatures are more dependable. It should, however, be noted that these temperatures are well above the range usually explored with classical molecular dynamics.

The molecular dynamics snapshots in Fig. 1.12 demonstrate one significant mechanism, which was not observed in full-density polymer shock studies. We see significant vaporization and ejecta within the collapsing void. This vaporized material is produced at pressures and densities well below the onset of dissociation in dense polymer, because the temperatures in the collapsing voids are high enough to break bonds. Once the material is volatized, it can travel ballistically ahead of the shock front. In closed foams, these ejected particles impact the far side of the void, preheating the un-shocked material. In an open foam, the vaporized particles can travel far ahead of the shock front through the interconnected void space. The ejected particles travel at velocities as high as twice the bulk particle velocity. This observation explains features seen in the wave profiles. In both simulation and experiment, shock fronts are broader than expected, and the rise time of the front appears to be correlated to the void size.

These results indicate that classical molecular dynamics can be used effectively to model Z machine experiments in polymer shock compression. Specifically, with the right potentials, quantitative agreement can be made with DFT and experimental

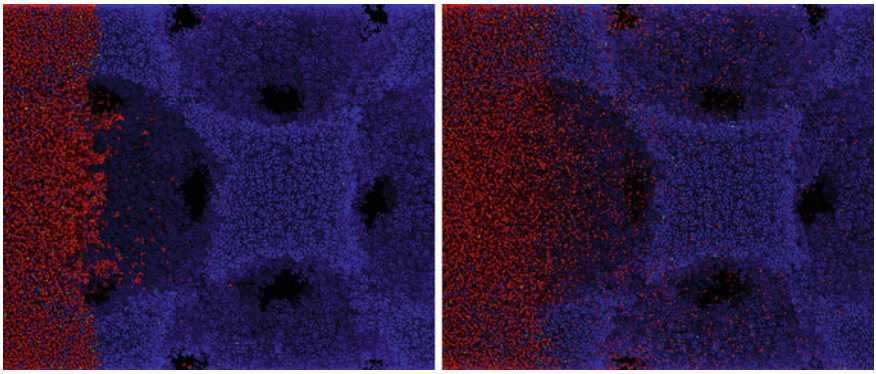


Fig. 1.12 Snapshots of a propagating shock front showing vaporized ejecta from the collapsing void. Atoms are colored by velocity in the shock direction. (left) piston velocity of 10 km/s, giving pressure of ~40 GPa (right), piston velocity of 25 km/s, giving pressure of ~240 GPa. Dissociated polymer is seen extending ahead of the shock front [31]

results over a useful range of compression. Moreover, because MD allows for the simulation of much larger features than those which can be captured with DFT, it can explore the mechanisms and processes associated with larger length scale structure in polymers (i.e., long-range order, defects, or voids). Moreover, aspects of the compression which are difficult or impossible to observe experimentally, including void collapse and temperature production, are readily studied with MD. For these reasons, classical molecular dynamics has proven to be a useful tool to augment our analysis and understanding of Z machine experiments.

1.4 Z Experiments on Polymer Materials

In the 1990s, Sandia researchers developed the Z machine as a pulsed power source for generating X-rays for radiation–material interaction studies and as an X-ray drive for experiments relevant for inertial confinement fusion studies [40, 52]. In 2007, the Z machine was refurbished [51] increasing the maximum current delivered to a target to approximately 20 MA in a pulse length of 100–600 ns making Z capable of producing magnetic pressures up to 600 GPa. Early on, researchers [15] determined that the current pulse could be used to drive a smoothly increasing compression wave into a target load where the time-dependent magnitude of the pressure loading is given as

$$P(t) = \frac{B^2(t)}{2\mu_0} = \mu_0 \frac{J^2(t)}{2} = \frac{\mu_0 I^2(t)}{2W(t)^2}$$

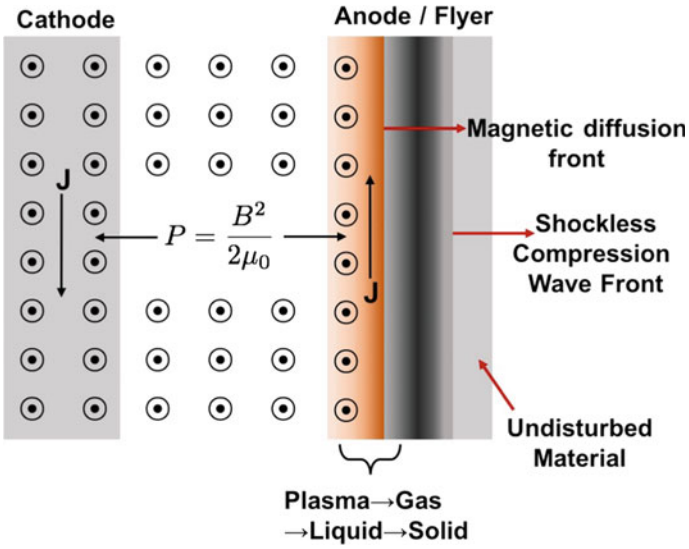


Fig. 1.13 Illustration of the current drive on the anode. The magnetic field penetrates only a fraction of the flyer, the shockless compression front outruns the magnetic field

where P is the pressure, $B(t)$ is the time-dependent magnetic field, $J(t)$ is the time-dependent current density in units of (amps/unit length), $I(t)$ is the time-dependent current, and $W(t)$ is a time-dependent scaling factor that depends on the load geometry. This led to the first shockless compression experiments to measure quasi-isentropes in materials [15]. From prior work at Sandia on the hyper-velocity launcher (HVL) gun using a shockless compression wave to launch flyer plates [8], the idea to use the Z machine to launch flyer plates was quickly conceived [16, 24].

Figure 1.13 shows the basic concept for either shockless compression or flyer plate experiments on Z. Current is driven in a loop along the anode and cathode through a shorting cap on the target's top. This current loop generates a magnetic field between the cathode and the anode. The combination of current and magnetic field creates a Lorentz force ($\vec{F} = \vec{J} \times \vec{B}$) that drives a shockless compression wave through the anode. A sample directly mounted to the anode will be shocklessly compressed as the compression wave transits into it. If the anode is left with a free surface, just as in the HVL, when the shockless compression wave reaches the anode free surface, the anode is accelerated outward from the cathode. Using a shockless compression wave minimizes the temperature rise in the anode, however, the high current causes Joule heating on the inner surface of the anode. This causes the conductivity to drop and the current begins to diffuse into the anode, which creates a magnetic diffusion front that propagates behind the compression wave front. The Joule heating melts and vaporizes the inner surface of the anode [10].

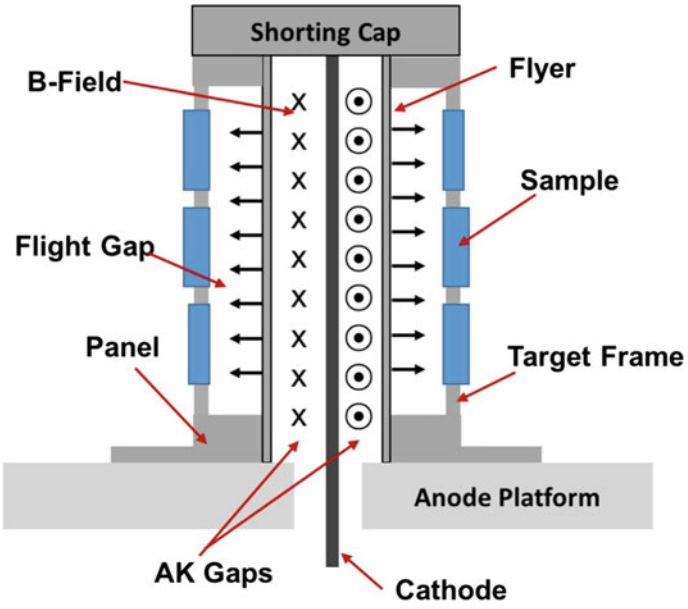


Fig. 1.14 Illustration of a Z coaxial geometry flyer plate experiment. The load is about 5 cm tall and 1 cm wide

The ability to shape the current pulse along with developments in 1D and 2D magnetohydrodynamics (MHD) simulations greatly improved the capabilities to use the Z machine for quasi-isentropic compression experiments [10] and for flyer plate experiments [32, 33]. By proper tailoring of the current pulse, the flyer plate is shocklessly accelerated to the desired velocity. Experiments are designed to ensure a steady shock through the sample and to avoid contamination in the measurement from release waves reflecting from Joule heated, melted portion of the flyer. Further refinements to the Z target geometry and improvements in MHD and Z circuit modeling have led to the successful launching of aluminum flyer plates up to around 40 km/s [34].

Figure 1.14 illustrates a rectangular target geometry used for flyer plate experiments. In this geometry, rectangular flyer plates are placed on opposite sides of the cathode. The flyer plates are typically aluminum, but Z has also used Al/Cu layered flyers where Cu is the impactor. The distance between the flyer plates and cathode, called the AK gap, is asymmetric—one side has a larger AK gap than the other side. This results in two different magnetic pressures that create two different flyer velocities on each side of the target, and thus, two different Hugoniot states in one experiment. MHD simulations are used to determine the flight gap so that the flyer plate reaches a terminal velocity prior to impact with the samples with a few hundred microns of solid density on the impact face [32, 33]. The target frame is designed to hold as many samples as possible to maximize the data return.

The Z machine has been used to study the high-pressure response of the hydrocarbon polymer: poly(4-methyl-1-pentene) plastics up to 985 GPa [50]. Poly(4-methyl-1-pentene) plastic (PMP) is a CH-based plastic in which the carbon atoms are sp^3-hybridized. PMP is often referred to by its trade name TPX® (Mitsui Chemicals, Inc.) and is available in several different variations. PMP is possible to machine and is transparent, which makes it a valuable window material for shock-release studies on the Z machine. For the work in [50], the PMP samples were the DX845 variant of TPX® with a density of 0.833 g/cm^3 and a melting temperature of 232 °C. The index of refraction of PMP is relatively constant over a large range of wavelengths. For the Z experiments, we use a 532 nm wavelength for our laser interferometry system. At that wavelength, PMP has an index of 1.461 ± 0.001.

The laser interferometry system at Z is a velocity interferometer system for any reflector (VISAR) that was originally developed at Sandia National Laboratories in the late 1960s and early 1970s. Details of the VISAR technique for shock wave research are found in [2, 3, 14, 17]. The VISAR measures flyer plate velocities and shock velocities during the Z experiment.

Figure 1.15 shows a schematic illustration of the experimental configuration along with two sample VISAR measurements. An aluminum flyer plate is accelerated toward the PMP samples. Two different sample configurations were used for measuring the high-pressure response of PMP (*inset)*: the top configuration is a thick sample of PMP used to measure the Hugoniot. The bottom configuration is a PMP/quartz window stack used to measure the Hugoniot and the reshock state in PMP. In the experiment, VISAR tracks the flyer plate trajectory, directly measuring the flyer plate velocity up to impact. The flyer plates are shocklessly accelerated to the desired velocity. At impact, a shock is produced in the PMP sample. The shock strength is high enough that the shock front in the PMP becomes reflective and the VISAR laser reflects off the shock front. This allows for a direct, precise measurement of the shock velocity as the shock wave travels through the sample.

Figure 1.15 (red) shows the shock through the PMP is steady for 10 s of ns. At late times in the single PMP sample, the velocity begins to decrease. The velocity decrease is caused by reflected release waves from the melted portion of the flyer (Fig. 1.13). In the quartz backed PMP configuration, a steady wave is observed in the quartz for 10 s of ns as well. At later times, a second shock is observed in the quartz. This is caused by the reflected shock from the PMP/quartz interface and the PMP/aluminum impactor interface. Again, release waves from the melted portion of the flyer cause the decrease in the shock velocity.

The advantage to using the Z machine is that it launches solid density flyers at the targets. Thus, the flyer plate is well characterized at impact. Combined with the direct measurement of the shock velocity in the PMP, the Hugoniot state is easily and accurately determined through impedance matching [6]. A *Monte Carlo* impedance matching (MCIM) method [50] is used to calculate the Hugoniot state and uncertainty. The MCIM method accounts for uncertainty in the measurements, initial densities, and the uncertainty in the Hugoniot of the flyer plate. Figure 1.16 shows the compiled Hugoniot data for PMP along with the QMD calculated Hugoniot for PMP discussed earlier in the chapter.

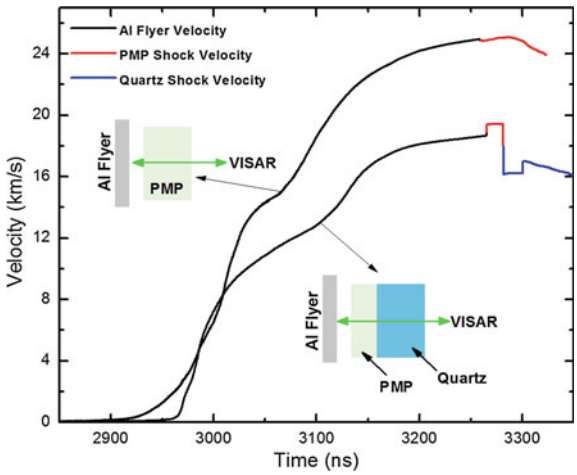


Fig. 1.15 VISAR measured velocity traces from two different experiments on PMP plastics showing the flyer velocity, PMP shock velocity, and the quartz shock velocity. Inset: Configuration for experiments using the direct shock experiment (top sample) and the shock—reshock experiment (bottom sample)

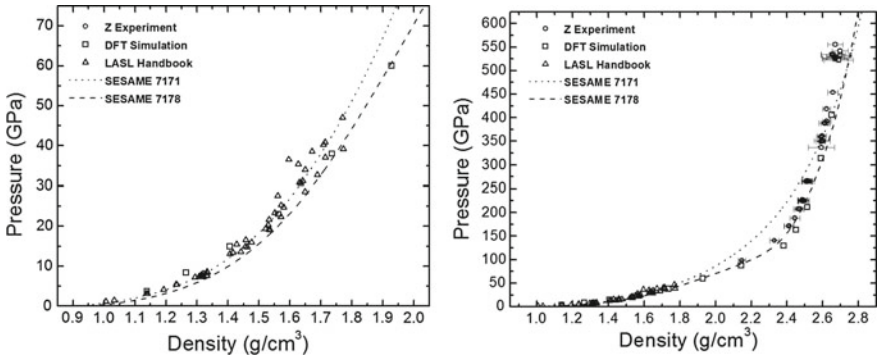


Fig. 1.16 Hugoniot of PMP plastic in density–pressure space. The left panel shows the low-pressure range where there was previous data. The QMD simulations are in excellent agreement with data over a large range

Figure 1.16 shows the Z experimental data range from 100 to 550 GPa along the principal Hugoniot. The QMD calculations are in good agreement up to 400 GPa. Starting at about 150 GPa, the principal Hugoniot steepens dramatically compared to the Hugoniot below 100 GPa. The sharp upturn suggests dissociation of the molecular system into an atomic system. This similar behavior along the Hugoniot was observed in the molecular systems of CO_2 [48, 49] and ethane (C_2H_6) [35].

The bond tracking analysis method discussed earlier was used to assess dissociation along the Hugoniot. Figure 1.17 shows the results of the bond tracking analysis. At low shock pressures of approximately 20 GPa, the polymer chain begins to break

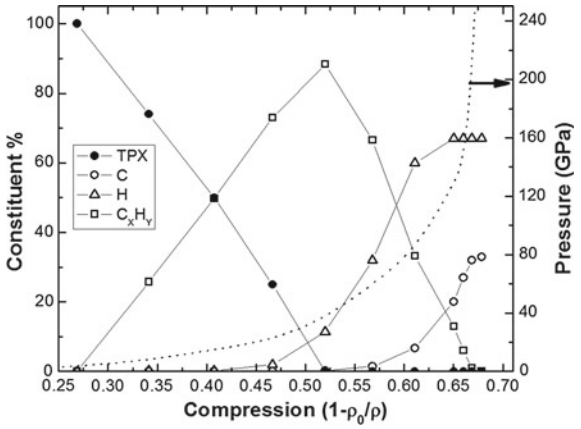


Fig. 1.17 Analysis of the atomic species of shocked PMP using the bond tracking analysis method

down and complex C_xH_y constituents are observed. Increasing the shock pressure increased the amount of free hydrogen. At a shock pressure of approximately 150 GPa, all H atoms have dissociated from the C chains and the number of free C atoms begins to increase. Complete dissociation into C and H atoms requires a shock pressure of approximately 315 GPa.

At 315 GPa, the compression factor $\mu = 1 - \rho_0/\rho$ is $\mu = 0.68$. At 160 GPa where the system is primarily atomic C and H with a small percentage (~6%) of C–C clusters, $\mu = 0.66$. Interestingly, this compression factor of $\mu = 0.66$–0.68 for complete dissociation is consistent with other C–H systems with sp^3-hybridized bonded carbon. Complete dissociation for ethane (C_2H_6) occurs at $\mu = 0.69$ [35], for polyethylene $\mu = 0.62$ [9], and polystyrene $\mu = 0.62$ [54]. For each system, the pressure and temperature at complete dissociation are different, but the compression factors are all similar. This suggests that the C–C bonds are more affected by compression than pressure or temperature since compression reduces the gap between bonding and anti-bonding orbitals making it easier to occupy the anti-bonding states in the C.

1.5 Summary and Conclusions

Hydrocarbon polymers are commonly used in shock physics research, and their behavior is complex. Strong compression of hydrocarbon polymers involves a series of physical mechanisms, together resulting in the Hugoniot response shown in Fig. 1.2. Under progressively stronger shock loading, the hydrocarbon polymers discussed in this chapter (linear polyethylene and polymethylpentene with side chains) first respond softly as the void space between polymer chains is compressed resisted only by the relatively weak inter-molecular Van der Waals interaction. During this phase, the covalent bonds of the polymer remain intact (Fig. 1.7). Once the voids

space has been removed, the atomic repulsion becomes progressively stronger, the temperature continues to increase, and the polymer melts. Upon further compression, dissociation of the polymer occurs over a range in density, pressure, and temperature (Fig. 1.7). The breaking of covalent bonds cost significant energy, resulting in a distinct plateau in the Hugoniot (Figs. 1.2 and 1.16). As the pressure, temperature, and density increase even further, dissociation is complete with the system now being comprised of free atoms. Beyond dissociation, the pressure and temperature of the final shock state rise very steeply with increased shock strength and the state turns into a dense strongly coupled plasma of hydrogen and carbon.

In this chapter, we have shown how to utilize different simulation techniques to interrogate and understand the behavior of two polymers under shock compression. The results from simulations are compared to experimental data from Sandia's Z machine and found to be in qualitative and quantitative agreement. First, molecular dynamics based on density functional theory is used to calculate thermodynamic final states that are in quantitative agreement with high-precision experiments executed on Sandia's Z machine; second, the DFT/QMD simulations are employed to analyze how the molecular structure changes with increasing shock pressure; third DFT/QMD simulations are used to determine the range of validity of model potentials commonly used in classical MD simulations; and fourth classical MD simulations are applied to nonequilibrium conditions to explore the effect of structure like voids.

We conclude that it is possible to model shock compression of polymers with high fidelity for a range of shock strengths when using DFT/QMD and that classical MD simulations provide valuable insights into the shock compression dynamics of mesoscale materials like polymer foams.

Acknowledgements We sincerely thank our many collaborators and colleagues who have inspired or contributed to the EOS work presented in this chapter, including Gary Grest, Mike Desjarlais, Ronald Redmer, Marcus Knudson, Tracy Vogler, Tom Haill, Ray Lemke, Aidan Thompson, Flint Pierce, Diana Schroen, Rudy Magyar, Ann Mattsson-Wills, John Carpenter, Luke Shulenburger, Dan Dolan, Martin French, Mark Herrmann, John Benage, and Dawn Flicker.
We also thank the large team of staff and management operating the Z machine and target fabrication. The work was supported by the NNSA Science Campaigns. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

References

1. Allen, Tildesley (1990) Computer simulations of liquid. Cambridge University Press
2. Barker LM, Hollenbach RE (1972) J Appl Phys 43:4669
3. Barker LM, Schuler KW (1974) J Appl Phys 45:3692
4. Borodin O, Smith GD (2006) J Phys Chem B 110:6279
5. Borodin O, Smith GD, Bedrov D (2002) J Phys Chem B 106:9912

6. Boslough MB, Asay JR (1993) In: Asay JR, Shahinpoor M (eds) High-pressure shock compression of solids. Springer, New York, p 7
7. Chenoweth K, van Duin ACT, Goddard WA III (2008) J Phys Chem A 112:1040
8. Chhabildas LC, Kmetyk LN, Reinhart WD, Hall CA (1995) Int J Impact Eng 17:183
9. Cochrane Kyle R, Desjarlais Michael, Mattsson Thomas R (2012) AIP Conf Proc 1426:1271
10. Davis J-P, Deeney C, Knudson MD, Lemke RW, Pointon TD, Bliss DE (2005) Phys Plasmas 12:056310
11. Desjarlais MP (2003) Phys Rev B 68:064204
12. Dowell F (1982) Technical report no LA-9559-MS, Los Alamos National Laboratory, Los Alamos, NM
13. Fröhlich MG, Sewell TD, Thompson DL (2014) J Chem Phys 140:024902
14. Goosman DR (1975) J Appl Phys 46:3516
15. Hall CA (2000) Phys Plasmas 7:2069
16. Hall CA, Knudson MD, Asay JR, Lemke RW, Oliver B (2001) Int J Impact Eng 26:275
17. Hemsing WF (1979) Rev Sci Instrum 50:73
18. Hohenberg P, Kohn W (1964) Phys Rev 136:B864
19. Holian KS (1984) Technical report no LA-10160-MS, Los Alamos National Laboratory, Los Alamos, NM
20. Holian BL, Lomdahl PS (1998) Science 280:2085
21. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225
22. Kadau K, Germann TC, Lomdahl PS, Holian BL (2002) Science 296:1681
23. Knudson MD et al (2001) Phys Rev Lett 87:225501
24. Knudson MD, Hall CA, Lemke RW, Deeney C, Asay JR (2003) Int J Impact Eng 29:377
25. Knudson MD, Desjarlais MP, Dolan DH (2008) Science 322:1822
26. Knudson MD, Desjarlais MP (2009) Phys Rev Lett 118:035501
27. Knudson MD, Desjarlais MP (2009) Phys Rev Lett 103:225501
28. Knudson MD et al (2012) Phys Rev Lett 108:091102
29. Kohn W, Sham LJ (1964) Phys Rev 140:A1133
30. Lane JMD, Grest GS, Thompson AP, Cochrane KR, Desjarlais MP, Mattsson TR (2012) AIP Conf Proc **1426**, 1435
31. Lane JMD, Grest GS, Mattsson TR (2013) Comput Mater Sci 79:873
32. Lemke RW, Knudson MD, Hall CA, Haill TA, Desjarlais MP, Asay JR, Mehlhorn TA (2003) Phys Plasmas 10:1092
33. Lemke RW, Knudson MD, Bliss DE, Cochrane K, Davis J-P, Giunta AA, Harjes HC, Slutz SA (2005) J Appl Phys 98:073530
34. Lemke RW, Knudson MD, Davis J-P (2011) Int J Impact Eng 38:480
35. Magyar RJ, Root S, Cochrane K, Mattsson TR, Flicker DG (2015) Phys Rev B 91:134109
36. Marsh SP (ed) (1980) LASL shock Hugoniot data University of California Press, Berkeley
37. Mattsson AE, Schultz PA, Desjarlais MP, Mattsson TR, Leung K (2005) Model Simul Mater Sci Eng 13:R1
38. Mattsson TR, Lane JMD, Cochrane KR, Desjarlais MP, Thompson AP, Pierce F, Grest GS (2010) Phys Rev B 81:054103
39. Mattsson TR et al (2014) Phys Rev B 90:184105
40. Matzen MK (1997) Phys Plasmas 4:1519
41. Matzen MK et al (2005) Phys Plasmas 12:055503
42. Mermin N (1965) Phys Rev 137:A1441
43. O'Connor TC, Andzelm J, Robbins MO (2015) J Chem Phys 142:024903
44. Plimpton S (1995) J Comp Phys **117**(1). http://lammps.sandia.gov
45. Prisbrey ST et al (2012) Phys Plasmas 19:056311
46. Ravelo R, Holian BL, Germann TC, Lomdahl PS (2004) Phys Rev B 70:014103
47. Root S et al (2010) Phys Rev Lett 105:085501
48. Root S, Cochrane KR, Carpenter JH, Mattsson TR (2013) Phys Rev B 87:224102
49. Root S, Haill TA, Lane JMD, Thompson AP, Grest GS, Schroen DG, Mattsson TR (2013) J Appl Phys 114:103502

50. Root S, Mattsson TR, Cochrane K, Lemke RW, Knudson MD (2015) J Applied Physics 118:205901
51. Savage ME et al (2007) In: Proceedings of the IEEE pulsed power conference, vol 1–4, p 979
52. Spielman RB et al (1996) In: Proceedings of the 11th international conference on high-power particle beams, vol 150
53. Stuart SJ, Tutein AB, Harrison JA (2000) J Chem Phys 112:6472
54. Wang C, He XT, Zhang P (2011) Phys Plasmas 18:082707

Chapter 2
Computational Discovery of New High-Nitrogen Energetic Materials

Brad A. Steele and Ivan I. Oleynik

Abstract High-nitrogen-content energetic compounds containing multiple N–N bonds are an attractive candidate for new generation of environmentally friendly, and more powerful energetic materials. High-N content translates into much higher heat of formation resulting in much larger energy output, detonation pressure, and velocity upon conversion to large amounts of non-toxic, strongly bonded N_2 gas. This chapter describes recent advances in the computational discovery of a new family of polynitrogen pentazolate compounds using powerful first-principles evolutionary crystal structure prediction methods. After description of the methodology of the first-principles crystal structure prediction, several new high-nitrogen-content energetic compounds are described. In addition to providing information on structure and chemical composition, theory/simulations also suggests specific precursors, and experimental conditions that are required for experimental synthesis of such high-N pentazolate energetic materials. To aid in experimental detection of newly synthesized compounds, XRD patterns and corresponding Raman spectra are calculated for several candidate structures. The ultimate success was achieved in joint theoretical and experimental discovery of cesium pentazolate, which was synthesized by compressing and heating cesium azide CsN_3 and N_2 precursors in diamond anvil cell. This success story highlights the key role of first-principles structure prediction simulations in guiding experimental exploration of new high-N energetic materials.

2.1 Introduction

Energetic materials (EMs) are condensed phase compounds that are used as explosives, propellants and pyrotechnics [1]. Although their development can be traced back to ancient times, commercial large-scale production of EMs began only in nineteenth century when Alfred Nobel introduced nitroglycerin-based explosive

B. A. Steele · I. I. Oleynik (✉)
Department of Physics, University of South Florida, 4202 E. Fowler Ave, Tampa, FL 33620, USA
e-mail: oleynik@usf.edu

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_2

Dynamite in 1867 [2]. Since then, several more powerful EMs were developed at the end of nineteenth and in twentieth centuries that are currently used in munitions, rocket propulsion, mining, construction, and demolition [1]. These EMs are molecular crystals such as trinitrotoluene (TNT), pentaerythritol tetranitrate (PETN), cyclotrimethylenetrinitramine (RDX), cyclotetramethylene tetranitramine (HMX), and triaminotrinitrobenzene (TATB) [1]. They consist of organic C–H–N–O molecules, which combine both fuel (C–H backbone) and oxidizer (nitro (NO_2) or nitrate (NO_3) groups within a single molecule. Upon initiation by impact or heating, such compounds react violently and release energy due to exothermic conversion of the constituent molecules in condensed phase to gas phase products such as CO/CO_2, H_2O, N_2 and solid carbon. The amount of energy released depends on the effectiveness of the C–H oxidation by NO_2 and NO_3 and the amount of nitrogen in constituent molecules. In particular, the conversion of nitrogen to very strong triply bonded N_2 results in an enormous release of energy. These factors are quantified by the notions of oxygen balance and nitrogen content.

Development of traditional C–H–N–O energetic materials faces a stumbling block as their poor oxygen balance and low nitrogen content cannot be further improved [1]. Moreover, their high initiation sensitivity resulting to accidental explosions and their toxicity are of great concern. High-nitrogen-content (high-N) energetic compounds with multiple N–N bonds are attractive alternative towards developing new generation of environmentally friendly, and more powerful EMs. High-N content translates into much higher heat of formation resulting in much larger energy output, detonation pressure and velocity upon conversion to large amounts of non-toxic, strongly bonded N_2 gas.

Compared to carbon, nitrogen is a unique element: N–N double bond is more than twice strong as single N–N bond and triple bond—more than three times strong as single N–N bond, which is not the case for the C–C bonds. As a result, carbon prefers to form extended compounds with mostly single C–C bonds, whereas nitrogen prefers to be in molecular triply bonded N≡N form. Due to this energetic preference, most of double- and single-bonded high-nitrogen compounds, if they exist, should be metastable at ambient pressure and temperature, i.e., their structure is locked in one of the energy minima separated from N≡N global minimum by an appreciable energy barrier. Application of high pressure and temperature (high-P-T) facilitates the efficient synthesis of such high-N compounds, by providing additional stimuli for breaking the strong intramolecular bonds of the original molecular precursors and forming new N–N bonds of the high-N products.

These single- and double-bonded metastable forms of pure nitrogen, if synthesized, are ultimate energetic materials as their conversion to N_2 would result in up to ten times increase in detonation pressure [3]. In 1992 Mailhiot et al. predicted the existence of such single-bonded pure nitrogen cubic gauche (cg-N) crystal structure and showed it is thermodynamically stable over N_2 molecular solid above 50 GPa [4]. Motivated by this theoretical prediction, in 2004 Eremets et al. successfully synthesized the cg-N at 110 GPa and 2,000 K [5]. Unfortunately, the cubic cg-N phase of nitrogen reverts to the molecular form around 40 GPa [6]. Subsequent attempts to recover cg-N at ambient conditions were unsuccessful [6–8].

One of the alternative routes towards synthesis of metastable poly-nitrogen EMs includes stabilization of single and double N–N bonding via addition of small amounts of atoms of other elements into pure nitrogen subsystem to cause the redistribution of electrons in the system resulting in change in character of chemical bonding. The competition between ionic and covalent bonding might promote new nitrogen phases other than triply bonded N_2. Then, the fundamental questions are whether such addition would result in stabilization of novel forms of single and double-bonded nitrogen in condensed phase, what is the role of pressure in overcoming energy barriers associated with such transformations, and what are the suitable precursors and optimal conditions for their synthesis?

To answer these questions, an extensive exploration of bonding and structure is required as a function of pressure, temperature, and chemical composition. A purely experimental approach would be extremely time-consuming and less certain as there are many unknowns at the molecular and atom-scale levels that cannot be uncovered due to limitation of experimental capabilities. Although theory/simulation does have its own limitations, in most cases it can answer the questions that are difficult or even impossible to address using experimental techniques especially those dealing with the atomic scale. Therefore, computational exploration of novel high-N compounds is an attractive alternative that allows for systematic investigation of these new compounds at the atomic scale. The ultimate goal of theory/modeling is to guide experiment by predicting composition-structure phase diagrams of novel high-N compounds, characterize the most energetically preferred materials and suggest specific synthesis routes towards realization of high-PT synthesis. However, to be successful in this formidable endeavor efficient methods for crystal structure prediction are required to make meaningful predictions.

This chapter describes recent advances that were made recently in computational discovery of a new family of polynitrogen alkali and hydronitrogen pentazolate compounds, all of them containing pentazolate N_5^- anion, the last all-nitrogen member of the azole series. Such discovery was made possible through application of powerful first-principles evolutionary crystal structure prediction methods. Importantly, theory/simulations also suggested specific precursors, and experimental conditions that are required for experimental synthesis of high-N pentazolate EMs. To aid in experimental detection of newly synthesized compounds, XRD patterns and corresponding Raman spectra were calculated for several candidate structures. The ultimate success was achieved in joint theoretical and experimental discovery of cesium pentazolate, which was synthesized by compressing and heating cesium azide CsN_3 and N_2 precursors in diamond anvil cell. This success story highlights the key role of first-principles structure prediction simulations in guiding experimental exploration of new high-N energetic materials.

2.2 First-Principles Crystal Structure Prediction

The crystal structure prediction problem is formulated as follows: given a set of specific chemical elements, determine all stable crystal phases and stoichiometries of all chemically stable compounds consisting of these elemental species at specified pressure and temperature. Mathematically, such a problem involves determination of the local and global minima as a function of atomic coordinates of multi-dimensional free energy surface. The solution of this problem is being considered as a grand challenge of theoretical materials chemistry [9]. The brute-force approach to crystal structure prediction would involve generation of all possible atomic arrangements in all possible crystals, optimization of geometrical parameters to get the free energy for each structure, and then determination of the structure at the lowest energy minimum. It is possible to show that the computational expense of such calculation scales exponentially with the number of atoms in the system. For example, finding the energy minimum of the system consisting of 10 atoms would involve 10^{11} combinations of different arrangements of the atoms, which will require the total computational time $\sim 10^3$ years, thus making this approach impractical [10].

In spite of the enormous complexity of the problem, there are several key features of the energy landscape of a complex system of atoms arranged in the crystal structure that make theoretical structure prediction possible [11, 12]. The simplification of the structure search problem occurs when each crystal is optimized, i.e., brought to a local minimum. Such optimization introduces chemical constraints on bond lengths, bond angles, resulting in avoidance of unfavorable contacts between atoms and overall reduction of dimensionality of the potential energy surface.

Another important feature simplifying the global energy landscape is clustering of several structures into basins of attraction (i.e., the same energy minimum) upon their geometry optimization [11]. The basins are typically arranged such that if one hops from one basin to another one, it is more likely the neighboring basin will have a lower energy and the energy barrier between them is small. It turned out that the low-energy basins tend to occur near each other, although they can be widely separated in clumps of low-energy basins. These basics facts tremendously simplify the complexity, thus making the crystal structure prediction possible.

There are several popular state-of-the-art structure prediction codes including random structure searching AIRSS [11], particle swamp optimization code CALYPSO [13] and universal structure predictor method USPEX [10]. The research on computational prediction of high-N EMs reviewed in this chapter has been performed using USPEX, which covers basic functionalities of the first two search methods, AIRSS and CALYPSO. The efficiency of USPEX is based on its carefully designed set of variation operators that allow to produce increasingly lower energy structures in subsequent generations. The USPEX method has been successfully applied to predict several novel crystal structures at high pressures that have been confirmed by experiment such as layered polymeric nitrogen [14, 15], a sodium chloride compound with unexpected Na_3Cl stoichiometry [16], superconducting H_3S [17, 18], and an ionic form of boron [19], to mention a few.

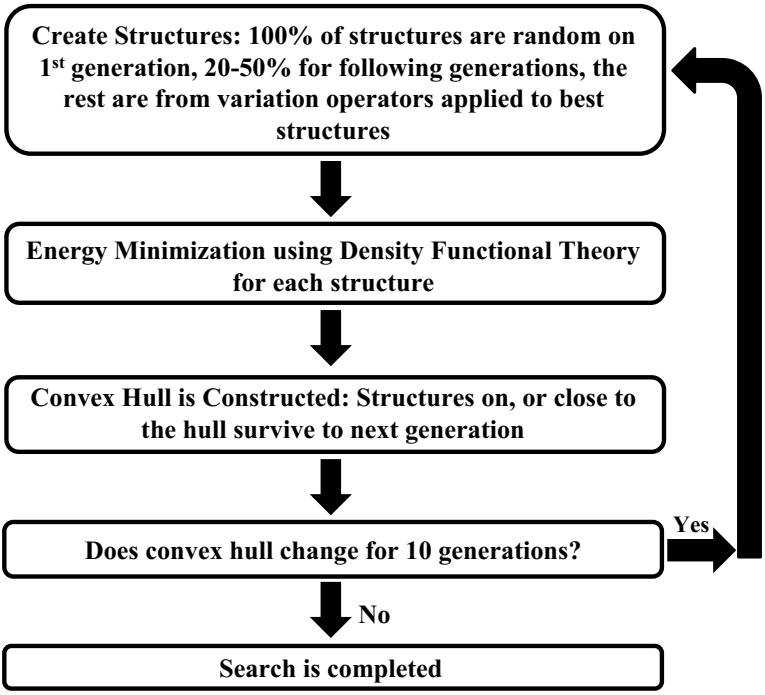


Fig. 2.1 Calculation flow diagram of the crystal structure prediction method USPEX

In the USPEX method, the structure search algorithm deals with generations—a set of crystal structures (individuals), that evolve from generation to generation. The overall schematic of the crystal structure search as implemented in USPEX code is given in Fig. 2.1. The search begins by creating the first generation consisting of random crystal structures with random atomic positions and lattice parameters, followed by the structure optimization of each individual from the first generation. This involves optimization of the atomic positions and lattice parameters at a given pressure of each individual to achieve the minimum of its total enthalpy at a given pressure. The total energy and enthalpy are calculated using first-principles density functional theory.

The computational time is saved by imposing distance constraints that prevent the atoms from being nonphysically close to each other. Constraints are also imposed on the minimum length of all the lattice parameters to prevent them from being too small. In addition, an initial guess for the volume is made based on values of the atomic volumes at a given pressure. The population size, i.e., a fixed number of individuals in each generation, is an important parameter, which is chosen to ensure structural and chemical diversity in each generation. Otherwise, the structure search will get trapped. The number and chemical nature of the atoms in each generated crystal are randomly sampled within a specified range as well as a given set of specific elements.

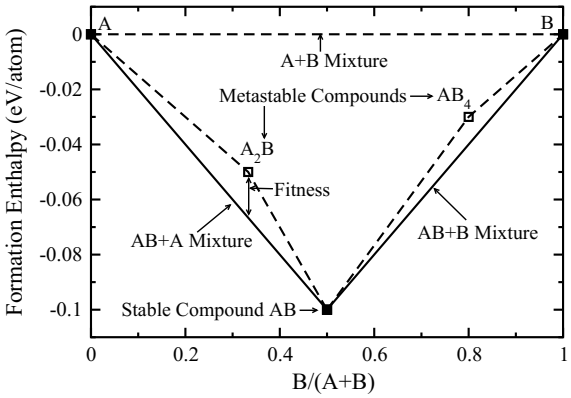


Fig. 2.2 Schematic of a convex hull for a binary system that consists of elements A and B. There are two metastable compounds A_2B and AB_4, their fitness being defined by the vertical distance from the convex hull. Both A_2B and AB_4 are thermodynamically unstable as they may decompose into a mixture of AB+A and AB+B compounds respectively

The complexity, which is defined as a total number of possible structures in a fixed volume, increases rapidly with the number of atoms. This is why using more atoms in the unit cell requires much larger population size.

USPEX can be used to search for binary (A_xB_y), ternary ($A_xB_yC_z$), and higher compositional dimensionality structures, by varying the number of atoms of type A, B, C, … in the unit cell. In the case of a binary search A_xB_y with the number of atoms restricted to be within the range of $\sim 8-20$ atoms per unit cell, a combinatorial calculation gives the estimate for the total number of possible chemical compositions ~ 150. Typical structure search involves about two to three structures per composition on the first generation (random structures only in the first generation and about one random structure per composition in each subsequent generation). By including additional random structures in each generation, trapping in one of the local potential energy minima is avoided.

The algorithm of structure search at a given pressure requires construction of the so-called convex hull for each generation. The convex hull is the convex envelope in the stoichiometry-formation enthalpy space, which is constructed by plotting data points corresponding to each optimized individual of the generation, see Fig. 2.2. In the case of a binary search A_xB_y, the lowest enthalpy data points are connected by line segments, starting from the pure elemental structure A all the way to B, avoiding the stoichiometries that would violate the convex nature of this envelope, see Fig. 2.2.

The formation enthalpy of A_xB_y compound is defined as $H_f(A_xB_y) = \left(H(A_xB_y) - xH(A) - yH(B)\right)/(x+y)$, where H() is the enthalpy of the corresponding material. The convex hull represents the set of *thermodynamically* stable compounds (A_xB_y) at a given pressure and temperature, Fig. 2.2. Lines connecting each pair of binary compounds represent the mixture of the two compounds: the formation enthalpy of each mixture represented by a point on a line is calculated using formation enthalpies of constituent compounds assuming that the interaction energy between two phases is negligible. If a compound is represented by a data point above the convex hull line, then it is thermodynamically unstable and will decompose into the two compounds at the ends of the corresponding line segment of the convex

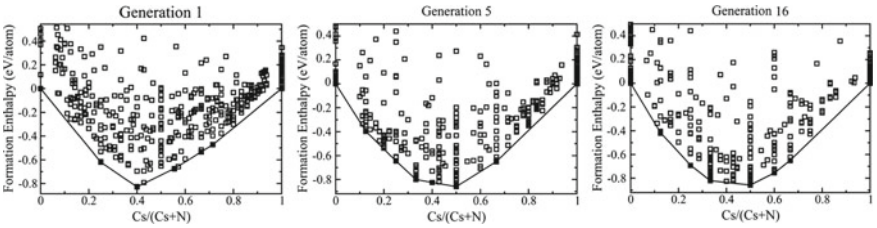


Fig. 2.3 Typical evolution of the calculated convex hull from the first generation to the last generation. The filled squares represent stable compounds on the convex hull. They change from generation to generation until the convex hull is converged

hull. For example, thermodynamically unstable compounds A_2B and AB_4 in Fig. 2.2 may decompose into a mixture of AB+A and AB+B compounds respectively. The compounds represented by the vertices of the convex hull constitute the full set of thermodynamically stable compounds at a given pressure.

To build the next generation, the structures in the previous generation are optimized by DFT and then ranked based on their fitness, which is referred to as the distance from the convex hull line (for binary systems) or from the convex plane (for ternary systems). The fittness is the major descriptor of the evolutionary crystal structure search. In the case of a single element compound, the structures are ranked by their enthalpy per atom (from the lowest to the highest). Then, a fraction of the best structures (typically 60%) from previous generation is selected and variation operators are applied to produce a subset of structures for a new generation. The kind of variation operators varies but it usually involves displacing the atoms and changing lattice parameters by a random value, merging two structures together, swapping atoms, transmutation of atoms into other atoms, etc. In addition, a prescribed fraction (typically 20–50%) of randomly generated structures are also added to every generation. Once the new generation is created, the search cycle is repeated by optimizing every individual of the new generation, constructing the convex hull, ranking individuals and building a new generation, see Fig. 2.1.

A typical evolution of the convex hull from the first generation to the last is shown in Fig. 2.3. With each generations new structures and stoichiometries may appear on the hull. The structures with stoichiometries that lie closer to the hull (a small fitness value) are given higher priority in the evolutionary algorithm designed by USPEX and variation operators are applied to them more often. One desirable feature that can bee seen in Fig. 2.3 is that the first generation is not dramatically different than the last generation, which implies that the first generation has enough individuals to produce a diverse and somewhat high-quality hull. In fact some of the compounds on the hull in the first generation are still on the hull in the final generation, see Fig. 2.3.

The structure search stops under assumption that the entire energy landscape is covered. In practice, it runs until it does not produce any new structures for several generations. Then, it is advisable to rerun the search again with differently generated random structures in the first generation, to make sure that the same structures are

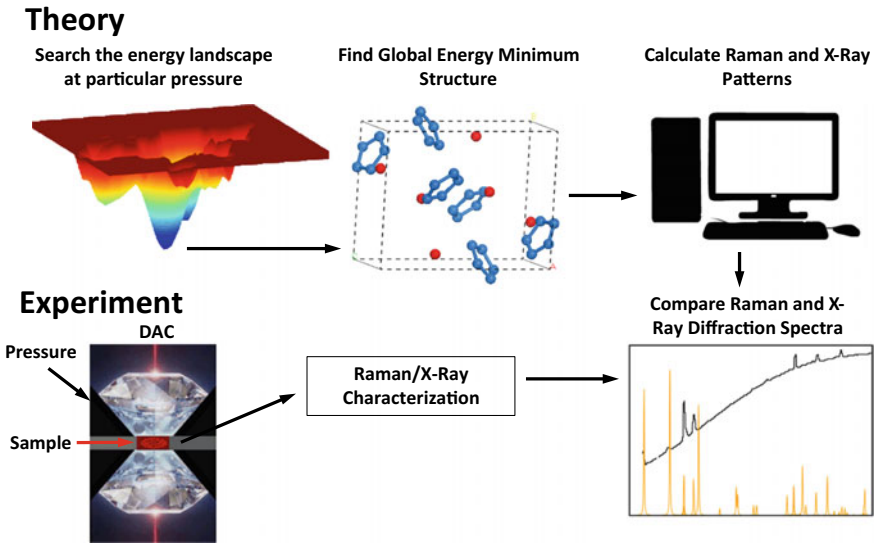


Fig. 2.4 Schematic of combined experimental and theoretical approaches for materials discovery at high pressures. The theoretical thrust involves searching the energy landscape, identifying new crystalline materials as the minima of the multi-dimensional potential energy surface, and calculating corresponding XRD patterns and Raman spectra. Experimental thrust includes compressing and heating the material in a diamond anvil cell (DAC), followed by characterizing the reaction products using Raman spectra and X-Ray diffraction patterns. The two methods come together when comparing the calculated and experimental spectra

found. An additional validation of the search involves predicting a crystal that is already known to exist from experiment: if the structure search finds this structure without any experimental input then the search may be considered to be performed adequately. Care must be taken when the experimental structure is complex (i.e. consists of a large number of atoms), as it may not be found due to system size limitations of DFT calculations.

The typical research project focusing on a discovery of a new high-N EM involves the following stages, see Fig. 2.4. At the very beginning, the elemental composition is specified (A, B, C, …), followed by USPEX variable stoichiometry crystal structure search at several pressures covering the interval of experimentally accessible pressures. The goal is to search a large fraction of the full energy landscape and find the global energy minimum structures for each stouchiometry, as well as energetically competitive local minima. Once the structure search at each pressure is completed, the final convex hulls, representing all thermodynamically stable compounds at each pressure, are constructed. Several structures at the vicinity of the convex hull might also be retained to represent possible candidates of metastable compounds. The degree of metastability is defined by the distance from the convex hull for each dynamically stable compound. The dynamical stability is assessed by calculating the phonon spectrum and making sure that there are no imaginary frequency modes

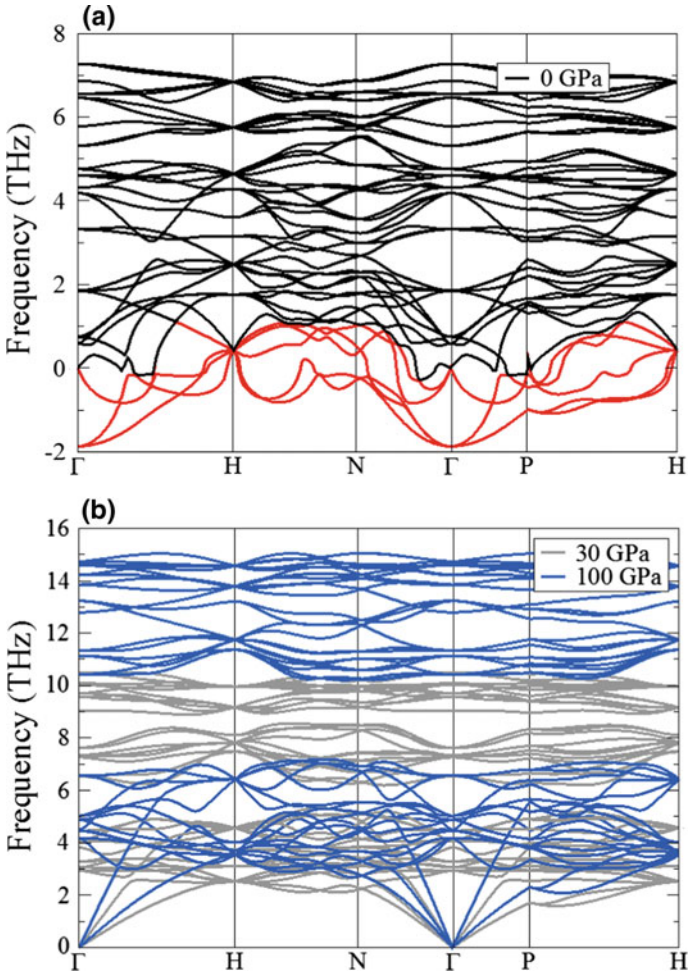


Fig. 2.5 Dynamical stability from phonon spectra: **a** negative/imaginary bands depicted by the red curve signify that this compound is dynamically unstable and is actually a saddle point on the potential energy surface at 0 GPa; **b** phonon spectrum does not have negative/imaginary bands, therefore, this compound is dynamically stable at 50 and 100 GPa

present in the spectrum, see Fig. 2.5. An additional dynamical stability test can be performed by running molecular dynamics simulations at sufficiently-high temperatures and checking whether chemical decomposition of constitutive molecules is observed or not.

Once the stable and metastable compounds are found, their pressure-dependent evolution is investigated to determine the conditions when each of them becomes stable, metastable, or unstable. Then, the possible synthesis routes and mechanisms are explored by identifying specific precursor compounds found during the search

that can be combined in corresponding proportions to produce a specific target compound. To understand whether synthesis can be accomplished by compressing and heating precursors to initiate the high-PT transformations, the evolution of the Gibbs free energy difference between reactants and products is followed as a function of applied pressure to determine whether there is a thermodynamic driving force towards such transformation and at what pressure the target compound becomes energetically preferred in respect to a mixture of the precursor compounds.

The final step of first-principles structure prediction involves calculation of powder X-ray diffraction (XRD) patterns and Raman spectra of the predicted high-N compounds. The XRD patterns (intensity versus 2θ plots) are obtained from known crystal structure by using standard crystallographic XRD software such as POWDER CELL [20] or VESTA [21]. The calculation of Raman spectra is more involved: off-resonant Raman frequencies are obtained within frozen phonon approximation by calculating phonons at the gamma point using DFT, and their intensities are obtained by calculating the derivatives of macroscopic dielectric polarizability tensor along the normal mode eigenvectors.

The crystal structure prediction can only be successful if the results of the search are validated by experimental discovery of the predicted compounds. A combined experimental and theoretical approach involves two interdependent thrusts, see Fig. 2.4. The theory/simulation thrust focuses on searching the energy landscape and eventual prediction of the new stable and metastable compounds. Experimental thrust focuses on high pressure/temperature synthesis of these new compounds by compressing and heating precursors suggested by theory in a diamond anvil cell (DAC) and characterizing the resultant materials using Raman and X-Ray diffraction measurements. The two methods come together by comparing the calculated and experimental XRD patterns and Raman spectra, thus confirming synthesis of new compounds.

2.3 Computational Discovery of High-N Pentazolate Energetic Materials

Although synthesis of cyclo-N_5 pentazolate, an energetic nitrogen oligomer, was attempted since the end of nineteenth century, it was discovered only in the mid-1950s as a part of an aryl pentazole molecule [22, 23] using synthetic organic chemistry methods. This last elusive member of azole series was shown to be aromatic with N–N bond lengths 1.3−1.35 Å, intermediate between single (hydrazine, 1.45 Å) and double (*trans*-diimine, 1.25 Å) bonds [24]. The pentazole (HN_5) or the N_5^- anion [25–30] were not isolated until Vij et al. produced N_5^- in the gas phase in 2002 by cleaving the C–N bond in substituted phenylpentazoles [31, 32]. These pentazolates are important components of high-N energetic materials as they can release a large amount of energy upon conversion of the single-double N–N bonds in the N_5^- aromatic ring to triple N–N bonds in the N_2 molecule. This makes pentazolates

potentially important components for the development of new, green, high-energy-density materials.

Although there exist several metastable nitrogen species, such as linear N_3^- anions [33–35], gas phase pentazolate N_5^- anions [31], N_5^+ chain cations [36], and gas phase N_4 [37], the only other experimentally observed all-nitrogen compounds are covalently bonded non-molecular crystalline phases of nitrogen, which were synthesized at very high pressures (>120 GPa) and temperatures (>2000 K) [4, 5, 15]. We have recently predicted several condensed phase pentazolates [38–40] in an attempt to extend the range of metastability of high-N compounds with the ultimate goal of their recovery at ambient conditions.

Here we review the work directed towards predicting the new pentazolate compounds and suggesting viable transformation pathways of their synthesis from precursor mixtures at high pressures in order to guide experimental synthesis efforts. The idea we pursued in this work is to explore additions to pure nitrogen system, such as alkali metals or hydronitrogens, to achieve enhanced stability and metastability of pentazolates in the solid phase. It was found that the electron-donating species such as alkali metals, ammonium, and hydrogen transfer appreciable negative charge to the N_5 rings, thus enabling both aromaticity in the isolated N_5^- and ionic bonding between non-N cations and pentazolate N_5^- anions within the crystalline environment.

2.3.1 Sodium Pentazolates

Sodium is one of the alkali metals that might be effective in transferring electrons from metallic atoms to nitrogen oligomers. In addition, there is an experimental indication that new high-N compounds containing Na are formed at high pressures and temperatures. In fact, Eremets et al. in 2004 performed experiments by compressing and laser heating sodium azide (NaN_3) precursor to high pressures [35], and observed new peaks in the Raman spectra in the 700–800 cm^{-1} and 1,000–1,300 cm^{-1} range, which can not be attributed to molecular vibrations of the initial azide precursor [41]. They assumed that these new peaks originate from either polymeric nitrogen or a compound containing nitrogen molecular clusters [35]. However, it was impossible to reach an unambiguous conclusion on the nature of the Raman peaks without knowing the structure and composition.

Being motivated by previous experimental attempts, a systematic search of Na_xN_y materials has been attempted using structure prediction method USPEX [38]. The number of atoms in the unit cell in the structure search was varied between 6–16 atoms. To evaluate enthalpies of individuals in USPEX generations, first-principles calculations were performed using the Perdew–Burke–Ernzerhof (PBE) generalized gradient approximation (GGA) functional [42] within density functional theory (DFT) implemented in VASP [43]. The PBE functional was shown to give reliable results for sodium azide [41]. The calculated enthalpies of the predicted structures are used to determine the most stable compounds at a given pressure, thus allowing us

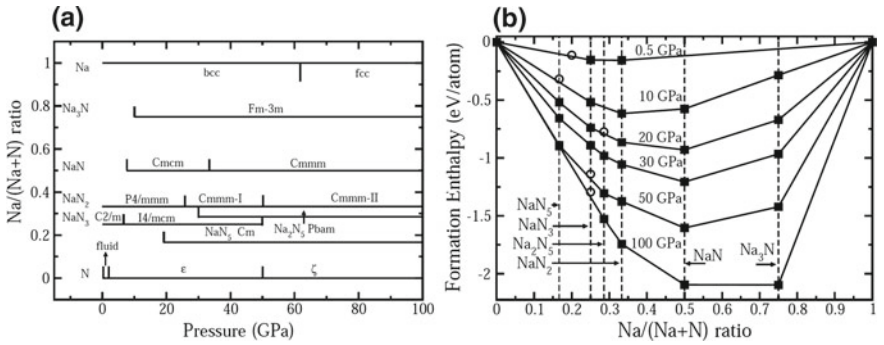


Fig. 2.6 **a** Pressure-composition phase diagram of new Na–N crystal phases discovered in simulations. **b** Convex hull diagram at pressures 0.5, 10, 20, 30, 50, and 100 GPa. Solid squares represent thermodynamically stable phases, open circles—metastable phases. Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

to construct the phase diagram of the materials of varying stoichiometry. The crystal structure search methodology is validated by predicting known phases of sodium azide (NaN_3) without any input from experiment. The α-phase of NaN_3 with the symmetry C2/m is found to be the most stable at 0.5 GPa in agreement with experiment, while at 30 GPa the I4/mcm polymorph of NaN_3 appear [44, 45]. At 60 GPa, the P6/m-Na_2N_6 structure containing N_6 rings is found to be the lowest enthalpy phase with a 1:3 sodium to nitrogen ratio.

The convex hull at a range of pressures up to 100 GPa for the Na_xN_y system is given in Fig. 2.6a. The reference structures for the convex hull are α-N_2, ϵ-N_2, and cg-N for nitrogen and bcc-Na and fcc-Na for sodium each taken at corresponding pressure of their stability. Snapshots of several crystal structures found during the search are given in Fig. 2.7.

In order to further justify the adequate accuracy of calculated formation enthalpies using the PBE functional, the convex hull at 50 GPa has been calculated using both PBE functional and the hybrid HSE06 functional [46], see comparison in Fig. 2.8. Overall, both HSE06 and PBE convex hulls are very similar, although the HSE06 formation enthalpies are slightly higher than those calculated using PBE functional, with exception of Na_3N. The HSE06 functional is considered to be state-of-the-art and gives formation enthalpies and atomization energies close to experiment across a wide range of molecules and crystals [47]. Therefore, the similarity of the two curves demonstrates a good accuracy of the PBE calculations of the pentazolate systems reviewed in this chapter.

One interesting result from the structure search was that NaN_3 no longer resides on the convex hull above 50 GPa, thus implying it is thermodynamically unstable beyond this pressure. This results is also reproduced with the HSE06 functional, see Fig. 2.8. Therefore, upon compression in a diamond anvil cell (DAC) above 50 GPa, NaN_3 will transform into a combination of Na_xN_y phases.

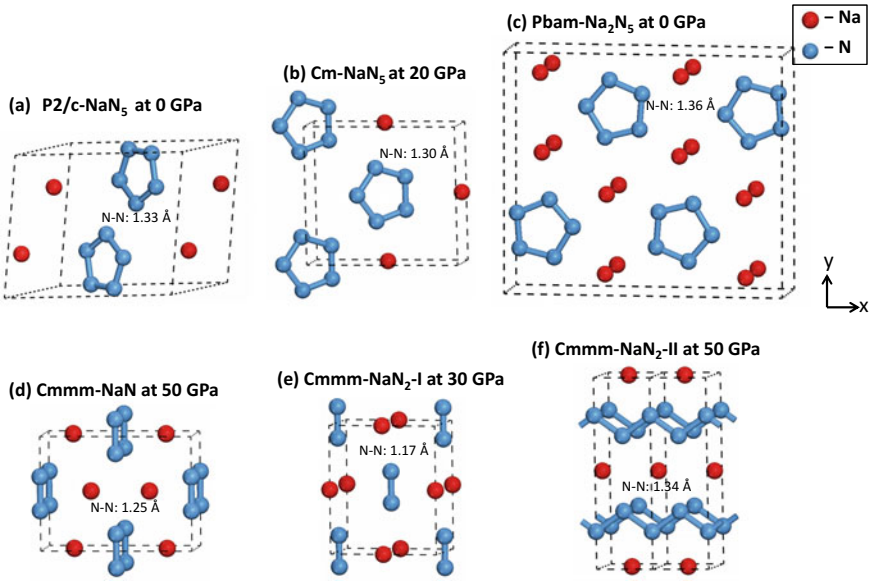


Fig. 2.7 Several newly predicted Na_xN_y crystals at pressures corresponding to where they are thermodynamically stable, or in the case of P2/c-NaN_5 and Pbam-Na_2N_5 metastable at 0 GPa. The space group composition of each structure is shown as well as the N bond length. Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

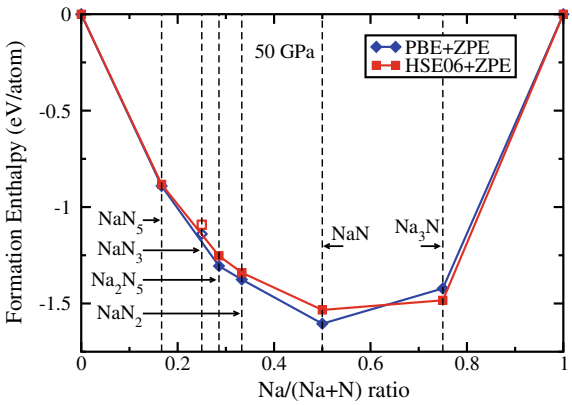


Fig. 2.8 A comparison between the convex hull's calculated with PBE and HSE06 at 50 GPa with zero-point energies (ZPE) included. Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

Our search found several high-N crystals containing cyclo-N_5 pentazolate: P2/c-NaN_5, Cm-NaN_5, and Pbam-Na_2N_5, shown in Fig. 2.7a–c. They are found to be on the convex hull from 20 GPa for NaN_5 and 30 GPa for Na_2N_5 up to 100 GPa see Fig. 2.6b. These newly discovered materials can be potential sources of unidentified Raman peaks in experiments on compression of NaN_3 at high pressures [35]. We predict that NaN_5 may be synthesized by compressing NaN_3 above 50 GPa, since NaN_3 becomes unstable above this pressure. Alternatively, direct synthesis of NaN_5

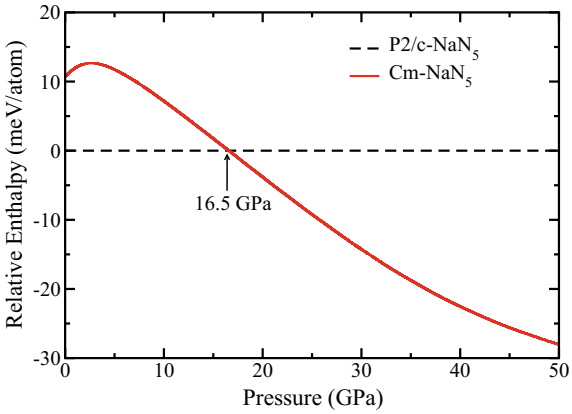


Fig. 2.9 The relative enthalpy of the two phases with NaN_5 stoichiometry. Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

can be facilitated by heating and compressing NaN_3 in a DAC using a nitrogen-rich pressure medium via the following chemical reaction, $N_2 + NaN_3 \rightarrow NaN_5$.

To investigate the metastability of new EM materials at ambient conditions, the most energetically preferred polymorphs of NaN_5 and Na_2N_5 are determined by performing fixed composition search at 0 GPa. A new polymorph, P2/c-NaN_5 is found to be the lowest enthalpy structure compared to Cm-NaN_5, see Fig. 2.7a, while Pbam-Na_2N_5, see Fig. 2.7c, remains the only polymorph of Na_2N_5, metastable at ambient conditions as well as at higher pressures. The transition pressure between the high-pressure Cm phase of NaN_5 (Cm-NaN_5) and the zero pressure phase (P2/c-NaN_5) is calculated to be 16.5 GPa, see Fig. 2.9.

To be useful as emergent EMs, the predicted sodium pentazolates must be metastable at ambient conditions. Metastable high-N structures are those that are slightly above the convex hull and dynamically stable, see Sect. 2.2. The dynamical stability of the crystal phases P2/c-NaN_5 and Pbam-Na_2N_5 is determined by checking the absence of any imaginary frequencies in the phonon spectra at 0 GPa in the entire Brillouin zone. Independent check of the dynamical stability of these two phases at 0 GPa is performed by running DFT molecular dynamics simulations at 1,000 K in the NVT ensemble, which displayed no chemical decomposition in the system, thus confirming that both of these phases are dynamically stable at 0 GPa.

To identify the appearance of these newly predicted phases during their synthesis, the Raman and infrared spectroscopy characterization is performed, which includes calculation of the phonon frequencies at the gamma point, followed by the mode assignments. Both structures contain modes in the interval 80–250 cm^{-1} that are lattice modes as well as pentazole librational modes in both phases. Bending modes of the pentazolate molecules have frequencies at 635 cm^{-1} in Pbam-Na_2N_5 and 766 cm^{-1} in P2/c-NaN_5. Deformational modes of the pentazolate molecule have frequencies 900 cm^{-1} in Pbam-Na_2N_5 and 990 cm^{-1} in P2/c-NaN_5. The frequency of the symmetric stretching modes of the pentazolate molecule is at 1,050 cm^{-1} in Pbam-Na_2N_5 and 1,170 cm^{-1} in P2/c-NaN_5.

Table 2.1 Calculated Mulliken charges and Mayer Bond Orders for the nitrogen atoms and bonds in each new crystal at 0 GPa. Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

Structure	N-cluster	Charge	Bond order
P2/c-NaN_5	N_5	−0.166	1.42
Pbam-Na_2N_5	N_5	−0.306	1.20
P4/mmm-NaN_2	N_2	−0.340	2.09
Cmmm-NaN_2-II	N-chain	−0.320	1.17
Cmcm-NaN	N_2	−0.630	1.46

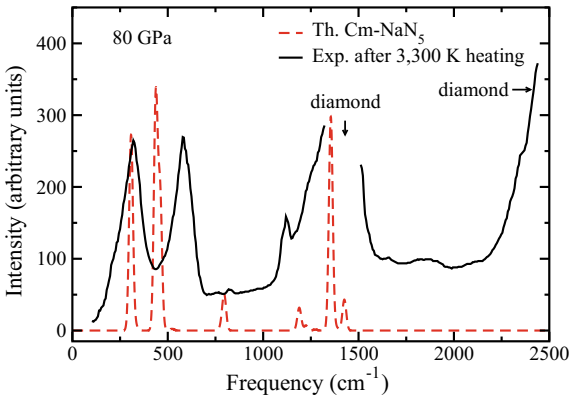


Fig. 2.10 Comparison of theoretical (red dashed line) and experimental (black solid line) [35] Raman spectra of Cm-NaN_5. Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

To determine the character of chemical bonding, bond lengths and bond orders are calculated. It was found that the pentazole anions in the NaN_5 and Na_2N_5 crystal phases are aromatic. The N–N bond lengths in the pentazoles are between the N–N single bond (1.449 Å as in hydrazine [48]) and double bond (1.252 Å as in trans-diimine, [49]); see Fig. 2.7a–c. The calculated bond orders are also between the single (1.0) and double (2.0) bond; see Table 2.1. The charge (−0.83 e) and the bond order (1.42) are also close to those found in the gas phase N_5^- anion (−1 e and 1.45 respectively), which displays the structural and chemical similarity of the N_5^- anion in both gas phase and crystalline NaN_5 environments. The calculated band structure demonstrates that P2/c-NaN_5 is an insulator with a band gap of 5 eV while Pbam-Na_2N_5 is metallic.

It is likely that the new sodium pentazolate structures might appear in experiments by Eremets et al. [35]. At pressures above 80 GPa, and a temperature of about 3,000 K, the peaks in the Raman spectrum associated with sodium azide were found to disappear. This is in agreement with the theoretical prediction that sodium azide becomes thermodynamically unstable at high pressures. In addition, the Raman spectrum of the newly predicted sodium pentazolate phase Cm-NaN_5 shows good agreement with that in experiment; see Fig. 2.10. In particular, the agreement is best near 760 cm^{-1} (the pentazole bending mode) and near 1,150 cm^{-1} (the penta-

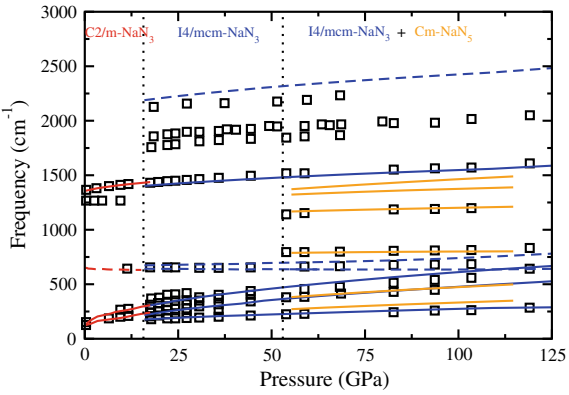


Fig. 2.11 Theoretical NaN_5 and NaN_3 Raman active (solid lines) and IR active (dashed lines) modes as a function of pressure compared to those measured in experiment [35] (black open squares). Reprinted with the permission from [38]. Copyright 2016, with permission from Elsevier

zole deformation mode). Also, in agreement with experiment, the Cm-NaN_5 phase has two lattice modes with appreciable intensity at 307 and 440 cm^{-1} compared to the experimental frequencies at 320 and 580 cm^{-1}. The relatively large difference between the experimental and theoretical frequencies of the second peak may be due the non-hydrostatic effects at 80 GPa which are not considered in our calculations. In addition, it is highly unlikely that NaN_3 will be fully converted into the pure NaN_5 compound. Therefore a one-to-one correspondence between theory and experiment is not expected due to possible appearance of unidentified nitrogen-containing species as well as bulk sodium upon compression and heating of NaN_3.

The theoretical and experimental frequencies of the Raman-active modes of NaN_5 at room temperature as a function of pressure also agree remarkably well, see Fig. 2.11. As shown in Fig. 2.11, the internal Raman-active modes of the pentazole in NaN_5 with frequencies 760 and 1,150 cm^{-1} appear in the experimental spectra above 50 GPa that are not from the initial azide compound. This is in agreement with the prediction that NaN_5 becomes stable above 20 GPa. Since these measurements are performed by compressing NaN_3 at room temperature without additional heating, it expected that complete conversion of NaN_3 into NaN_5 is not achieved. That is why the peaks from sodium azide (NaN_3) are still present in the experimental spectrum up to 100 GPa, see Fig. 2.11. Overall, the agreement between theory and experiment suggests the synthesis of the predicted NaN_5 compound.

2.3.2 Cesium Pentazolates

Although the discovery of NaN_5 consisting of N_5^- was interesting, the experimental evidence for its existence was solely relying on the Raman spectra, which does not directly probe the crystal structure. In order to provide a more definitive conclusion whether alkali pentazolates exist at high pressures, another structure search was performed for cesium polynitrogen compounds at high pressures [39]. The cesium atoms scatter X-rays much more strongly than sodium atoms, therefore it is expected

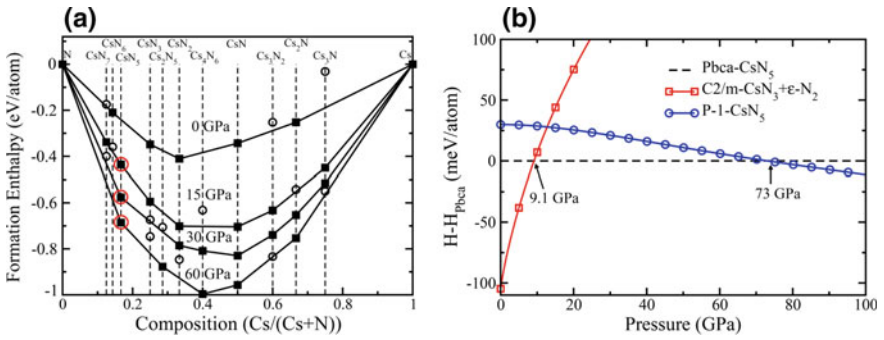


Fig. 2.12 **a** Convex hull diagram at pressures up to 60 GPa. Solid squares represent stable phases, open circles—metastable phases, large red circles—stable CsN_5 phase on the convex hulls. **b** Relative enthalpy difference between two cesium pentazolate (CsN_5) polymorphs and the ε-N_2 phase of solid nitrogen plus the C2/m-CsN_3 phase of cesium azide as a function of pressure. Reprinted with permission from [39]. Copyright (2017) American Chemical Society

that the XRD diffraction pattern would be of much higher quality, thus allowing to unambiguously confirm the experimental synthesis of the new compound.

Therefore, variable composition structure search was performed using the USPEX code for a system consisting of a variable number of cesium and nitrogen atoms (Cs_xN_y) at several pressures up to 60 GPa [39]. The convex hull is first constructed during variable composition search at several pressures, 0, 30, and 60 GPa, followed by fixed stoichiometry crystal structure searches using up to 8 formula units per unit cell. The dispersive correction due to Grimme [50] is added to DFT energy and forces to take into account the long-range van der Waals forces, which are significant in the system under investigation due to the large polarizability of Cs atoms.

The convex hull for the Cs_xN_y system is given in Fig. 2.12a. The reference structures for calculating the formation enthalpy are the α-N_2, ϵ-N_2, and cg-N [4, 7] crystal phases for nitrogen, and the bcc-Cs and Cmca-Cs (Phase V) phases for Cs [51], each taken at its corresponding pressure of stability. The calculations demonstrate that cesium pentazolate salt CsN_5, consisting of pentazolate anion rings (N_5^-) and cesium cations, is on the convex hull at just 15 GPa (Fig. 2.12a). The relative enthalpy difference between the mixture of cesium azide plus dinitrogen ($CsN_3 + N_2$) and reference pentazolate phase Pbca-CsN_5, plotted in Fig. 2.12b, indicates that CsN_5 is energetically preferred above a relatively low pressure of 9.1 GPa. The latter pressure is about 41 GPa lower than the predicted transition pressure of molecular nitrogen to single-bonded threefold coordinated cubic gauche phase of nitrogen (cg-N) in condensed phase [4]. This indicates that the synthesis of CsN_5 requires much lower pressure and temperature stimuli than is needed to synthesize cg-N (over 100 GPa and 2,000 K) [5].

As in case of sodium, discussed in Sect. 2.3.1, CsN_5 consists of isolated N_5 pentazolate anions and Cs cations, see predicted crystal structures in Fig. 2.13. The other compounds discovered during the search are: Cs_2N_5 also consisting of pentazolate molecules; CsN_3 consisting of N_3 azides at pressures from 0–50 GPa and then N_6

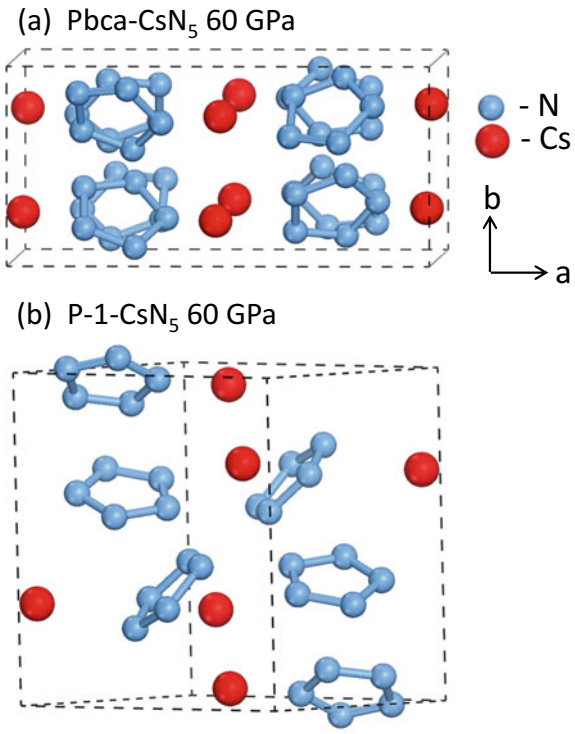


Fig. 2.13 Two energetically competitive cesium pentazolate (CsN_5) polymorphs at 60 GPa discovered during the structure search: **a** Pbca-CsN_5; **b** P-1-CsN_5. Reprinted with permission from [39]. Copyright (2017) American Chemical Society

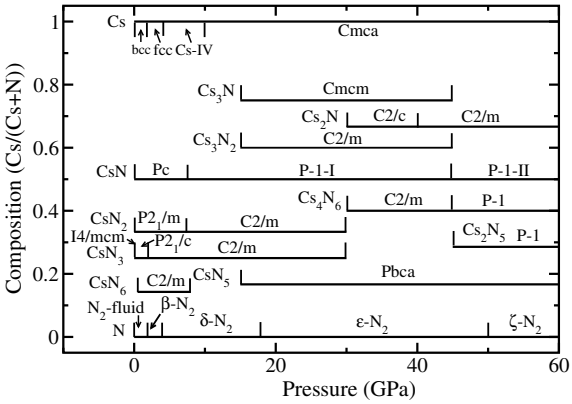


Fig. 2.14 Predicted phase diagram for Cs_xN_y system which contains several novel crystal structures. Reprinted with permission from [39]. Copyright (2017) American Chemical Society

rings at pressures above 50 GPa; CsN_2 consisting of diatomic N_2 anions from 0–40 GPa and then infinite chains of nitrogen above 40 GPa, Cs_4N_6 which consist of N_6 rings from 30 to 60 GPa, CsN which also consist of diatomic N_2 anions from 0–40 GPa and then N_4 chains above 40 GPa; Cs_3N_2 and Cs_2N also containing N_2 anions; and finally Cs_3N consisting of of isolated nitrogen atoms. The summary of the stability pressures for each compound is given in the phase diagram in Fig. 2.14. Note

that although Cs_2N_5 and CsN_3 are not on the hull at 60 GPa, see Fig. 2.12, the R_2N_5 stoichiometry is on the hull for R=Na, and RN_3 is on the hull at lower pressures for both R=Cs and R=Na as shown in Fig. 2.6 [38, 39]. Similar poly-nitrogen compounds have been predicted for other group-I alkali metals [38, 40, 52–54], which implies that nitrogen rings and chains are energetically favorable at high pressures when doped with an alkali metal.

2.3.3 *Pentazole and Ammonium Pentazolate*

Several group-I alkali pentazolates (RN_5) were predicted to be thermodynamically stable at high pressures, but one material that had yet to be found is pentazole HN_5. Hydrogen is unique among group-I elements because it is a gas at ambient conditions and does not become metallic until much higher pressures. In addition, hydrogen can covalently bond to nitrogen to create a rich variety of compounds including ammonia (NH_3), hydrazine (N_2H_4), diimine (N_2H_2), triazene (N_3H_3), tetrazene (N_4H_4), hydrogen azide (N_3H), and ammonium azide (NH_4)(N_3) [55]. The latter compound contains an ammonium cation NH_4^+, which is chemically similar to heavy alkali cations [55–57]. In an analogy with alkali pentazolates, one can also envision the existence of ammonium pentazolate (NH_4)(N_5). Metallic ammonium consisting of NH_4^+ cations glued together by the sea of free electrons has been hypothesized to exist at high pressures inside giant planets such as Uranus and Neptune [56, 57]. It is therefore possible that the high-pressure chemistry of hydronitrogens may be much different than alkali poly-nitrogen compounds.

The high-pressure chemistry of hydronitrogen systems has been recently in the focus of experimental and theoretical investigations [58–64]. The experiments using the mixture of standard precursors, such as molecular N_2 and H_2, with nitrogen content ranging from 5 to 80% suggest the formation of hydronitrogen compounds as evidenced by the disappearance of N_2/H_2 vibrons and the simultaneous appearance of N–H stretching modes [58, 59]. However, a conclusive identification of the type of new nitrogen oligomers or extended networks as well as the determination of the crystal structure have not been made. A hydronitrogen precursor ammonium azide (NH_4)(N_3), containing both H and N, has not shown any signs of chemical transformation upon compression up to 70 GPa [62, 65]. Although theoretical calculations indicate the appearance of polymeric hydronitrogen compounds at a relatively low pressure in a variety of H–N stoichiometries [60, 63, 64], no such transformation was observed in experiment when hydrogen azide is compressed to high pressures [66].

To bridge the bap between experiment and theory, novel hydronitrogen crystalline compounds have been searched at high pressure using first-principle crystal structure prediction [40]. The goal is to understand whether pentazolate compounds such as pentazole N_5H and ammonium pentazolate (NH_4)(N_5) do exist and to identify specific conditions needed for their synthesis. Variable composition USPEX calculations are performed at 30 and 60 GPa. After the variable composition search is completed, fixed-compositions structure searches with larger number of atoms

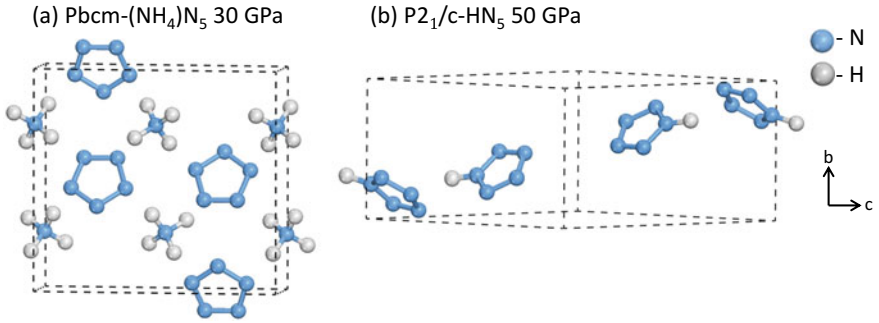


Fig. 2.15 New hydronitrogen pentazolate compounds discovered during N_xH structure search: **a** ammonium pentazolate $(NH_4)(N_5)$, and **b** pentazole N_5H. Reprinted with permission from [40]. Copyright (2017) American Chemical Society

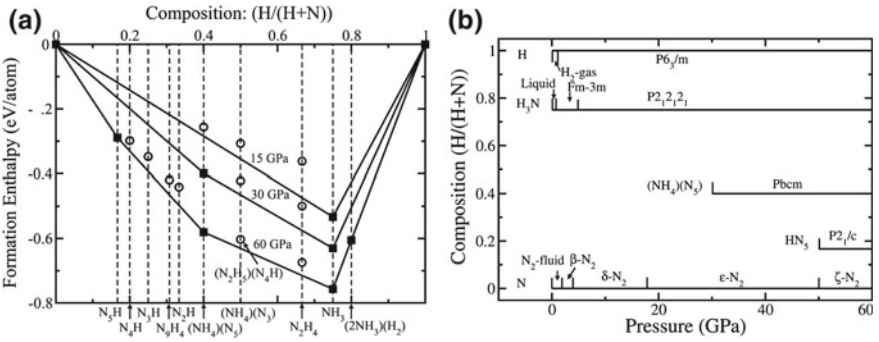


Fig. 2.16 Hydronitrogen phase diagram: **a** formation enthalpy-composition convex hull, and **b** the crystal structure/pressure stability diagram. At 15 and 30 GPa the elemental N structure is ε-N_2 and at 60 GPa—cg-N. For all pressures the elemental hydrogen crystal structure has space group $P6_3/m$ [67]. Reprinted with permission from [40]. Copyright (2017) American Chemical Society

(up to 40 atoms/unit cell) are performed to find the lowest enthalpy structure for each composition.

Two new crystalline materials, ammonium pentazolate $(NH_4)(N_5)$, and pentazole N_5H, are found, featuring all-nitrogen cyclic pentazoles, see Fig. 2.15a, b. The first crystal, $(NH_4)(N_5)$, containing isolated ammonium cations NH_4^+ and pentazolate anions N_5^- (Fig. 2.15a), appears on the convex hull at 30 GPa (Fig. 2.16). The second crystal, N_5H, consists of cyclo-N_5 covalently bonded to a H atom (Fig. 2.15b), and appears on the convex hull at 50 GPa. Both structures are predicted to be dynamically stable at 50 GPa for N_5H and 30 GPa for $(NH_4)(N_5)$ as they lack any imaginary frequencies in the phonon spectrum in the entire Brillioun zone.

The appearance of these two new compounds dramatically modifies the convex hulls at lower pressures. In particular, the only thermodynamically stable crystals at 60 GPa are pentazole N_5H, ammonium pentazolate $(NH_4)(N_5)$, ammonia NH_3, and a crystal containing a mixture of NH_3 and H_2 molecules with net stoichiometry NH_4

(Fig. 2.16a). The formation enthalpy of hydrazine N_2H_4 [68] is also calculated to be above the convex hull, therefore it is also metastable. At a lower pressure of 30 GPa, ammonium pentazolate $(NH_4)(N_5)$ is still thermodynamically stable, whereas pentazole N_5H is not. At 15 GPa, $(NH_4)(N_5)$ is marginally metastable, leaving only one stable compound—ammonia (NH_3) (Fig. 2.16a). Another known metastable compound, ammonium azide $(NH_4)(N_3)$ with symmetry P2/c [69] (and stoichiometry N_4H_4) is also displayed at 15 and 30 GPa. Ammonium azide is predicted to undergo the phase transition to trans-tetrazene (TTZ) at 42 GPa [62], followed by another transition at higher pressure to a crystal that consists of N_2H_5 molecules and infinite nitrogen chains (N_4H) [64] therefore, the lowest enthalpy polymorph of metastable (N_4H_4) with P1 space group is shown at 60 GPa as well.

Both pentazole and ammonium pentazolate crystals contain aromatic cyclo-N_5 with the N–N bond lengths 1.30–1.35 Å, which are intermediate between single N–N (1.45 Å) and double N=N (1.25 Å) bonds. The strength of the aromatic N–N bond in the N_5^- ring in the predicted crystals $P2_1/c$-N_5H and Pbcm-$(NH_4)(N_5)$ is approximately the same as that in the gas phase. To quantify this conclusion, the total Mulliken charges, Mayer bond orders, and bond lengths in the N_5^- ring in both the crystalline environment and the gas phase are calculated at 0 GPa and reported in Fig. 2.17. In the case of N_5H this comparison is straightforward as the structural unit

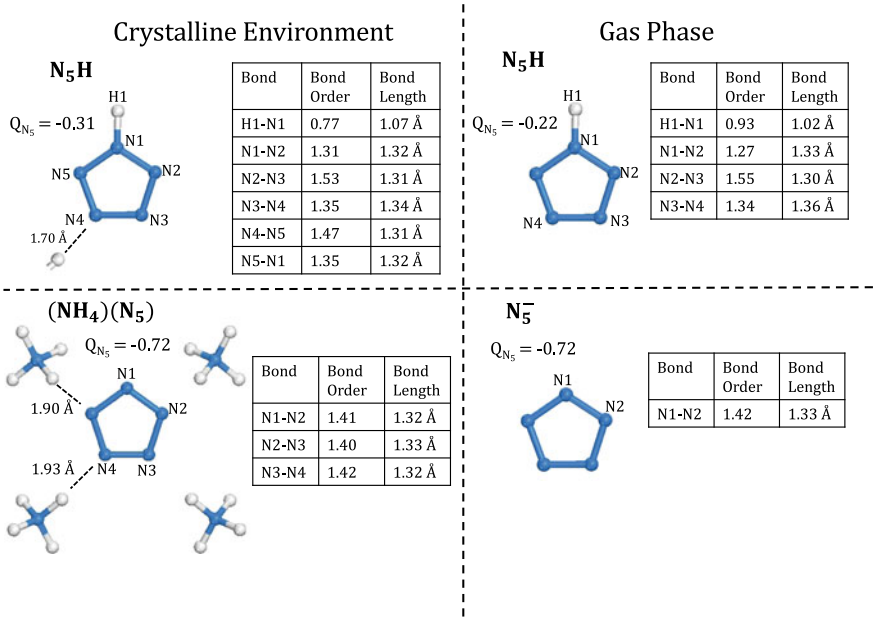


Fig. 2.17 Total Mulliken charges, Mayer bond orders, and bond lengths in the N_5^- ring in $P2_1/c$-N_5H and Pbcm-$(NH_4)(N_5)$ crystals compared to those in gas phase N_5H and N_5. The N···H, hydrogen bonds are also shown in both crystalline environments. Reprinted with permission from [40]. Copyright (2017) American Chemical Society

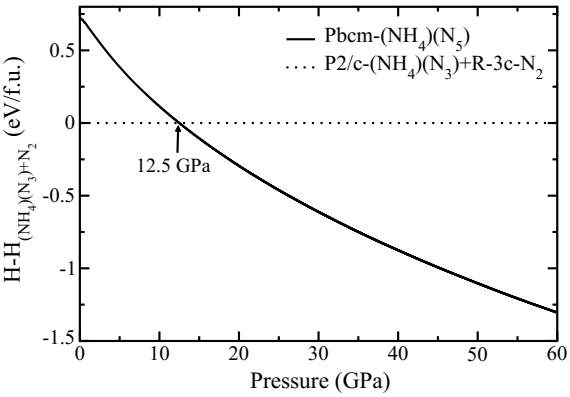


Fig. 2.18 The enthalpy difference between the predicted ammonium pentazolate ($Pbcm$-$(NH_4)(N_5)$) crystal and ammonium azide ($P2/c$-$(NH_4)(N_3)$) plus dinitrogen (R–$3c$–N_2). The predicted transition pressure is 12.5 GPa. Reprinted with permission from [40]. Copyright (2017) American Chemical Society

N_5H is electrically neutral. The N_5 bond lengths and bond orders are mostly the same in both the gas phase and in the crystalline environments, except for bonds N4–N5 and N5–N1 which have a larger bond order compared to the gas phase, see Fig. 2.17. The bond lengths of both crystalline and gas phase N_5H are in a good agreement with those calculated by Ferris and Bartlett [25].

To make a unique comparison of bond orders of the N_5 ring in crystalline $(NH_4)(N_5)$ with those in the gas phase N_5, its negative charge (-0.72 e) was fixed to that in the crystalline $(NH_4)(N_5)$. The bond lengths ($\approx$1.33 Å) and bond orders ($\approx$1.42) in both environments are very close, the gas phase N_5^- possesses fivefold D_{5h} symmetry so each bond length and bond order are the same.

Ammonium pentazolate was predicted to become thermodynamically stable at lower pressure (30 GPa) than pentazole N_5H (50 GPa). To achieve its synthesis, the stoichiometrically balanced mixture of ammonium azide $(NH_4)(N_3)$ and diatomic nitrogen (N_2) should be compressed to high pressures in a diamond anvil cell to activate the chemical reaction $(NH_4)(N_3) + N_2 \rightarrow (NH_4)(N_5)$. The pressure-dependent enthalpy difference between the products and reactants for this transformation shown in Fig. 2.18 displays the possibility of the phase transformation at relatively low pressure of 12.5 GPa. Ammonium azide alone has been reported to be chemically stable upon compression up to 70 GPa at room temperature [62]. However, setting up the right stoichiometry in the diamond anvil cell by adding N_2 might activate the conversion of the azide (N_3^-) anions and N_2 molecules to the N_5^- ring at pressures above 12.5 GPa. The higher pressures and temperatures seem to be required to overcome the significant energy barrier associated with this transformation.

2.4 Comparison of Pentazolate Crystals

The predicted crystal structures, phase transition pressures, and pressure range of stability of crystal phase pentazolates RN_5 are summarized in Table 2.2. Their crystals structures are different: all but a couple possess a different space group

Table 2.2 Comparison of various pentazolates RN_5: the predicted crystallographic space groups, phase transition pressures, range of stability, charges and bond orders

Stoichiometry	Space groups	Pressure range of stability (GPa)	N_5 Charge	N_5 bond order
HN_5	$P2_1/c$ [40]	50–70	−0.31	1.31
$(NH_4)N_5$	Pbcm [40]	30–60	−0.72	1.41
LiN_5	$P2_1/c$ 65 GPa → C2/c [54]	15–100	–	–
	$P2_1$ 20 GPa → $P2_1/m$ [52]	10–100	–	1.42
NaN_5	Cm: 20–100 GPa [38]	20–100	−0.83	1.42
CsN_5	Pbca 73 GPa → P-1 [39]	15–100	−0.85	1.41
	$Cmc2_1$ [53]	14–100	−0.85	1.37–1.46

symmetry. Although LiN_5 and HN_5 crystals have the same space group symmetry $P2_1/c$, molecular constituents are different: N_5^- in LiN_5 and HN_5 in HN_5 crystal. One possible reason for the diversity of crystal structures is that the steric repulsion between cations of different alkali atoms varies greatly when moving down the column of the periodic table. Therefore, as the atomic radius gets larger, the packing and orientation of the N_5 molecules may change resulting in different crystal volumes and different molecular arrangements within the unit cell, especially at high pressures when the PV term in the enthalpy becomes significant.

The charge transfer from cations to N_5^- ring anions enhances the stability of crystal phase pentazolates as it enables aromaticity and increases the electrostatic attraction between ions. The calculated charges and bond orders in the N_5 ring are almost identical among different pentazolate salts, see Table 2.2. The main exception is pentazole (HN_5) which has a smaller amount of negative charge on the N_5 ring, see Table 2.2. An interesting observation is that there is an inverse relationship between the amount of charge transfer to N_5 ring and the stability pressure of pentazolate crystals, see Table 2.2. HN_5 has the smallest charge transfer which causes the pressure of stability to increase up to 50 GPa, while $(NH_4)N_5$ only has slightly less charge transfer than the alkali pentazolates so the pressure of stability is reduced to 30 GPa. In contrast, the pressure of stability decreases down to 10–20 GPa for alkali-metal pentazolates which also have the largest charge transfer to N_5, see Table 2.2. The small charge transfer to N_5 in HN_5 also lowers the bond order slightly in comparison to alkali pentazolates as shown in Table 2.2, which might also play a role in increasing the stability pressure.

All the alkali pentazolates reviewed in this paper have been predicted to be metastable at ambient conditions with no imaginary frequencies and no decomposition observed at high temperature MD simulations [38, 52]. However, pentazole (HN_5) and ammonium pentazolate $(NH_4)N_5$ are not dynamically stable at ambient conditions [40], which seems to be due to reduced charge transfer to N_5.

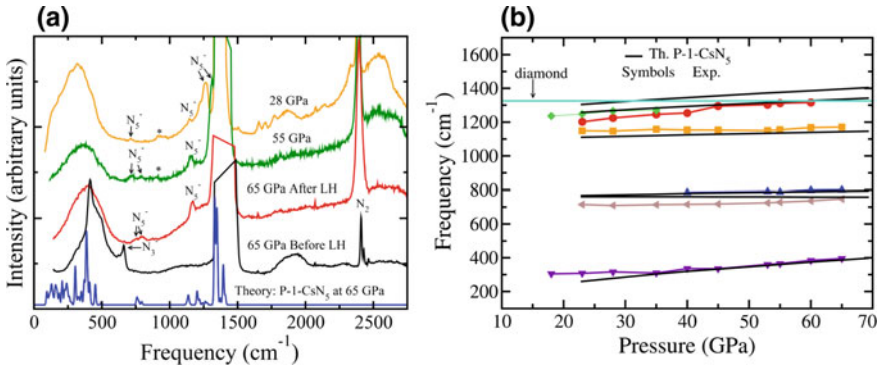


Fig. 2.19 **a** Experimental Raman spectra of the $CsN_3 + N_2$ mixture at 65 GPa before laser heating (LH) (black curve), and of CsN_5 after LH (red curve) then upon decompression to 55 GPa (green curve) and 28 GPa (amber curve). Theoretical Raman spectrum of P-1-CsN_5 polymorph (blue curve) is also shown for comparison. Arrows indicate N_5^- peaks. **b** Pressure dependence of experimental and theoretical frequencies of Raman-active modes of CsN_5: theory—black lines, experiment—colored symbols connected by different colored lines for different modes. The light blue line represents the lower frequency of the first-order Raman scattering of diamond. Reprinted with permission from [39]. Copyright (2017) American Chemical Society

2.5 Synthesis of Cesium Pentazolate (CsN_5)

Theoretical prediction of CsN_5 and its enhanced stability at relatively low pressures stimulated the experimental effort to synthesize this material. It was predicted that compressing cesium azide in a nitrogen-rich environment might cause simple reaction $CsN_3 + N_2 \rightarrow CsN_5$ which is calculated to be energetically favorable above a pressure of just 9.1 GPa, see Fig. 2.12 [39]. Therefore, experiment was set up to compress CsN_3 in a diamond anvil cell (DAC) filled cryogenically with nitrogen N_2, the latter serving both as a pressure transmitting medium and reagent.

It was observed during the experiment that laser heating (LH) of CsN_3 and N_2 mixture at 65 GPa results in disappearance of the azide modes and appearance of the N_5 pentazolate modes in the Raman spectra, see Fig. 2.19a. The azide ν_1 mode is a shoulder of the diamond first-order peak at 1,490 cm^{-1} at 65 GPa, which disappears after LH. The N_5^- breathing mode is also very close to the diamond peak but is visible in the spectrum at 55 GPa at frequency 1,320 cm^{-1}. The other two modes associated with the azide anion N_3^- also disappear: sharp peaks ν_2 at 656 cm^{-1} and the broad peak in the 1,750–2,000 cm^{-1} frequency range see Fig. 2.19a. New Raman peaks also appear between 700 and 850 cm^{-1}, and at 1,170 and 1,320 cm^{-1} associated with N_5. Both the relative intensities and the frequencies of the experimentally observed Raman peaks are in agreement with those of the theoretically predicted Raman modes of P-1-CsN_5, see Fig. 2.19a. The observed Raman modes at 55 GPa were assigned as follows: 1,320 cm^{-1} – N_5^- ring breathing; 1,170 cm^{-1} – N_5^- antisymmetric N–N breathing and angle deformations, and 700–850 cm^{-1} – N_5^- ring bending and

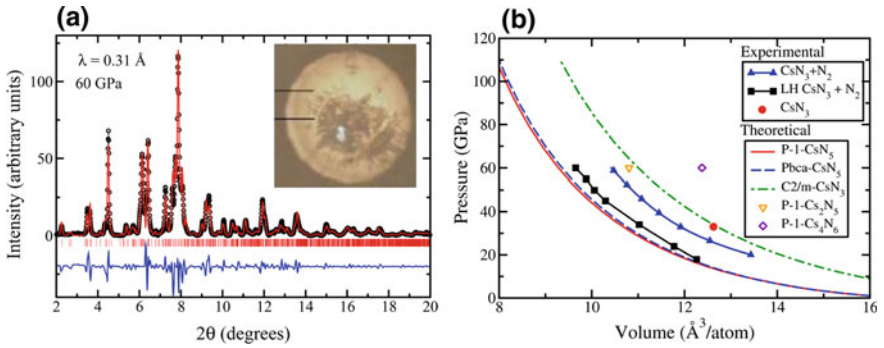


Fig. 2.20 **a** Measured X-Ray diffraction patterns (black circles) after laser heating against the Le Bail fit (red solid line) using predicted crystal structure of P-1-CsN_5. Vertical red ticks mark positions of Bragg peaks. The blue line is the difference between the measured and fitted intensities. The inset shows a microphotograph of the sample after heating in transmitted light, indicating the transparency of the synthesized phase. **b** Comparison of the measured and calculated pressure versus volume equation of state (EOS) for the synthesized compound: experimental data are plotted with solid symbols, whereas theoretical predictions—by open symbols and lines. Also shown are the experimental (P, V) points corresponding to CsN_3 at 33 GPa and the EOS of $CsN_3 + N_2$ mixture before laser heating (LH). Reprinted with permission from [39]. Copyright (2017) American Chemical Society

librational modes. Upon pressure release, the intensity of the Raman modes of CsN_5 dropped substantially, and only the most intense internal N_5 mode (1,236 cm^{-1}) and the intense broad feature at the low frequency interval (0–550 cm^{-1}) were observed below 20 GPa, see Fig. 2.19b. Upon further release of pressure below 18 GPa, all Raman modes other than the one corresponding to high frequency nitrogen vibrons disappear. The N_2 vibrons were always present because nitrogen was the pressure transmitting medium.

Additional confirmation of CsN_5 synthesis comes from XRD measurements during the compression experiments which reveal the appearance of new intense and narrow Bragg peaks with a simultaneous disappearance of the broad CsN_3 peaks. The Bragg peaks of the new phase are well indexed using the theoretically predicted P-1-CsN_5 structure. The resultant Le Bail-fitted and experimental XRD patterns at 60 GPa, shown in Fig. 2.20a, are well-matched. The lattice parameters obtained from the Le Bail refinement are close to those predicted by theory. The comparison between the experimental pressure versus volume equation of state (EOS) with the theoretical EOS of CsN_5 and various candidate structures obtained during the evolutionary structure search is shown in Fig. 2.20b. The experimental EOS matches the theoretical EOSs well for both P-1-CsN_5 and Pbca-CsN_5 phases in the full range of applied pressures. In addition, the experimental volume of $CsN_3 + N_2$ after laser heating is 10% less than the sum of volumes of CsN_3 and N_2 before laser heating at a given pressure. This provides a strong indication of major atomic rearrangements to form N_5 rings during the conversion of $CsN_3 + N_2$ to CsN_5.

2.6 Conclusions

This review of a novel family of poly-nitrogen compounds demonstrates the power of first-principles crystal structure search methods for the discovery of new materials. The novelty of this work is in framing the theoretical effort as an integral part of joint experimental and theoretical exploration of materials chemistry. Theory/simulation guidance not only involves the prediction of chemical composition and crystal structure of new compounds, but it also suggests specific precursors, and experimental conditions for experimental synthesis of high-N pentazolate EMs. The ultimate success was achieved in joint theoretical and experimental discovery of cesium pentazolate, which was synthesized by compressing and heating cesium azide CsN_3 and N_2 precursors in diamond anvil cell.

This success story paves the way for synthesis of novel high-N energetic materials with the promise of quenching their metastable phases at ambient conditions. In fact, the theoretical work on prediction of new pentazolate compounds inspired a renewed attempt to synthesize of cyclo-N_5^- compounds using traditional methods of organic synthesis [70–73]. A breakthrough has been achieved very recently in synthesis of N_5^- compounds by three independent groups both at ambient pressure and at high pressures [39, 70–73]. The ambient pressure synthesis involved the cleavage of the C–N bond in arylpentazoles [72] followed by stabilizing the N_5^- anion with various cations such as hydronium, ammonium, sodium, chloride, cobalt, iron, or manganese [70–72, 74]. The computational discovery of alkali and hydronitrogen pentazolates demonstrates the diversity of metastable high-N compounds, thus opening up new routes for their synthesis and recovery at ambient conditions.

References

1. Klapötke TM (2015) Chemistry of high-energy materials. De Gruyter, Berlin, München, Boston
2. Nobel A (1868) Improved explosive compound. US Patent 78,317
3. Christe K (2007) Propellants Explos Pyrotech 32(3):194
4. Mailhiot C, Yang L, McMahan A (1992) Phys Rev B 46(22):14419
5. Eremets MI, Gavriliuk AG, Trojan IA, Dzivenko DA, Boehler R (2004) Nat Mat 3(8):558
6. Eremets MI, Gavriliuk AG, Trojan IA (2007) Appl Phys Lett 90(17):171904
7. Gregoryanz E, Goncharov AF, Sanloup C, Somayazulu M, Mao Hk, Hemley RJ (2007) J Chem Phys 126(18):184505
8. Lipp M, Klepeis J, Baer B, Cynn H, Evans W, Iota V, Yoo CS (2007) Phys Rev B 76(1):014113
9. Maddox J (1988) Nature 335(6187):201
10. Oganov AR, Glass CW (2006) J Chem Phys 124(24):244704
11. Pickard CJ, Needs RJ (2011) J Phys Cond Matt 23(5):053201
12. Zhu Q, Oganov AR, Zhou XF (2014) Top Curr Chem 345:223
13. Wang Y, Lv J, Zhu L, Ma Y (2012) Comput Phys Commun 183(10):2063
14. Ma Y, Oganov A, Li Z, Xie Y, Kotakoski J (2009) Phys Rev Lett 102(6):100
15. Tomasino D, Kim M, Smith J, Yoo CS (2014) Phys Rev Lett 113:205502
16. Zhang W, Oganov AR, Goncharov AF, Zhu Q, Boulfelfel SE, Lyakhov AO, Stavrou E, Somayazulu M, Prakapenka VB, Konôpková Z (2013) Science 342(6165):1502

17. Duan D, Liu Y, Tian F, Li D, Huang X, Zhao Z, Yu H, Liu B, Tian W, Cui T (2014) Sci Rep 4:6968
18. Drozdov AP, Eremets MI, Troyan IA, Ksenofontov V, Shylin SI (2015) Nature 525(7567):73
19. Oganov AR, Chen J, Gatti C, Ma Y, Ma Y, Glass CW, Liu Z, Yu T, Kurakevych OO, Solozhenko VL (2009) Nature 457
20. Kraus W, Nolze G (1996) J Appl Crystallogr 29(3):301
21. Momma K, Izumi F (2011) J Appl Crystallogr 44(6):1272
22. Huisgen R, Ugi I (1957) Chem Ber 90(12):2914
23. Clusius K, Hurzeler H (1954) Helvetica Chimica Acta 37(3):798
24. Wallis JD, Dunitz JD (1983) J Chem Soc Chem Commun (16):910
25. Ferris KF, Bartlett RJ (1992) J Am Chem Soc 114(21):8302
26. Butler RN (1996) In: Katritzky AR, Rees CW, Scriven EF (eds) Comprehensive heterocyclic chemistry II. Elsevier, New York, pp 897–904
27. Ugi I (1984) In: Katritzky AR, Ree (eds) Comprehensive heterocyclic chemistry I. Elsevier, New York, pp 839–845
28. Burke LA, Butler RN, Stephens JC (2001) J Chem Soc Perkin Trans 2(9):1679
29. Janoschek R (1993) Angew Chem Int Ed 32(2):230
30. Nguyen MT, Ha TK (2001) Chem Phys Lett 335:311
31. Vij A, Pavlovich JG, Wilson WW, Vij V, Christe KO (2002) Angew Chem 141(16):3177
32. Östmark H, Wallin S, Brinck T, Carlqvist P, Claridge R, Hedlund E, Yudina L (2003) Chem Phys Lett 379(5–6):539
33. Evans BL, Yoffe AD, Gray P (1959) Chem Rev 59(4):515
34. Peiris SM, Russell TP (2003) J Phys Chem A 107:944
35. Eremets MI, Popov MY, Trojan IA, Denisov VN, Boehler R, Hemley RJ (2004) J Chem Phys 120(22):10618
36. Christe KO, Wilson WW, Sheehy JA, Boatz JA (1999) Angew Chem Int Ed 40(16):2947
37. Cacace F, de Petris G, Troiani A (2002) Science 295:480
38. Steele BA, Oleynik II (2016) Chem Phys Lett 643:21
39. Steele BA, Stavrou E, Crowhurst JC, Zaug JM, Prakapenka VB, Oleynik II (2017) Chem Mater 29(2):735
40. Steele BA, Oleynik II (2017) J Phys Chem A 121:1808
41. Steele BA, Landerville AC, Oleynik II (2014) J Phys Conf Ser 500(16):162005
42. Perdew J, Burke K, Ernzerhof M (1996) Phys Rev Lett 77(18):3865
43. Kresse G, Furthmiiller J (1996) Comput Mat Sci 6(1):15
44. Zhang M, Yin K, Zhang X, Wang H, Li Q, Wu Z (2013) Solid State Commun 161:13
45. Millar DIA, Barry C, Marshall WG, Pulham CR (2014) Z Kristallogr 229(3):259
46. Heyd J, Scuseria GE, Ernzerhof M (2003) J Chem Phys 118(18):8207
47. Schimka L, Harl J, Kresse G (2011) J Chem Phys 134(2):024116
48. Morino Y, Ijima T, Murata Y (1960) Bull Chem Soc Jpn 33(1):46
49. Carlotti M, Johns JWC, Trombetti A (1974) Can J Phys 52(4):340
50. Grimme S (2006) J Comput Chem 27(15):1787
51. Takemura K, Christensen N, Novikov D, Syassen K, Schwarz U, Hanfland M (2000) Phys Rev B 61(21):14399
52. Peng F, Yao Y, Liu H, Ma Y (2015) J Phys Chem Lett 6(12):2363
53. Peng F, Han Y, Liu H, Yao Y (2015) Sci Rep 5:16902
54. Shen Y, Oganov AR, Qian G, Zhang J, Dong H, Zhu Q, Zhou Z (2015) Sci Rep 5:14204
55. Wiberg E (2001) Inorganic chemistry. Academic Press
56. Stevenson DJ (1975) Nature 258(5532):222
57. Bernal MFM, Massey HSW (1954) Mon Not Roy Astron Soc 114(2):172
58. Goncharov AF, Holtgrewe N, Qian G, Hu C, Oganov AR, Somayazulu M, Stavrou E, Pickard CJ, Berlie A, Yen F, Mahmood M, Lobanov SS, Konôpková Z, Prakapenka VB (2015) J Chem Phys 142(21):214308
59. Wang H, Eremets MI, Troyan I, Liu H, Ma Y, Vereecken L (2015) Sci Rep 5:13239
60. Yin K, Wang Y, Liu H, Peng F, Zhang L (2015) J Mater Chem A 3(8):4188

61. Spaulding DK, Weck G, Loubeyre P, Datchi F, Dumas P, Hanfland M (2014) Nat Commun 5:5739
62. Crowhurst JC, Zaug JM, Radousky HB, Steele BA, Landerville AC, Oleynik II (2014) J Phys Chem A 118(38):8695
63. Hu A, Zhang F (2011) J Phys Cond Matt 23(2):022203
64. Qian GR, Niu H, Hu CH, Oganov AR, Zeng Q, Zhou HY (2016) Sci Rep 6:25947
65. Medvedev S, Eremets M, Evers J, Klapotke T, Palasyuk T, Trojan I (2011) Chem Phys 386(1–3):41
66. Evers J, Gobel M, Krumm B, Martin F, Medvedyev S, Oehlinger G, Steemann FX, Troyan I, Klap TM, Eremets MI (2011) J Am Chem Soc 133:12100
67. Pickard CJ, Needs RJ (2007) Nat Phys 3(7):473
68. Chellappa R, Dattelbaum D, Daemen L, Liu Z (2014) J Phys Conf Ser 500:052008
69. Yu H, Duan D, Tian F, Liu H, Li D, Huang X, Liu Y, Liu B, Cui T (2015) J Phys Chem C 119(45):25268
70. Bazanov B, Geiger U, Carmieli R, Grinstein D, Welner S, Haas Y (2016) Ang Chem 1–4
71. Xu Y, Wang Q, Shen C, Lin Q, Wang P, Lu M (2017) Nature 549(7670):78
72. Zhang C, Sun C, Hu B, Yu C, Lu M (2017) Science 355(6323)
73. Zhang C, Yang C, Hu B, Yu C, Zheng Z, Sun C (2017) Angew Chem Int Ed 56(16):4512
74. Zhang C, Sun C, Hu B, Lu M (2017) Science 355:374

Chapter 3
Accelerated Molecular Dynamics Simulations of Shock-Induced Chemistry: Application to Liquid Benzene

E. Martínez, E. M. Kober and M. J. Cawkwell

Abstract Shock-induced phenomena in materials occur on timescales that while short may still be beyond the reach of traditional molecular dynamics simulations. The shock-induced chemistry of liquid benzene provides an excellent example of the importance of timescale in shock experiments; reactions are seen at about 13.3 GPa on microsecond timescales in plate impact experiments but it appears inert at up to 20 GPa over 100s of picoseconds during laser-driven shock experiments. We have studied the shock-induced chemistry of liquid benzene using a semiempirical reactive interatomic potential at timescales beyond those routinely accessible to traditional molecular dynamics simulations. We have applied replica-based accelerated molecular dynamics to this system because the initial chemical reactions themselves can be viewed as rare, state-to-state transitions that take place under thermal activation. Replica-based accelerated molecular dynamics enables us to parallelize the simulations in time with no loss of accuracy, provided that transitions (reactions) can be detected reliably. We have simulated the shocked chemical dynamics of benzene on timescales up to 7.7 ns with high parallel efficiency. The simulations show the formation dimers through Diels–Alder condensation. The dimers subsequently condense into larger polymeric structures, in good accord with experiments and quantum chemical data.

E. Martínez · E. M. Kober · M. J. Cawkwell (✉)
Los Alamos National Laboratory, Los Alamos, NM 87545, USA
e-mail: cawkwell@lanl.gov

E. Martínez
e-mail: enriquem@lanl.gov

E. M. Kober
e-mail: emk@lanl.gov

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_3

3.1 Introduction

The shock compression of materials generates thermodynamic states of elevated temperature and stress. Out-of-equilibrium vibrational populations and stress states in the vicinity of the shock front are responsible for a rich set of material responses that include intense plastic flow and phase transformations. The combined high temperature, high-pressure states can also give rise to the making and breaking of covalent bonds in nonmetallic materials. The most notable example of such shock-driven chemical changes in materials is the cascade of reactions that occur on the path to detonation in energetic materials [11]. The topic of shock-driven chemistry is broader than just detonation. Reactions that lead to an increase in density are of course favored by the compression imparted by the shock, and shock heating can provide the necessary thermal activation for the reactions to occur on the required timescales.

Molecular dynamics (MD) simulations have been crucial for identifying and characterizing the atomic-scale phenomena that occur in materials under shock compression for close to 50 years [20–22, 47]. The MD simulation of shock-induced phenomena such as plasticity and phase transformations in crystalline materials has been successful mainly because they occur on timescales commensurate with those accessible to classical MD. The timescales accessible to classical MD simulations are controlled principally by the size of the time step used in the integration of nuclear degrees of freedom. In turn, the size of the time step is determined by period of the fastest vibrational mode in the material under study [2]. High- performance computing has made significant advances in increasing the size of simulations in terms of the number of atoms and/or the sophistication of the interatomic potentials. However, despite the advances in computational power over the 50-year history of the MD simulation of shock waves, the accessible timescales have not kept pace and have remained fairly static. MD simulations are inherently serial whereby each time step is calculated in sequence. Parallel processing has undeniably led to simulations on ever larger systems, but in traditional, classical MD, it does not affect the serial propagation of the nuclear coordinates in time.

Despite the undisputed success that MD simulations have had in furthering our understanding of shock-induced processes in materials there remains much work to be done. One of the primary limitations of MD remains the timescale problem, especially for processes characterized by rare transitions between different states. The application of MD to shock-induced reactions in prototypical or chemically realistic systems has been pursued heavily [4, 31, 46, 51]. However, the initial thermodynamic conditions for this type of simulation are typically chosen so that the reactions of interest will occur on the timescale of the simulation. This has limited applications to strongly overdriven regimes where reactions occur promptly behind the shock front, that is, within a few picoseconds.

Since the rates of chemical reactions depend on both temperature and pressure it is not guaranteed that the chemical reactions seen in the overdriven regime are the same as those seen under the milder conditions achieved in gas gun or laser-driven

shock compression experiments. In order to better resolve and understand the chemical reactions that occur in laboratory experiments and to quantify reaction rates for continuum-scale models, it is necessary to greatly extend the timescales achievable in reactive MD. Here we report the application of the parallel replica dynamics method [48], one of the accelerated molecular dynamics (AMD) formalisms developed originally by Voter, to shock-induced chemistry. Parallel replica dynamics brings parallel processing to bear on the propagation of nuclear degrees of freedom. It is capable of extending the timescales achievable in MD simulations of infrequent events by orders of magnitude while keeping the chemical fidelity of the underlying simulation and interatomic potential intact [38]. In this case, the infrequent events we seek to accelerate are changes in the intramolecular covalent bonding in a shock-compressed organic liquid.

Parallel replica dynamics has been applied most frequently to the simulation of diffusional processes in crystalline materials where the thermally activated migration of defects is the accelerated rare event. These systems are well suited to the parallel replica method since diffusional processes can be considered rare events with an intrinsic frequency much slower than atomic vibrations and the diffusional events themselves can be detected with little ambiguity. Nevertheless, several groups have applied the method to "soft" chemical systems including the conformational changes of molecules and the pyrolysis of organic compounds [24, 27]. These soft systems present additional challenges owing to the practical difficulties of defining states and detecting events.

In this work, the parallel replica dynamics method has been applied to the simulation of shock-induced reactions in liquid benzene on timescales beyond those routinely accessible to traditional MD. This has enabled us to examine chemistry under thermodynamic conditions that can be compared directly with experiments instead of the heavily overdriven regime. Liquid benzene was selected as a test case owing to its relative simplicity and the wealth of data available in the literature from the both μs gas gun and sub-ns laser-driven shock experiments. Interatomic interactions in the liquid benzene were described using an accurate and transferable density functional tight binding (DFTB) parameterization for hydrocarbons.

3.2 Shock Compression of Liquid Benzene

The response of benzene, C_6H_6, to static high-pressure and high-temperature conditions and shock compression is rich and has been studied intensively both experimentally and theoretically. Benzene melts at $-3\,^{\circ}C$ at ambient pressure and its equilibrium pressure–temperature phase diagram shows three molecular crystalline phases in addition to an amorphous carbon–hydrogen phase and a potential polymeric structure under high-pressure and high-temperature regimes [8]. Ciabini et al. showed that a transition from crystalline polymorphs to an amorphous carbon–hydrogen-based structure is triggered by the length of intermolecular C–C bonds decreasing to a critical value of 2.6 Å [8].

The single shock Hugoniot of liquid benzene was measured up to a pressure of 43.2 GPa by Dick [12] and Walsh and Rice [49] using explosively driven flyer plates. Cusps in the principal Hugoniot were reported at 13.3 and 19.4 GPa. The abrupt decrease in slope of the shock velocity, U_s, versus particle velocity, U_p, Hugoniot at 13.3 GPa is accompanied to a volume decrease of about 16%. In the original work, Dick suggested that the cusp in the Hugoniot at 13.3 GPa corresponds to the onset of an instantaneous (sub-μs) volume-decreasing transformation. The possibility of a freezing transition was discounted owing the (over)estimate of the shocked state at 13.3 GPa of $T = 2300$ K. As a result, Dick suggested that a polymerization reaction takes place at the cusp. A more recent estimate of the shock heating in liquid benzene at 13.3 GPa by Lacina and Gupta that accounts for the temperature dependence of the heat capacity gives instead a temperature of $T = 1485$ K [28]. More recently, Dattelbaum and co-workers performed light gas gun experiments with embedded electromagnetic gauges to resolve the temporal evolution of the wave structure in reacting, shocked liquid benzene [10].

Changes in optical absorption have been measured in liquid benzene singly shocked up to around first transition pressure. Yakusheva et al. [52] and Holmes et al. [23] attributed significant losses of light transmission to the formation of carbon particles upon the decomposition of the benzene molecules. However, Akin and Chau showed that liquid benzene becomes 100% absorbing at 523 nm within 300 ns of shock compression even when it is shocked to 12 GPa, i.e., below the first transformation at 13.3 GPa [1]. Hence, the changes in the optical properties of benzene upon shock compression may not depend on dramatic changes in interatomic bonding such as polymerization or decomposition but instead may rise from interactions between delocalized π orbitals on neighboring molecules. Moreover, laser-driven shock experiments by Dang et al. detect no changes in the absorption spectra of liquid benzene in spectral region from 440 to 780 nm over the first 350 ps during impacts up to 20 GPa [9]. While Dang et al. derive traditional wave speed data from their experiment, it is difficult to resolve from these results whether cusps exist in the principal Hugoniot.

Reactive molecular dynamics simulations of the shock compression of liquid benzene have yielded contradictory results. Molecular dynamics simulations by Bickham, Collins, and Kress using a semiempirical, orthogonal, non-charge-dependent tight binding model showed polymerization reactions within 7 ps of simulation time for shock pressures in excess of about 24.6 GPa and temperatures exceeding 2840 K [3]. Maillet and Pineau performed molecular dynamics simulations of benzene decomposition in the canonical (constant NVT) ensemble using an empirical ReaxFF force field at selected states on the experimental principal Hugoniot [30]. Simulations at a volumetric compression of $v/v_0 \approx 0.44$, where v and v_0 are the shocked and initial volumes, respectively, and a temperature of 2350 K, corresponding to a Rankine–Hugoniot shock pressure $p \approx 19$ GPa showed the decomposition of benzene molecules, the formation of polymeric structures, and a decrease in pressure. Maillet and Pineau did not present data on possible reactions in benzene under thermodynamic conditions corresponding to the first cusp in the principal Hugoniot at 13.3 GPa. First principles molecular dynamics simulations of benzene by

Wang and Zhang examined Hugoniot conditions in the range $4.3 \leq p \leq 67.2$ GPa and $598 \leq T \leq 4334$ K and revealed decomposition reactions starting at $p = 11$ GPa [50]. However, while first principles-based descriptions of interatomic bonding are generally more accurate than empirical or semiempirical models, the fidelity of the NVT molecular dynamics simulations presented by the authors is questionable owing to their use of a time step for the integration of the equations of motion for the nuclei of 2 fs. This is a factor of 8 and 20 larger than those used by Bickham et al. [3] and Maillet and Pineau [30], respectively. Indeed, a time step in the range 0.4–0.5 fs is typically around the upper limit to capture C–H stretch vibrational models at about 3000 cm^{-1} even under ambient conditions. Hence, the decomposition reactions reported in [50] at relatively low pressures might be viewed with caution.

3.3 Density Functional Tight Binding for Hydrocarbons

We have represented the interatomic bonding in benzene using the semiempirical density functional tight binding model developed by Krishnapriyan et al. [26]. DFTB is a fast, parameterized electronic structure model that captures the formation of covalent bonds and charge transfer between elements of different electronegativity [13, 14, 18, 19, 25, 44]. The potential energy in DFTB theory is,

$$u = 2\text{Tr}\left[(P - P_0)H^0\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\gamma_{ij}q_iq_j + E_{\text{pair}}, \tag{3.1}$$

where $\text{Tr}[X]$ denotes the trace of matrix X, P the self-consistent density matrix, P_0 the density matrix for noninteracting atoms, H^0 the charge-independent, two center Slater–Koster tight binding Hamiltonian [45], γ_{ij} a screened electrostatic potential, and q_i an atom-centered Mulliken charge. E_{pair} a sum of atom-centered pair potentials that provide mainly short-range repulsion, i labels atoms, and N is the total number of atoms. The density matrix, P, is computed from the electronic occupancy and the DFTB Hamiltonian

$$H = H^0 + H^1, \tag{3.2}$$

where H^1 depends on the electrostatic potential due to charge transfer,

$$H^1_{i\alpha,j\beta} = \frac{1}{2}S_{i\alpha,j\beta}(V_i + V_j). \tag{3.3}$$

Here, S is the overlap matrix, $\alpha = s$, p_x, etc., labels orbitals, and,

$$V_i = U_{(i)}q_i + \sum_{k\neq i=1}^{N}\gamma_{ik}q_k, \tag{3.4}$$

where $U_{(i)}$ is the Hubbard U for the element at site i. The electrostatic potentials, (3.4), are evaluated by Ewald summation. Since the DFTB Hamiltonian depends on the density matrix through the Mulliken charges,

$$q_i = \sum_{\alpha \in i} \left(S_{i\alpha, j\beta} P_{j\beta, i\alpha} + P_{i\alpha, j\beta} S_{j\beta, i\alpha} \right) - n_i^e, \tag{3.5}$$

where n_i^e is the number of valence electrons on the neutral atom; the density matrix must be solved self-consistently.

Our DFTB model for hydrocarbons uses a minimal basis of one s and three p orbitals on carbon and one s orbital on hydrogen. The dependences of Hamiltonian and overlap matrix elements on the bond angles are determined by the angular character of the valence orbitals and are prescribed. The radial dependences of the pair potentials that contribute to E_{pair} and the bond and overlap integrals are parameterized with the aim that the complete DFTB model reproduces experimental and/or ab initio calculated data as accurately as possible. The radial dependences of the bond and overlap matrix elements, $h(R)$ and $s(R)$, respectively, are represented by an exponential function,

$$h_{ll'\tau}(R) = h_{ll'\tau}(R_0)\zeta_{ll'\tau}(R, R_0) \tag{3.6}$$

$$s_{ll'\tau}(R) = s_{ll'\tau}(R_0)\zeta_{ll'\tau}(R, R_0) \tag{3.7}$$

$$\zeta(R, R_0) = \prod_{k=1}^{M} \exp(A_k(R - R_0)^k), \tag{3.8}$$

where l labels the orbital angular momentum ($l = s, p$) and τ the bond type ($\tau = \sigma, \pi$). R_0 is a reference bond distance, and $h(R_0)$, $s(R_0)$, $\{A\}$ are adjustable parameters for each integral [7, 26]. The product extends to $M = 2$ for the bond integrals and $M = 4$ for the overlap integrals. The radial dependence of the pair potentials, $\Phi(R)$, where

$$E_{\text{pair}} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \Phi(R_{ij}), \tag{3.9}$$

take a similar form to the bond and overlap integrals where

$$\Phi(R) = \Phi_0 \prod_{k=1}^{4} \exp(B_k R^k), \tag{3.10}$$

and Φ_0 and $\{B\}$ are adjustable parameters. The remaining terms requiring parameterization are the energies of the valence orbitals on noninteracting, neutral atoms, ε_l, and the Hubbard Us for each element.

The DFTB parameters for hydrocarbons were determined by a series of numerical optimizations that minimized the value of an objective function that measures the error in the predicted atomization energies and interatomic forces of small, distorted molecules with respect to identical density functional theory calculations at the B3LYP/cc-pVTZ level. The optimized parameter set is presented in [26]. Careful testing of the DFTB model showed that it is both accurate and transferable since the root mean square errors in the bond lengths, $\Delta d_{\mathrm{CC}} = 0.01$ and $\Delta d_{\mathrm{CH}} = 0.002$ Å, and atomization energies, $\Delta E = 0.016$ eV/atom, computed from the model for both the molecules used to train the model and a separate set of test molecules are almost indistinguishable.

3.4 Extended Lagrangian Born–Oppenheimer Molecular Dynamics

The MD trajectories that underpin the parallel replica AMD simulations were computed using the extended Lagrangian Born–Oppenheimer MD (XLBOMD) formalism of Niklasson and co-workers [34–37]. The XLBOMD method is an elegant and powerful solution to the systematic energy drift seen in microcanonical trajectories (constant number of particles, volume, and total energy, NVE) in traditional Born–Oppenheimer MD. It was demonstrated that broken time-reversal symmetry due to the approximate, incomplete self-consistent field (SCF) optimization of the electronic degrees of freedom in traditional Born–Oppenheimer MD, in combination with ad hoc extrapolation of the electronic degrees of freedom from one time step to the next, were responsible for systematic energy drifts during NVE dynamics [39]. The electronic degrees of freedom are propagated with a time-reversible integrator like the nuclear degrees of freedom in XLBOMD which restores time- reversal symmetry and provides precise, long-term energy conservation. The XLBOMD propagation of the electronic degrees of freedom has the additional and significant benefit of speeding up the calculation of the trajectories by about an order of magnitude. The integration of the electronic degrees of freedom generates initial guesses for the SCF procedure that are very close to the self-consistent ground state. Hence, in XLBOMD stable integration can be obtained with only one SCF cycle per MD time step rather than the 5–10 cycles that might be required in traditional BOMD.

The XLBOMD method is applied with DFTB-based MD to propagate the set of Mulliken charges and the chemical potential [6, 7, 53]. Our simulations of shock-compressed liquid benzene each used a time step for the integration of the equation of motion of $\delta t = 0.25$ fs. The making and breaking of covalent bonds can lead to numerical instabilities that manifest themselves as spurious forces and poor energy conservation. In order to avoid these events in our reactive simulations, we thermally smeared the occupation of states around the chemical potential when calculating the density matrix using the Fermi–Dirac distribution with an electronic temperature corresponding to 0.2 eV. Furthermore, to avoid SCF instabilities, we increased the

number of SCF cycles to 2 per time step with a linear mixing coefficient for the Mulliken charges of 0.1. Finally, the parallel replica dynamics simulations were performed in the canonical ensemble using a Langevin thermostat with 1 ps^{-1} damping.

3.5 Parallel Replica Dynamics

Parallel replica dynamics [38, 48] exploits the fact that for many physical processes the system trajectory executes transitions from state to state on timescales orders of magnitude larger than the atomic vibrations; i.e., the dynamics of the system are dominated by infrequent events. It can be shown that in these cases the process can be parallelized in time, in a manner complementary to the spatial parallelization of MD, with full accuracy [5, 48]. Information about exactly where the trajectory is inside each state is lost (although it can be recovered if desired), but the state-to-state trajectory can be made as accurate as in direct MD.

The parallel replica method is derived from the exponential distribution of escape times, $p_x(t)$, from a given state, x,

$$p_x(t) = k_x \exp(-k_x t), \tag{3.11}$$

where k_x is the escape rate over all possible escape paths, y, i.e., $k_x = \sum_y k_{xy}$. By running N_r statistically independent replicas of the trajectory, the probability that any of them escape in the interval $t \rightarrow t + \mathrm{d}t$ is,

$$p_x^\mathrm{r}(t)\mathrm{d}t = N_\mathrm{r} k_x \exp(-N_\mathrm{r} k_x t)\mathrm{d}t. \tag{3.12}$$

Defining the time accumulated over the N_r replicas as $t_\mathrm{r} = N_\mathrm{r} t$, we obtain,

$$p_x(t_\mathrm{r})\mathrm{d}t_\mathrm{r} = k_x \exp(-k_x t_\mathrm{r})\mathrm{d}t_\mathrm{r}. \tag{3.13}$$

Hence, the probability of any of the replicas escaping state x in time t is the same as that for the serial, nonparallel system in time $t_\mathrm{r} = N_\mathrm{r} t$ [48].

In practice, the boost factor obtained from parallel replica dynamics can be considerably less than N_r owing to overheads that are inherent to the method. Since the replicas must be statistically independent, each must be carefully dephased before the parallel simulation time is accumulated. In this work, the replicas were dephased by coupling to a stochastic, Langevin thermostat [32] with different random number sequences. Each replica was dephased for 5 ps before accumulating simulation time. We assume that the system is in a quasi-stationary distribution after each dephasing time. Another overhead in the parallel replica dynamics method arises from monitoring those replicas where a transition has occurred for the correlation time. Parallel time is not accumulated over this interval. Since a transition to a new state might be transient, either because it quickly returns to its original state or undergoes a series of further escapes, by ensuring that the system remains in a new state for at least the

correlation time any subsequent states are uncorrelated and we return to a system of rare-event dynamics. Upon establishing that a replica has settled into a new state, its configuration generates a new set of replicas, each of which is dephased for the process to start again. A system that undergoes frequent transitions owing to high rates and/or large numbers of parallel replicas will have low parallel efficiency since the overheads arising from dephasing and the correlation time are linked with the escapes.

State-to-state transitions in parallel replica dynamics have been detected traditionally by periodically 'quenching' each trajectory through a molecular statics relaxation after which the atomic positions are compared to those at the start of the trajectory [48]. While this approach is reliable for crystalline materials, its efficacy for soft materials and liquids is very limited. We instead interrogate the network of inter- and intramolecular covalent bonding in liquid benzene directly during the trajectories. The nearest neighbor bonds were reevaluated every 1 ps and without quenching, and changes in the nearest neighbor bonding were used to identify chemical events. The scission of a C–C bond is defined to have occurred when the following condition is satisfied:

$$\frac{d_{\text{new}} - d_{\text{old}}}{d_{\text{old}}} > f_{\text{cut}} = 0.3, \tag{3.14}$$

where d_{new} and d_{old} are the current and reference bond lengths in the hot state, respectively. A new C–C is defined to have formed if carbon atoms come within 2 Å for at least 1 ps. The correlation time was taken as 5 ps. We ignore changes in the C–H and H–H bonding in the simulations since the dynamics of these transitions are fast compared to the rate of transitions in the C–C bonding. The parallel efficiency of our simulations would be greatly reduced if events involving H were tracked explicitly.

3.6 Calculation of the Unreacted Hugoniot of Liquid Benzene

Since DFTB is computationally expensive compared with empirical reactive or unreactive force fields, it is still not feasible to employ it in direct simulations of the shock compression of materials. Instead, the thermodynamic conditions that correspond to states on the principal, or single shock Hugoniot have been computed through the use of a simple Hugoniostat [17, 29, 41, 42]. Those material states were used as the initial conditions for a series of AMD trajectories. Hence, we prepare systems on the principal unreacted Hugoniot and through MD or AMD simulations allow them to evolve in time.

Hugoniostats enable the computation of shocked states by self-consistently updating the temperature and pressure or density until the Hugoniot equation is satisfied,

$$E - E_0 = \frac{1}{2}(p + p_0)(v_0 - v), \tag{3.15}$$

where E is the total energy, p the pressure, and v the system volume [17]. The subscript 0 refers to those quantities of the system in its initial, pre-shocked state. The initial conditions for our simulations comprised a cubic, periodic simulation cell containing 20 benzene molecules at temperature of 300 K at its experimental density, $\rho_0 = 0.876$ g/cm^3. For a series of target pressures, $2 < p < 30$ GPa, the specific volume v was updated incrementally using a simple volume rescaling barostat and the temperature by updating the target temperature of a Langevin thermostat to satisfy (3.15). The initial conditions for the subsequent AMD simulations were determined by time averaging the temperature and specific volume once the system reached a fluctuating steady state. The calculated temperatures along the unreacted Hugoniot as a function of specific volume and pressure are presented in Figs. 3.1 and 3.2, respectively. The temperatures predicted by the DFTB model using (3.15) underestimate those calculated by Lacina and Gupta. For instance, at 13.3 GPa our calculations give a temperature of 1161 K on the unreacted Hugoniot whereas the equations of state of Root [43] and Lacina and Gupta [28] give 1315 and 1485 K, respectively. The shock temperatures predicted by Nellis et al. are marginally lower than those of Root and hence are in better accord with our predictions [28, 33]. Nevertheless, our predicted shock temperatures appear to be in much better agreement with estimates derived from equations of state than those obtained from earlier MD simulations [3].

The unreacted Hugoniot of liquid benzene in the particle velocity-shock velocity and pressure-specific volume planes calculated with the DFTB model for hydrocarbons are presented with experimental data from Dick in Figs. 3.3 and 3.4, respectively [12]. Figure 3.4 illustrates clearly that the DFTB model overestimates the compressibility of liquid benzene at low shock pressures. The discrepancies in the experimental and predicted Hugoniots at higher compressions are due to the volume-decreasing reactions that are not captured in the nonreactive MD simulations. The gradient of the calculated Hugoniot in the $U_s - U_p$ plane is in excellent accord with experiment. The predicted and experimental unreacted Hugoniots, i.e., including experimental data up to 13 GPa, are offset by about 0.5 km/s. This result implies that the DFTB model underestimates the sound velocity of liquid benzene under ambient conditions by about the same amount.

3.7 Results

Five states on the unreacted Hugoniot of liquid benzene from 20 to 28 GPa were used as the initial conditions for the parallel replica dynamics. The corresponding specific volumes and temperatures are summarized in Table 3.1. Each simulation used 60 replicas where each DFTB MD simulation ran on 16 cores of one compute node for a total of 960 CPU cores per simulation.

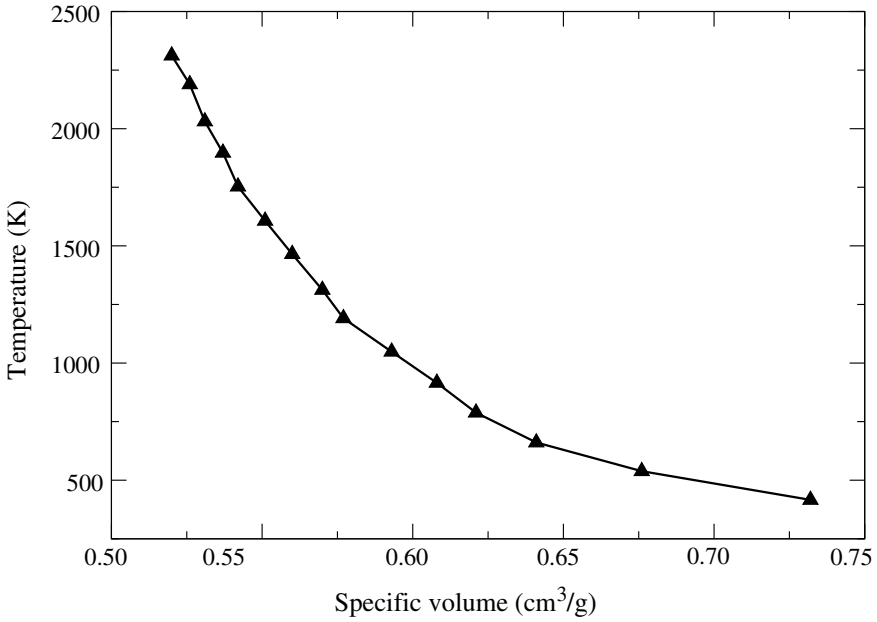


Fig. 3.1 Locus of the temperatures and specific volumes on the unreacted principal Hugoniot for liquid benzene computed from DFTB-based MD

Table 3.1 Initial conditions on the liquid benzene principal Hugoniot for the AMD simulations

Pressure (GPa)	Temperature (K)	Specific volume (cm^3/g)
20	1613	0.551
22	1751	0.542
24	1897	0.537
26	2031	0.526
28	2207	0.520

Dimer- or polymerization reactions between benzene molecules were seen for all of the initial conditions. The reactions were more frequent at higher shock pressures leading to lower parallel efficiencies owing to the overheads arising from the dephasing and correlation times. The accumulated simulations times and efficiencies are presented in Table 3.2. Our simulations at 20, 22, and 24 GPa each exceeded 1 ns of simulation time. To put these values in better context, on identical hardware a serial DFTB MD simulation of the same systems yields only 40 ps per 24 hours.

The dimerization reactions seen at 20, 22, and 24 GPa are consistent with Diels–Alder reactions between benzene molecules. The formation of benzene dimers by a Diels–Alder reaction is endothermic but it has a low barrier and is promoted by pressure. Detailed calculations of the reaction pathways were presented recently by Quenneville and Germann [40]. Similar Diels–Alder reactions have been inferred in other shocked aromatic compounds [15, 16].

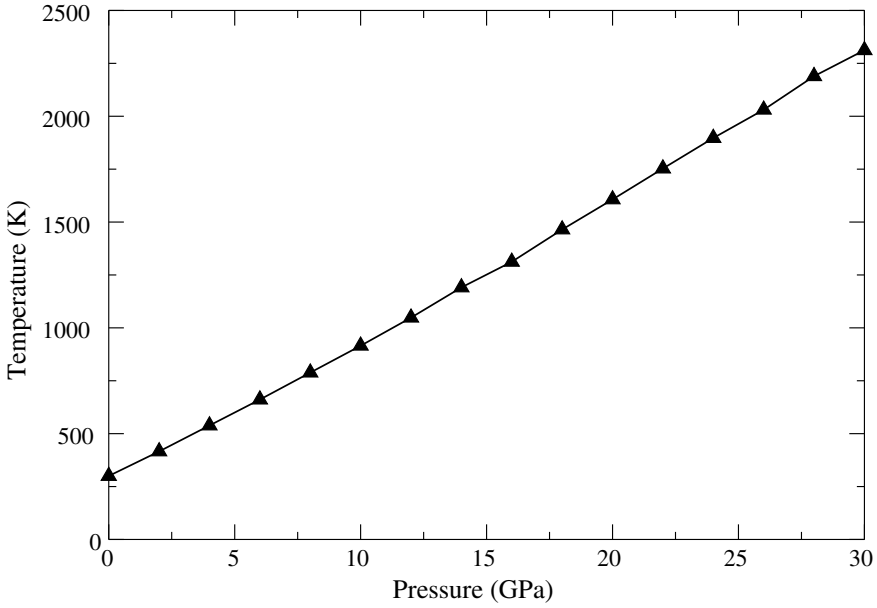


Fig. 3.2 Calculated temperatures on the unreacted Hugoniot of liquid benzene as a function of shock pressure

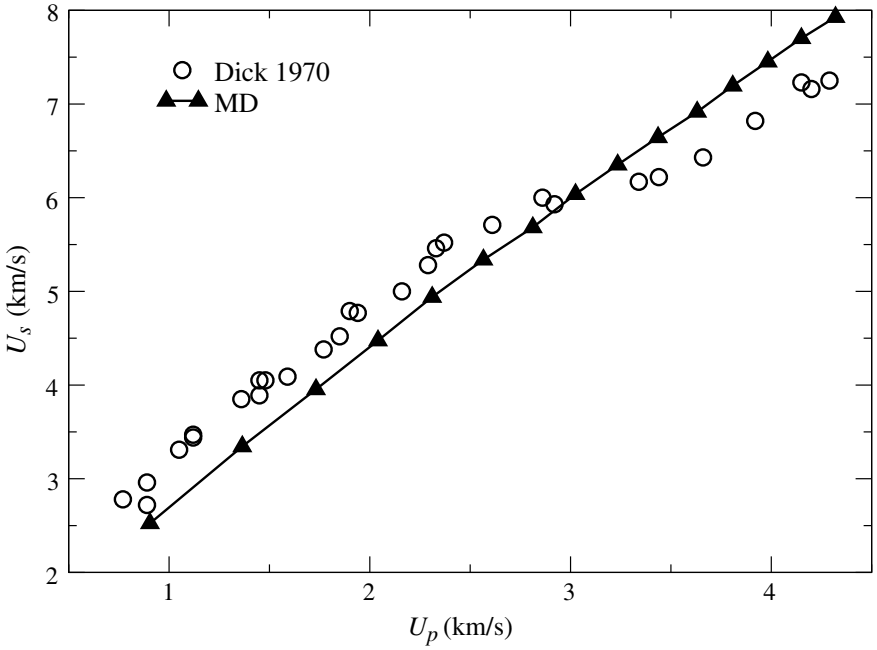


Fig. 3.3 Shock velocity versus particle velocity Hugoniot of liquid benzene from DFTB-based MD and experiment [12]. Experimental data for $U_p < 2.8$ km/s correspond to unreacted material

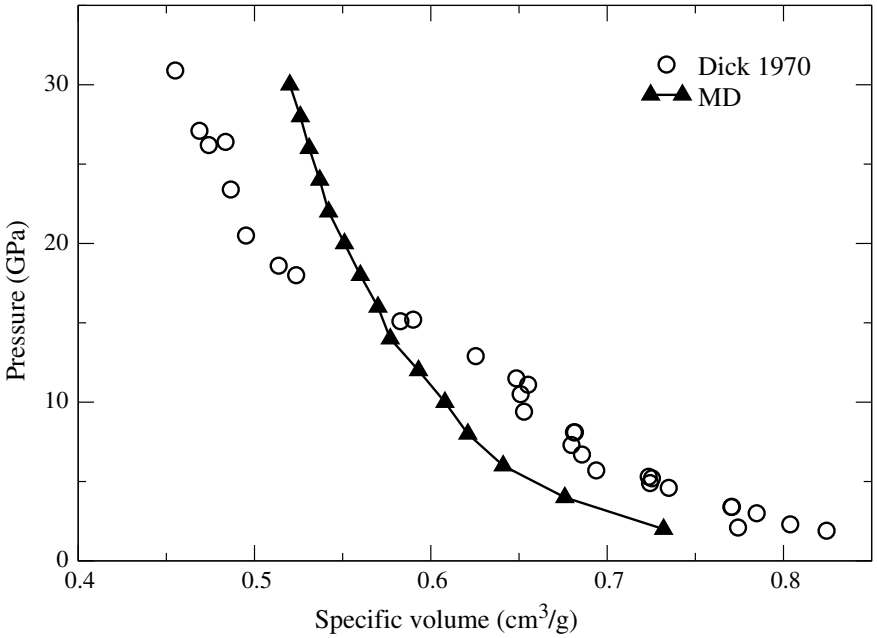


Fig. 3.4 Locus of pressure and specific volume on the principal Hugoniot for liquid benzene from DFTB-based MD simulations and experiment [12]. Experimental data at specific volumes greater than 0.6 cm^3/g correspond to unreacted material

The reactions were analyzed by employing simple cutoff distances to determine whether pairs of atoms are bonded together or not. The values used in this study were: C–C 2.00 Å, C–H 1.45 Å, and H–H 1.00 Å. The atoms were also classified according to their coordination geometries as a means of tracking the reaction types. For example, benzene contains C atoms that are bonded to one H and two other C atoms: C[HCC]. For the Diels–Alder condensation reaction, four C atoms (two each from the two benzenes) will be transformed into C atoms that are bonded to one H and three other C atoms: C[HCCC]. The condensations observed here are all

Table 3.2 Summary of the accumulated time, parallel efficiency, and reactions seen in the AMD simulations as a function of the initial pressure

Pressure (GPa)	Reaction	Stability	Total time (ns)	Parallel efficiency (%)
20	Diels–Alder dimer	Metastable	7.70	75.9
22	Diels–Alder dimer	Metastable	4.08	57.0
24	Diels–Alder dimer	Metastable	2.08	32.4
26	Polymerization	Stable	1.27	25.4
28	Polymerization	Stable	0.27	15.2

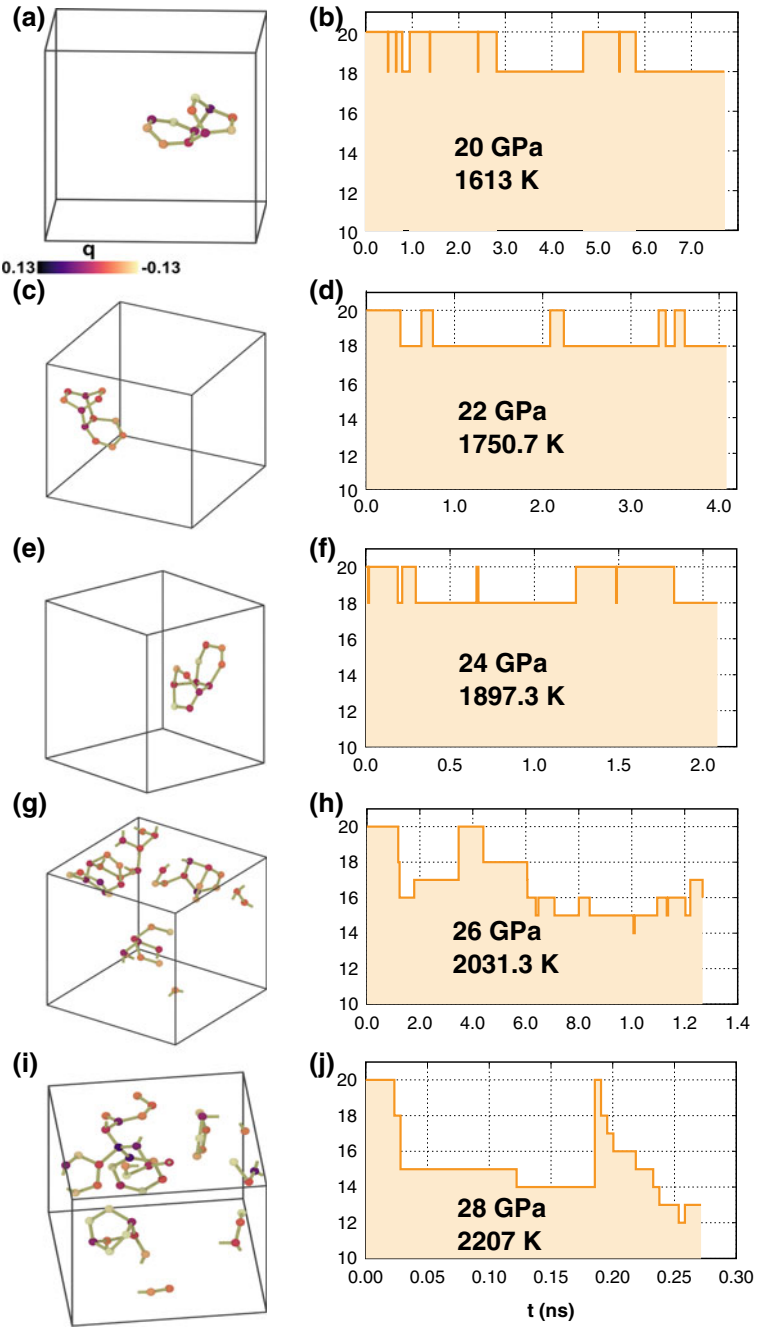


Fig. 3.5 Snapshots of polymerization reactions in benzene and the time history number of non-reacted benzene molecules as a function of temperature and pressure. **a** and **b** 20 GPa and 1613 K, **c** and **d** 22 GPa and 1751 K, **e** and **f** 24 GPa and 1897 K, **g** and **h** 26 GPa and 2031 K, and **i** & **j** 28 GPa and 2207 K. Note that the timescale is different in each plot

concerted (four C atoms change their coordination environments simultaneously), which is expected for the ground state reactions. Quenneville and Germann had also considered stepwise reactions from the ground and excited states, but these were not observed here [40].

The covalent dimers seen at 20, 22, and 24 GPa are metastable and decompose to benzene during the AMD simulations. The formation and decomposition of the Diels–Alder dimers are illustrated in Fig. 3.5 where we plot the time history of the number of benzene molecules in the simulation cell. The number of benzene molecules alternates between 20 (no dimers) and 18 (one dimer) in the AMD simulations performed at 20, 22, and 24 GPa. The dimer- and polymerization reactions observed in our AMD simulations at 26 and 28 GPa are more extensive. Figure 3.5 shows clearly that more benzene molecules take part in the reactions. At later times in those two cases, the transfer of H from the tertiary carbons over to a secondary C is also observed, where this is presumably accompanied by shifting the isolated double bond over to the tertiary C. Such reactions should likely render the condensation reactions irreversible, forming stable oligomers that could undergo further rearrangements or decomposition.

3.8 Discussion and Summary

The application of parallel replica dynamics with an accurate, reactive interatomic potential based explicitly on electronic structure has enabled us to identify and characterize shock-induced chemical reactions at timescales far in excess of those that can be obtained practically in traditional MD. We obtained nearly 80% parallel efficiency for our simulations at 20 GPa and 1613 K. Owing to the high efficiency, further increases in the number of replicas are both feasible and desirable so that chemistry can be studied at longer timescales and under weaker shocks.

The reactions observed in shock-compressed benzene are consistent with endothermic, pressure-assisted Diels–Alder reactions. The product dimers appear to be metastable at the lower shock pressures studied in this work. However, by extending the duration of the simulations further it may be possible to observe additional polymerization steps which stabilize the products further. Crucially, we observe reactions that lead to an increase in density, which is fully consistent with the results of shock compression experiments.

The μs-duration explosively driven flyer plate experiments of Dick identified a cusp in the principal Hugoniot at about 13 GPa [12]. The more recent laser-driven shock experiment of Dang and co-workers detected no reactions up to about 20 GPa [9]. Their observations are fully consistent with our AMD simulations since at 20 GPa we find only a handful of reactions over a timescale of nearly 8 ns. Hence, the reaction rates at 20 GPa are such that very little chemistry can occur over the 300 ps duration of the experiments. Similarly, the much longer duration of gas gun and explosively driven flyer experiments find evidence of significant densification reactions down to 13 GPa.

In summary, we have applied an accelerated MD method to a relatively simple system, albeit one with rich chemistry, that has enabled the analysis of chemical reactions under input conditions that can be compared directly with experiments. These techniques provide a tractable route for the study of shock-induced chemistry in organic materials under less overdriven conditions than have been used to date. The dividing surfaces that characterize events in hot organic materials should be defined with care and reliable event detection is an important consideration. Nevertheless, there is no practical reason why AMD methods cannot be applied to study the relatively slow, endothermic reactions that occur on route to the rapid, exothermic chemistry that occurs during detonation.

Acknowledgements This work was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory. We thank Josh Coe, Dana Dattelbaum, Shawn McGrane, Anders Niklasson, Danny Perez, Kyle Ramos, and Art Voter for many illuminating discussions over the course of this work.

References

1. Akin MC, Chau R (2013) Observations of shock induced chemsity of cyclohexane. J Chem Phys 139:024502
2. Allen MP, Tildesley DJ (1989) Computer simulation of liquids. Oxford University Press, Oxford
3. Bickham SR, Kress JD, Collins LA (2000) Molecular dynamics simulations of shocked benzene. J Chem Phys 112:9695
4. Brenner DW, Robertson DH, Elert ML, White CT (1993) Detonations at nanometer resolution using molecular-dynamics. Phys Rev Lett 70:2174
5. Bris CL, Leliévre T, Luskin M, Perez D (2012) A mathematical formalization of the parallel replica dynamics. Monte Carlo Methods Appl 18:119
6. Cawkwell MJ, Niklasson AMN (2012) Energy conserving, linear scaling born-oppenheimer molecular dynamics. J Chem Phys 137:134105
7. Cawkwell MJ, Niklasson AMN, Dattelbaum DM (2015) Extended Lagrangian Born-Oppenheimer molecular dynamics simulations of the shock-induced chemistry of phenylacetylene. J Chem Phys 142:064512
8. Ciabini L, Santoro M, Gorelli FA, Bini R, Schettino V, Raugei S (2007) Triggering dynamics of the high-pressure benzene amorphization. Nat Mater 6:39
9. Dang NC, Bolme CA, Moore DS, McGrane SD (2012) Shock induced chemistry in liquids studied with ultrafast dynamic ellipsometry and visible transient absorption spectroscopy. J Phys Chem A 116:10301
10. Dattelbaum DM, Sheffield SA, Coe JD (2017) Shock-driven chemistry and reactive wave dynamics in liquid benzene. AIP Conf Proc 1793:040020
11. Davis WC (1981) High explosives: the interaction of chemistry and mechanics. Los Alamos Sci 2:48
12. Dick RD (1970) Shock wave compression of benzene, carbon disulfide, carbon tetrachloride, and liquid nitrogen. J Chem Phys 52:6021
13. Elstner M, Poresag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys Rev B 58:7260
14. Elstner M, Seifert G (2014) Density functional tight binding. Philos Trans R Soc A 372:20120483

15. Engelke R, Blais NC (1994) Chemical dimerization of crystalline anthracene produced by transient high pressure. J Chem Phys 101:10961
16. Engelke R, Blais NC, Sheffield SA, Sander RK (2001) Production of a chemically-bound dimer of 2,4,6-tnt by transient high pressure. J Phys Chem A 105:6955–6964
17. Erpenbeck JJ (1992) Molecular dynamics of detonation. I. Equation of state and Hugoniot curve for a simple reactive fluid. Phys Rev A 46:6404
18. Frauenheim T, Seifert G, Elstner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S, Scholz R (2000) A self-consistent charge density functional based tight binding method for predictive materials simulations in physics, chemistry, and biology. Phys Stat Sol B 217:41
19. Gaus M, Cui Q, Elstner M (2014) Density functional tight binding: application to organic and biological molecules. WIREs Comput Mol Sci 4:49–61
20. Holian BL (1995) Atomistic computer-simulations of shock-waves. Shock Waves 5:149
21. Holian BL, Lomdahl PS (1998) Plasticity induced by shock waves in nonequilibrium molecular dynamics simulations. Science 280:2085
22. Holian BL, Straub GK (1979) Molecular-dynamics of shock-waves in 3-dimensional solids—transition from nonsteady to steady waves in perfect crystals and implications for the Rankine–Hugoniot conditions. Phys Rev Lett 43:1598–1600
23. Holmes NC, Otani G, McCandless P, Rice SF (1990) Absorption spectroscopy of shocked benzene. In: Proceedings of the ninth international symposium on detonation, p 190
24. Joshi KL, Raman S, van Duin ACT (2013) Connectivity-based parallel replica dynamics for chemically reactive systems: from femtoseconds to microseconds. J Phys Chem Lett 4:3792–3797
25. Koskinen P, Mäkinen V (2009) Density-functional tight-binding for beginners. Comput Mater Sci 47:237
26. Krishnapriyan A, Yang P, Niklasson AMN, Cawkwell MJ (2017) Numerical optimization of density functional tight binding models: application to molecules containing carbon, hydrogen, nitrogen, and oxygen. J Chem Theory Comput, p 6191
27. Kum O, Dickson BR, Stuart SJ, Uberuaga BP, Voter AF (2004) Parallel replica dynamics with a heterogeneous distribution of barriers: application to n-hexadecane pyrolysis. J Chem Phys 121:9808
28. Lacina D, Gupta YM (2013) Temperature measurements and an improved equation of state for shocked liquid benzene. J Chem Phys 138:174506
29. Maillet JB, Mareschal M, Soulard L, Ravelo R, Lomdahl PS, Germann TC, Holian BL (2001) Uniaxial Hugoniostat: a method for atomistic simulations of shocked materials. Phys Rev E 63:016121
30. Maillet JB, Pineau N (2008) Thermodynamic properties of benzene under shock conditions. J Chem Phys 128:224502
31. Manaa MR, Fried LE, Reed EJ (2003) Explosive chemistry: simulating the chemistry of energetic materials at extreme conditions. J Comput Aided Mater Des 10:75–97
32. Martínez E, Cawkwell MJ, Voter AF, Niklasson AMN (2015) Thermostating extended Lagrangian Born-Oppenheimer molecular dynamics. J Chem Phys 142:154120
33. Nellis WJ, Ree FH, Trainor RJ, Mitchell AC, Boslough MB (1984) Equation of state and optical luminosity of benzene, polybutene, and polyethylene shocked to 210 GPa (2.1 mbar). J Chem Phys 80:2789
34. Niklasson AMN (2008) Extended Born-Oppenheimer molecular dynamics. Phys Rev Lett 100:123004
35. Niklasson AMN, Steneteg P, Odell A, Bock N, Challacombe M, Tymczak CJ, Holmström E, Zheng G, Weber V (2009) Time-reversible Born-Oppenheimer molecularu dynamics with dissipation. J Chem Phys 130:214109
36. Niklasson AMN, Tymczak CJ, Challacombe M (2006) Time-reversible Born-Oppenheimer molecular dynamics. Phys Rev Lett 97:123001
37. Niklasson AMN, Tymczak CJ, Challacombe M (2007) Time-reversible ab initio molecular dynamics. J Chem Phys 126:144103

38. Perez D, Uberuaga BP, Voter AF (2015) The parallel replica dynamics method—coming of age. Comput Mater Sci 100:90–103
39. Pulay P, Fogarasi G (2004) Fock matrix dynamics. Chem Phys Lett 386:272
40. Quenneville J, Germann TC (2009) A quantum chemistry study of Diels-Alder dimerizations in benzene and anthracene. J Chem Phys 131:024313
41. Ravelo R, Holian BL, Germann TC, Lomdahl PS (2004) Constant stress Hugoniostat. Phys Rev B 70:014103
42. Reed EJ, Fried LE, Joannopoulos JD (2003) A method for tractable dynamical studies of single and double shock compression. Phys Rev Lett 90:235503
43. Root S, Gupta YM (2009) Chemical changes in liquid benzene multiply shock compressed to 25 GPa. J Phys Chem A 113:1268
44. Seifert G, Joswig JO (2012) Density-functional tight binding-an approximate density-functional theory method. WIREs Comput Mol Sci 2:456–465
45. Slater JC, Koster GF (1954) Simplified LCAO method for the periodic potential problem. Phys Rev 94:1498
46. Strachan A, van Duin ACT, Chakraborty D, Dasgupta S, Goddard WA (2003) Shock waves in high-energy materials: the initial chemical events in nitramine RDX. Phys Rev Lett 91:098301
47. Tsai DH, Beckett CW (1966) Shock wave propagation in cubic lattices. J Geophys Res 71:2601
48. Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57:13985–13988
49. Walsh JM, Rice MH (1957) Dynamic compression of liquids from measurements on strong shock waves. J Chem Phys 26:815
50. Wang C, Zhang P (2010) The equation of state and nonmetal-metal transition in benzene under shock compression. J Appl Phys 107:083502
51. Wood MA, Cherukara MJ, Kober EM, Strachan A (2015) Ultrafast chemistry under nonequilibrium conditions and the shock to deflagration transition at the nanoscale. J Phys Chem C 119:22008–22015
52. Yakusheva OB, Yakushev VV, Dremin AN (1971) The opacity mechanism of shock-compressed organic liquids. High Temp. High Press 3:261
53. Zheng G, Niklasson AMN, Karplus M (2011) Lagrangian formulation with dissipation of Born-Oppenheimer molecular dynamics using the density-functional tight-binding method. J Chem Phys 135:044122

Chapter 4
Force Matching Approaches to Extend Density Functional Theory to Large Time and Length Scales

Rebecca K. Lindsey, Matthew P. Kroonblawd, Laurence E. Fried and Nir Goldman

Abstract We present methods for the creation of semi-empirical quantum approaches and reactive force fields through force matching to quantum simulation data for materials under reactive conditions. Our methodologies overcome the extreme computational cost of standard Kohn–Sham Density Functional Theory (DFT) by mapping DFT computed simulation data onto functional forms with linear dependence on their parameters. This allows for quick parameterization of our models by avoiding the nonlinear fitting bottlenecks associated with most molecular dynamics model development. We illustrate our approach with two different systems: (i) determination of density functional tight binding models for aqueous glycine dimerization, and (ii) determination of the Chebyshev Interactional Model for Efficient Simulation (ChIMES) reactive force field for metallic liquid carbon. In each case, we observe that our approach is easy to parametrize and yields a model that is orders of magnitude faster than DFT while largely retaining its accuracy. Overall, our methods have potential use for studying complex long time and length scale chemical reactivity at extreme conditions, where there is a significant need for computationally efficient atomistic simulations methods to aid in the interpretation and design of experiments.

R. K. Lindsey · M. P. Kroonblawd · L. E. Fried · N. Goldman (✉)
Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
e-mail: ngoldman@llnl.gov

R. K. Lindsey
e-mail: lindsey11@llnl.gov

M. P. Kroonblawd
e-mail: kroonblawd1@llnl.gov

L. E. Fried
e-mail: lfried@llnl.gov

N. Goldman
Department of Chemical Engineering, University of California, Davis, CA 95616, USA

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_4

4.1 Introduction

Elucidating the chemistry of reactive materials under extreme conditions in the laboratory can require the investigation of a great number of permutations of different starting materials, thermodynamic conditions, and catalysts. The number of possible combinations can frequently be too numerous and costly to address with experimental trial and error alone. Experiments frequently rely on theoretical studies to elucidate measured data, including kinetic and spectroscopic properties. Nonetheless, in many cases, insufficient data exists for the equation of state and chemical reactivity of these materials under the extreme pressures attained during experiments [1]. For example, studies on carbon-rich materials under pressures up to 30 GPa (1 GPa = 10 kbar) suggest that slow chemical kinetics can extend beyond the timescales of nanosecond laser-driven compression experiments, even at temperatures of thousands of Kelvin [2]. Reported experimental temperatures can contain large uncertainties, making it difficult to adequately constrain the equation of state on the basis of experiment alone [3, 4]. Furthermore, low-resolution mesoscale simulations frequently require chemical kinetic input parameters that are challenging to obtain experimentally (e.g., due to difficulties obtaining atomic-scale resolution in experiments on bulk material synthesis) but can be readily computed via higher-resolution particle-based simulations [5–7].

Computer simulations such as molecular dynamics (MD) hold promise as an independent route to determining the equation of state and chemical reactivity during materials synthesis. Such studies can provide simple chemical pictures of ionized intermediates and reaction mechanisms, and can help identify atomic-scale properties that determine observed macroscopic kinetics (e.g., descriptors). These types of results can make experiments more tractable by aiding in their interpretation, and helping to narrow the number of different materials and reactive conditions to investigate.

Accurate modeling of the breaking and forming of bonds in condensed phases frequently requires the determination of many nuclear-body effects [8, 9], which are included in quantum theories such as Density Functional Theory (DFT). DFT remains one of the most popular and widely used modeling methods in condensed matter physics, computational chemistry, and materials science for predicting material properties and chemical reactivity. It has been shown to accurately reproduce the phase boundaries and thermal decomposition of many materials [10–13], particularly at extreme thermodynamic conditions such as planetary interiors [10, 11, 13], where long-range effects such as dispersion are less important. DFT-MD simulations, though, require immense computational effort per simulation time step and consequently are usually limited to picosecond time scales and nanometer system sizes. In contrast, many chemical events can occur over nanosecond timescales or longer [2, 11], and experiments can probe micron length scales or beyond [14–16].

Difficulty thus generally arises in determining models for chemical bonding that are both accurate and computationally efficient. Empirical models (e.g., [17]), where atomic forces are computed from parameterized potential energy surfaces, are

generally fit to properties of specific chemical reactants over a narrow range of thermodynamic conditions. These approaches are usually highly computationally efficient, but tend to contain numerous nonlinear parameters that can be time-consuming to optimize, and can also suffer from poor transferability. In contrast, semi-empirical quantum theories such as density functional tight binding with self-consistent charges (DFTB) retain some level of computational efficiency and tend to have a higher degree of transferability to different starting materials and reaction conditions. However, standard forms of these models are parameterized to gas phase quantum chemical calculations, and consequently can result in overbinding of reactants, e.g., reaction enthalpies that are up 20 kcal/mol too high for simple hydrogenation reactions [18] and somewhat poor representation of condensed phase reactivity [19]. Thus, there exists a great need to explore novel methods for atomistic simulations of reactive materials that are computational efficient while retaining the accuracy of higher order quantum methods.

In this chapter, we demonstrate how force matching can be used to determine both classical reactive MD and DFTB models through use of functional forms that are strictly linear in their parameters. Force matching maximizes the data set that can be obtained from DFT by fitting parameters of a potential energy function to each individual atomic force in an MD trajectory [20, 21], yielding a large quantity of data points for determining atomistic models (i.e., $3N$ data points per configuration, where N is the number of atoms in the system). It thus has potential to systematically determine MD models that yield highly accurate predictions of chemical reactivity for a given material or mixture and set of thermodynamic conditions. The use of linearly parametrized models allows for use of linear least-squares fitting, whereby optimum parameter values are rapidly determined in a single step. This removes the need for direct gradient minimization or iterative techniques (e.g., Levenberg–Marquardt) which can become trapped in a local minimum, or computationally intensive global energy minimum searches (e.g., simulated annealing). In this work, we discuss two application areas for this method: parameterization of a semi-empirical quantum model for glycine dimerization under ambient aqueous conditions, and generation of a reactive force field for molten carbon. Our intent is to provide a general overview of how these classes of MD models can be rapidly determined for reactive materials under a broad range of conditions.

4.2 Force Matching Overview

The force matching method (FM) was first developed by [20] for generation of an aluminum force field based on a repository of atomic configurations and corresponding ab initio forces [20]. This framework usually requires the generation of an MD training data set for given material at a specific set of state points, often through the use of DFT simulations. A force field functional form is then selected, e.g., $\phi(x; \{p\})$, which depends on interatomic distances x, and is subject to a set of parameters $\{p\}$.

Model parameters are then obtained by minimizing the following objective function:

$$\mathrm{RMSE} = \sqrt{\frac{1}{3MN}\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{\alpha=1}^{3}\left(F_{i_{m_\alpha},\mathrm{DFT}} - F_{i_{m_\alpha},\phi\{p\}}\right)^2}, \tag{4.1}$$

where RMSE is the root-mean-squared error, the subscript i_{m_α} indicates the αth Cartesian component of the force on atom i in MD configuration m, and F_{DFT} are the forces to which model $\phi(x;\{p\})$ is to be fit.

As mentioned previously, generation of fitted parameters can be time consuming for complex target models due to the need for slow to converge nonlinear approaches such as Levenberg–Marquardt [22, 23] or global search methods such as simulated annealing [24]. If the target model is *linear* in fitted parameters, however, the FM optimization problem reduces to the following matrix equation: $\mathbf{Mp} = \mathbf{F}_{\mathrm{train}}$. The elements of $\mathbf{M}$ are given as $M_{rc} = \partial F_{r,\phi\{p\}}/\partial p_c$, where r represents the combined index over Cartesian components of the force on each atom in the training trajectory (i.e., matrix rows), and c the combined index over coefficients in the model (i.e., matrix columns), while p and F_{train} are the vector of target model parameters and vector of training forces, respectively. Standard methods such as Singular Value Decomposition [25], ridge [26, 27], or LASSO regression [28] can then be leveraged to solve for a unique set of model parameters in a matter of minutes. Well-developed procedures exist for determining the robustness of regression problems, such as cross-validation [29].

Though force matching was originally developed for generation of atomic potentials based on DFT data, we note that the approach is flexible and can be expanded to include other features including the system stress tensor (to tune model pressure) and energy. Furthermore, the force matching approach can be used to generate coarse-grained models from all-atom force fields (e.g., [30]) by mapping forces arising from several atoms to specific coarse grain interaction sites.

4.3 DFTB: Rapidly Tunable Models for High Throughput Quantum Molecular Dynamics

The density functional tight binding method (DFTB) [18, 31, 32] is a semi-empirical quantum simulation approach that yields a high degree of computational efficiency while potentially retaining the accuracy of the computationally intensive Kohn–Sham density functional theory (DFT). The formalism for DFTB with self-consistent charges (SCC) has been discussed in detail elsewhere [32–37, 40]. 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

$$E_{\mathrm{DFTB}} = E_{\mathrm{BS}} + E_{\mathrm{Coul}} + E_{\mathrm{Rep}}. \tag{4.2}$$

The first term, E_{BS}, corresponds to the band structure energy computed via 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, where both the electronic wave functions and electron density are subjected to separate confining or compression potentials. The compression potentials force the wavefunction/electron density to zero at relatively large distances from the nuclei, which has been shown to improve transferability of the Slater–Koster tabulations [31]. The precalculation of matrix elements in DFTB yields several orders of magnitude increase in computational efficiency over most DFT implementations, which project the electronic density onto a fine grid, and then use Fourier transforms to compute the Hartree potential and other terms. However, standard forms of DFTB still require calculation of eigenstates, which scales as $\mathcal{O}(N^3)$, where N is equal to the number of orbitals in the system. This generally remains the bottleneck of DFTB calculations, limiting the approach to system sizes of several hundred to thousands, e.g., far below the capabilities of classical MD codes.

The second term in (4.2), E_{Coul}, corresponds to a charge fluctuation term which is computed self-consistently. The third term, E_{Rep} (the repulsive energy), corresponds to ion–ion repulsions, as well as Hartree and exchange-correlation double counting terms. In practice, E_{Rep} is expressed as a short-ranged empirical function whose parameters are fit to reproduce DFT or experimental data, and can be either pairwise [13, 40] or contain multicenter interactions [37, 38]. This balance in DFTB of approximate quantum mechanics and short-ranged empiricism allows for the simultaneous interrogation of electronic states and ion dynamics over temporal scales that can approach those of many high temperature–pressure experiments (e.g., [39]). We note that the short-ranged nature of E_{Rep} generally allows for strong transferability of DFTB models (applicability to different materials and thermodynamic conditions), since the bulk of the interaction energy is computed though the quantum-mechanical E_{BS} and E_{Coul} terms.

Determination of an optimum E_{Rep} requires training data generation, which is a relatively straightforward endeavor for systems exhibiting rapid chemistry. Extracting uniformly spaced frames from an unbiased DFT-MD trajectory of a highly reactive system is likely to yield a training set with configurations corresponding to reactants, products, and various intermediates and transition states. In fact, this approach was shown to improve DFTB predictions for small species production during the rapid combustion of phenolic polymer resin [40]. However, in cases where chemistry is slow to either initiate or equilibrate, short unbiased DFT-MD simulations might not sample relevant intermediate or product configurations. Accelerated MD techniques such as umbrella sampling [41] or metadynamics [42] provide alternative approaches to training trajectory generation. In each of these methods, a biasing potential is applied to help escape local free energy minima and thus enhance sampling of configurational space. The biasing potential is typically applied to one or more "collective variables" that mathematically describe characteristic changes in system structure as the target process progresses. Optimal choice of collective variables is highly dependent on the situation and numerous kinds have been used

including bond separation vectors and/or coordination numbers [43], crystalline order parameters [44], and path variables [45] to name only a few.

Given these issues, force matching provides a systematic way to generate DFTB E_{Rep} parameters that can recover DFT-level accuracy for systems and states not considered in standard DFTB parameterizations. For this approach, the objective function is given as follows:

$$\text{RMSE} = \sqrt{\frac{1}{3MN}\sum_{m=1}^{N}\sum_{i=1}^{M}\sum_{\alpha=1}^{3}\left[F_{i_{m_\alpha},\text{DFT}} - \left(F_{i_{m_\alpha},E_{\text{Rep}}\{p\}} + F_{i_{m_\alpha},\text{BS}} + F_{i_{m_\alpha},\text{Coul}}\right)\right]^2}. \tag{4.3}$$

In this case, $F_{i_{m_\alpha},E_{\text{Rep}}\{p\}}$ (the forces arising from the repulsive energy term E_{Rep} subject to a set of model parameters $\{p\}$) are tuned to the values from the DFT training set with the DFTB forces from E_{BS} and E_{Coul} precomputed and subtracted out. We model the repulsive potential as a pairwise interaction between atom i of element e_i and atom j of element e_j that takes the form of the power series:

$$E_{\text{Rep}}^{e_i e_j}(r_{ij}) = \begin{cases} \sum_{n=2}^{9} c_n^{e_i e_j}\left(r_{\text{c}}^{e_i e_j} - r_{ij}\right)^n & r_{ij} \le r_{\text{c}}^{e_i e_j}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

Here, r_{ij} is the interatomic separation distance, $r_{\text{c}}^{e_i e_j}$ is a radial cutoff distance, and $c_n^{e_i e_j}$ are parameters to be determined. Summing over all-atom pairs in the system gives the total repulsive energy E_{Rep}. Because (4.4) is linear in $c_n^{e_i e_j}$, the parameters can be determined from a simultaneous least-squares fit over all $e_i e_j$ pair types. Furthermore, selected $e_i e_j$ pair types can be excluded from the fit and substituted with some other interaction potential if desired, provided that corresponding contributions to the atomic forces are subtracted from the F_{DFT} values in (4.3).

4.3.1 Application to Glycine Dimerization: Computational Details

As an example of creating a DFTB for model of long timescale chemistry, we investigate the condensation reaction 2Gly → Gly-Gly + H_2O in which two glycine molecules react to form diglycine and water. We begin by generating DFT training data via umbrella sampling along path collective variables, which are well-suited because the reactant and product species are known. Two path collective variables,

$$s(t) = \frac{\sum_k k \exp(-\lambda D[\mathbf{R}(t), \mathbf{R}_k])}{\sum_k \exp(-\lambda D[\mathbf{R}(t), \mathbf{R}_k])}, \tag{4.5}$$

and

$$z(t) = -\frac{1}{\lambda} \ln \left\{ \sum_k \exp(-\lambda D[\mathbf{R}(t), \mathbf{R}_k]) \right\}, \tag{4.6}$$

were used that respectively measure the progress along and distance from the path defined by reference configurations indexed by k at time t [45]. The variable $D[\mathbf{R}(t), \mathbf{R}_k]$ is the "distance" from the kth reference configuration and λ is a constant to be determined later. The path was defined with respect to two reference configurations, namely $k = 1$ for the two glycine reactant molecules and $k = 2$ for the single diglycine product molecule. We used the "distance" metric developed by Pietrucci and Saitta [46]:

$$D[\mathbf{R}(t), \mathbf{R}_k] = \sum_{i,e_i} [C_{i,e_i}(t) - C^k_{i,e_i}]^2, \tag{4.7}$$

where $C_{i,e_i}(t)$ is the number of atoms of type e_i coordinated to atom i at time t and C^k_{i,e_i} is the corresponding value in reference frame k. We tracked the coordination of the four C and two N atoms that form the diglycine backbone to all C, N, O, and H atoms in the system. The specific equations and parameters for computing $C_{i,e_i}(t)$ were taken from [46] and the reference C^k_{i,e_i} were obtained from short DFTB-MD simulations of neutral glycine and diglycine in explicit H_2O solvent. Following [46], we chose λ so that $\lambda D[\mathbf{R}_k, \mathbf{R}_{k+1}] \approx 2.3$, which for the present case was $\lambda = 0.70$. One particular advantage of using path collective variables to generate a DFT training set is that the resolution for sampling the reaction path can be controlled, which facilitates extracting important intermediate configurations and forces. The free energy surface $F(s, z)$ was computed as a function of s and z using umbrella sampling [41] and the weighted histogram analysis method (WHAM) [47, 48].

Simulations of glycine condensation were performed at $T = 300$ K for two glycine molecules solvated by 55 H_2O molecules, yielding a total of 185 atoms and a density of $\rho = 1$ g cm^{-3}. Deuterium masses were substituted for hydrogen masses to allow for a longer time step. No empirical dispersion corrections were applied in our DFT-MD or DFTB-MD simulations of glycine. Umbrella sampling simulations were performed using the PLUMED 1.3 plugin [49, 50].

Car–Parrinello [51] DFT-MD simulations were performed using the Quantum Espresso *ab initio* simulation package [52, 53]. We closely followed the simulation protocol of [43] for simulations of glycine condensation. Three differences between [43] and the present work are the DFT software package, the choice of collective variables, and that the former used metadynamics [42] rather than umbrella sampling. Trajectories were integrated with a 0.145 fs (6.0 au) time step and electron mass of 700 au with a Nosé-Hoover thermostat [54, 55] applied to both the ionic and electronic degrees of freedom. Reducing the time step by a factor of four was necessary to conserve the Hamiltonian in some umbrella sampling simulations due to the strong bias potentials that were required to ensure adequate sampling of s and z space. The electronic structure was evaluated using the Perdew–Burke-Ernzerhof (PBE) functional [56] and ultrasoft pseudopotentials [57] with a planewave cutoff of 25 Ry.

Born–Oppenheimer DFTB-MD simulations were performed using LAMMPS [58, 59] with forces and stresses evaluated by the DFTB+ code [33, 60]. Isothermal-isochoric (NVT) trajectories were integrated with a 0.20 fs time step and a Nosé-Hoover-style thermostat [54, 55]. The electronic structure was computed with a minimal basis and with Fermi–Dirac thermal smearing with the electronic temperature set equal to the instantaneous ionic kinetic temperature at each time step. Electronic degrees of freedom were propagated using an extended Lagrangian approach [61–64] with four self-consistent charge cycles per time step. DFTB simulations were performed using the mio-1-1 parameter set (available at http://www.dftb.org), which is a typical off-the-shelf parameter set for CNHO systems. Selected mio-1-1 $E_{\text{Rep}}^{\alpha\beta}(r_{ij})$ potentials were replaced with fitted ones for the force-matched DFTB simulations, discussed below.

4.3.2 E_{Rep} Parameterization and Benchmarking

We begin with generation of the DFT training trajectory via umbrella sampling along "path progress" collective variable s, at fixed "path deviation" collective variable, z. Collective variables are constrained by the harmonic bias potential $V_{\text{Bias}}(s, z) = \frac{1}{2}K_s(s - s_0)^2 + \frac{1}{2}K_z(z - z_0)^2$, where z_0 is fixed at -0.10, corresponding to the location of the two reference configurations and their associated local free energy minima, in z-space. A total of 37 umbrellas are utilized, evenly spaced between $1.050 \leq s \leq 1.950$, by increments of 0.025, where initial configurations were taken from 5 ps long unbiased simulations at $s = 1.950$. We note that the s coordinates for glycine and diglycine minima are approximately 1.1 and 1.9, respectively. Each of the 37 umbrella simulations were run for 2 ps, where the first picosecond was reserved for construction of the training trajectory, and the second half, for free energy calculation. The training trajectory was extracted by taking configurations at $t = 250$ and 500 fs from each umbrella, resulting in a total of 74 frames, and DFT forces were evaluated at the electronic ground state.

Because the ratio of C, N, O, and H atoms was highly uneven (4:2:59:120 for C:N:O:H, respectively), target training forces and design matrix elements $M_{r_i,c}$ for atom i were weighted by $1/N_{e_i}$ where the subscripts r_i and c of the design matrix represents the combined index over Cartesian values of the force on each atom and the combined index over model coefficients, respectively, and N_{e_i} is the number of atoms of type e_i. This reduces the contribution from interactions between the solvent H_2O molecules, which significantly outnumber the C–N, C–O, and N–H interactions that are more important in the condensation reaction. Repulsive potentials of the form in (4.4) were obtained for C–C, C–N, C–O, C–H, N–H, and O–H interactions using the weighted DFT training set with radial cutoffs $r_c^{e_i,e_j}$ set to mio-1-1 defaults. The N–N, N–O, O–O, and H–H repulsive potentials were taken from the mio-1-1 parameter set because those interactions were not well-sampled within their corresponding cutoffs.

In order to evaluate performance of the resulting model, simulations analogous to those described above (except using a 10 ps unbiased simulation and 20 ps umbrella

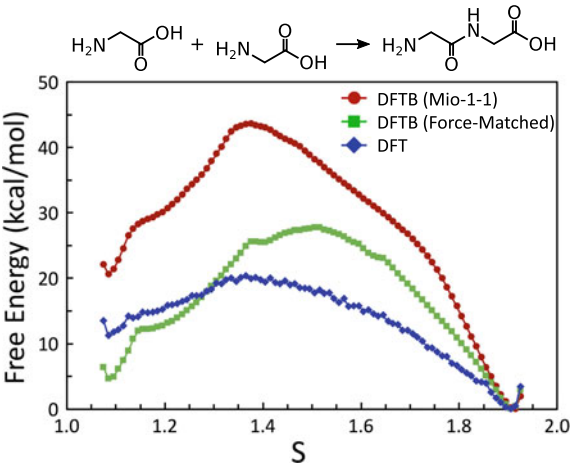


Fig. 4.1 Glycine condensation free energy surfaces $F(s, z = -0.10)$ predicted by DFTB with mio-1-1 and force-matched parameters, and DFT

simulations) were run for the resulting force-matched DFTB model, as well as for the standard mio-1-1 DFTB parameter set. A slice of the free energy surface $F(s, z = -0.10)$ was computed from histograms of s and z extracted from the last 10 ps of the force-matched and mio-1-1 DFTB and the last 1 ps of DFT using the two-dimensional WHAM equations (Fig. 4.1).

Beginning with results obtained from mio-1-1 DFTB calculations, we find the lowest free energy configuration in all three independent simulations is diglycine, so we chose those minima as the reference energy. We performed three independent sets of umbrella sampling simulations with this model as a means to assess the uncertainty in our calculations. The relative free energy differences from two unreacted, solvated glycine molecules were 12, 22, and 27 kcal mol^{-1} across the three independent simulations, with maximum barrier heights of 41, 44, and 48 kcal mol^{-1}, respectively. The WHAM solution computed using histograms from all three independent runs combined yields a free energy of reaction of $\Delta F = 21$ kcal mol^{-1} and an approximate free energy barrier of $\Delta F^{\text{barrier}} = 44$ kcal mol^{-1} (for the back reaction). Determination of a free energy of activation ($\Delta F^{\ddagger}$) and transition state would require a committor-style trajectory analysis [65], which is beyond the scope of the current study. We compute the uncertainties in ΔF and $\Delta F^{\text{barrier}}$ to be 8 and 3 kcal mol^{-1}, based on the standard deviation of these calculations.

Figure 4.1 also provides the free energy slices for the presently force-matched DFTB model as well as DFT. We note that DFT results should be viewed as an estimate, as the free energy calculations are not necessarily converged after one picosecond of averaging. We observe that the force-matched DFTB model predicts a significantly lower barrier height ($\Delta F^{\text{barrier}} = 28$ kcal mol^{-1}) than mio-1-1, which is much closer to the DFT prediction ($\Delta F^{\text{barrier}} = 20$ kcal mol^{-1}). It is unclear from this comparison whether ΔF is improved through force matching. The mio-1-1 and force-matched DFTB results respectively predict $\Delta F = 21$ and 5 kcal mol^{-1}, both of which are close to the DFT prediction $\Delta F = 11$ kcal mol^{-1} within uncertainty. The topology of $F(s, z = -0.10)$ is qualitatively different between the DFTB and DFT

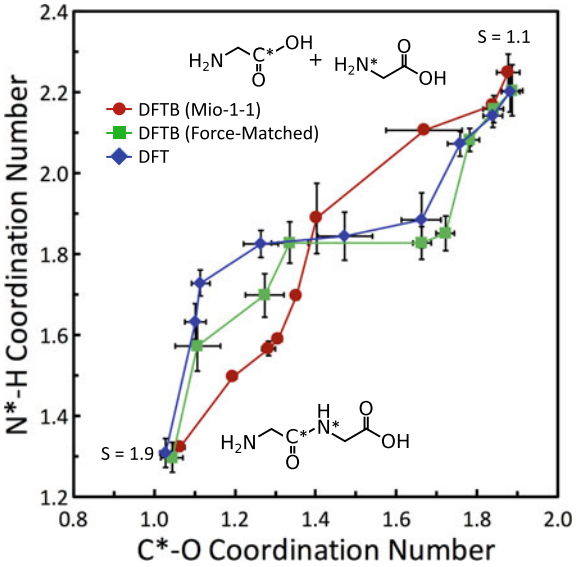


Fig. 4.2 Reaction paths expressed in terms of time-averaged coordination numbers for the dipeptide bond atoms C* and N* extracted from the umbrella sampling simulations used to compute the free energy surfaces in Fig. 4.1. Data points correspond to independent umbrella trajectories with $1.1 \leq s_0 \leq 1.9$ spaced in 0.1 increments. Uncertainties for the DFT and force-matched DFTB paths are standard deviations of the respective coordination numbers and uncertainties for the mio-1-1 DFTB path are standard errors computed over the paths predicted by three independent realizations

results, but the two DFTB predictions exhibit significant similarities aside from the numerical values for ΔF and $\Delta F^{\text{barrier}}$. One possibility is that topological features for $F(s, z)$ are largely determined through water solvation interactions, for which some second-order DFTB parameter sets are known to have significant errors [66].

All three models yield free energy surfaces that are in qualitative agreement with the DFT results for glycine condensation by Schreiner et al. [43] who predicted $\Delta F = 20$ kcal mol^{-1} and $\Delta F^{\text{barrier}} = 33$ kcal mol^{-1} at the same density and temperature. Quantitative differences between the Schreiner et al. and present (unconverged) DFT results are possibly due to the previously discussed differences in the free energy calculation protocols. Experiments [67, 68] and equation of state models [68, 69] suggest that glycine is slightly lower in free energy than diglycine by approximately 3 kcal mol^{-1}, which is in stark contrast to the mio-1-1 results and the results of Schreiner et al. The nominal equivalence in free energy for the reactants and products predicted by force-matched DFTB is generally more consistent with experiments.

To further assess the accuracy of our force-matched DFTB models for glycine condensation chemistry, we recast the reaction path in terms of the coordination between the atoms that form the dipeptide bond, namely the C* to O atoms and N* to H atoms with the same coordination number functions used to compute the matrix elements of (4.7). Reaction paths expressed in terms of these coordination numbers are shown in Fig. 4.2. We observe stark differences between mio-1-1 and DFT, with

agreement only at the end points. The mio-1-1 calculations yield a concerted mechanism with near simultaneous coordination of the O–H and H groups to C* and N*, resulting in a monotonically increasing linear form in its mapping. In contrast, DFT predicts a sequential process wherein H first coordinates to N* in diglycine followed by O–H coordinating to C*, resulting in the plateau shown in the middle of the plot. Force-matched DFTB also predicts a sequential process that is nearly identical to the DFT one within uncertainty. Similarities between the DFT and force-matched DFTB paths indicate that those two models likely have similar local minima in the regions of **R**-space that map to these particular path-dependent coordinates. The computed uncertainties in the three reaction paths presented here indicate that care should be exercised when interpreting chemistry from subtleties in path coordinates, as they are not necessarily able to differentiate between chemically distinct configurations. Nevertheless, force-matched DFTB provides a straightforward means of accelerating predictive simulations of long timescale chemistry when combined with enhanced sampling methods such as Umbrella Sampling, as shown here.

4.3.3 The Total Free Energy surface for Glycine Dimerization: Insights from Force-Matched DFTB

The computational efficiency of DFTB allows for a more thorough exploration and convergence of long timescale chemical reactivity than is currently practical with DFT. The z-dependent free energy surface was determined for both mio-1-1 and force-matched DFTB over an interval that captures neutral reactant and product species. A total of 259 independent 20 ps umbrella sampling simulations were performed to generate each surface, which corresponds to a combined simulation time of 5.18 ns each. As before, the s_0 were set in 0.025 increments and the z_0 were evenly spaced in 0.05 increments in the interval $-0.15 \leq z_0 \leq 0.15$. Note that zwitterionic configurations map to $z > 0.2$ with the chosen parameters and reference configurations for (4.7). Free energy surfaces were computed following the same protocol as previously discussed and are shown in Fig. 4.3.

The total changes in free energy predicted by mio-1-1 and the force-matched model (vs. the free energy slices, discussed previously) are respectively $\Delta F = 20.7$ and 4.5 kcal mol^{-1}, which are both within uncertainty of the results for the surface section $F(s, z = -0.10)$ shown in Fig. 4.1. It is apparent that the $F(s, z = -0.10)$ section does not capture the minimum free energy barrier for either model. The mio-1-1 results predict the reaction barrier to be 40.5 kcal mol^{-1}, compared to the force-matched model minimum barrier is 16.7 kcal mol^{-1}. Comparing the two surfaces reveals different z-coordinates for the minimum barrier, with the location for mio-1-1 being at $z = 0.02$ and for the force-matched model at $z = 0.17$. The force-matched model minimum barrier is located much closer to the zwitterionic region of the surface ($z > 0.20$). We note that the variation in $F(s, z)$ with z is generally greater for the force-matched model than for mio-1-1. Significant z-dependency highlights

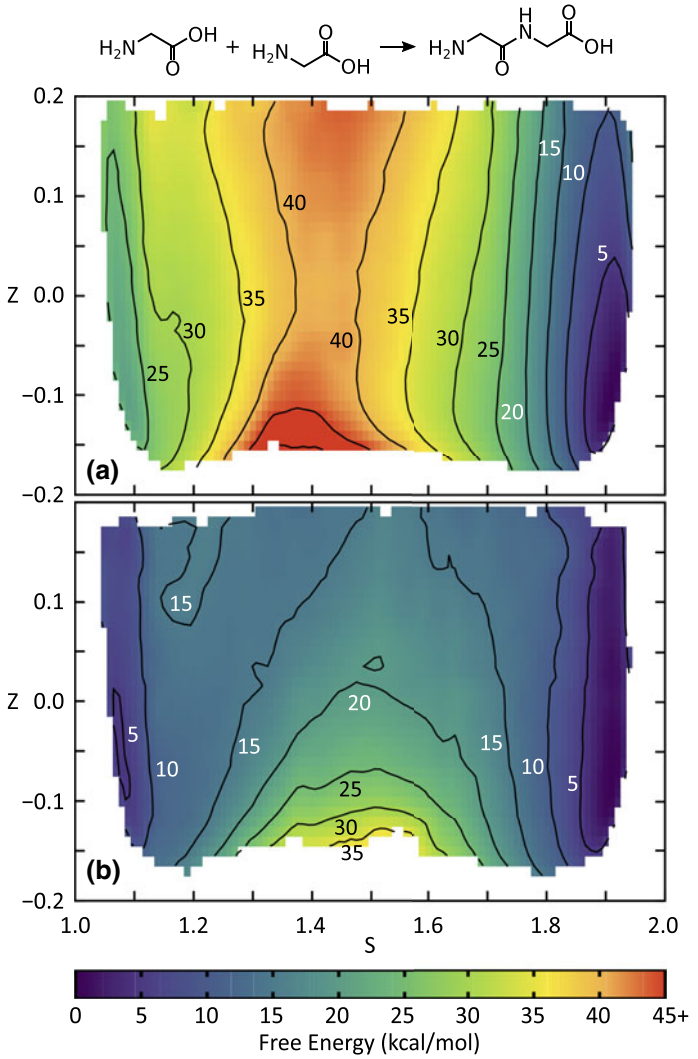


Fig. 4.3 Glycine condensation free energy surface $F(s, z)$ predicted by DFTB with **a** mio-1-1 and **b** force-matched repulsive potentials

the importance of extensively exploring the collective variable space when estimating the free energy and underscores the necessity for extending simulation times beyond those that can be practically reached with DFT.

4.3.4 DFTB Force Matching: Outlook

Accessing long timescales and performing large-scale ensemble studies beyond those that can be practically performed with DFT is critical for accurate determination of many condensed phase chemical reactions [40, 70]. We have shown that force-matched DFTB models hold promise as one avenue to extend quantum simulations to more realistic timescales while retaining much of the accuracy of DFT [40]. In particular, combining force-matched DFTB with accelerated sampling allows for the rapid generation of converged and accurate free energy surfaces for chemical reactions (discussed in further detail in recently published work [19]). This significantly expands the range of states that can be feasibly considered within the scope of a single study and allows for accurate quantum simulations that approach experimental time scales for reactivity. Our approach can help in the design and interpretation of laboratory synthesis studies, where there is a strong need for knowledge of the kinetics of chemical bond formation and breaking over extended periods of time.

4.4 ChIMES: Fast, Scalable Machine-Learned Reactive Force Fields

In this section, we discuss the problem of how to leverage force matching for ground-up construction of classical reactive MD models based on the Chebyshev Interaction Model for Efficient Simulation (ChIMES). ChIMES models are computationally efficient, fast to parameterize due to linear parameterization, and have the potential to maintain the accuracy of first principles methods across the large range of temperatures and pressures of interest for materials under extreme conditions. ChIMES does not include any explicit quantum-mechanical components, which affords significant gains in computational efficiency but can simultaneously complicate the development of transferable models relative to DFTB. Nonetheless, ChIMES can allow for large-scale million atom simulations where precise knowledge of electronic states might not be necessary for a specific research problem. We have applied ChIMES to water under ambient and high pressure–temperature conditions [71, 72], and metallic liquid carbon [73]. Here, we will focus on simple ChIMES force field development for metallic liquid carbon near the graphite–diamond–liquid triple point (e.g., ∼15 GPa, 5000 K) as an illustrative example.

4.4.1 The ChIMES Force Field

There are numerous force fields designed to describe the physics governing chemical reactivity and metallic nature. For example, there are the ReaxFF [17], REBO [74], and COMB [75] potentials that leverage a reactive bond order description for bond

formation and breaking, the more general GAP [76, 77] and SNAP [78], methods which describe atomic environments through use of bispectrum components, and the EAM [79], and MEAM [80] embedded atom models. While highly successful in their intended applications, these models are not available for a large variety of systems or conditions, and thus frequently require reparameterization. Unfortunately, the complex nature of these reactive force fields combined with the large *ab initio* training data they frequently require can make their tailored parametrization for specific systems a challenging task.

ChIMES models are comprised of explicit n-body interatomic energy terms, which themselves, are constructed from linear combinations of Chebyshev polynomials. Resulting models are highly flexible and thus well suited for materials under extreme conditions. This allows for rapid parameterization of any number of new MD models through linear least-squares approaches like the singular value decomposition (SVD) [25].

To date, the ChIMES total energy of interaction has been described as the following sum of two- and three-body terms:

$$E_{\text{ChIMES}} = \sum_{i}^{N} \sum_{j>i} E_{ij} + \sum_{i}^{N} \sum_{j>i} \sum_{k>j} E_{ijk}. \tag{4.8}$$

Here, N is the total number of atoms in the system, E_{ij} is the pairwise interaction energy, and E_{ijk} is the energy between triplets of atoms. The two-body energy is expressed as follows:

$$E_{ij} = f_{\text{p}}^{e_i e_j}\left(r_{ij}\right) + f_{\text{c}}^{e_i e_j}\left(r_{ij}\right) \sum_{n=1}^{\mathscr{O}_2} C_n^{e_i e_j} T_n(s_{ij}). \tag{4.9}$$

In this case, $T_n\left(s_{ij}\right)$ represents a Chebyshev polynomial of order n, s_{ij} is the pair distance transformed over the interval $[-1, 1]$ (discussed below), and $C_n^{e_i e_j}$ is the corresponding coefficient for the interaction between atom types e_i and e_j. Permutational invariance of the polynomials is enforced for all interactions, e.g., $C_n^{e_i e_j} \equiv C_n^{e_j e_i}$. The term $f_{\text{c}}^{e_i e_j}(r_{ij})$ is a smooth cutoff function which is set to zero beyond a maximum distance defined for a given atom pair type, i.e., $f_{\text{c}}^{e_i e_j}(r_{ij}) = (1 - r_{ij}/r_{\text{max}}^{e_i e_j})^3$. In order to prevent sampling of r_{ij} distances below what is present in our DFT training set, we introduce a penalty function $f_{\text{p}}^{e_i e_j}(r_{ij})$, which we define as follows [71]:

$$f_{\text{p}}^{e_i e_j} = A_{\text{p}} \left(r_{\text{p}}^{e_i e_j}\right)^3 \tag{4.10}$$

$$r_{\text{p}}^{e_i e_j} = \begin{cases} r_{\text{min}}^{e_i e_j} + d_{\text{p}} - r_{ij}, & r_{ij} - d_{\text{p}} < r_{\text{min}}^{e_i e_j} \\ 0, & \text{otherwise.} \end{cases} \tag{4.11}$$

The parameter A_{p} is a penalty function scaling factor and d_{p} is the penalty distance, which are preset to specific values. This allows for atoms to be "pushed" to larger

distances to avoid unphysical regions of the potential. We note that the penalty function was not sampled for any of the MD calculations presented here.

We map the interatomic distances over the interval of [−1,1] by first transforming the internuclear distance r_{ij} to the Morse variable, $x_{ij} = \exp(-r_{ij}/\lambda^{e_i e_j})$, where the λ parameter is the Morse variable range parameter [81–83], defined individually for each type of atomic pair interaction. The Morse variables lead to a natural decrease in the interaction strength as distance is increased. As a rule of thumb we set these values to correspond approximately to the nearest neighbor distance for each atom pair type from our fitting set, though in general we find that the results of our fit are relatively insensitive to these values. We then define the variable s_{ij} to be within the range [−1, 1] through the operation $s_{ij} \equiv (x_{ij} - x_{\text{avg}}^{e_i e_j})/x_{\text{diff}}^{e_i e_j}$ where:

$$x_{\text{avg}}^{e_i e_j} = 0.5(x_{\text{max}}^{e_i e_j} + x_{\text{min}}^{e_i e_j}) \tag{4.12}$$

$$x_{\text{diff}}^{e_i e_j} = 0.5(x_{\text{max}}^{e_i e_j} - x_{\text{min}}^{e_i e_j}) \tag{4.13}$$

$$x_{\text{max}}^{e_i e_j} = \exp(-r_{e_i e_j}^{\text{min}}/\lambda^{e_i e_j}) \tag{4.14}$$

$$x_{\text{min}}^{e_i e_j} = \exp(-r_{e_i e_j}^{\text{max}}/\lambda^{e_i e_j}). \tag{4.15}$$

In this work, we will focus on development and application of a ChIMES model for two-body interactions, only, for the sake of simplicity. However, for completeness we include discussion of higher-body ChIMES terms as well. Similar to the two-body representation, the three-body energy is given as the product of Chebyshev polynomials for each of the three constituent atomic pairs:

$$E_{ijk} = f_{\text{c}}^{e_i e_j}\left(r_{ij}\right) f_{\text{c}}^{e_i e_k}\left(r_{ik}\right) f_{\text{c}}^{e_j e_k}\left(r_{jk}\right) {\sum_{m=0}^{\mathcal{O}_3}}' {\sum_{p=0}^{\mathcal{O}_3}}' {\sum_{q=0}^{\mathcal{O}_3}}' C_{mpq}^{e_i e_j e_k} T_m\left(s_{ij}\right) T_p\left(s_{ik}\right) T_q\left(s_{jk}\right). \tag{4.16}$$

The single sum given for the two-body energy is now replaced with a triple sum for the ij, ik, and jk polynomials, yielding a single permutationally invariant coefficient for each set of powers and atom types, $C_{mpq}^{e_i e_j e_k}$. The primed sum indicates that only terms for which two or more of the m, p, q indicies are greater than zero are included, which guarantees that three bodies i, j, k enter into the expression. The expression for E_{ijk} also contains the f_{c} smoothly varying cutoff functions for each constituent pair distance. Penalty functions are not included for three-body interactions and are instead handled by the two-body E_{ChIMES} terms. We note that the ChIMES functional form can readily be extended to higher-body terms, e.g., four-body energies, by expanding (4.16) to include a multiplication of all $\binom{n}{2}$ pairwise interactions for that set (i.e., six total for four-body terms). Determination of greater than three-body ChIMES models is the subject of future work.

4.4.2 Application to Molten Carbon: Computational Details

In our efforts discussed here, ChIMES models can be viewed as "cooked to order," where transferability is not considered an inherent feature of the force field. Instead, models are generated for a target region of state point space by force matching to one or more short DFT simulations ($\mathscr{O}$ 5–10 ps) for the system of interest. Self-consistent refinement can be used for systems expected to undergo significant changes (i.e., as for formation of new phases) [73]. Here, we generate 2-body ChIMES models for molten carbon at 5000 K and 2.43 g/cm^3. This system is particularly well-suited for demonstrative purposes as it is monoatomic and exhibits fast dynamics. Model sensitivity to user-specified parameters including polynomial order, outer cutoff, and Morse transformation variable will be discussed, and results will be compared against those from popular reactive carbon models LCBOP [84] and REBO [85].

We obtain a NVT-MD trajectory for 256 carbon atoms at our target state point via DFT-MD simulations utilizing the VASP software package [86–89]. The Perdew–Burke–Ernzerhof generalized gradient approximation functional (PBE) [90, 91] and projector-augmented wave pseudopotentials [92, 93] were used, along with a planewave cutoff of 1000 eV, a 0.5 fs time step, and a global Nose–Hoover thermostat [54, 55]. A total of 20 evenly spaced frames are selected from the resulting DFT trajectory, resulting in 250 fs between frames. Chebyshev polynomial coefficients are then obtained by minimizing the objective function shown in (4.1). We note that this step generally takes from a few seconds to several minutes to yield optimal parameter values.

Simulations using the presently developed ChIMES models are run with in-house developed MD software and a 0.25 fs time step, whereas LCBOP and REBO simulations are run with the LAMMPS suite [59] and a 0.5 fs time step. An atomic configuration from our DFT-MD training trajectory was used as the starting point for all simulations. All classical MD simulations are in the canonical ensemble at 5000 K and 2.43 g/cm^3, and utilize a global Nose–Hoover thermostat.

4.4.3 ChIMES Sensitivity to User-Specified Parameters

A total of nine ChIMES models were fit, which varied by polynomial order $\mathscr{O}_{2B} =$ 4, 8, or 12, outer cutoff distance $r_{max}^{CC} = 2.00$, 3.15, 4.25, or 5.00 Å, and Morse variable $\lambda^{CC} = 1.25$, 2.25, and 3.25 Å. The remaining model parameters, A_p^{CC}, d_p^{CC} and r_{min}^{CC} were fixed at values of 10^8 kcal/mol $\cdot$ Å^3, 0.01 Å, and 1.0 Å, respectively, where the inner cutoff was set to the minimum distance observed in the training configurations, and A_p^{CC} and d_p^{CC} were selected to yield a penalty function strong enough to prevent the system from exploring distances smaller than $r_{c,in}^{CC}$, while otherwise having minimal influence on dynamics. Models have a total of $n_{usrpar} + \mathscr{O}_{2B}$ parameters, where n_{usrpar} is the number of user defined parameters, and for the

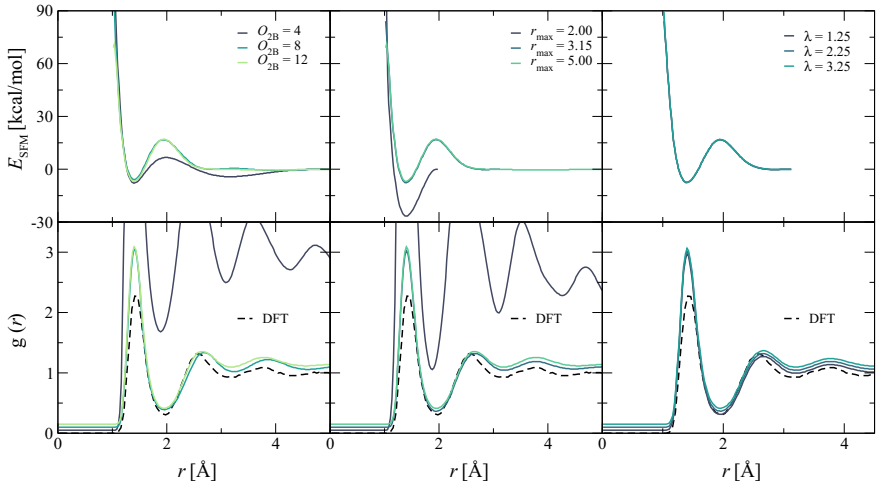


Fig. 4.4 Sensitivity of system structure to user-defined model parameters. Radial distribution functions have been offset by 0.05 Å for clarity. E_{SFM} is the 2-body potential function derived from force matching

present work, encompasses the 6 parameters $\{\mathscr{O}_{2B}, r_{min}^{CC}, r_{max}^{CC}, \lambda^{CC}, A_p^{CC}, A_p^{CC}\}$. Thus, a model with $\mathscr{O}_{2B} = 10$ would have 16 parameters.

Figure 4.4 demonstrates how the potential energy surfaces (PESs) and subsequent radial distribution functions (RDFs) are altered as the Chebyshev polynomial order, outer cutoff distances, and Morse variable are increased. The results indicate that for both order and outer cutoff, some threshold value must be surpassed to obtain reasonable results, but beyond that value, the predicted system structures are indistinguishable from one another. Specifically, simulations using models with either $\mathscr{O} = 4$ or an r_{max}^{CC} of 2.0 solidify while all remaining simulations (using larger order and r_{max}^{CC} values) yield RDFs that are in agreement with one another. Decreasing polynomial order limits model flexibility, and, as shown for the $\mathscr{O} = 4$ case, under-structuring of the PES (i.e., as compared to $\mathscr{O} = 8$ and 12, between 1.5 and 2.25 Å). The case with $r_{max}^{CC} = 2.0$ Å, on the other hand, fails because the cutoff distance forces the potential to zero where it would otherwise contain a maximum. Furthermore, the data suggests that a reasonable choice for selection of r_{max}^{CC} would be the location of the second nonbonding minimum for a given $g(r_{max}^{CC})$. The data also indicate that the choice of λ^{CC} has little influence on resulting system structure. Nevertheless, a reasonable selection would be some characteristic "bonding" distance for the system, such as the location of the first peak in the pair RDF. An additional and pertinent user-specified aspect of the fit is the number of training frames utilized. For the present model problem, 20 frames were found to be sufficient for good fits, as including more did not yield significantly improved RMS errors or RDF predictions. In practice, one should consider the influence of training database size on a case-by-case basis, particularly

when increasing bodiedness of the ChIMES potential, and when targeting systems of multiple atom pair types and/or a variety of state points or phases.

4.4.4 *Molten Carbon: Comparing ChIMES and Existing Reactive Models*

Performance of the force-matched model with $\mathscr{O}_{2B}$, r_{max}, and λ = 12, 3.15, and 1.25 Å, respectively [73] is compared to two existing reactive bond order carbon models, LCBOP and the 2002 parameterization of REBO [84, 85]. Both of these models contain numerous nonlinear parameters that are fit by iterative adjustment to either experimental or first-principles reference data. Notably, both models are fit to solid phase carbon data (i.e., graphite and diamond), and are designed to reproduce these corresponding material properties. Figure 4.5 presents the distribution of forces predicted from DFT, the present force-matched model, LCBOP, and REBO for a given set of configurations. Not surprisingly, the force-matched model yields significantly better agreement with DFT across the range of predicted values. Both LCBOP and REBO exhibit poor reproduction of the DFT force distribution; REBO overestimates the probability of observing low-lying forces, while LCBOP yields too high of a distribution of large magnitude forces.

The improved description of forces presented by the ChIMES model leads to more accurate system structure and dynamics, as is shown in Fig. 4.6. Both LCBOP and REBO predict overly ordered systems, with the first minimum in each RDF at nearly zero. REBO yields a better estimate of first peak height and recovers the DFT structure by the third peak, while LCBOP maintains solid-like structure over the

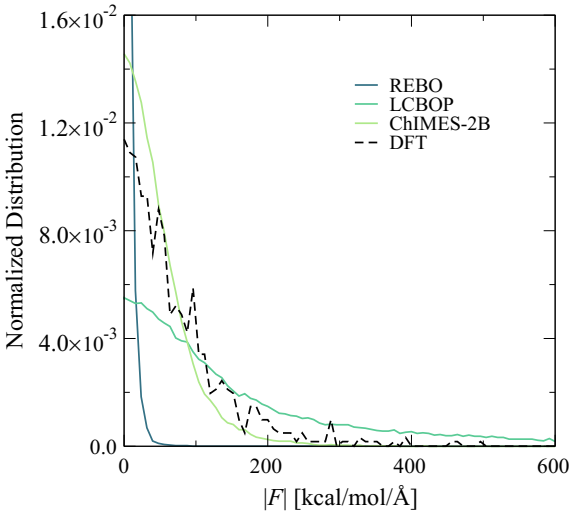


Fig. 4.5 Normalized distribution of forces acting on carbon at 5000 K and 2.43 g/cm^3

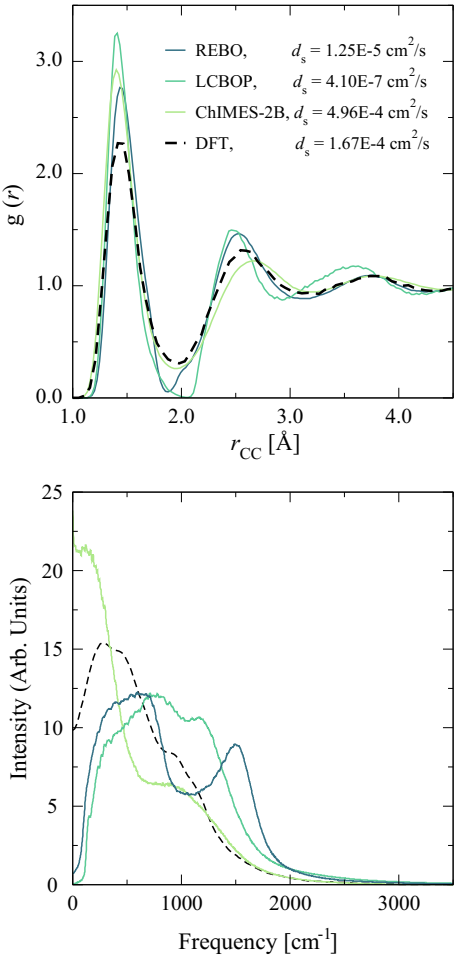


Fig. 4.6 Radial distribution functions (top) and vibrational power spectra (bottom) for carbon at 5000 K and 2.43 g/cm^3. Self-diffusion constants are given in the legend for each model

range of distances presented. In contrast, the ChIMES model exhibits a reasonable description of the structure predicted by DFT over the entire range of distances. The overly-solid nature of the LCBOP and REBO systems is also reflected in the power spectra and self-diffusion constants. Both of these models yield peak intensities of nearly zero at low frequency, and accordingly, and predict diffusion constants that are several orders of magnitude smaller than the DFT result. Furthermore, the LCBOP and REBO power spectra do not capture the shape predicted by DFT, resulting in a broad distribution of high-intensity frequencies between 250 and 1500 cm^{-1}. In contrast, the ChIMES model captures the overall DFT power spectrum shape (a peak near 500 cm^{-1} and a shoulder near 1000 cm^{-1}) and exhibits a significantly improved diffusion constant, which is of the same order of magnitude as that from DFT. These results are particularly illustrative of the capability of ChIMES, given the exceedingly small training set and low-order functional form used here. We note, however, that

LCBOP and REBO are likely to be accurate over a wider range of conditions than the simple 2-body ChIMES model considered here. Better transferability with ChIMES can be obtained by adding many-body interactions and using more diverse training sets [73].

4.4.5 ChIMES: Outlook

Further ChIMES refinements of all of these properties have been determined though creation of three-body force fields, discussed in detail in recently published work [73]. Many-body effects can additionally be included in a number of ways such as explicit polarizability, or over-bonding terms (as in [71]). Our model problem discussed here demonstrates that the ChIMES force field and fitting framework provides a convenient approach for rapid generation of accurate DFT-informed reactive force fields. The utility of ChIMES extends far beyond the present application; in addition to use as a standalone molecular mechanics force field, three-body ChIMES models has been successfully leveraged as the repulsive term in DFTB for surface corrosion studies on plutonium [36].

4.5 Conclusions

Coupled advances in computing hardware and algorithms have enabled particle-based simulations to play an increasingly critical role in reactive materials research. Nevertheless, computer experiments based on quantum-mechanical methods remain too costly for many problems of interest. In this chapter, we have discussed two methods to force match models capable of extending the predictive power of quantum mechanics to larger scales. In the first, the repulsive term in the semi-empirical DFTB method was force-matched to DFT, to study glycine dimerization at low T and p conditions. Our model was found to yield a strong refinement of standard DFTB parameterizations, allowing for accurate determination of the diglycine formation free energy surface. In the second, force matching was used to construct a fully reactive ChIMES force field for molten metallic carbon, which is 10^5 times more efficient than DFT while retaining most of its accuracy. In general, our force-matched models are capable of serving as a high-fidelity proxy for obtaining DFT-quality forces on large time and length scales. There are any number of research areas that require atomic level knowledge of material reactivity, including studies of planetary interiors [94, 95], organic polymers subject to radiation and heat damage [96, 97], and the synthesis of materials with tailored properties [70]. The computationally efficient methods discussed in this chapter will have particular impact in these types of research areas, where there is traditionally a reliance on expensive DFT calculations for interpretation of imaging and spectroscopy experiments, and prediction of properties to guide materials synthesis.

Acknowledgements This work was partially supported by NASA Astrobiology: Exobiology and Evolutionary Biology Program Element NNH14ZDA001N. Work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The project 16-LW-020 was funded by the Laboratory Directed Research and Development Program at LLNL with N.G. as principal investigator.

References

1. Barrios MA, Hicks DG, Boehly TR, Fratanduono DE, Eggert JH, Celliers PM, Collins GW, Meyerhofer DD (2010) Phys Plasmas 17:056307
2. Manaa MR, Reed EJ, Fried LE, Goldman N (2009) JACS 131(15):5483. https://doi.org/10.1021/ja808196e
3. Eggert J, Hicks D, Celliers P, Bradley D, McWilliams R, Jeanloz R, Miller J, Boehly T, Collins G (2010) Nat Phys 6(1):40
4. Goldman N, Reed EJ, Fried LE (2009) J Chem Phys 131(20):204103
5. Menikoff R, Sewell TD (2002) Combust Theory Model 6(1):103
6. Austin RA, Barton NR, Reaugh JE, Fried LE (2015) J Appl Phys 117(18):185902
7. Kroonblawd MP, Sewell TD (2016) Propellants. Explos Pyrotech 41(3):502
8. Goldman N, Fellers RS, Brown MG, Braly LB, Keoshian CJ, Leforestier C, Saykally RJ (2002) J Chem Phys 116(23):10148
9. Goldman N, Leforestier C, Saykally RJ (2005) Philos Trans R Soc A 363(1827):493
10. Goldman N, Reed EJ, Kuo IFW, Fried LE, Mundy CJ, Curioni A (2009) J Chem Phys 130:124517
11. Goldman N, Reed EJ, Fried LE, Kuo IFW, Maiti A (2010) Nat Chem 2:949
12. Cannella CB, Goldman N (2015) J Phys Chem C 119(37):21605
13. Goldman N, Fried LE (2012) J Phys Chem C 116(3):2198
14. Kenney JF, Kutcherov VA, Bendeliani NA, Alekseev VA (2002) Proc Natl Acad Sci USA 99(17):10976
15. Scott HP, Hemley RJ, Mao HK, Herschbach DR, Fried LE, Howard WM, Bastea S (2004) Proc Natl Acad Sci USA 101(39):14023
16. Kolesnikov A, Kutcherov VG, Goncharov AF (2009) Nat Geosci 2:566
17. Van Duin AC, Dasgupta S, Lorant F, Goddard WA (2001) J Phys Chem A 105(41):9396
18. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Phys Rev B 58:7260. https://doi.org/10.1103/PhysRevB.58.7260. https://link.aps.org/doi/10.1103/PhysRevB.58.7260
19. Kroonblawd MP, Pietrucci F, Saitta AM, Goldman N (2018) J Chem Theory Comput 14:2207
20. Ercolessi F, Adams JB (1994) Europhys Lett 26(8):583
21. Liu XY, Adams JB, Ercolessi F, Moriarty JA (1996) Model Simul Mater Sci Eng 4(3):293
22. Levenberg K (1944) Quart Appl Math 2:164
23. Marquardt D, SIAM (1963) J Appl Math 11:431
24. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Science 220(4598):671
25. Press WH, Tuekolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing, 3rd edn. Cambridge University Press, Cambridge, UK
26. Tikhonov AN, Arsenin VY (1997) Solutions of Ill-posed problems, 3rd edn. Wiley, New York. Translated from Russian
27. Hoerl AE, Kennard RW (1970) Technometrics 12(1):55
28. Tibshirani R (1996) J R Stat Soc Ser B (Methodological) 267–288
29. Behler J (2015) Int J Quantum Chem 115:1032
30. Izvekov S, Rice BM (2014) J Chem Phys 140(10):104104

31. Porezag D, Frauenheim T, Köhler T, Seifert G, Kaschner R (1995) Phys Rev B 51:12947. https://doi.org/10.1103/PhysRevB.51.12947. https://link.aps.org/doi/10.1103/PhysRevB.51.12947
32. Koskinen P, Mäkinen V (2009) Comput Mat Sci 47(1):237. https://doi.org/10.1016/j.commatsci.2009.07.013. http://www.sciencedirect.com/science/article/pii/S0927025609003036
33. Aradi B, Hourahine B, Frauenheim T (2007) J Phys Chem A 111(26):5678. https://doi.org/10.1021/jp070186p. http://dx.doi.org/10.1021/jp070186p. PMID: 17567110
34. Gaus M, Cui Q, Elstner M (2011) J Chem Theory Comput 7:931
35. Goldman N (2015) Chem Phys Lett 622:128
36. Goldman N, Aradi B, Lindsey RK, Fried LE (2018) J Chem Theory Comput 14:2652
37. Goldman N, Srinivasan SG, Hamel S, Fried LE, Gaus M, Elstner M (2013) J Phys Chem C 117:7885
38. Srinivasan SG, Goldman N, Tamblyn I, Hamel S, Gaus M (2014) J Phys Chem A 118:5520
39. Armstrong MR, Zaug JM, Goldman N, Kuo IFW, Crowhurst JC, Howard WM, Carter JA, Kashgarian M, Chesser JM, Barbee TW, Bastea S (2013) J Phys Chem A 117:13051. https://doi.org/10.1021/jp407595u
40. Goldman N, Fried LE, Koziol L (2015) J Chem Theory Comput 11(10):4530. https://doi.org/10.1021/acs.jctc.5b00742. http://dx.doi.org/10.1021/acs.jctc.5b00742. PMID: 26574245
41. Torrie GM, Valleau JP (1974) Chem Phys Lett 28(4):578. https://doi.org/10.1016/0009-2614(74)80109-0. http://www.sciencedirect.com/science/article/pii/0009261474801090
42. Laio A, Parrinello M (2002) Proc Nat Acad Sci 99(20):12562. https://doi.org/10.1073/pnas.202427399
43. Schreiner E, Nair NN, Marx D (2009) J Am Chem Soc 131(38):13668. https://doi.org/10.1021/ja9032742. http://dx.doi.org/10.1021/ja9032742 PMID:19725519
44. Khaliullin RZ, Eshet H, Kühne TD, Behler J, Parrinello M (2011) Nat Mat 10:693. https://doi.org/10.1038/nmat3078. http://dx.doi.org/10.1038/nmat3078
45. Branduardi D, Gervasio FL, Parrinello M (2007) J Chem Phys 126(5):054103. https://doi.org/10.1063/1.2432340. http://dx.doi.org/10.1063/1.2432340
46. Pietrucci F, Saitta AM (2015) Proc Nat Acad Sci 112(49):15030. https://doi.org/10.1073/pnas.1512486112. http://www.pnas.org/content/112/49/15030.abstract
47. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) J Comput Chem 13(8):1011. https://doi.org/10.1002/jcc.540130812. http://dx.doi.org/10.1002/jcc.540130812
48. Roux B (1995) Comput Phys Commun 91(1):275. https://doi.org/10.1016/0010-4655(95)00053-I. http://www.sciencedirect.com/science/article/pii/001046559500053I
49. Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA, Parrinello M (2009) Comput Phys Commun 180(10):1961. https://doi.org/10.1016/j.cpc.2009.05.011. http://www.sciencedirect.com/science/article/pii/S001046550900157X
50. Plumed is available at http://www.plumed.org/
51. Car R, Parrinello M (1985) Phys Rev Lett 55:2471. https://doi.org/10.1103/PhysRevLett.55.2471. https://link.aps.org/doi/10.1103/PhysRevLett.55.2471
52. Giannozzi P, Baroni S, Bonini N, Calandra M, Car R, Cavazzoni C, Ceresoli D, Chiarotti GL, Cococcioni M, Dabo I, Corso AD, de Gironcoli S, Fabris S, Fratesi G, Gebauer R, Gerstmann U, Gougoussis C, Kokalj A, Lazzeri M, Martin-Samos L, Marzari N, Mauri F, Mazzarello R, Paolini S, Pasquarello A, Paulatto L, Sbraccia C, Scandolo S, Sclauzero G, Seitsonen AP, Smogunov A, Umari P, Wentzcovitch RM (2009) J Phys Condens Matter 21(39), 395502 (2009). http://stacks.iop.org/0953-8984/21/i=39/a=395502
53. Quantum espresso is available at http://www.quantum-espresso.org/
54. Nosé S (1984) J Chem Phys 81(1):511. https://doi.org/10.1063/1.447334. http://dx.doi.org/10.1063/1.447334
55. Hoover WG (1985) Phys Rev A 31:1695. https://doi.org/10.1103/PhysRevA.31.1695. https://link.aps.org/doi/10.1103/PhysRevA.31.1695
56. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865. https://doi.org/10.1103/PhysRevLett.77.3865. https://link.aps.org/doi/10.1103/PhysRevLett.77.3865

57. Vanderbilt D (1990) Phys Rev B 41:7892. https://doi.org/10.1103/PhysRevB.41.7892. https://link.aps.org/doi/10.1103/PhysRevB.41.7892
58. Plimpton S (1995) J Comput Phys 117(1):1. https://doi.org/10.1006/jcph.1995.1039. http://www.sciencedirect.com/science/article/pii/S002199918571039X
59. Plimpton S (1995) J Phys Chem C 117(1):1
60. DFTB+ is available at https://www.dftb-plus.info/
61. Niklasson AMN, Tymczak CJ, Challacombe M (2006) Phys Rev Lett 97:123001. https://doi.org/10.1103/PhysRevLett.97.123001. https://link.aps.org/doi/10.1103/PhysRevLett.97.123001
62. Niklasson AMN (2008) Phys Rev Lett 100:123004. https://doi.org/10.1103/PhysRevLett.100.123004. https://link.aps.org/doi/10.1103/PhysRevLett.100.123004
63. Niklasson AMN, Steneteg P, Odell A, Bock N, Challacombe M, Tymczak CJ, Holmström E, Zheng G, Weber V (2009) J Chem Phys 130(21):214109. https://doi.org/10.1063/1.3148075. http://dx.doi.org/10.1063/1.3148075
64. Zheng G, Niklasson AMN, Karplus M (2011) J Chem Phys 135(4):044122. https://doi.org/10.1063/1.3605303. http://dx.doi.org/10.1063/1.3605303
65. Mullen RG, Shea JE, Peters B (2014) J Chem Theory Comput 10(2):659
66. Goyal P, Qian HJ, Irle S, Lu X, Roston D, Mori T, Elstner M, Cui Q (2014) J Phys Chem B 118(38):11007
67. Borsook H (1953) Adv Protein Chem 8:127. https://doi.org/10.1016/S0065-3233(08)60092-3. http://www.sciencedirect.com/science/article/pii/S0065323308600923
68. Kitadai N (2014) J Mol Evol 78(3):171. https://doi.org/10.1007/s00239-014-9616-1. https://doi.org/10.1007/s00239-014-9616-1
69. Shock EL (1992) Geochim et Cosmochim Acta 56(9):3481. https://doi.org/10.1016/0016-7037(92)90392-V. http://www.sciencedirect.com/science/article/pii/001670379290392V
70. Kroonblawd MP, Goldman N (2018) Phys Rev B 97(18):184106
71. Koziol L, Fried LE, Goldman N (2017) J Chem Theory Comput 13:135
72. Lindsey RK, Fried LE, and Goldman N (2019) J Chem Theory Comput 15:436
73. Lindsey RK, Fried LE, Goldman N (2017) J Chem Theory Comput 13:6222
74. Brenner DW (1990) Phys Rev B 42(15):9458
75. Yu J, Sinnott SB, Phillpot SR (2007) Phys Rev B 75(8):085311
76. Bartók AP, Payne MC, Kondor R, Csányi G (2010) Phys Rev Lett 104(13):136403
77. Bartók AP, Kondor R, Csányi G (2013) Phys Rev B 87(18):184115
78. Thompson AP, Swiler LP, Trott CR, Foiles SM, Tucker GJ (2015) J Comput Phys 285:316
79. Daw MS, Baskes MI (1984) Phys Rev B 29(12):6443
80. Baskes M (1992) Phys Rev B 46(5):2727
81. Wang Y, Shepler BC, Braams BJ, Bowman JM (2009) J Chem Phys 131(5):054511
82. Wang Y, Huang X, Shepler BC, Braams BJ, Bowman JM (2011) J Chem Phys 134(9):094509
83. Braams BJ, Bowman JM (2009) Int Rev Phys Chem 28(4):577
84. Los J, Fasolino A (2003) Phys Rev B Condens Matter Mater Phys 68(2):024107
85. Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) J Phys Condens Matter 14(4):783
86. Kresse G, Hafner J (1993) Phys Rev B Condens Matter Mater Phys 47(1):558
87. Kresse G, Hafner J (1994) Phys Rev B Condens Matter Mater Phys 49(20):14251
88. Kresse G, Furthmüller J (1996) Comput Mater Sci 6(1):15
89. Kresse G, Furthmüller J (1996) Phys Rev B Condens Matter Mater Phys 54(16):11169
90. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77(18):3865
91. Perdew JP, Burke K, Ernzerhof M (1997) Phys Rev Lett 78:1396
92. Blöchl PE (1994) Phys Rev B Condens Matter Mater Phys 50(24):17953
93. Kresse G, Joubert D (1999) Phys Rev B Condens Matter Mater Phys 59(3):1758
94. French M, Mattsson TR, Nettelman N, Redmer R (2009) Phys Rev B 97:054107
95. Knudson MD, Desjarlais MP, Lemke RW, Mattsson TR, French M, Nettelmann N, Redmer R (2012) Phys Rev Lett 108:091102
96. Chenoweth K, Cheung S, van Duin ACT, Goddard WA, Kober EM (2005) J Am Chem Soc 127:7192
97. Melo LGA, Hitchcock AP, Susac D, Stumper J, Berejnov V (2018) Phys Chem Chem Phys 20:16625

Chapter 5
Free Energy Calculations of Electric Field-Induced Chemistry

Giuseppe Cassone, Fabio Pietrucci, Franz Saija and A. Marco Saitta

Abstract The old and challenging problem of dealing with the interaction between condensed matter systems and intense external electric fields are currently evolving in an impressive way. In fact, the growth of the computational resources allows for accurate first-principles numerical calculations showing unprecedented predictive power. We review the phenomenological evidence that has recently emerged from state-of-the-art ab initio molecular dynamics simulations in describing how static electric fields can be exploited to manipulate matter and possibly design novel compounds or materials, obtain new exotic properties, and achieve more efficient reaction yields. In particular, we show the microscopic behavior of simple molecular liquids (water, methanol, and simple mixtures), under the action of static and homogeneous electric fields, showing different shades of the effects produced by the application of the latter. In addition, ab initio molecular dynamics approaches are coupled with advanced free energy methods, that currently represents a unique technique for adequately treating, reproducing, and predicting both molecular mechanisms and chemical reaction networks triggered when matter is exposed to the action of intense electric fields.

G. Cassone
Institute of Biophysics, Czech Academy of Sciences, Královopolská 135, 61265 Brno, Czech Republic
e-mail: cassone@ibp.cz

F. Pietrucci · A. M. Saitta (✉)
Sorbonne Université, CNRS, MNHN, IRD, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Unité Mixte de Recherche 7590, 75005 Paris, France
e-mail: marco.saitta@sorbonne-universite.fr

F. Pietrucci
e-mail: fabio.pietrucci@sorbonne-universite.fr

F. Saija
CNR-IPCF, Viale Ferdinando Stagno d'Alcontres 37, 98158 Messina, Italy
e-mail: franz.saija@cnr.it

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_5

5.1 Introduction

Among the many possible "extreme conditions" that can drive or facilitate chemical reactions and transformations of matter, such as pressure, temperature, pH, confinement, bombardments, and so forth and so on, intense electric fields have received so far little experimental attention and almost no theoretical at all. Thermal effects induced on matter by electric fields are well-known and largely understood [1]. However, the same is not true for nonthermal effects and for the related microscopic mechanisms that are triggered by intense electric fields [2, 3]. Our ignorance on this matter is manifest over different scales. In fact, nonthermal effects are poorly understood both at a sub-atomic, quantum, level as well as at molecular and macroscopic levels. An account of the current state of the art in this specific field has recently appeared in the literature [4]. Aragones et al. have given the first experimental evidence that an electric field can control chemical reactions, showing that the field strength, as well as its polarity, can actually drive, enhance, or even inhibit a given reaction (e.g., a Diels–Alder reaction). In order to explain such a "delay" in investigating and clarifying phenomenological aspects of materials which, in some respects, would sound logically plausible—if not absolutely trivial—even to a non-expert, we should consider some experimental and theoretical boundary conditions on the nature of the problem. First of all, very intense local electric fields are necessary in order to induce a rearrangement of covalent bonds, definitely not an easy task on the experimental side. Second, only very recently reliable theoretical approaches have been implemented and developed, which are capable of treating, in the framework of a quantum description of matter, field-induced effects. Such tools, thanks to the fast growth of available computing power, have eventually allowed for the discovery of some fundamental "pieces" of physics and chemistry when matter is being irradiated with electric fields.

The basic interaction between a stationary, spatially homogeneous, electric field and one single molecule in the vacuum is well understood over a wide range of intensities. By resorting to the theoretical tools provided by quantum mechanics and Density Functional Theory (DFT), in conjunction with advanced computing techniques, it is possible to simulate in a realistic way and, thus, finely describe all the relevant degrees of freedom of such a simple model system. However, the situation becomes much more intricate and complicated when dealing with condensed matter systems, in which also complex interactions between molecules have to be taken into account. The subtle interplay between field-induced and neighbor-induced polarization effects, in addition to the role of thermodynamic conditions, makes the task of the computational physicist much harder. But an even more fundamental reason has represented, on this side, a tough hurdle when attempting at modeling and simulating realistic material samples under the effect of static electric fields. Indeed, typical numerical experiments carried out over homogeneous samples are ordinarily performed on a cell with Periodic Boundary Conditions; in this way, one avoids, to a large extent, the effects induced by the artificial spatial confinement of the elementary interacting constituents, which would obviously inhibit any significant

comparison of the numerical results with the properties of a real-life homogeneous material [5]. In classical Molecular Dynamics (MD) simulations the electric field is implemented as an additional force acting on each particle. However, the serious problem in dealing with such force fields resides in the effective intermolecular potential which is not able to account, in a reliable way, for polarization effects as well as, more importantly, for the very quantum nature of matter.

Since, as previously pointed out, electric fields are able to induce the cleavage and the formation of (intrinsically quantum) covalent bonds, it is clear that classical simulations fail to be predictive in this specific context in that they do not provide a complete physical representation of the involved phenomena. In order to model and reproduce the whole spectrum of effects induced by the application of an electric field, ab initio molecular dynamics (AIMD) simulations are mandatory for a twofold reason. First, they are able to reproduce correctly intra- and intermolecular structural properties in many disparate systems because of the appropriate, more fundamental description of the material (see [6] and references contained therein). Second, AIMD simulations are, by their own nature, intrinsically able to treat, in principle, every kind of covalent bond rearrangement and, hence to deal with chemical reactions.

Unfortunately, because of the nonperiodic nature of the quantum position operator, the implementation of electric fields in ab initio simulation codes are other than easy. Indeed, only in 2002 Umari and Pasquarello [7] discussed and implemented the first operational theoretical framework which, upon exploiting Berry's phases and the modern theory of polarization [8–10], was able to manage static electric fields in first-principles simulations under Periodic Boundary Conditions.

In this chapter, we will describe a few recent works and results, capable of underlying how valuable electric fields can be used as unexplored tools to manipulate matter and possibly design novel compounds or materials, obtain new exotic properties, achieve more efficient yields. In particular, we will review the microscopic phenomenology of liquid solutions of simple molecular systems under the action of a static and homogeneous electric field, showing different shades of the effects produced by the application of the field. Schematically, chemical bonds within these systems might have fully covalent, partial covalent, electrostatic, H-bond, dispersed, and polarized character [11], which can be dramatically modified by an external electric field. Hence, with such multifaceted properties of electronic matrix, their treatment by means of ab initio techniques is particularly suited. In fact, molecular dissociations can be induced for high enough field intensities; correspondingly, related simple and complex chemical reactions can be modeled and investigated. These latter ones usually have barriers higher than the thermal or thermal+electric energy. Hence, from a theoretical perspective, AIMD simulations, combined with advanced free energy methods (metadynamics, committor analysis, umbrella sampling) currently represent a unique technique for adequately treating, reproducing, and predicting many microscopic and macroscopic phenomena which occur when matter is exposed to the action of intense electric fields.

The chapter is structured as follows. In the next sections, various effects manifested for increasing field strengths will be treated with a gradual approach: in Sect. 5.1 the theoretical framework of ab initio calculations in the presence of external electric

fields and the basic concepts of the free energy methods will be presented; in Sect. 5.2 a few emblematic and barrierless spontaneous chemical reactions induced by very intense electric fields will be analyzed by AIMD; finally, in Sect. 5.3 more complex chemical reactions will be studied through ab initio free energy methods.

5.2 Methods

5.2.1 *Static and Homogeneous Electric Fields in Ab Initio Codes*

The theoretical development and the implementation of electric fields in ab initio codes is all other than trivial. A wide literature exists in this field [7, 9, 10, 12–14] and it is impossible to exhaustively resume all the conceptual steps which started since the precise definition of polarization.

One of the key points that one has to face in dealing with this delicate aspect is the ubiquitous presence of Periodic Boundary Conditions. The first problem that arises is the inherent discontinuity carried by the infinite replication of the simulation box. In fact, at the edges of the box(es), infinite electric field strengths appear when the linear electrostatic potential is implemented.

More seriously, the periodicity in the presence of a macroscopic electric field $\mathcal{E}$ leads to a change in the electron potential within each replica of the simulation box. The origin of this latter problem resides indeed in the nonperiodic nature of the position operator. In particular, the electric field changes by a factor of $e\mathcal{E} \cdot \mathbf{R}$ under a translation by a lattice vector $\mathbf{R}$, and even a small field varies the nature of the energy eigenstates. Moreover, because the potential is unbounded from below, the ground state is ill-defined [12, 15, 16].

Many perturbative treatments of the application of an electric field have been proposed but only within the Modern Theory of Polarization and Berry's phases this delicate issue can be efficiently tackled [8–10]. Few of them [12] are founded on a Wannier-function-based solution to the finite-field problem that was not very useful in practice [9, 10]. By means of the use of Bloch's functions, Nunes and Gonze [13] showed how the common perturbative treatments could be directly obtained from a variational principle based on minimizing an energy functional F of the following form

$$F = E_{\mathrm{KS}}(\{\psi_{\mathbf{k}n}\}) - \mathcal{E} \cdot \mathbf{P}(\{\psi_{\mathbf{k}n}\})\,. \tag{5.1}$$

In the latter, $E_{\mathrm{KS}}(\{\psi_{\mathbf{k}n}\})$ represents the Kohn–Sham (KS) energy per unit volume as a function of all occupied Bloch's functions and $\mathbf{P}(\{\psi_{\mathbf{k}n}\})$ is the usual zero-field Berry's phase expression for the electronic polarization. In practical applications, this equation has to be minimized with respect to all $\{\psi_{\mathbf{k}n}\}$ in the presence of the field. This way also the Bloch's functions will become functions of $\mathcal{E}$, and F develops into a function that implicitly depends on the field.

Within the KS scheme, the Berry's phase polarization of the noninteracting KS system is not—in general—the correct one [17, 18]. One way to circumvent this problem is to consider a generalized Hohenberg–Kohn theorem in which $n(\mathbf{r})$ and $\mathbf{P}(\mathbf{r})$ uniquely determine any ground state property [17, 18]. Moreover, in the corresponding KS scheme one has to find an effective periodic potential and an effective electric field $\mathcal{E}_{\mathbf{KS}}$ that yield for a noninteracting system the same $n(\mathbf{r})$ and $\mathbf{P}(\mathbf{r})$. In this theory [17, 18], the polarization is the correct one by construction, whereas a correction $\mathcal{E}_{\mathbf{KS}} - \mathcal{E}$ has to be introduced. It is thus clear that introducing a sort of "exchange-correlation electric field" is far from being a satisfactory solution for a theoretically well-defined problem where a similar approximation holds also for the energy.

Only in 2002, Umari and Pasquarello [7] demonstrated that the functional (5.1) is exploitable as energy functional for a variational approach to the finite-field problem as well. The proof is not obvious since the occupied wavefunctions $\{\psi_{\mathbf{k}n}\}$ are not eigenstates of the Hamiltonian. Hence, the problem of computing the polarization $\mathbf{P}$ in an electric field provides the solution to the problem of computing any property of an insulator in a finite homogeneous electric field; in fact, it is the introduction of the Berry's phase [8] polarization into the functional F of (5.1) that solves the overall problem. In particular, by considering an electric field along the x direction, the following variational energy functional can be built

$$E^{\mathcal{E}}[\{\psi_i\}] = E^0[\{\psi_i\}] - \mathcal{E} \cdot P[\{\psi_i\}]\,, \tag{5.2}$$

where $E^0[\{\psi_i\}]$ is the well-known energy functional in the zero-field system and $P[\{\psi_i\}]$ is the polarization defined by Resta [14]:

$$P[\{\psi_i\}] = -\frac{L}{\pi} Im(\ln det\, S[\{\psi_i\}])\,, \tag{5.3}$$

where, in turn, L is the periodicity of the cell and $S[\{\psi_i\}]$ is the following matrix

$$S_{i,j} = \langle \psi_i | e^{2\pi i x/L} | \psi_i \rangle \tag{5.4}$$

for the set of doubly occupied wavefunctions. By comparing their results with other perturbative approaches [13], Umari and Pasquarello conclude that their method is applicable "even when the periodic symmetry is relaxed" [7]. In addition, this formulation can be extended to yield at the same time the atomic forces. Indeed by adding the following term to the functional (5.2)

$$E^{\mathcal{E}}_{\text{ion}} = -\mathcal{E} \cdot P_{\text{ion}}\,, \quad P_{\text{ion}} = \sum_{i=1}^{N_{\text{ion}}} Z_i \cdot R_i\,, \tag{5.5}$$

where P_{ion} is the ionic polarization, R_i is the position coordinate in the field direction and Z_i is the charge of the ionic core, this definition leads to an extra-term on

the force acting on the ith atom equal to $F_i = \mathcal{E} Z_i$. This powerful and reliable methodology developed by Umari and Pasquarello [7] has been thoroughly employed in the simulations presented in this chapter when an electric field has been applied.

5.2.2 *Free Energy Landscapes: Collective Variables and Metadynamics*

As stated before, ab initio methods are by their nature extremely powerful tools in predicting and analyzing chemical reactions. One of the most important physical quantities in describing activated processes is, of course, the free energy surface (FES). This latter is represented by a function that typically depends on a number of variables (i.e., collective variables (CV)) which can be either one or several depending on the specific reaction and on the chosen representation.

Let us consider a system in the canonical (NVT) ensemble; let us introduce a collective variable $q(\mathbf{R})$, a function of the atomic coordinates, that is able to distinguish the relevant metastable states of the system (i.e., reactants and products). The probability of finding the system in a specific configuration characterized by the reaction coordinate s is given by

$$P(s) = \frac{1}{Q} \int e^{-\frac{U(\mathbf{R})}{k_B T}} \delta(s(\mathbf{R}) - s) \mathrm{d}^N \mathbf{R} \,, \tag{5.6}$$

with Q is the partition function. The free energy is related to this quantity by

$$F(s) = -k_B T \ln P(s) \,, \tag{5.7}$$

and thus $P(s) \propto e^{-\frac{F(s)}{k_B T}}$. As a consequence, if it were be possible to explore the entire configuration space of a system, what is in principle feasible by means of an extremely long equilibrium MD trajectory, it would be straightforward to reconstruct its FES. Unfortunately, the rates at which chemical reactions evolve are easily smaller than 1 h^{-1} making infeasible the application of even classical MD techniques in most cases. However, enhanced sampling algorithms, reviewed in [19], allow overcoming this hurdle and to reconstruct the FES; among the most employed methods, metadynamics (MetD) [20] holds a prominent place. This approach biases the potential energy along a CV to enhance the sampling of the corresponding free energy landscape. The bias is represented by a history-dependent function which, as the dynamics evolves, decreases the probability of visiting already explored configurations. In particular, by depositing (typically Gaussian) hills of potential energy centered in the visited points of the CV space, MetD is able to fill and escape the local free energy minima and, at the end of the simulation, to reconstruct the underlying FES. At time t, the bias at s is

$$W(s,t) = \sum_{i=1}^{N_{\text{hills}}(t)} h e^{-\frac{1}{2}\left(\frac{s-s(t_i)}{\sigma}\right)^2}, \; t_i = i \cdot \tau \,, \tag{5.8}$$

where h is the barrier height and τ is the inverse of the hills deposition rate (i.e., the MetD time step). Step-by-step, the "valleys" characterizing the FES are filled, and this latter is flattened, so that—under favorable conditions—the simulation achieves convergence. In this case:

$$-\lim_{t\to\infty} W(s,t) \simeq F(s) + \text{const.} \tag{5.9}$$

The statistical convergence of MetD calculations is a key aspect for the accurate reconstruction of the FES, that depends critically upon the choice of CV and the available computer resources. Although expression (5.9) features an infinite limit, for practical purposes stationary conditions, if achieved, are observed after a finite time, when the free energy profile shows a parallel growth of the FES (flattening). This behavior occurs when fixed-height Gaussian hills of potential are deposited; on the other hand, it exists also the possibility to employ the well-tempered MetD variant [21]. In this approach, potential hills of progressively reduced heights are deposited when the simulation proceeds, becoming almost negligible once convergence is approaching.

In some cases, MetD is more effective for the rapid exploration of reaction pathways than for the efficient reconstruction of converged FES. Hence, a convenient strategy consists in using in tandem MetD and umbrella sampling (US) [22]: the latter, starting from configurations along transition pathways discovered with MetD, exploits multiple simulations, each exhaustively sampling a limited portion of configuration space. To this aim, restraining (static) bias potentials are employed, usually of parabolic form as a function of the CV. A global and accurate FES is finally reconstructed combining together the probability estimates of all trajectories with, e.g., the weighted histogram analysis [23–26].

Finally, the necessity of the a priori choice of a CV driving the reaction is the main limitation to the applicability of all free energy calculation techniques based on a bias potential, like MetD and US. In principle, to each reaction $A \rightleftharpoons B$ corresponds an ideal reaction coordinate defined as the committor function $P_{\text{B}}(x)$, i.e., the probability to reach B before A for an unbiased trajectory starting from x with velocities randomly drawn from a canonical distribution [19]. In practice, this huge amount of information is unavailable before (or even after) the simulation, and one is forced to postulate—more or less heuristically—one or a few CVs that attempt to capture the slow degrees of freedom contained in the ideal reaction coordinate. Moreover, since the volume of the CV space grows exponentially with dimensionality, reconstructing a FES as a function of more than 2 CV is typically too expensive.

5.2.3 A New Metric for Free Energy Calculations

When some information is available about the dynamical evolution of a given chemical reaction—i.e., its trajectory—the problem of the predetermination of a set of CV becomes easier. Branduardi et al. [27] introduced the so-called "path-CV", which are constructed from a tentative reaction path $\{\mathbf{R}_k\}_{k=1,\ldots,N_\mathrm{f}}$ connecting the reactants and the products states, according to the following definition:

$$s(t) = \frac{\sum_{k=1}^{N_\mathrm{f}} k e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_k)}}{\sum_{k'=1}^{N_\mathrm{f}} e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_{k'})}} \tag{5.10}$$

$$z(t) = -\frac{1}{\lambda} \log\left(\sum_{k=1}^{N_\mathrm{f}} e^{-\lambda D(\mathbf{R}(t),\mathbf{R}_k)}\right), \tag{5.11}$$

where s represents the progress along a pathway defined through an ordered sequence of atomic configurations $\mathbf{R}_1$, $\mathbf{R}_2$, …, $\mathbf{R}_{N_\mathrm{f}}$, while z is a measure of the distance from the pathway itself. This class of variables proved to be crucial to obtain free energy landscapes of a variety of processes including gas phase chemical reactions with concerted mechanisms [28–30], and transformations of carbon nanostructures [31]. The key ingredient of path collective variables, determining their effectiveness in a given problem, is the definition of distance $D(\mathbf{R}(t), \mathbf{R}_k)$ between the atomic configuration at time t and the kth reference structure. Pietrucci and Saitta [32] introduced thus the following metric:

$$D\left(\mathbf{R}(t), \mathbf{R}_k\right) = \sum_{\mathrm{IS}} \left(C_{\mathrm{IS}}(t) - C_{\mathrm{IS}}^k\right)^2 \tag{5.12}$$

where C_{IS} is the coordination number between atom I of species (element) S' and all atoms J of species S, defined by means of a smooth switching function:

$$C_{\mathrm{IS}}(t) = \sum_{J\in S} \left[1 - \left(R_{\mathrm{IJ}}(t)/R_{\mathrm{SS}'}^0\right)^N\right] \Big/ \left[1 - \left(R_{\mathrm{IJ}}(t)/R_{\mathrm{SS}'}^0\right)^M\right]. \tag{5.13}$$

The parameter $R_{\mathrm{SS}'}^0$ depends on the two species, since, e.g., a C–H bond is shorter than a C–C bond. We remark that, apart from the latter parameters, the sole input needed to construct the CV are the coordination patterns of the reactant and product species (and possibly other intermediates) as in Fig. 5.1. We performed and thoroughly tested MetD both in the fixed Gaussian height [32, 33] and in the well-tempered variants [21, 32, 33]. We estimated that reconstructed free energy surfaces have a statistical precision, typically, of about ±2 kcal/mol by comparing bias profiles at different times and across independent simulations.

A particularly convenient feature of the present MetD formulation is that the detailed geometry of N_f reference structures needs not to be explicitly given. As we

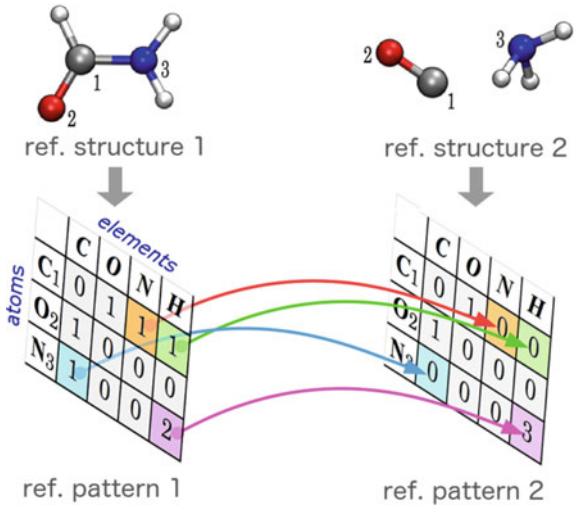


Fig. 5.1 Construction principle of topological path collective variables. The connectivity patterns of reactants and products are represented by tables having individual, non-hydrogen atoms on rows and atomic species (the set of all atoms of a given element) on columns. Arrows indicate changes of coordination numbers, however, all other matrix elements are free to change as well thanks to the flexibility of path collective variables (see text). Adapted from [32]

verified in several cases, it is sufficient to provide two reference states (the reactant and putative product) in the simple form of coordination matrices, and the simulation is able to discover transition pathways, including possible intermediates, as well as off-pathway states. The coordination patterns, in turn, can be directly written starting simply from structural formulas (Fig. 5.1). In practice, employing the $C_{\mathrm{IS}}^{\mathrm{k}}$ entries obtained from a short MD simulation of the end states can improve the resolution of intermediate and transition states in the free energy landscape. Another crucial feature of the new CV space is that it allows to simulate both gas phase and solution reactions (including possible interfaces with solid surfaces, etc.) within a unified formalism, as it will be clear from the examples in the following sections.

Note that coordination numbers have been repeatedly employed as an effective collective variable to study, e.g., (de)protonation reactions in water [34], since they are well suited to describe the active participation of the solvent. On the other hand, a typical reaction may require *a set* of several coordination numbers to be faithfully described, to track the creation/dissociation of different chemical bonds, with all the resulting difficulties in reconstructing high-dimensional free energy landscapes. Combining path-CV with a high-dimensional space of coordination numbers [32] allows exploiting the best of both worlds, providing a handy two-dimensional space able to track complex transformations, leaving the system the freedom to depart from the proposed path, and describing in a seamless way both gas- and condensed phase chemistry [32, 33].

5.2.4 Computational Details

As for the technical specific details of the simulations carried out in presence of progressively increasing electric field strengths and/or under the action of MetD, the interested reader can refer to the available original literature, e.g., [32, 33, 35–37]. For the sake of completeness, however, here we report on some of the technical aspects that unite such kind of apparently disparate investigations.

With the exception of the seminal work where a new MetD approach has been presented [32] in conjunction with typical Born–Oppenheimer Molecular Dynamics (BOMD), most part of the remainders works has been carried out by means of Car–Parrinello Molecular Dynamics (CPMD), sometimes along with the assistance of MetD in order to reconstruct the free energy landscape underlying a given process. In addition, since the statistical convergence of MetD calculations is sometimes cumbersome in AIMD calculations, we exploited for the first time the MetD approach presented in [32] coupled with Umbrella Sampling [22] in order to obtain well-converged and hence accurate FESs [33].

In general, the results were obtained using the density functional theory (DFT)-based software suite Quantum ESPRESSO [38], where the implementation of an external electric field is achieved by exploiting the modern theory of polarization and Berry's phases [8] (see, e.g., [7]). As usual, the structures were always replicated in space by using Periodic Boundary Conditions.

Before starting each simulation, equilibration runs of at least 5 ns were always performed by means of specific classical force fields in order to prepare suitable initial atomic configurations for AIMD. In general, when chemical reactions were observed to occur "spontaneously" in presence of intense electric fields, several tests have been performed in independent (parallel) additional CPMD and/or BOMD where the external electric field was abruptly switched off [33, 36, 37]. This way, rough but educative information have been gathered on the specific "stability" of each synthesized species.

As for the simulations where MetD calculations have been performed as the main calculation [32, 37] or as a coadjuvant to the results stemming from unbiased AIMD runs [33], MetD simulations have been performed by using the plugin Plumed-1.3 [39], where some of us [32] implemented our new approach (freely available on demand and soon in Plumed-1.3 and Plumed-2.0).

In all cases, the core electronic interaction has been depicted through ultra-soft pseudopotentials (USPP) generated via the Rappe-Rabe-Kaxiras-Joannopoulos (RRKJ) method [40]. As an approximation of the exchange and correlation functional, the Perdew–Burke–Ernzerhof (PBE) functional [41] has been adopted, which belongs to the generalized gradient approximation (GGA) but several tests and additional simulations have been also performed by exploiting the BLYP functional [42], sometimes in their dispersion-corrected version (i.e., by means of the Grimme D2 method) [43]. The dynamics of ions were simulated classically within a constant number, volume, and temperature (NVT) ensemble, using the Verlet algorithm and Nosé-Hoover or Bussi thermostats.

In order to characterize the transition states of the observed reactions, both from biased and unbiased calculations, committor analyses [44] have been performed [32, 33, 36, 37]. By choosing dozens of structures that were considered as plausible candidates for being the transition state of each simulation, more than 50 unbiased trajectories have been performed starting from each candidate and differing by the initial random velocities extracted from a Maxwell–Boltzmann distribution at the given temperature for which the chemical reaction was observed. A configuration has been identified as belonging to the transition state ensemble when it is committed to the reactants or products basin with a probability of $50 \pm 10\%$.

Finally, with the aim of tracing the bonds behavior of some of the chemical reactions here presented, also the Maximally Localized Wannier Functions (MLWF) [45] as well as their charge centers have been determined [33, 36, 37].

5.3 Chemical Reactions Under Intense Electric Fields

5.3.1 Ab Initio Miller Experiments

Origins of life studies represent an exciting and inherently multidisciplinary research field that incorporates contributions from biologists, chemists, geologists, physicists, mathematicians, chemists and computer scientists, among others. In 1953 in a crucial experiment, Miller reported [46] the surprising results he had achieved by the application of an electric discharge on a simple mixture of gases that simulated the atmosphere of the primordial Earth, based on the Oparin-Haldane hypothesis. This way, he had observed the spontaneous formation of aminoacids. By all means, this finding gave birth to a new multidisciplinary research field: prebiotic chemistry. Hence, the role played by electric fields in increasing the reactivity of atoms in matter had been already ascertained more than 60 years ago. However, the chemical reactions leading to the formation of aminoacids, as those occurring in Miller's experiments, have never been studied at the atomic level before the advent of the extremely powerful computing techniques that have become available in more recent times. Again, ab initio methods have proven to be particularly suited for this purpose: in 2014 the first in silico Miller experiments have been reported [35], showing for the first time the active role of molecules such as formic acid and formamide as direct precursors of the simplest aminoacid glycine.

A key aspect of the historical Miller experiments was the formation of hydrogen cyanide (HCN), aldehydes and ketones, and aminoacids, suggesting that the experimentally-observed products were obtained through a Strecker-Cyanohydrin reaction [47]. In the theoretical ab initio study [35], the authors chose to decompose the study of the full Strecker chemical reactions into more elementary steps, including at each one of them the products of the previous one. On this basis, Saitta and Saija set up three corresponding simulation boxes. In particular, the end products of the reaction were glycine (NH_2CH_2COOH) and NH_3 in equal proportions, while the

Table 5.1 Composition of the three Miller–Strecker simulation boxes, and average (over a 5 ps dynamics) potential energy, relative to the MS03, and expressed in kcal/mol per glycine + ammonia unit, (i.e. 2:2:2:8 O:C:N:H atoms). Adapted from [35]

System	H_2O	NH_3	CH_4	CO	N_2	H_2CO	HCN	glycine	ΔE
MS01	8	8	8	10	5	0	0	0	40.4
MS02	9	9	0	0	0	9	9	0	54.3
MS03	0	9	0	0	0	0	0	9	0.0

intermediate reactants were H_2O, NH_3, formaldehyde (H_2CO), and HCN, in equal proportions. In other words, both the intermediate and the end products had the same proportion 1:1:1:4 of O:C:N:H atoms. The authors thus built up a starting Miller's sample box (named MS01) containing a mixture of molecules similar to the most recent experiments performed by Miller and coworkers [48], and bearing the same atomic species ratios as the two other boxes, that is 8 H_2O, 8 NH_3, 8 CH_4, 10 CO, and 5 N_2 molecules, corresponding to 18:18:18:72 O:C:N:H atoms. Similarly, the intermediate Miller–Strecker (named MS02) box contained 9 H_2O, 9 NH_3, 9 H_2CO, and 9 HCN, while the end Miller–Strecker (named MS03) box contained 9 glycines, and 9 NH_3. This way, one could compare in a judicious manner the potential energies relative to the end products, (i.e., MS03), as shown in Table 5.1.

The first important result was that the MS02 system, containing the supposedly intermediate reactants, i.e., water, ammonia, formaldehyde, and hydrogen cyanide had a higher internal energy (enthalpy) than both the MS03 and, especially, the starting MS01 system. In other words, the formation of formaldehyde and HCN from the Miller molecules are thermodynamically unfavored, and thus very unlikely to occur in small simulation boxes over short picosecond timescales. Interestingly, while no chemical reactions were observed in MS01 and MS03, at the end of the zero-field evolution of MS02, the authors observed that a H_2CO and an ammonia molecule spontaneously combined to yield a formamide molecule ($HCONH_2$).

The MS01 system was then put under a finite electric field, in order to possibly observe the spontaneous formation of formaldehyde and/or HCN, and thus validate the suggested Strecker mechanism. However, once the electric field was switched on, up to 0.25 V/Å the system only exhibited occasional water/ammonia proton dissociations/jumps/recombinations. At fields around 0.35 V/Å, however, more interesting chemical reactions occurred. In particular, within 2.1 ps of a 10 ps-long trajectory, the field-induced formation of two formic acid molecules has been detected. Even more interestingly, at about 2.4 ps a formamide ion $HCONH_3^+$, later evolving into a neutral formamide molecule, formed via the simultaneous association of a Grotthuss proton and an ammonia molecule on the two available bonds of a carbon atom in a CO molecule, thus creating the simplest possible O–C–N backbone.

The authors noted that all those reactions, like the ones described in the following, were not simply due to the presence of reactive charged species in solution, and that the field is the indispensable driving force. This point was checked by running zero-field simulations of the same initial set of Miller molecules, but replacing all

water and ammonia molecules with their ionic OH^- and NH_4^+ counterparts [35]. Although the proportion of ionic species was thus much larger than any instantaneous one ever observed at finite-field, no reactions other than proton jumps until the full neutralization of all molecules occurred, thus proving that the presence of ionic species is not sufficient to promote the Miller reactions without an external field. In a way, this also confirmed that the electric field is not just a mere source of chemical energy, and that it plays an "order-maker" role that inherently favors the assembling of larger chemical units from smaller ones, and thus complexity.

On the basis of the above results and of the absence of Strecker intermediate products in the MS01 system and, on the contrary, of the spontaneous formation of formamide in the MS02 system even in the absence of field, the exploration of a different path to explain the Miller experiments was conducted. In particular, another system was conceived (Miller-Formamide (MF01)), based on the MS01 one, but containing from the start 8 molecules of each chemically relevant molecular species so far identified, i.e., formic acid, formamide, water, ammonia, and carbon monoxide, while the inert nitrogen and methane were discarded.

At the intense field strength of ~0.5 V/Å, the most interesting reactions occurred. In particular, all along the trajectory, the formation of formamide was observed through the combination of ammonia and carbon monoxide. Formamide then either formed larger and more complex molecules, or broke down into water and hydrogen cyanide, possibly after converting into its formimidic acid tautomer, as follows:

$$NH_3 + CO \rightleftharpoons O{=}CH{-}NH_2, \tag{5.14}$$

$$O{=}CH{-}NH_2 \rightleftharpoons H_2O + HCN. \tag{5.15}$$

At the same time, a similar yet less frequent reaction occurred between water/CO and formic acid:

$$HCOOH \rightleftharpoons H_2O + CO. \tag{5.16}$$

Typically, a carbon monoxide then combined with a formamide molecule, a complex which dissociated into a hydrogen cyanide and a formate anion, which combined with a just-formed formamide-proton cation to yield a α-hydroxyglycine, as shown in the following reaction:

$$[NH_2CHOH]^+ + COOH^- \rightharpoonup NH_2{-}CHOH{-}COOH. \tag{5.17}$$

Even more interestingly, within a short time (0.03 ps) the carbonylic oxygen bonded to the central α carbon atom, after losing a proton, moved (in the opposite direction with respect to the field) onto the carboxylic β carbon, to finally bond a proton moving along the field direction, and dissociate from the rest of the molecule, thus leaving behind a NH...=CH...−COOH molecule. This molecule is known as dehydroglycine, an α-aminoacid, which does not belong to the family of the 23 proteinogenic aminoacids, 20 of which are known as the "standard" aminoacids.

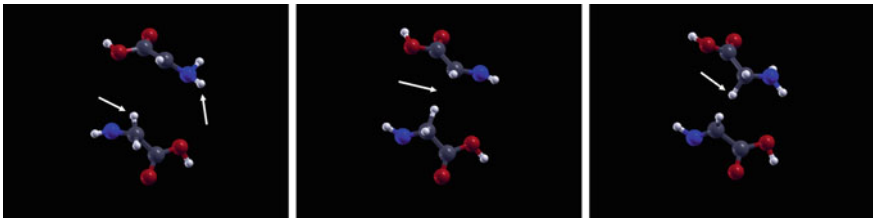


Fig. 5.2 Trajectory snapshots of the electric-field-driven evolution of dehydroglycine into glycine. At first, a H_2 molecule splits, the H atoms bonding with the α-carbon of one dehydroglycine molecule and with the nitrogen of a nearby one (left panel, shown by arrows). After about 90 fs the two α-carbon of both molecules approach (middle panel, indicated by the arrow), and the extra proton jumps from one of the two molecules onto the other one, yielding glycine (right panel, indicated by the arrow). Adapted from [35]

In order to attempt to find glycine at the end of the Miller chemical path, the authors first checked that the average energy of a simulation box containing glycine and water molecules (named Gly) was about 32 kcal/mol lower than the same box containing equal amounts of dehydroglycine, H_2, and water molecules (named DHGW). They set up then a simulation box containing 10 dehydroglycine, 10 H_2, and 10 formic acids molecules (labeled DHGF). At a field strength of $\sim$0.5 V/Å, within less than a picosecond one favorable event was recorded, namely the simultaneous attack of two hydrogen atoms on a α-carbon atom of one dehydroglycine, and a N atom of an adjacent one. In addition, within about 100 fs, the two α-carbons approached, and the hydrogen atom finally jumped from one to another, thus effectively yielding glycine, as shown in Fig. 5.2.

One of the main aspects emerged from this study is the revelation of a chemical pathway leading to glycine which is different from the standard interpretation of Miller experiments. Moreover, it proved that the electric field acts as an order-maker, promoting the spontaneous assembling of simple Miller molecules into more complex ones of biochemical interest.

5.3.2 Ab Initio Miller-Like Experiments and the Synthesis of Simple Sugars

One of the crucial aspects related to the origins of life and thus to the onset of informational polymers on early Earth—such as RNA—is represented by the formation of the sugar composing this latter macromolecule (i.e., ribose). However, the formation of the first C–C bonds from very simple molecules such as formaldehyde represents the rate-limiting step of the so-called formose reaction (i.e., the autocatalytic reactions leading to the synthesis of sugars from aldehydes). Only very recently, the free energy surface associated with such a process has been determined in condensed phase [37]. In fact, by means of AIMD and MetD techniques, the free

energy landscape underlying the synthesis of glycolaldehyde from a formaldehyde aqueous solution has been reported [37]. The latter work quantitatively explained why the very first step of the formose reaction is kinetically inhibited.

Moreover, in the same work [37], electric fields proved to be dramatically important in the complexification of the investigated systems. In fact, numerical samples of formaldehyde (both neat and in water solution) and of glycolaldehyde (both neat and in aqueous solution) were exposed to intense electric fields. The application of electrostatic gradients strongly prevented the formaldehyde umpolung and catalyzed the formation of C–O-bonded polymers in formaldehyde-containing samples. Notwithstanding this latter finding, when the field was applied on glycolaldehyde aqueous solutions, new C–C bonds have been formed and (D)-erythrose has been synthesized. This way, a numerical Miller-like experiment led to the formation of a prebiotically relevant (D)-tetrose—i.e., a direct precursor to ribose—from ubiquitarious molecules such as glycolaldehyde and water [37].

The just described field-driven chemical reactions are the top of the iceberg of the full chemical landscape that the application of a static electric field can produce when it is applied to a material. As an example, even in the methanol case, molecular dissociations are not the only chemical reactions which take place. Indeed, depending on the field strength, a progressive complexification of the system has been recently emphasized [33, 36], as it will be laid out in the following section.

5.3.3 Field-Induced Chemical Reactions in Energy-Related Research

The catalytic effects afforded by the application of intense electric fields on condensed matter are witnessed by the discovery of novel chemical routes also in the research for alternative "green" combustion fuels. In this respect, methanol, the simplest alcohol, and dimethyl ether (DME), the simplest ether, are central compounds. However, despite a massive amount of research in this field, the synthesis of DME from liquid methanol has just been recently reported [36]. In particular, the conversion of methanol to DME is traditionally achieved at high temperatures (i.e., in the gas phase) and in presence of very specific catalysts [49–52]. Of course, the importance of liquid methanol resides inter alia in the fact that it is preferable over highly volatile and potentially explosive materials (e.g., dihydrogen) for energy storage and transportation [53]. Based on unbiased AIMD, it has been indicated [36] a novel synthesis route to methanol dehydration—leading thus to the DME synthesis—through the application of strong electric fields at ambient temperature.

At relatively moderate field strengths (i.e., 0.30 V/Å) it is possible to induce the cleavage of the O–H covalent bond which results in the concept of generalization of *p*H in methanol [54]. This way, molecular dissociations are activated and protons H^+ can migrate via a Grotthuss-like mechanism through the H-bond network. This process is thus assisted by a certain fraction of ionic species, such as the methanol

cation $CH_3OH_2^+$ (methyloxonium) and anion CH_3O^- (methoxide) which are responsible for the ionic charge flow in the system and for enhanced contributions to the external electrostatic potential. The intensity of the local electric fields—due to the ions—can be of the order of 1 V/Å [55–57], which strongly increases the overall molecular reactivity. For external electric field intensities stronger than 0.50 V/Å, several field-induced chemical reactions have been observed. Indeed, a field strength of 0.55 V/Å is able to induce the formation of formaldehyde molecules, as well as of methane and water, according to the following reaction [33].

$$2CH_3OH \xrightarrow{\mathbf{E}} H_2CO + CH_4 + H_2O\,. \tag{5.18}$$

Formaldehyde, being an extremely reactive compound, gives rise to a progressive complexification of the system. First, due to the water production, for field intensities stronger than 0.55 V/Å formaldehyde hydration takes place as in the following

$$H_2CO + H_2O \xrightarrow{\mathbf{E}} CH_2(OH)_2\,, \tag{5.19}$$

and formaldehyde monohydrate is transiently observed.

More interestingly, at the same field strengths, methanol dehydration also occurs leading to the synthesis of DME according to:

$$2CH_3OH \xrightarrow{\mathbf{E}} CH_3OCH_3 + H_2O\,. \tag{5.20}$$

As shown in Fig. 5.3, the involved field intensities are able to open several reaction channels that connect the simplest alcohol to the simplest aldehyde on one hand and, on the other, to the simplest ether. The most direct chemical transformations that join methanol with formaldehyde (reaction (5.18)) and, separately, with DME (reaction (5.20)), involve the cation $CH_3OH_2^+$ and the anion CH_3O^- which act in both cases as intermediate species. In addition, the produced formaldehyde can either undergo hydration or yield DME, which represents the most abundant compound that has been synthesized in the system, as displayed in Table 5.2. Indeed, for field strengths of 0.60 and 0.65 V/Å, the transiently formed formaldehyde molecules have been readily employed for the synthesis of DME and formaldehyde monohydrate, yielding the molecular fractions shown in Table 5.2. For higher field strengths—0.70 and 0.75 V/Å—the system is temporarily characterized by an elevated degree of mixing. However, after few ps of dynamics performed at these field intensities, only water, DME, and methane species—in order of decreasing amount—are observed. Instead, formaldehyde, the most reactive species among the neutral ones, ultimately contributes to the process of molecular complexification of the system. In the end, DME is the most abundant (nonaqueous) synthesized species which has been accumulated along the simulation. Incidentally, as shown in Fig. 5.3, the basin of DME in the field-induced reaction network of methanol is the only one that can be reached by following two different chemical pathways, and the only one displaying exclusively incoming "chemical fluxes" (i.e., DME is chemically inert).

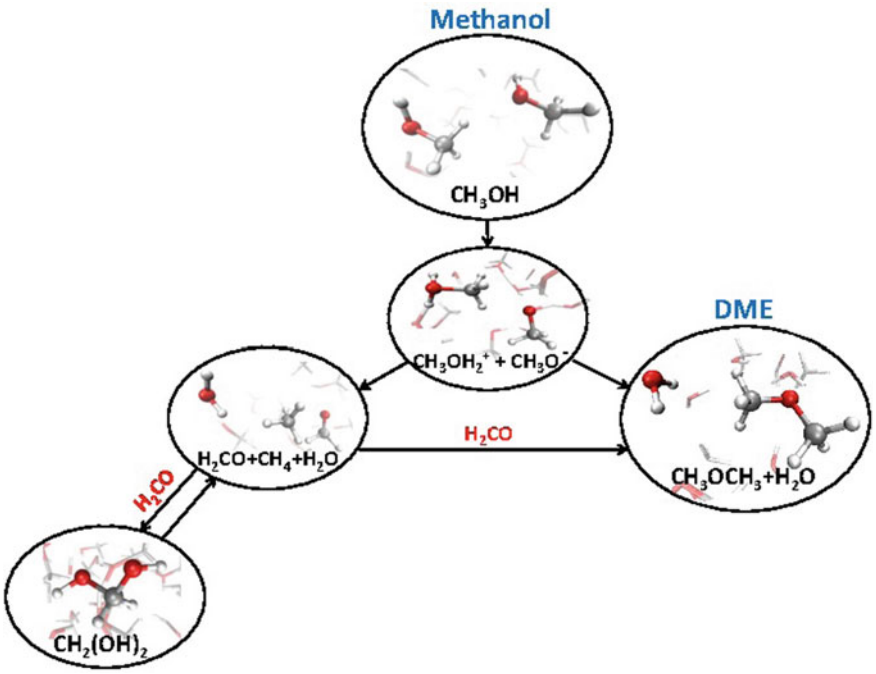


Fig. 5.3 Methanol reaction network in presence of a static electric field. The application of this latter dissociates some methanol molecules (second panel from the top). Strengths above 0.50 V/Å are able to recombine the formed methanol cations and anions both into formaldehyde, methane, and water, both into DME and water (third panel from the top). Formaldehyde molecules will be further employed in order to synthesize also formaldehyde monohydrate (which has a strong tendency to dissociate) and DME. Adapted from [36]

Table 5.2 Total populations (percent fractions) of the molecular species present in the system at the end of each simulation carried out at field strengths of 0.55, 0.65, and 0.75 V/Å. The fractions are determined as the ratio between the number of molecules of a given species and the number of total methanol molecules composing the original sample (i.e., 32). The lack of formaldehyde above the field threshold which is able to create it (i.e., 0.55 V/Å) is due to its extensive employment in diverse chemical reactions leading both to formaldehyde hydration both to DME synthesis. Formaldehyde acts as a reactive intermediate species which contributes to the complexification of the system

Molecular species	0.55 V/Å	0.65 V/Å	0.75 V/Å
Formaldehyde	1 (3%)	0	0
Methane	1 (3%)	2 (6%)	2 (6%)
Water	1 (3%)	3 (9%)	6 (19%)
Formaldehyde monohydrate	0	1 (3%)	0
Dimethyl ether (DME)	0	2 (6%)	4 (13%)

As a comparative mention on the involved field strengths, air ionizes at a field strength of "only" $3 \cdot 10^{-4}$ V/Å. Albeit field strengths of the order of 0.50−0.75 V/Å are clearly of extreme intensity from the macroscopic point of view, it is nowadays well established that field intensities of (or greater than) 1 V/Å are locally (and naturally!) found at the microscopic scale in disparate condensed systems in presence of simple solvated ionic species [55–57], or even at the proximity of the surface of clean apolar oxides such as MgO(001) [58]. In field emitter tips experiments, field strengths of 1−3 V/Å are recorded [59–61] and it has been proven that intensities of 0.30 V/Å are necessary in order to induce water dissociation [62–64] and, in general, to significantly shift the bonding electrons [65]. Very recently, fields within this order of magnitude have been again experimentally detected at the tip proximity [66]. All these evidences strongly suggest the experimental feasibility of the proposed reactions, by exploiting the high field capability of, e.g., field emitter tips.

5.3.4 The Role of Solvation in Presence of an Electric Field

As shown above, field strengths of 0.60 V/Å and higher lead to the synthesis of DME. The processes synthesizing new species in methanol are always characterized by instantaneous local molecular frustrations—with respect to the field direction—of the molecular/ionic configurations in the numerical samples [36]. Indeed, as shown in Fig. 5.4 (bottom panel), neither methoxide—which has its dipolar axis perpendicular to the field direction—nor methyloxonium—which, as the result of a proton transfer, is oriented with the excess proton in opposition to the field—are arranged as expected for free dipoles in an electrostatic field. In addition, all the mechanisms giving rise to the onset of DME species from pure liquid methanol clearly show that the local environment plays a major role in assisting the chemical reactions by acting inter alia as a sort of *reservoir* of proton H^+ and hydride H^- acceptor/donor sites.

The main intermediate state of the DME synthesis and the basic mechanism of formation of formaldehyde are characterized by the presence of the two counterions of methanol (Fig. 5.4). However, it is clear that two very different reaction channels can be undertaken by the system depending on the mutual orientation of nearby methyloxonium and methoxide ions. At intense field strengths, the head-to-tail arrangement $^{+}H_2O–CH_3 \cdots {}^{-}O–CH_3$ leads to DME and water formation (i.e., reaction (5.20)). On the other hand, if the adjacent parts of the methanol counterions are the respective methyl groups, the reaction evolves towards the synthesis of formaldehyde, methane, and water (i.e., reaction (5.18)) [33]. These two reaction channels are shown in Fig. 5.4. By definition, the identification of a transition state splitting the probability to reach reactants or products imply the existence of a residual free energy barrier even under the intense electric field triggering the reactions. It is known, however, that condensed phase reactions involving ionic states (e.g., proton transfer) may be described in terms of dynamical barriers, that are modulated in time by the fluctuations of the environment-induced electrostatic field [67, 68].

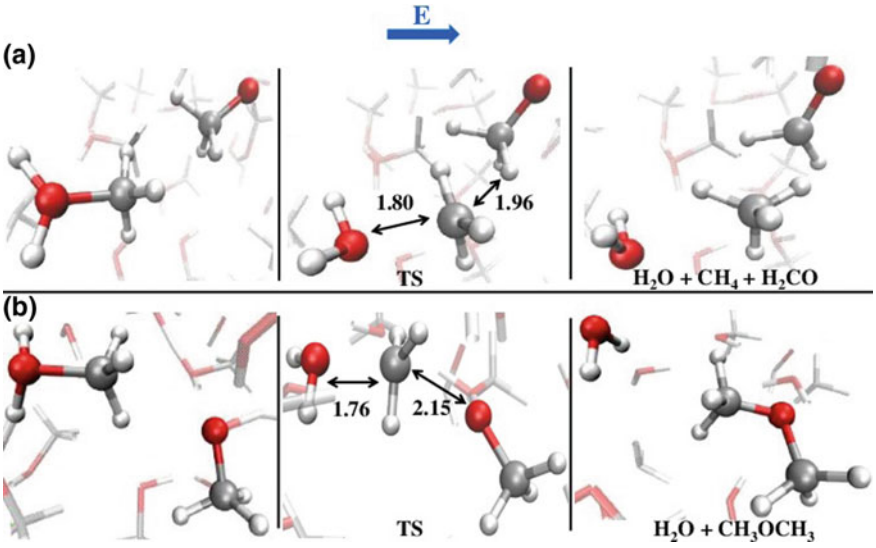


Fig. 5.4 A twofold reaction channel. Depending on the mutual orientation of methoxide CH_3O^- and methyloxonium $CH_3OH_2^+$, the reaction proceeds yielding either formaldehyde, methane, and water (**a**) or DME and water (**b**). In the central panels, the transition states of both reactions, obtained by means of a committor analysis performed at 300 K and in presence of field strengths of 0.55 V/Å (**a**) and 0.60 V/Å (**b**), are shown. Adapted from [36]

Since the observed reaction mechanisms are intimately related to the multifaceted and collective character of the medium, it is to be expected that the chemical pathways undertaken by the system in condensed phase are entirely different from any reproduced gas phase reactions. Indeed, by performing additional simulations starting from the ionic intermediate configurations shown in the left panels of Fig. 5.4 but placed in the gas phase—i.e., by removing the solvent—, it has been observed that both reactions (5.18) and (5.20) are barrierless and spontaneously proceed even in absence of any electrical perturbation [36].

Finally, the authors demonstrated that if, on one hand, the local solvation environment hampers highly polarized—and thus reactive—molecular states, on the other hand it plays a key role in shaping the final steps of a reaction, orienting it towards specific channels and hence products, as also shown in Fig. 5.4.

Hence, the field plays a crucial role as it affects the local ionic concentration of a solution and, as it is well-known, it can create and stabilize ionic transition states [65]. Moreover, as shown in the central panels of Fig. 5.4, the transition states characterizing the main reactions here presented are resemblant to the reactants configurations, proving (e.g., following transition state theory) the decisive energetic contribution carried by the application of the field (i.e., 30−35 kcal/mol [33]) and its own natural consequences (i.e., molecular ionizations). Nevertheless, the fate of a given reaction is dictated by local environment circumstances. In fact, once

strong local electrostatic contributions are generated by the field action, molecular correlations are decisive in pushing a given intermediate state into a specific reaction product (Fig. 5.4).

5.4 Ab Initio Free Energy Methods for Chemical Reactions in Solution

5.4.1 Free Energy Landscapes of Prebiotic Chemical Reactions

Understanding the thermodynamics (free energy differences) and kinetics implied in reactions in solution is a crucial issue in chemistry. For example, it is key to assess the likelihood of the different prebiotic scenarios put forward in the literature, and mentioned in the previous section. Computational approaches are a formidable complement to experiments in this field, as they exploit the fundamental laws of quantum mechanics to study chemical reactions, interpret experimental results and predict novel mechanisms.

However, contemporary computational physical chemistry is dominated by gas phase calculations at zero temperature, with effects due to temperature, pressure and chemical environment relegated to approximated extrapolations. For instance, our knowledge of reaction dynamics in condensed phases is far from complete [69, 70], despite the fact that water is a polyvalent molecule, known to participate also in formamide chemistry under different roles, including as a stabilizer through hydrogen bonds, as an efficient acid–base bifunctional catalyst, and as a co-reactant [71]. Additional effects should also be considered, including vibrational energy dissipation upon birth of exothermic products, or solute trapping into finite-lifetime cages affecting its diffusion and reactivity [70, 72, 73]. The large number of possible configurations (already including few water molecules [71]) together with the strong anharmonicity of liquids naturally calls for methods like MD that include from the start the finite-temperature dynamics. Particularly in a prebiotic perspective, it is necessary a comparative understanding of reaction networks in different environments (gas or condensed phase, with different solvents and also interfaces with minerals) and at different conditions (T, P, irradiation, shock waves, etc.), eventually embracing also nonequilibrium scenarios, for their role in the emergence of life.

The overwhelming gap between the (long) time scale of reactive events and the (short) time scale of ab initio MD simulations can be effectively overcome employing the available enhanced sampling techniques, including MetD [20] and transition path sampling [44]. Despite of this, the widespread adoption of MD simulations in the study of chemical reactions has so far been slowed down by the lack of standard, general purpose formulations of reaction coordinates. In particular, it is challenging to design coordinates that fully include the important role of the solvent degrees of freedom [74–76] and that are general enough to be applied to a palette of diverse reaction mechanisms.

As laid out in the "Methods" section, novel free energy calculation approaches, able to address in a general way a wide range of chemical reaction mechanisms in solution, are now available, allowing to unveil reaction networks of remarkable complexity. As an example, this has been applied to the study of the formamide decomposition channels in aqueous solutions [32]. In this study, the authors investigated, in a first set of simulations, the emblematic decarbonylation reactions and their inverse in formamide:

$$\mathrm{HCONH_2} \rightleftharpoons \mathrm{CO} + \mathrm{NH_3}\,. \tag{5.21}$$

Within a single gas phase simulation of 170 ps the trajectory explored 7 forward and 7 backward reactions. In the resulting free energy landscape (Fig. 5.5, top panel) two basins corresponding to reactants and products were readily identified, differing by less than 1 kcal/mol in stability. Reactants and products were separated by a free energy barrier $\Delta F^* = 79$ kcal/mol, very similar to the barrier of 80.5 kcal/mol found in zero-temperature calculations at CCSD(T)/CBS+ZPE level in [77].

Compared to gas phase, the landscape of the solution resulted significantly modified in both quantitative and qualitative features (Fig. 5.5, bottom panel). In particular, besides the reference states used to build the collective variables (formamide and $\mathrm{CO} + \mathrm{NH_3}$), the simulation explored a third basin featuring formic acid plus ammonia. This finding was remarkable since the chosen space of collective variables was not explicitly constructed to take into account formic acid, as shown from the higher values of the z coordinate with respect to the reference species. Formamide, $\mathrm{CO} + \mathrm{NH_3}$ and $\mathrm{HCOOH} + \mathrm{NH_3}$ showed a similar stability within our statistical uncertainty (± 2 kcal/mol). Inspection of the landscape resulted also in a barrier of 35 kcal/mol for the reaction $\mathrm{HCONH_2} \rightarrow \mathrm{HCOOH} + \mathrm{NH_3}$ and of 40 kcal/mol for the reaction $\mathrm{HCONH_2} \rightarrow \mathrm{CO} + \mathrm{NH_3}$.

The two decomposition channels, decarbonylation and hydrolysis, have thus quite similar barriers, confirming the effectiveness of topological path collective variables in discovering competitive pathways and even relevant chemical species. Additionally, ionic forms $\mathrm{NH_4^+}$ and $\mathrm{HCOO^-}$ were also explored, as expected from the basicity of ammonia and acidity of formic acid. An advantage of this MD approach is to have built-in anharmonic temperature effects, so that one can directly break down ΔF into energetic and entropic contributions: ΔE is directly obtained as the average energy observed in a relatively long equilibrium (unbiased) MD simulation performed in each metastable state, whereas $T\Delta S = \Delta E - \Delta F$.

More in detail, all observed reactions began with water donating a proton to the amino group of formamide as the first step (leaving an $\mathrm{OH^-}$ species in solution), overcoming a barrier $\Delta F^* \approx 35$ kcal/mol. The resulting cation, that was identified as a relevant intermediate for this same reaction in [35], is locally stable only within a time scale of less than 1 ps, as verified with hundreds of unbiased trajectories started from 10 different configurations featuring this species. Experimentally, formamide is a weak base, with the protonated species having a standard free energy almost equivalent (only 0.1 kcal/mol higher) to the neutral species [78], however, the carbonyl is expected to be the most probable protonation site [79], in agreement with

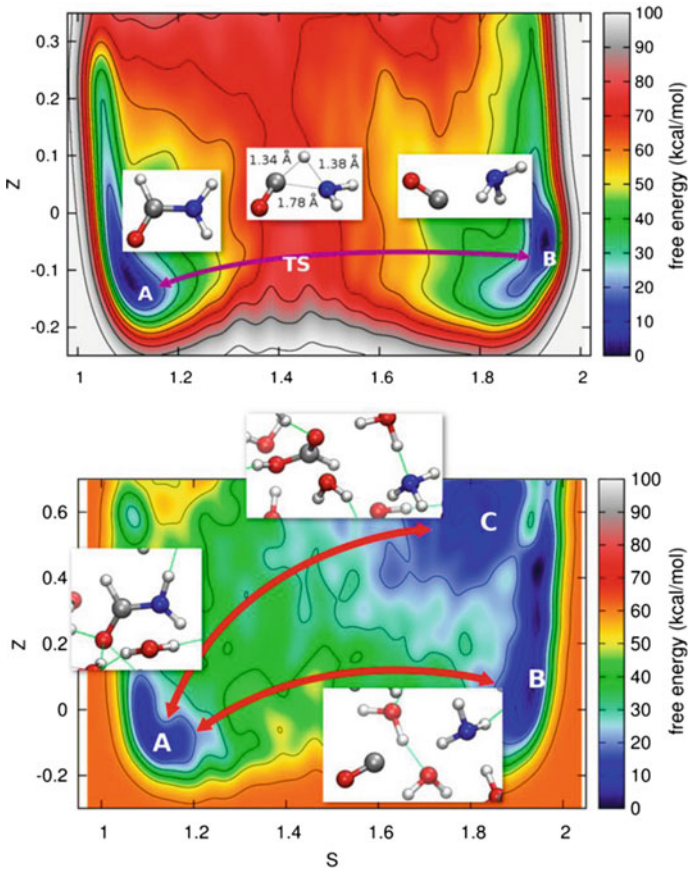


Fig. 5.5 Top panel: free energy lansdcape for interconversion of formamide (**a**) and CO + NH_3 (**b**) in the gas phase. Representative atomic configurations of free energy minima and transition states are shown as insets. Bottom panel: free energy lansdcape for reactions in aqueous solution between formamide (**a**), CO + NH_3 (**b**), and HCOOH + NH_3 (**c**). Representative atomic configurations are shown as insets. Adapted from [32]

the transient nature of the protonated nitrogen in these simulations. Due to the transient nature and very short lifetime of the $HCONH_3^+$ intermediate, the mechanism could also be approximately considered as one-step. From $HCONH_3^+$ the system can evolve either towards carbon monoxide and ammonia or towards formic acid and ammonia: the authors focused there on the former channel.

The dissociation of the C-N bond was accompanied by the donation of the carbon-bound proton to water. The mechanism appeared fully reversible in these simulations, with backward transitions (formamide formation) featuring a barrier of 40 kcal/mol. The barrier can be flattened by a strong enough electric field as demonstrated in [35]. The mechanism was validated and the transition state structures assigned by performing a committor analysis.

This study showed that a step forward in the study of chemical reactivity, in neutral conditions or under intense electric fields, treating gas phase and liquid phase on the same footing, is now possible through the definition of simple, intuitive, and transferable reaction coordinates (see also the "Methods" section). In combination with state-of-the-art free energy calculation methods such as MetD, the new coordinates allow to discover relevant reaction mechanisms and reconstruct the corresponding free energy landscapes, which can be directly compared between phases and/or at different thermodynamical conditions. Finally, detailed inspection of trajectories and committor analysis allow to compellingly identify transition states and reaction pathways and mechanisms.

This new approach was applied to the emblematic case of formamide, a centerpiece of many prebiotic scenarios recently put forward to explain the chemical origins of life on Earth, and observed in the ab initio simulation of Miller experiment under electric field [35]. On one hand, this method is capable of quantitatively reproducing existing gas phase results. Much more importantly, liquid phase calculations of all three fundamental reaction channels, i.e., dehydration, decarbonylation, and hydrolysis, provided novel results of prebiotic significance and implications.

5.4.2 Free Energy of Electric-Field-Driven Chemical Reactions

Although we have already mentioned somehow *en passant* the direct conversion of methanol to formaldehyde, the one-pot synthesis of formaldehyde and methane stemming from methanol has been just very recently indicated for the first time [33]. In particular, based on AIMD and free energy methods, the simultaneous oxidation and reduction (i.e., the disproportionation) of liquid methanol into methane and formaldehyde has been obtained at ambient temperature through the application of static electric fields. It has been proven—by means of advanced MetD techniques [32]—that the chemical pathway connecting methanol to the detected products in the bulk liquid phase is very different from its reproduced gas phase counterpart. Moreover, with the same techniques, it has been demonstrated that switching on an external electric field drastically modifies the reaction network of methanol, lowering some activation barriers, stabilizing some reaction products, and opening otherwise difficult-to-achieve chemical routes.

In order to assess the energetic contribution of the field to the reaction and the field-induced changes in the thermodynamic paths undertaken by the newly revealed reaction, our own recent method [32] of unbiased exploration of chemical reaction networks has been employed, capable of revealing unexpected pathways and microscopic mechanisms and, at the same time to provide the free energy landscape of those reactions, fully including the effect of the explicit solvent and of the thermodynamic conditions. This way, the real energetic contribution carried by intense

external electric fields in assisting a simple chemical reaction has been estimated to be 30−35 kcal/mol [33] at finite-temperature (i.e., 300 K).

As laid out before, if in neat water, at ambient conditions, fields of ∼0.50 V/Å can give rise only to an improved proton transfer [80], the reactivity of carbon atoms in methanol leads to more complex chemical transformations. In fact, it has been observed that a field strength equal to 0.55 V/Å, assisted by local molecular advantageous circumstances (i.e., nearest neighbors arrangements), is able to break one of the covalent bonds in some methyl groups of methoxide molecules and to create formaldehyde. The formation of this species is naturally accompanied by the release of water and methane (see reaction (5.18)).

In condensed phase, at a given instant, several molecules are arranged in such a way that the most electronegative "pole" is oriented towards the field direction (e.g., see the methoxide CH_3O^- in Fig. 5.6a). Moreover, also as a consequence of the activated proton transfer, some methyloxonium cations may be oriented with the excess proton in opposition to the electric field direction, as shown in Fig. 5.6a, e. In particular, during the proton migration process, an oxygen to oxygen intermolecular distance of ∼2.3 Å with an evenly shared proton can be transiently observed. This represents the formation of a transient Zundel-like ion [54], shown in Fig. 5.6a, e for two field intensities (i.e., the $[C_2H_6O_2H_3]^+$ complex), that readily leads to the release of a proton from one donor methanol molecule to an acceptor one. This way, the typical CO bond length of the just formed methyloxonium cation (Fig. 5.6b, f) becomes sensibly higher than the one characterizing the neutral molecular state in which this species lies before accepting the proton from the solvent. The stretch of the CO bond, if assisted by the local presence of a methoxide anion with its methyl group as first neighbor, leads to a visible decrease of the CO bond strength (see Fig. 5.6b/c and f/g). Indeed in few dozens of fs the CO covalent bond of $CH_3OH_2^+$ is broken whereas the distance between the carbon atom of the newly formed methenium ion CH_3^+ (see Fig. 5.6c, g) and the closest methyl hydrogen atom of CH_3O^- readily approaches to a very short value. The final step of this concerted reaction is represented by the ultrafast release of a hydride H^- stemming from the methoxide anion which recombines with the methenium just after the "umbrella inversion" of the latter (Fig. 5.6d, h). The transformation of methoxide into formaldehyde is characterized by a drastic reduction of the relative CO bond length (i.e., ∼1.2 Å) which is the manifestation of its peculiar double bond formation. The mechanistic pictures described in Fig. 5.6 are in practice stackable and a recurrent pattern can be recognized in the whole process: the proton transfer that triggers the reaction and the subsequent electrostatic approach of the two counterions which are arranged in a specific orientation with respect to each other and with respect to the field.

5.4.2.1 Free Energy Landscapes in the Liquid Phase

Despite the importance of elucidating the microscopic process that characterizes the above discussed chemical reaction (reaction (5.18)), more fundamental thermodynamic considerations are in order, beyond its mechanistic description. Enhanced

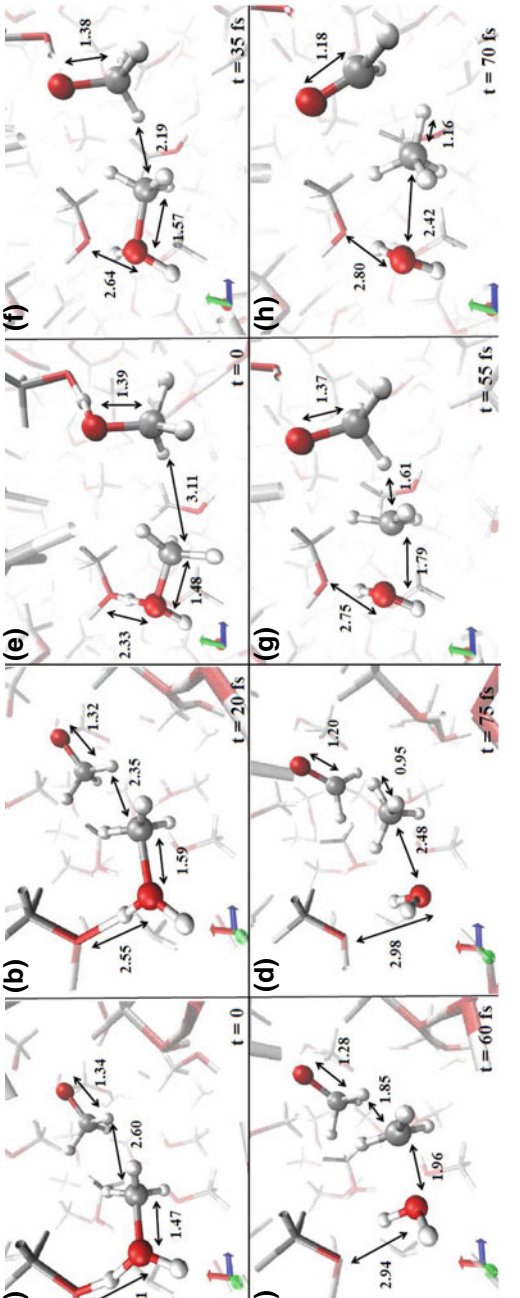


Fig. 5.6 Formaldehyde and methane formation mechanism in the presence of a static electric field oriented along the positive *z*-axis (i.e., the blue cartesian axis direction) with strengths of 0.55 V/Å (left panel) and 0.60 V/Å (right panel). Red, silver, and white coloring refers to oxygen, carbon, and hydrogen atoms, respectively. Four typical distances (shown in Å) have been chosen in order to better follow the reaction: the oxygen to oxygen separation between one solvent molecule and the (future, in panel a) $CH_3OH_2^+$, the CO bond length of the latter, the distance between the carbon of methyloxonium and the closest methyl hydrogen atom of CH_3O^-, and the CO bond length of the latter. A proton transfer event (**a**, **e**) leads to the elongation of the CO bond of the newly formed methyloxonium cation (**b**, **f**), i.e., from 1.47 to 1.59 Å (left panel) and from 1.48 to 1.57 Å (right panel). Simultaneously, electrostatic effects further shorten the CH distance between the carbon of $CH_3OH_2^+$ and its closest extramolecular hydrogen (**b**, **f**). This way the methyloxonium species undergoes the cleavage of its characteristic CO bond (i.e., CO distance of 1.96 Å (left panel) and 1.79 Å (right panel)) and a methenium ion CH_3^+ is thus transiently formed (**c**, **g**). The ultrafast recombination of the latter with a hydride H^- stemming from the methoxide ion leads to the formation of water, methane, and formaldehyde (**d**, **h**). Adapted from [33]

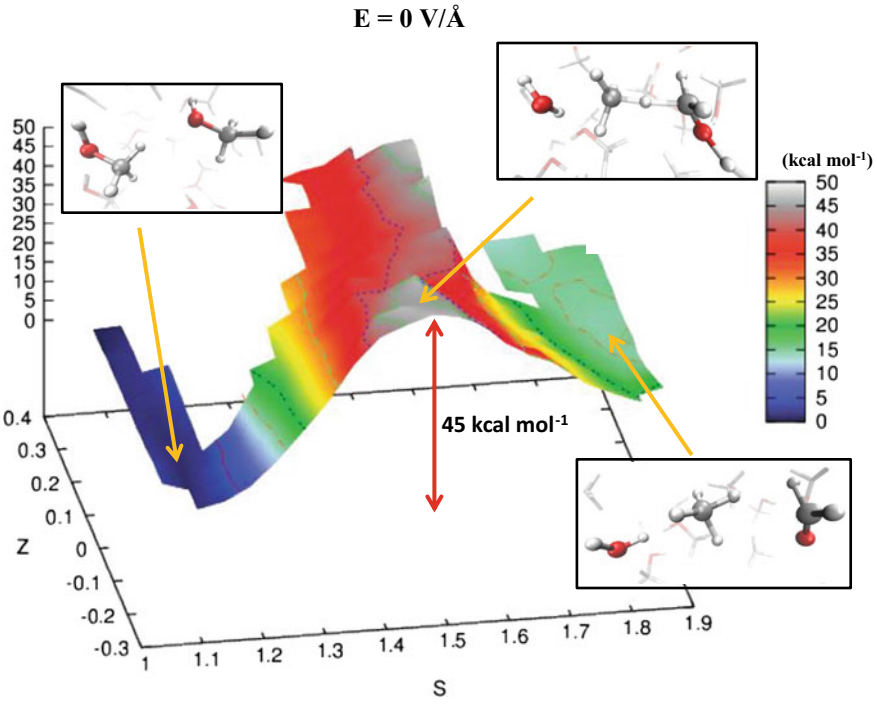


Fig. 5.7 FES of reaction (5.18) in the zero-field regime. The energy scale (depth) is in kcal mol^{-1} whereas the S-axis and the Z-axis represent the progress along the reaction and a sort of distance from its ideal path, respectively. Low values of S characterize a system of pure liquid methanol whereas high values of this parameter describe a sample composed by a formaldehyde, a methane, and a water molecule in a bath of methanol molecules. In the picture frames are shown the local structures of reactants, transition state, and products corresponding to their relative location on the FES in the space spanned by the CV. Adapted from [33]

sampling techniques [20] allow for the evaluation of the FES of disparate processes and, recently, a path collective variables (path-CV) MetD method has been developed [32] which is particularly useful for condensed phase reactions. Within this technique, the portion of the phase space close to the reactants (i.e., 2 neutral methanol molecules in a bath of methanol) and the products (i.e., a formaldehyde, a methane, and a water molecule surrounded by methanol molecules) basins have been at first explored and consequently sampled in solution and in the gas phase. This way, the FES of reaction (5.18) has been reconstructed in the zero-field regime and under the action of a field strength of 0.30 V/Å. The choice of sampling the system under a lower field intensity than that which renders barrierless the reaction (0.55 V/Å) is dictated by the necessity of monitoring and quantifying the field-induced changes on the free energy landscape for field intensities that are not large enough to induce the barrierless formation of formaldehyde, methane, and water. In Fig. 5.7, the free energy landscape of reaction (5.18) in the absence of the electric field is shown. The

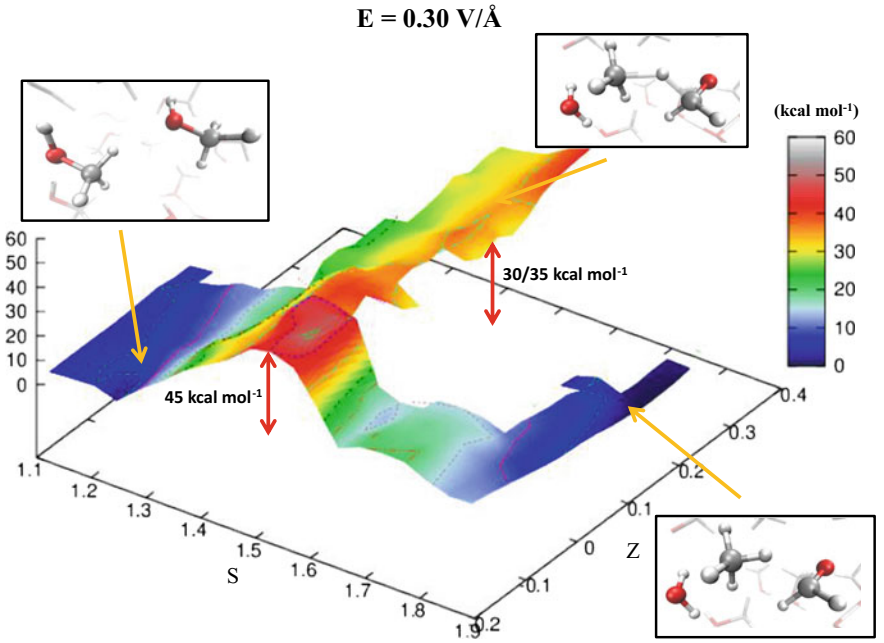


Fig. 5.8 FES of reaction (5.18) under a relatively moderate field strength of 0.30 V/Å. The energy scale (depth) is in kcal mol^{-1} whereas the S-axis and the Z-axis represent the progress along the reaction and a sort of distance from its ideal path, respectively. Low values of S characterize a system of pure liquid methanol whereas high values of this parameter describe a sample composed by a formaldehyde, a methane, and a water molecule in a bath of methanol molecules and its ionic equivalents. In the picture frames are shown the local structures of reactants, transition state, and products corresponding to their relative location on the FES in the space spanned by the CV. For $(S; Z) \sim (1.45; 0.3 - 0.4)$, where is located the observed transition state in presence of a field of 0.55 V/Å, the free energy is $\sim$30−35 kcal mol^{-1}. This value is not negligibly lower than that observed for the fieldless case (i.e., $\sim$45 kcal mol^{-1}) (see Fig. 5.7). Adapted from [33]

chemical transformation is inhibited by the presence of a high free energy barrier of 45 kcal mol^{-1}. The catalysis effects of the external electric field are manifest from the induced changes on the FES of reaction (5.18), as shown in Fig. 5.8. Even in presence of a field strength lower than that capable to render barrierless the reaction, the products are stabilized: a comparable free energy estimate characterizes the initial and the final states. Another important aspect concerns the field-induced changes in the transition region ($S \sim 1.45$). By comparing the FES evaluated in the zero-field regime (Fig. 5.7) with that in the presence of a field strength of 0.30 V/Å (Fig. 5.8), an evident modification of the shape can be recognized. In fact, a field strength that is roughly a half (0.30 V/Å) of that which is able to overcome the reaction barrier (0.55 V/Å), definitely lowers a portion of the barrier of approximately a third with respect to that of the zero-field case, i.e., $\sim$30 kcal mol^{-1} versus $\sim$45 kcal mol^{-1}, respectively. This finding is not so surprising if one takes into account that the

solvent is fundamental in assisting this kind of reaction. Indeed, the field strength of 0.30 V/Å represents the known methanol dissociation threshold [54] and the formation of ionic species such as $CH_3OH_2^+$ and CH_3O^- produces intense local fields, as it is usual when dealing with solvated ions [55–57]. Moreover, the region of the free energy barrier that is significantly modified by the field action corresponds to the effectively recorded passage in our unbiased AIMD simulation (i.e., it is the natural reaction pathway when a strong electric field is applied). Hence the external electric field is able to open reaction pathways that were avoided in the zero-field regime.

5.4.2.2 Free Energy Landscapes in the Gas Phase

A gas phase variant of reaction (5.18) has been also investigated by exploiting BOMD joined with the previously described MetD protocol. First, a well-tempered path-CV MetD approach [32] has been carried out.

Generally speaking, the interactions that characterize a bimolecular reaction in condensed phase are ultimately different with respect to the gas phase counter-

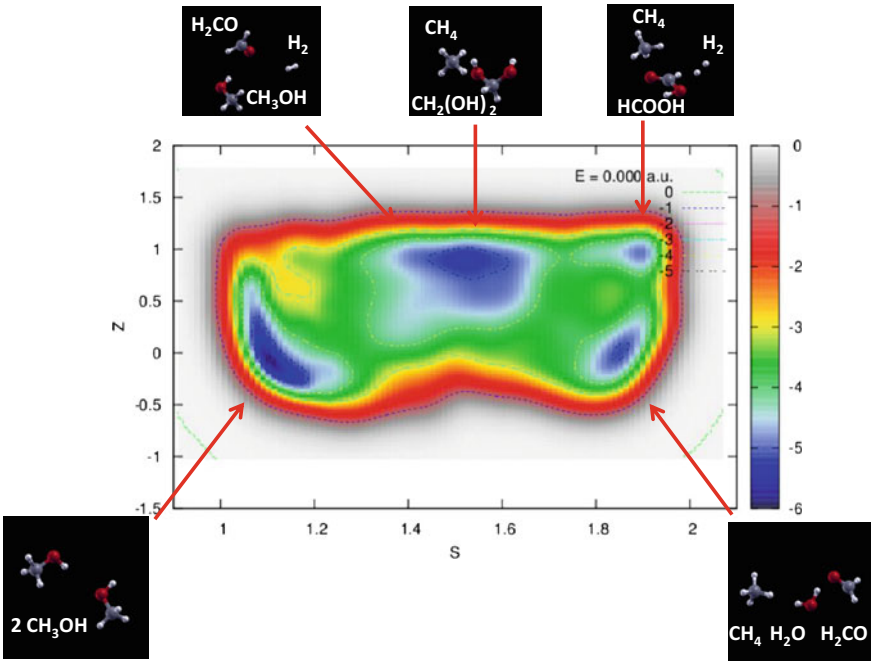


Fig. 5.9 "Exploratory" (not converged) FES projected onto the path-CV space of reaction (5.18) performed in the gas phase and in the zero-field regime. Three basins separate the reactants and the products. In ascending order of *S* (progress along the reaction path) we have: (1) reactants: two methanol molecules; (2) a methanol, a formaldehyde, and a hydrogen molecule; (3) a formaldehyde monohydrate and a methane molecule; (4) a methane, a hydrogen, and a formic acid molecule; (5) products: a formaldehdye, a methane, and a water molecule. Adapted from [33]

part [81]. Indeed the molecules participating in a liquid phase reaction experience solute–solvent interactions which occur in dozens of femtoseconds [70]. Nevertheless, the mere study of the features characterizing the field-induced changes in the gas phase reactions is per se an interesting work station. The first interesting evidence of the application of an external electric field in the gas is the induced manifest enhancement of the selectivity of the reaction. Indeed in absence of the field, an exploratory MetD simulation displays the presence of at least three different intermediate states between the reactants and the products basins. By following the reaction progress (i.e., going in ascending order of the parameter S) the explored metastable basins that the sample explores are associated with formaldehyde, dihydrogen, and methanol at first, then with methane and formaldehyde monohydrate and, finally, with methane, dihydrogen, and formic acid (Fig. 5.9).

When a field intensity of 0.25 V/Å is applied, only one of these three basins survives, as shown in Fig. 5.10. Hence the application of an external electric field to a gas phase methanol sample induces an enhancement of the selective character of the reaction.

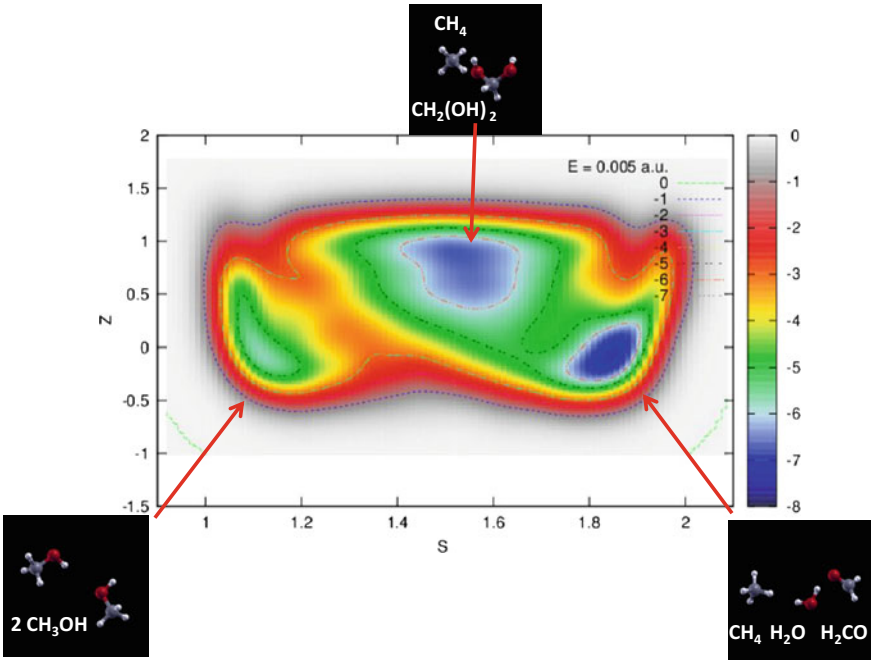


Fig. 5.10 "Exploratory" (not converged) FES of reaction (5.18) performed in the gas phase under the effect of a field strength of 0.25 V/Å. Two of the three basins previously observed in the absence of the electrical perturbation are not accessible under such a circumstance. In ascending order of S we have: (1) reactants: two methanol molecules; (2) a formaldehyde monohydrate and a methane molecule; (3) products: a formaldehdye, a methane, and a water molecule. Adapted from [33]

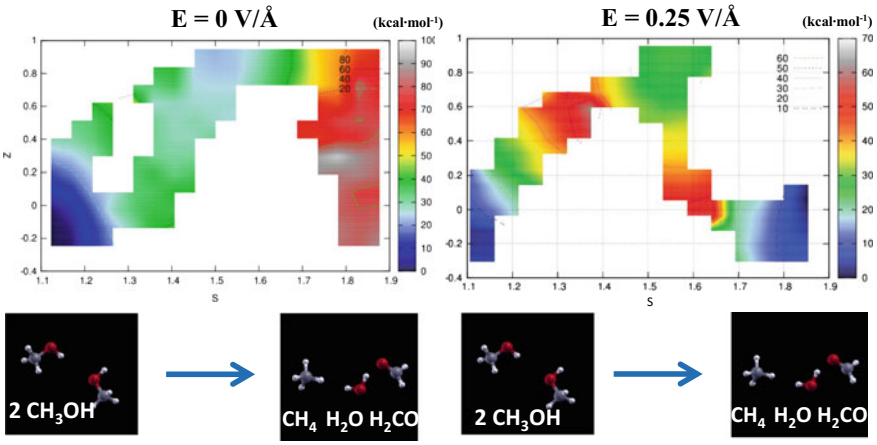


Fig. 5.11 FESs of the gas phase variant of reaction (5.18) in the zero-field regime (left) and in presence of a field strength 0.25 V/Å (right). The energy scale (depth) is in kcal mol^{-1} whereas the S-axis and the Z-axis represent the progress along the reaction and a sort of distance from its ideal path, respectively. Low values of S characterize a system of two methanol molecules whereas high values of this parameter describe a sample composed by a formaldehyde, a methane, and a water molecule. Adapted from [33]

The characterization of the overall shape of the FES can give information on the selectivity of the reaction but, in order to accurately evaluate the relative energetics and compare them also with the condensed phase results, a direct sampling of the relevant regions of the CV space is mandatory. To this aim a series of Umbrella Sampling [22] simulations partially covering the reaction pathways has been carried out. By recollecting the independent data by means of the Weighted Histogram Analysis Method (WHAM) [23–26], estimates of the free energies have been obtained. The results are shown in Fig. 5.11. Again, the application of a moderate field accounts for the stabilization of the reaction products since the basin associated with them shows similar free energies to those of the reactants basin. Finally, it is straightforward to note that in both cases the involved free energies are sensibly higher than those characterizing the liquid phase reaction reported in the previous section, highlighting again a posteriori the irreplaceable role played by the strong intermolecular correlations in assisting such a kind of simple chemistry in condensed phase.

5.5 Conclusions

In this chapter, we have reviewed the phenomenological evidence that has recently emerged from state-of-the-art ab initio simulations on the chemical transformations of molecular systems under the action of a static and homogeneous electric field, and on the formulation of advanced free energy schemes capable to deal with the

complexity of the underlying chemical spaces. In particular, we have shown that the combination of the latter methods could potentially be of enormous help in systematically making use of electric fields to develop new chemistries and chemical routes to the synthesis of novel systems and materials.

References

1. English NJ, Waldron CJ (2015) Phys Chem Chem Phys 17:12407
2. de Pomerai D et al (2003) FEBS Lett 543:93
3. Porcelli M et al (1997) FEBS Lett 402:102
4. Aragones AC et al (2016) Nature 531:88
5. Futera CJ, English NJ (2017) J Chem Phys 147:031102
6. Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge
7. Umari P, Pasquarello A (2002) Phys Rev Lett 89:157602
8. Berry MV (1994) Proc R Soc Lond A 392:45
9. King-Smith RD, Vanderbilt D (1993) Phys Rev B 47:1651
10. Resta R (1994) Rev Mod Phys 66:899
11. Desiraju G, Vittal J, Ramanan A (2011) Crystal engigneering: a textbook. World Scientific, New Jersey, London
12. Nunes RW, Vanderbilt D (1994) Phys Rev Lett 73:712
13. Nunes RW, Gonze X (2001) Phys Rev B 63:155107
14. Resta R (1998) Phys Rev Lett 80:1800
15. Wannier GH (1960) Phys Rev 117:432
16. Nenciu G (1991) Rev Mod Phys 63:91
17. Gonze X et al (1995) Phys Rev Lett 74:4035
18. Gonze X et al (1997) Phys Rev Lett 78:294
19. Pietrucci F (2017) Rev Phys 2:32
20. Laio A, Parrinello M (2002) Proc Natl Acad Sci USA 99:12562
21. Barducci A et al (2008) Phys Rev Lett 100:020603
22. Torrie GM, Valleau JP (1977) J Comput Phys 23:187
23. Kumar S et al (1992) J Comput Chem 13(13):1011
24. Ferrenberg AM, Swendsen RH (1989) Phys Rev Lett 63:1195
25. Bennet CJ (1976) J Comput Phys 22:245
26. Shirts MR, Chodera JD (2008) J Chem Phys 129:124105
27. Branduardi D et al (2007) J Chem Phys 126:054103
28. Branduardi D et al (2011) J Chem Theory Comput 7:539
29. Gallet G et al (2012) J Chem Theory Comput 8:4029
30. Saitta AM et al (2015) Proc Natl Acad Sci USA 112:E343–E344
31. Pietrucci F, Andreoni W (2014) J Chem Theory Comput 10:913
32. Pietrucci F, Saitta AM (2015) Proc Natl Acad Sci USA 112:15030
33. Cassone G et al (2017) Chem Sci 8:2329
34. Sprik M (2000) Chem Phys 258:139
35. Saitta AM, Saija F (2014) Proc Natl Acad Sci USA 111:13768
36. Cassone G et al (2017) Sci Rep 7:6901
37. Cassone G et al (2018) Chem Commun 54:3211–3214
38. Giannozzi P et al (2009) J Phys Condens Matter 21:395502
39. Bonomi M et al (2009) Comput Phys Commun 180:1961
40. Rappe AM et al (1990) Phys Rev B 44:13175
41. Perdew JP et al (1997) Phys Rev Lett 77:3865; Ibidem Phys Rev Lett 78:1396

42. Becke AD (1988) Phys Rev A 38:3098; Lee C et al (1988) Phys Rev B 37:785
43. Grimme S (2006) J Comput Chem 27:1787
44. Bolhuis PG et al (2002) Ann Rev Phys Chem 53:291
45. Marzari N et al (2012) Rev Mod Phys 84:1419
46. Miller SL (1953) Science 117:528
47. Bada JL (2004) Earth Plan Sci Lett 226:1
48. Miyakawa S et al (2002) Proc Natl Acad Sci USA 99:14628
49. Kim JH et al (2004) Appl Catal A General 264:37
50. Yaripour F et al (2005) Catal Commun 6:542
51. Yaripour F et al (2005) Catal Commun 6:147
52. Song W et al (2002) J Am Chem Soc 124:3844
53. Olah GA et al (2009) J Org Chem 74:487
54. Cassone G et al (2015) J Chem Phys 142:054502
55. Sellner B et al (2013) J Phys Chem B 117:10869
56. Reischl B et al (2009) Mol Phys 107:495
57. Bronstein Y et al (2016) Phys Rev B 93:024104
58. Laporte S et al (2015) Phys Chem Chem Phys 17:20382
59. Price D, Halley JW (1983) J Electroanal Chem 159:347
60. Kreuzer J (1991) Surf Sci 246:336
61. Schmickler W (1995) Surf Sci 335:416
62. Stuve EM (2012) Chem Phys Lett 519–520:1
63. Hammadi Z et al (2012) Appl Phys Lett 101:243110
64. Lee WK et al (2013) Nano Res 6:767
65. Shaik S et al (2016) Nat Chem 8:1091
66. Balke N et al (2017) Nanotechnology 28:065704
67. Geissler PL et al (2001) Science 291:2121
68. Olsson MHM et al (2006) Phil Trans R Soc B 361:1417
69. Nitzan A (2006) Chemical dynamics in condensed phases. Oxford University Press
70. Orr-Ewing AJ (2014) J Chem Phys 140:090901
71. Nguyen VS et al (2013) J Phys Chem A 117:2543
72. Guido CA et al (2012) J Chem Theory Comput 9:28
73. Ma C et al (2014) J Phys Chem Lett 5:1672–1677
74. Geissler PL et al (1999) J Phys Chem B 103:3706
75. Ensing B et al (2006) Acc Chem Res 39:73
76. Prasad BR et al (2013) J Phys Chem B 117:153
77. Nguyen VS et al (2011) J Phys Chem A 115:841
78. Iglesias E, Montenegro L (1996) J Chem Soc Faraday T 92:1205
79. Chaudhuri C et al (2001) J Phys Chem A 105:8906
80. Saitta AM et al (2012) Phys Rev Lett 108:207801
81. Crim FF (2012) Faraday Discuss 157:9

Chapter 6
Force Field Development and Nanoreactor Chemistry

Lee-Ping Wang

Abstract The application of theory and computation to understand reactivity at high pressures is beset by several challenges: (1) the nontrivial changes in electronic structure that take place during the reaction, (2) the many possible initial configurations of reacting species, and (3) the simulation timescales needed for reaction events to occur. In this chapter, we will discuss two methods for meeting these challenges. *The development of accurate molecular mechanics force fields* is needed to sample initial configurations of reactants. This chapter provides a perspective on the functional forms and parameterization strategies of modern force fields. In particular, we highlight the ForceBalance parameterization method for optimizing force fields systematically and reproducibly using a free and open-source code. *The* ab initio *nanoreactor* is a new simulation method for rapidly discovering new reaction pathways from first-principles molecular dynamics. The main components of the nanoreactor approach include an external time-dependent potential that induces high-velocity molecular collisions, a trajectory analysis and visualization tool for identifying and extracting individual reaction events, and a reaction path optimization workflow for estimating the reaction energies and barrier heights from a reaction event.

The Born-Oppenheimer potential energy surfaces (PES) of systems exhibiting reactivity depend on the nuclear positions in a nontrivial way that cannot be concisely described by simple functions of the molecular structure, as opposed to nonreactive systems that interact principally through their intermolecular interactions. The most predictive methods for reactivity under high-pressure, shock conditions, or combustion conditions are in the realm of ab initio quantum chemistry [35, 78, 106], though reactive force fields have also been used with varying degrees of success. This reactivity often occurs in disordered media, which presents a major challenge to modeling the atomic structure of the initial state. Moreover, the timescales of reactivity might be nanoseconds, microseconds, or even longer, which are out of reach

L.-P. Wang (✉)
Department of Chemistry, University of California at Davis, Davis, CA 95616, USA
e-mail: leeping@ucdavis.edu

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_6

for traditional ab initio molecular dynamics (AIMD) approaches. These challenges make it apparent that innovative simulation methods are needed to sample the initial configurations and discover the possible reaction pathways in a computationally affordable manner.

In this chapter, we will discuss two distinct yet interrelated methods for meeting the aforementioned challenges. (1) *The development of accurate molecular mechanics force fields* is an effective way of sampling initial reactive configurations and (2) *the* ab initio *nanoreactor* is a new method for discovering new reaction pathways.

6.1 Force Field Development

Molecular mechanics force fields, also called empirical potentials/force laws in various research fields, are inexpensive models of the potential energy surface; distinct from ab initio methods that typically involve solving for the electronic degrees of freedom, force fields directly approximate the energy as a function of the atomic positions. This is made possible through the use of chosen analytic functions (i.e., functional forms) that describe the shape of the potential surface in terms of atomic coordinates, as well as tunable parameters that are adjusted to make the potential energy surface or simulation results conform to some aspect of reality. The empirical aspect mainly lies in the tuning of these parameters when developing the force field. In addition, molecular mechanics simulations often proceed from an assumed atomic connectivity or topology; this is because the potential energy surface behaves very differently for intramolecular versus intermolecular degrees of freedom. Because chemical bonding is largely due to quantum mechanical behavior of the electrons, the assumption of a fixed topology allows one to model the potential energy surface using rather simple functional forms (the boundaries of this paradigm are somewhat blurred by reactive and polarizable force fields). The atomic positions are usually propagated using classical equations of motion similar to Born–Oppenheimer molecular dynamics; path integral methods such as ring polymer molecular dynamics may be used to model nuclear quantum effects.

The earliest examples of force fields were developed to model small organic molecules [52, 58], biomolecules [16, 105], and water [38, 43]. Following these initial developments, greater numbers of force fields were introduced for diverse and increasingly general applications such as hydrocarbons [4]; organic/inorganic molecules and polymers [87]; organic liquids [44]; and general small molecules [56, 97], even covering the whole periodic table [72].

Computational scientists today may access a vast literature of force fields that can be used to model almost any imaginable molecular or material substance. It is impossible to cover the whole field in this book chapter, and the reader is instead referred to reviews discussing force fields for several important classes of materials: molecular liquids [17], biomolecules [7, 69], disordered polymers [49, 65]. Despite the vastness of this field, it is instructive to discuss the general concepts of force fields using biomolecules and water as a representative example.

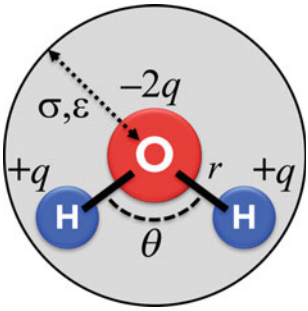


Fig. 6.1 Schematic of a three-point rigid water model showing the atomic sites and parameters. In the TIP3P model, the (rigid) geometric parameters are set to the gas phase structure as $r = 0.9572$ Å, $\theta = 104.52°$. The Lennard–Jones (LJ) parameters for oxygen are $\sigma = 3.15075$ Å, $\varepsilon = 0.63597$ kJ/mol; the H atoms do not interact via the LJ potential. The charge parameters are $q_H = 0.417\ e$, $q_O = -2q_H$

One of the simplest force field archetypes is the rigid three-point water model (Fig. 6.1) with solely intermolecular interactions; these terms are the Lennard–Jones (LJ, or 6–12) interaction representing dispersion, and the pairwise Coulomb interaction between point charges representing electrostatics. The effects of polarization are included in a mean-field sense, as the molecular dipole moments of these models (2.3–2.4 D) tend to be larger than the gas-phase dipole moment of water (1.85 D) [83]. Hydrogen bonding is modeled purely using point charge electrostatics and by omitting the LJ interactions on hydrogen. In the most widely used parameterizations of three-point water models (TIP3P and SPC/E), the geometric parameters were not optimized, and the charge/LJ parameters were hand-tuned in order to match simulated observables with experiment [9, 43]. The density and heat of vaporization were commonly used properties for parameter fitting. Also noteworthy are four-point water models (TIP4P, TIP4P-Ew, and TIP4P/2005) where the negative charge is not located on the O atom, but rather on a *virtual site* located on the HOH angle bisector on the same side as the H atoms. Because four-point models can describe the molecular quadrupole moment more accurately, they can yield improved agreement with experimental properties [1, 39].

The most important difference between water and macromolecules is the importance of intramolecular degrees of freedom. A typical macromolecular functional form is as follows:

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{nonbonded}},$$

$$E_{\text{bond}} = \sum_{i,j\in\text{bonds}} \frac{k^b_{ij}}{2}\left(r_{ij} - r^0_{ij}\right)^2, \quad E_{\text{angle}} = \sum_{i,j,k\in\text{angles}} \frac{k^\theta_{ijk}}{2}\left(\theta_{ijk} - \theta^0_{ijk}\right)^2,$$

$$E_{\text{dihedral}} = \sum_{i,j,k,l\in\text{dihedrals}} \sum_{n=1}^{6} k^\phi_{ijkl,n}\left(1 + \cos\left(n\phi_{ijkl} - \phi^0_{ijkl,n}\right)\right),$$

$$E_{\text{nonbonded}} = \sum_{i,j \in \text{nonbonded}} 4\varepsilon_{ij}\left[-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 + \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}\right] + \frac{q_i q_j}{r_{ij}} \tag{6.1}$$

E_{bond} and E_{angle} are harmonic energy terms for covalent bonds and bond angles; the equilibrium values and force constants are empirical parameters. Bonds involving hydrogen are typically constrained to their equilibrium lengths. E_{dihedral} is a periodic function of the dihedral angle formed by four atoms, and is parameterized as a Fourier series with adjustable linear coefficients and phase factors. *Proper* dihedrals involve four atoms connected in a line (*i*–*j*–*k*–*l*), and effectively describe steric and electrostatic contributions to the energy as molecular groups are rotated around the covalent bond *j*–*k*; out of all terms in the force field, these tend to be the most heavily parameterized [40, 53]. The total energy profile for rotating about the bond contains proper dihedral contributions from all groups bonded to *j* and *k*. By contrast, the atom ordering of *improper* dihedral terms does not follow the order of covalent bonds (for example, three atoms bonded to a central one). Improper terms are meant to describe the resistance of sp^2-hybridized atoms to out-of-plane bending motions. $E_{\text{nonbonded}}$ describes nonbonded interactions as a pairwise sum; these include all atom pairs on different molecules, as well as pairs within the same molecule separated by more than three chemical bonds. When calculating a pairwise nonbonded interaction, the atomic LJ parameters of both atoms are combined using a combining rule, such as the Lorentz–Berthelot rules which state $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$; $\varepsilon_{ij} = (\varepsilon_i * \varepsilon_j)^{\frac{1}{2}}$. The atomic partial charges are multiplied together according to Coulomb's law for point charges. In some force fields, LJ and electrostatic interactions between groups separated by exactly three bonds are scaled by a global factor and are called "1–4" interactions, (e.g., in AMBER force fields, the scale factors are 5/6 for electrostatics and 1/2 for Lennard–Jones).

Macromolecular force field parameters must be carefully organized to describe the diverse chemical environments of the atoms while keeping the number of fitting parameters within manageable limits. The organization of parameters varies between different force fields and software implementations, and parameterization strategies vary across different force fields and research groups; however, the overarching concepts are largely the same, and the following discussion uses the AMBER force field as an example. Central to the organizational framework is a list of *atom type* definitions; an atom type represents an atom within a specific chemical environment (e.g., a carbonyl oxygen atom). List of bond, angle, dihedral, and Lennard–Jones parameters are then defined; each parameter is specific to the combination of atom types involved in the interaction (e.g., equilibrium length and force constant parameters are defined for the carbonyl C=O bond).

The force field also contains residue definitions, which are templates for small molecules or pieces of macromolecules. Each residue definition lists the atoms in the residue and the atom pairs connected by chemical bonds. Each atom within the residue is assigned an atom type based on its chemical environment, enabling the assignment of bonded and Lennard–Jones parameters. The atomic partial charges are defined uniquely for each atom within a residue. When the force field and residue definitions

are loaded, the user may load a coordinate file containing the lists of residues to be simulated and their initial atomic positions. The residues in the coordinate file are matched to the local residue definitions to create the topology. Parameters for individual bond, angle, dihedral, and nonbonded interactions are assigned based on matching of atom types. After this preprocessing step, the system is specified by the complete list of particles and nonbonded interaction parameters, as well as the lists of all bonded interactions and parameters. The system and initial conditions are provided to initiate the simulation.

It remains to discuss how parameters are fitted. The electrostatic parameters are often derived by fitting the electrostatic potential from ab initio calculations. The HF/6-31G* level of theory is a popular choice because it is thought to over-polarize the system, thus effectively describing some amount of condensed phase polarization [23, 54]; it is also inexpensive, allowing new charges to be determined easily for simulating new molecules. The bond and angle parameters are fitted to reproduce equilibrium geometries and vibrational frequencies from experiment or ab initio calculations, and the Lennard–Jones parameters are fitted to densities and heats of vaporization of molecular liquids. The dihedral parameters occupy an intermediate space between bonded and nonbonded terms, and are often fitted to reproduce ab initio potential energy profiles. Due to the large amounts of effort involved in parameter fitting and the interdependence of parameters on the simulation results, modern force fields often contain subsets of parameters that were fitted at different times.

Given this conceptual framework, it is straightforward to implement higher order terms such as anharmonic bond and angle potentials, cross terms that couple multiple bonds/bond angles, and modified nonbonded functional forms. These terms can improve the force field accuracy in terms of fitting ab initio calculations or predicting experimental observables. Some force fields employ atom-centered multipole expansions, allowing the molecular electrostatic potential to be closely fit [28, 74]. Within the diverse ecosystem of force field functional forms, two families deserve special mention: polarizable models and reactive models.

Polarizable force fields describe the electronic response to electric fields, and can employ Drude particles/charges on springs [47, 108], point dipoles [5, 25], fluctuating charges [73, 76], or a combination of these. In these models, calculating the energy and forces in these models require relaxing the polarizable degrees of freedom at each MD step, or propagating them using an extended Lagrangian formalism [3, 90]. A major advantage of polarizable models is they are transferable between the gas phase and solvent phase; for example, they can be fitted to gas-phase ab initio data and predict condensed phase properties. However, the increased computational cost limits the applicability of these models in fields where nonpolarizable models are the norm. The reader is referred to several reviews discussing the current state-of-the-art of polarizable models [22, 70]; the AMOEBA model is a particular example where the ForceBalance software has been used to automatically optimize the model to improve agreement with ab initio data and experiment [48, 59, 71, 101].

Reactive force fields are designed to describe changes in chemical bonding without the use of quantum mechanical electrons. Because chemical bonding is a quantum mechanical phenomenon, reactive force fields are much more challenging to develop

than their nonreactive counterparts, and their accuracy is limited compared to ab initio and semiempirical methods described elsewhere in this book. Here, we list three successful strategies:

- *Bond order potentials*: This class of models originates in the modeling of solids, and includes the early models of Tersoff, Brenner, and others [32, 81, 85, 88]. In simulations employing bond order potentials, there are no predefined topologies. A common feature of the models is the calculation of an approximate bond order between each pair of atoms, a function of the interatomic distance. Similar to ab initio definitions of bond order, these functions are constant at small distances and decay rapidly to zero at distances corresponding to bond dissociation. Summing the bond orders involving the same atom as $v_i = \sum_j b_{ij}(r_{ij})$ yields the *valence* of that atom. The bond orders and valences are important intermediate quantities in computing the final energy; for example, the energy contribution from a carbon atom may have a minimum where $v = 4$ corresponding to four covalent bonds. Another prominent example of this class is ReaxFF, which has been applied to model many reactive systems including high-energy materials and combustion processes [20, 51, 91].
- *Empirical valence bond (EVB) potentials*: These models employ several predefined topologies and a means for switching between them. Computing the EVB energy involves diagonalizing an effective Hamiltonian matrix where the diagonal elements are MM potentials with distinct topologies and the off-diagonal elements are couplings [19, 104]. The couplings are modeled as a Gaussian function of position centered at the transition state structure. Because EVB potentials require specifying the topology for each chemical state, it is most useful for studying systems where there is a single well-defined reaction of interest. The multistate EVB model is a variant of EVB in which new topologies are automatically generated on-the-fly; this is useful for studying problems where the topologies are known but too numerous to represent simultaneously, such as the case of proton transfer in water [26, 80].
- *Machine learning (ML) potentials*: These models compute the potential energy using functional forms that does not represent physical interactions, opting instead for highly flexible models capable of fitting almost any input data. ML potentials may use a neural network architecture where the inputs are translationally and rotationally invariant functions of the molecular environments around each atom (i.e., feature vectors or descriptors), the intermediate variables are passed through several layers of interconnected nodes, and the final energy is computed as the output [8, 84, 110]. Kernel ridge regression is another ML technique that has been applied to build force fields [21].

The existence of many force fields for different molecules and applications speaks to the major challenge of *transferability*. Compared to ab initio methods, the accuracy of any given force field is not guaranteed when applied across wide regions of chemical and thermodynamic space; arguably this is intrinsic to the nature of empirical models. Thus, highly transferable and general *tools* are needed for systematic development of force fields for target applications.

Automated parameterization. The parameterization of force fields may incorporate training data from diverse experimental and ab initio theoretical data sources. During parameterization, the force field is used to simulate physical quantities that are directly compared to the reference data. The parameters are then adjusted to minimize the difference between simulation and reference, and the process repeated iteratively until convergence.

Any physical property that could be computed from simulations may be fitted to experimental values. In practice, only properties that are precisely measurable and easily computable are the ones used for fitting; specific examples for molecular liquids are the density at constant temperature and pressure, and the heat of vaporization. Thermodynamic response properties such as the thermal expansion coefficient, isothermal compressibility, isobaric heat capacity, and dielectric constant are also easily computable from equilibrium simulations and comparable to experiment. When developing force fields intended for classical Hamiltonian simulations, the size of nuclear quantum effects on different experimental properties must be considered; the enthalpy of vaporization and isobaric heat capacity have significant quantum effects requiring ad hoc corrections [9, 39]. These approximate corrections may involve modeling the high-frequency motions of the system using non-interacting quantum harmonic oscillators, and calculating the correction using the difference between the quantum and classical solutions. Path-integral methods simulate the quantum statistical mechanics directly, and do not require such corrections [18, 24, 37].

Theoretical data sources include ab initio calculated values of total potential energies, nuclear gradients [29], and interaction energies; electrostatic observables such as multipole moments and response properties such as vibrational frequencies may also be used. The approximations in the ab initio method, the empirical model and the classical approximation imply that the optimized model should deviate somewhat from the training data, and the deviations are expected to increase in size as the functional form becomes more approximate. Explicit polarization is important for quantitative comparisons to ab initio data in the gas phase; on the other hand, fixed-charge models are not expected to closely reproduce gas-phase interactions because their charges are parameterized to include average condensed-phase polarization. Successes in developing fixed-charge models have been achieved when performing the calculations in solvent, such as the adaptive force matching approach that uses a QM/MM environment [2, 50].

The training data, parameters being adjusted, and the optimization algorithm are three important choices in the parametrization procedure. Carrying out this procedure involves measuring, calculating or collecting the training data set, running the corresponding simulations, and tuning parameters; this task has many interconnected components that are arduous to carry out and difficult to reproduce. Automated scripts are often developed to glue the required components together; an early example is a *tcsh* script for simplex optimization [30]. More recently, several parameterization programs have been made available for further generality and reproducibility; these include ForceBalance [102], *potfit* [15], and Wolf(2)Pack [41]. We also note related research in the AMOEBA, AMBER, and CHARMM simulation communities that provide automated programs for parameterizing new molecules by following fixed

workflows [11, 75, 96]; these procedures are designed for consistency with the published parameters and are recommended for generating parameters that are "compatible" with the existing ones.

ForceBalance is a free and open-source software package for systematic and reproducible model parameterization that has been used to develop a series of force fields for water [48, 71, 100–102], organic molecules [59], lipid bilayers [57], and proteins [98]. Going beyond force fields, ForceBalance has also been used to parameterize atom-centered grids for approximate electronic structure calculations [45]. Three key abstractions are introduced to accommodate diverse model parameterization workflows:

- The *force field* represents a plain text or XML file containing numerical values to be optimized, and enables writing copies of the file with modified values. Importantly, the force field allows functional relationships between parameters, as well as constraints and rescaling factors; these are often needed for parameters with physical meanings and which may have very different orders of magnitude depending on the unit system.
- The *engine* is an interface to the simulation software package that implements the model and the desired property calculations, which can be done using APIs (when available) or the operating system. Engine implementations include OpenMM, AMBER, TINKER, Gromacs, and Psi4; interfaces to CHARMM, and LAMMPS are under development.
- The *target* represents an observable that can be calculated using the model and directly compared to a stored reference value; the *objective function* is a weighted sum of least squares errors from multiple targets, plus a regularization term that penalizes parameter overfitting.

In an optimization cycle (presented graphically in Fig. 6.2), the current values of optimization parameters are passed to the force field object to create a parameter file. The targets then call the engine functions (which call external codes) to evaluate the simulated observables and their parametric first derivatives, which are used to build the objective function, gradient, and approximate Hessian (the latter is obtained from the Gauss–Newton approximation). An optimization algorithm then predicts the next set of optimization parameters to minimize the objective function. ForceBalance implements several optimization algorithms; in practice, the best performance is obtained from a natively implemented trust-radius Newton–Raphson algorithm. Gradient-based optimizations have the possibility of getting "stuck" in local minima, but we have not found this to be a serious issue in practice, and at least one study has shown that disparate sets of starting values can converge to the same minimum [102].

In order to use the Newton–Raphson optimizer, ForceBalance requires first derivatives of all calculated properties with respect to the parameters being optimized. While derivatives of single-point properties (e.g., energies and gradients) are easily computed via finite difference, simulated thermodynamic properties are more challenging due to the high computational cost and inherent statistical uncertainty. ForceBalance implements semi-analytic expressions for efficiently obtaining parametric

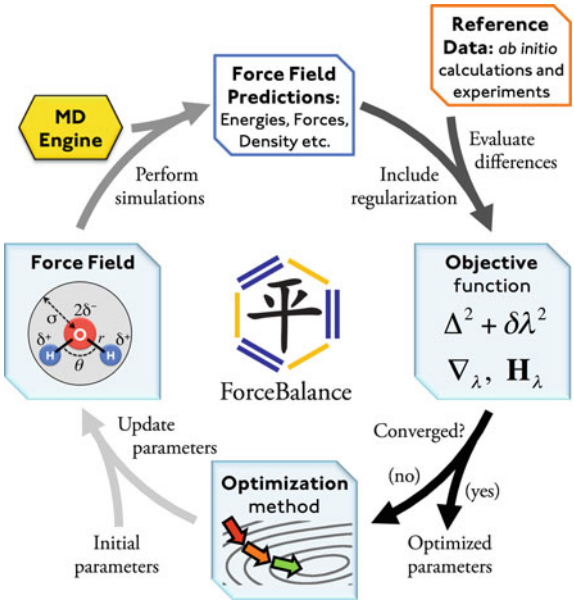


Fig. 6.2 *Steps of the ForceBalance optimization cycle.* The initial force field parameters (lower left) are used to perform simulations using molecular dynamics (MD) software (upper left). The objective function is computed as a least squares function of the differences between simulation results and reference data (upper right). The optimization method updates the parameters in order to minimize the objective function (bottom right)

derivatives of many thermodynamic properties without needing to run multiple simulations. A statistical mechanical fluctuation formula [14] provides the parametric derivatives of a general thermodynamic property A as

$$\frac{\partial \langle A \rangle}{\partial \lambda} = \left\langle \frac{\partial A}{\partial \lambda} \right\rangle - \frac{1}{k_B T} \left(\left\langle A \frac{\partial E}{\partial \lambda} \right\rangle - \langle A \rangle \left\langle \frac{\partial E}{\partial \lambda} \right\rangle \right) \tag{6.2}$$

where λ is the model parameter, $\langle \cdot \rangle$ the ensemble average using the current value of λ, and E the potential energy. Because A and E can be evaluated individually for trajectory frames in the simulation, the quantities on the RHS may be evaluated in a post-processing step by evaluating $\partial E/\partial \lambda$ numerically for the trajectory frames. This differentiation approach is highly effective for accurately fitting many thermodynamic properties; more recently, pure numerical derivatives have also been implemented where separate simulations are carried out to get the parametric derivatives of properties.

The least squares objective function being minimized is given by

$$L[\mathbf{x}] = \sum_{A \in \text{targets}} \frac{w_A}{\langle A_{ref} \rangle^2} \sum_{i \in \text{points}} \left(A_{i,\text{MM}}[\mathbf{y}(\mathbf{y}_0; \mathbf{M}\mathbf{x})] - A_{i,\text{ref}} \right)^2 + w_0 |\mathbf{x}|^2 \tag{6.3}$$

where $\mathbf{x}$ is the array of independent optimization variables, the sum is taken over the squared residuals of targets A weighted by w_A and data points i within each target, and the last term is the regularization function with its own weight w_0. $\mathbf{y}$ represents the array of physical parameters that are literally written to the force field file—these

are calculated from the initial values $\mathbf{y_0}$ and the rescaled optimization variables $\mathbf{Mx}$ where $\mathbf{M}$ is a diagonal matrix. A full specification of the calculation requires choosing the weights w, the rescaling factors (diagonal of $\mathbf{M}$), and the functional relationships between physical parameters encoded in $\mathbf{y}(\mathbf{y}_0; \mathbf{Mx})$.

Because each target term corresponds to a property with physical units, they are first normalized to dimensionless values. The normalization scheme generally depends on the target type, for example single-point energies and forces are automatically normalized by the variance of the ab initio energies/forces $\langle A_{\text{ref}} \rangle^2$. Equation (6.3) shows that the normalization is mathematically equivalent to an inverse weight. The dimensionless weights w_A have a default value of unity and may be further tuned, for example, if one target is deemed to be more important than others.

The rescaling factors (diagonal of $\mathbf{M}$) play a dual role in improving conditioning and controlling regularization. The physical parameters $\mathbf{y}$ are unsuitable as optimization variables because they are in the unit system of the underlying simulation code, they obey physically motivated user-specified constraints (e.g., conditions on the net charge, traceless quadrupoles, and spatial/geometric relationships), and their size may vary by several orders of magnitude which can lead to an ill-conditioned problem. This problem is addressed by using optimization variables $\mathbf{x}$ that are dimensionless and order 1; the mapping to $\mathbf{y}$ involves rescaling as $\mathbf{x}' = \mathbf{Mx}$, followed by enforcing the physical constraints by construction as $\mathbf{y} = \mathbf{f}[\mathbf{x}']$. In this scheme, all chain rule terms are automatically included when numerically differentiating A_{MM} with respect to $\mathbf{x}$.

The regularization term (penalty function) is isotropic in $\mathbf{x}$ with a default quadratic form given as $w_0|\mathbf{x}|^2$, though other regularizations (such as LASSO) may also be used. In a simple optimization with no relationships between parameters, the changes in physical parameters and optimization variables are related as $\Delta y_i = M_{ii} \Delta x_i$. The value of M_{ii} thus corresponds to the expected variability of y_i, which is why M_{ii} is also called a "prior width." Increasing/decreasing the rescaling factor for a parameter will loosen/tighten the variations of the physical parameter in the optimization, allowing the user to control the optimization results in a qualitative sense. In practice, rescaling factors are grouped by parameter *type* and specified using order of magnitude estimates, and changing these values by successive factors of ½ or 2 can further fine-tune the optimization results.

As an example, consider a force field model of HCl with three adjustable physical parameters—the bond force constant and the charges on the H and Cl atoms. The starting parameter values are $\mathbf{y_0} = [2.5\text{e}5\ \text{kJ mol}^{-1}\ \text{nm}^{-2}, +0.25\ e, -0.25\ e]$. The rescaling factors are chosen as $1.0\text{e}5\ \text{kJ mol}^{-1}\ \text{nm}^{-2}$ for bond force constants and $0.1\ e$ for charges. The current value of the optimization variables is $\mathbf{x} = [+0.6, -0.3]$. After rescaling, we have $\mathbf{Mx} = [0.6\text{e}5\ \text{kJ mol}^{-1}\ \text{nm}^{-2}, -0.03\ e]$. The charge neutrality constraint leads to the following mapping to physical parameters:

$$\mathbf{y} = \mathbf{y}_0 + \left[0.6e5\,\text{kJ mol}^{-1}\,\text{nm}^{-2},\ -0.03\,e, +0.03\,e\right];$$
$$\mathbf{y} = \left[3.1\text{e}5\,\text{kJ mol}^{-1}\,\text{nm}^{-2}, +0.22\,e, -0.22e\right].$$

Example application: The iAMOEBA water model was the first force field to be optimized in the ForceBalance framework. iAMOEBA uses the same functional form as the existing AMOEBA model, with the approximation that all interactions between induced dipoles are excluded. This so-called "direct induction" approximation reduces the many-body polarization energy expression to only three-body terms, and saves computational time because the self-consistent solution for N-body "mutual induction" is no longer needed. The iAMOEBA parameterization used the AMOEBA model as a starting point, and 19 of the model parameters were optimized (5 bonded, 2 vdW, 9 permanent multiple, and 3 polarizability). The number of model parameters is rather large, implying the need for a large data set. The parameterization data sets described below are organized into *target* folders and included with the ForceBalance distribution.

1. Six thermodynamic properties—the density, heat of vaporization, thermal expansion coefficient, isothermal compressibility, isobaric heat capacity, and dielectric constant were collected at 42 thermodynamic phase points; these include 32 points at atmospheric pressure and a temperature range of 249–373 K, and 10 points at 298 K and a pressure range of 20–8000 bar. This data is accurately measured and widely available due to the universal importance of water [95].
2. The dipole moment, quadrupole moment, and vibrational frequencies of the water molecule in gas phase were included.
3. Single-point potential energies and gradients were calculated at the MP2/aug-cc-pVTZ level of theory for 42,000 water cluster structures extracted from liquid-phase simulations using the original (mutual induction) AMOEBA model. The simulations spanned a temperature range of 249–373 K (1 atm pressure) and cluster sizes ranged from 2 to 22 molecules. For the convenience of computation, this data set was separated into 21 targets, one for each cluster size.
4. Interaction energies for water clusters were calculated at known critical points on the potential energy surface, with structures taken from the literature. For structures that correspond to minima on the ab initio potential surface, the energy is minimized using the force field before calculating the objective function, and the root-mean-square-deviations (RMSD) from the ab initio structure are included.

    a. The global minima of the water dimer, trimer, tetramer and pentamer, and 8 local minima of the hexamer. The minimum energy structures were optimized at the MP2/heavy-aug-cc-pVTZ level of theory followed by CCSD(T)/CBS energy calculations.
    b. Ten critical points of the water dimer (the "Smith" set) were included; the energy was calculated using CCSD(T)/CBS without further geometry optimization.
    c. Also included were local minima of 2 octamer structures, 5 11-mer structures, 5 16-mer structures, 2 17-mer structures, and 4 20-mer structures. These were also optimized using MP2/heavy-aug-cc-pVTZ, and energies were computed at the MP2/CBS level of theory.

The choice of rescaling factors and weights for the optimization are described in Wang et al. [101] and also distributed with the ForceBalance code.

The results of the iAMOEBA optimization are shown in Fig. 6.3. The starting AMOEBA model predicts qualitatively correct behavior of the liquid density with respect to temperature, with a density maximum around 293 K (experiment 277 K). If direct polarization is used without parameter optimization, the density dependence on temperature becomes qualitatively incorrect and there is no density maximum within the temperature range. The final iAMOEBA model with optimized parameters gives very close agreement with both experiment and ab initio calculations. The simulated density is within 0.1% of experiment across the whole temperature range, and the temperature of maximum density is correctly reproduced at 277 K. The level of agreement with the MP2 data also matches the quality of the original AMOEBA model. This gave a positive indication that the direct polarization approximation is sufficient for describing electronic polarization effects in water.

The data set for parameterizing iAMOEBA is easily applicable to develop other water models within the ForceBalance framework. In particular, a three-point water model (TIP3P-FB) was developed [102] that has the same functional form and computational cost of TIP3P, but with greatly improved thermodynamic, structural, and kinetic properties (Fig. 6.4). Although TIP3P-FB is not quite as accurate as iAMOEBA, it is 50–100 times faster due to its simple functional form; furthermore, recent studies show TIP3P-FB is compatible with modern biomolecular parameter

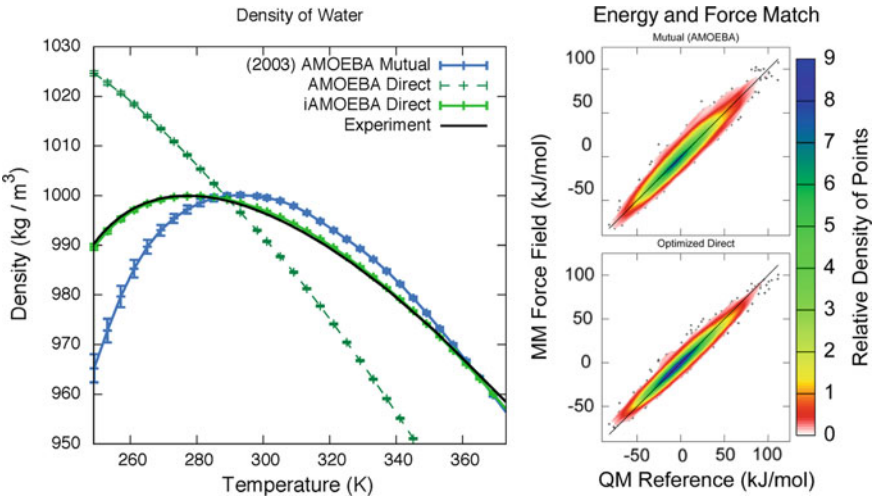


Fig. 6.3 *Results of iAMOEBA model optimization*. Left: density dependence on temperature, measured experimentally (black), calculated using the 2003 AMOEBA mutual induction model (blue), direct polarization without parameter optimization (dark green), and the optimized iAMOEBA model (light green). Right: scatter and density plot of ab initio vs. force field single-point energies for water 7-mer clusters, for the 2003 AMOEBA model (top) and 2013 iAMOEBA model (bottom). Error bars indicate one standard error

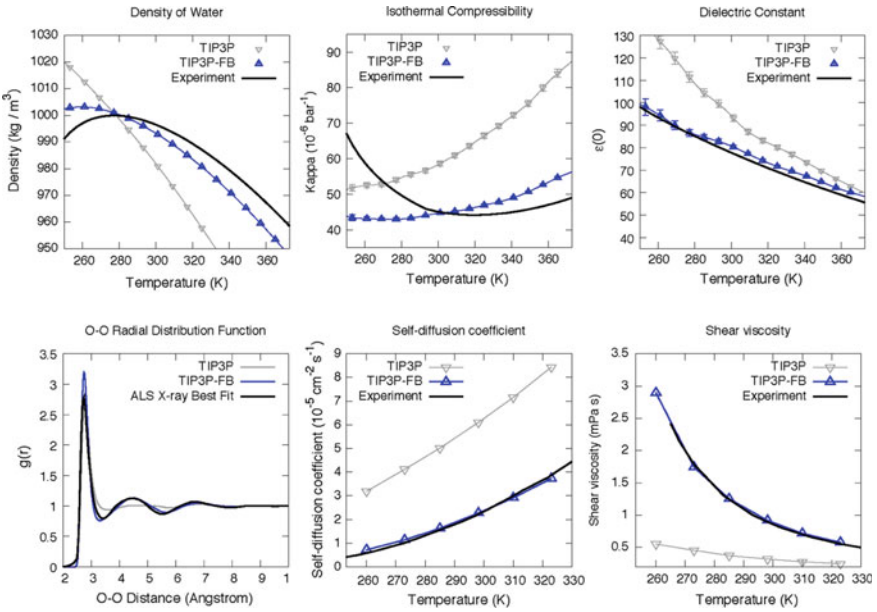


Fig. 6.4 *Results of TIP3P-FB model optimization*. Top row: the temperature dependence of thermodynamic properties for experiment (black), TIP3P (gray), and TIP3P-FB (blue), showing improvements in the quality of fit for TIP3P-FB. Bottom row: structural and kinetic properties not fitted during parameterization: O–O radial distribution function at 298.15 K (bottom left), diffusion coefficient (bottom middle) and shear viscosity (bottom right) as a function of temperature

sets [31, 109]. The good compatibility of TIP3P-FB may be due to the small changes in its parameters from TIP3P (all 5% or less). We expect that improved "simple" water models such as TIP3P-FB may play an important role in developing the next generation of biomolecular force fields.

Validation of force fields. The validation of a force field is an essential measure of a force field's predictive power, and involves assessing the model accuracy for known properties outside of the training data set. Knowledge gained from validation could also feed back into the force field development cycle at a high level—for instance, if evidence of overfitting is found, the model may be refitted using a smaller number of independent parameters or increasing the regularization strength. With that being said, validation studies should in principle be kept separate from parameterization.

Pure water is an ideal proving ground for force fields and force field development methods due to the plentiful experimental data available for training and validation. As mentioned previously, the easily computable/precisely known properties are often used for training; consequently, properties that are more challenging to compute or less precisely measured are typically reserved for validation.

An example of a validation property for water is the solid–liquid phase diagram, which is a rather challenging thermodynamic property to compute (Fig. 6.5). Water

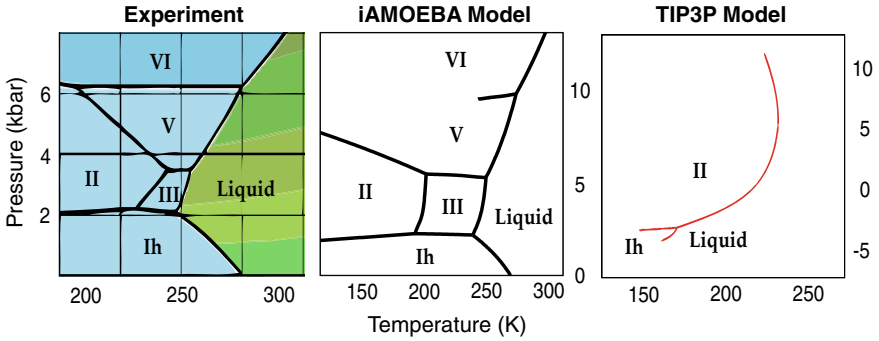


Fig. 6.5 *Solid–liquid phase diagram of water.* The experimental phase of water contains several distinct ice phases at high pressures (left). The iAMOEBA water model exhibits the same phase behavior in simulations, an important validation of the model (middle). The widely used TIP3P model yields qualitatively incorrect behavior; ordinary ice Ih is stable only at 150 K and negative pressures (right)

can be found in many distinct solid states such as ice Ih (ordinary ice), ice Ic, ice II, ice III, and ice V; many of these states are characterized by distinct crystalline orderings of the O atoms and random orientations of the hydrogen bonds consistent with the "ice rules" (each water molecule donates two hydrogen bonds and accepts two hydrogen bonds). The phase diagram may be constructed from the free energies of the water and ice phases in the relevant temperature and pressure range (150–350 K and 0–20 kbar). Alternatively, one may directly simulate the melting points of different ice phases at various pressures, then obtain the entire liquid–solid coexistence curve by means of Gibbs–Duhem integration. When two liquid–solid curves meet at a triple point, the same procedure may be applied to obtain the solid–solid coexistence curve, thereby tracing out the entire phase diagram. The final result is a positive one; iAMOEBA correctly predicts the experimental solid–liquid phase diagram. This qualitative agreement is also obtained by the TIP4P series of models; notably, the popular TIP3P model predicts a totally different (and qualitatively incorrect) diagram where most of the ice phases do not appear.

Conclusions and prospects. Force fields are essential tools for studying atomic and molecular behavior in the condensed phase. Historically, the major milestones in this field have arrived as new functional forms and parameter sets for describing different areas of chemical space. As the field has evolved, the number of functional forms and amount of experimental/computational data has greatly increased, posing a challenge for improving force fields due to the complexity of the parameterization problem. A new generation of methods and software (e.g., ForceBalance) addresses this modern challenge by allowing complex parameterization workflows to be carried out in an automatic and reproducible way.

Modern force field development software is expected to contribute to computational studies of reactivity in two ways: (1) Improved conventional force fields

(i.e., with fixed topologies) could provide more accurate starting configurations for reactive simulations employing ab initio or semiempirical methods. (2) Improved reactive force fields could potentially predict the reaction mechanisms in a complex chemical system. In what follows, we will describe a method for predicting reaction mechanisms without presumed reaction coordinates as a possible application for the reactive force fields of the future.

6.2 Nanoreactor Chemistry

The starting point of computational investigations into mechanistic studies has traditionally been a mechanistic hypothesis or reaction coordinate. In the simplest procedure, the reaction energy is characterized by energy minimization of the reactant and product structures. Reaction path optimization methods, such as the string method and nudged elastic band [82], provide estimates of the minimum energy path and the transition state (TS) structure. The barrier height is determined by a TS optimization calculation [12] that searches for a critical point on the potential surface with one imaginary frequency. Energy minimization along the directions of the imaginary mode lead back to the reactant and product [33]. These calculations can estimate the entropic contribution to the free energy using harmonic or quasi-harmonic approximations, and simple kinetic models such as transition state theory may be used to estimate reaction rates.

Many research efforts have gone beyond the single-point paradigm to provide improved estimates of the thermodynamics and kinetics of reactions. Accelerated sampling methods such as umbrella sampling, adiabatic free energy dynamics [77], metadynamics [46], and orthogonal space random walk [111] maximize the thermodynamic overlap along a chosen order parameter or reaction coordinate by applying biases to a molecular dynamics or Monte Carlo simulation. On the other hand, accurate rate calculations require statistical sampling of paths rather than configurations. The seminal research of Chandler and coworkers on transition path sampling [13] has led to many subsequent developments that further improve the calculation of reaction rates for rare events, including transition interface sampling [92, 93], aimless shooting [60, 67], and others. Path sampling methods use Monte Carlo methods to generate a distribution of paths in the local neighborhood of an initial pathway; the reaction rate is then computed as the fraction of the sampled paths that successfully complete the reaction.

State-of-the-art methods for generating reaction pathways. The above methods are designed to quantitatively characterize a reaction pathway, given an initial mechanistic hypothesis. More recently, computational methods have been developed that explore the possible reaction pathways starting from an initial structure. Broadly speaking, two kinds of methods have emerged: rules-based methods and dynamics-based methods. Rules-based methods work by combinatorial application of fundamental moves (such as bond breaking, bond forming, and constraints) followed by reaction path optimization to generate the reaction network. Examples of rules-based

methods include the ZStruct method [27, 66], the artificial force induced reaction method [34, 55], the Reaction Mechanism Generator [86], the graph-based reaction path sampling of Habershon [36], and recent work of Zubarev et al. [112] on the tricarboxylic acid cycle. Bergeler et al. [10] developed a heuristics-guided approach that applies conceptual electronic structure theory to identify likely reactive sites.

Methods based on molecular dynamics for exploring reactivity employ various methods for generating reaction events in the simulation trajectory. Examples of these methods and applications include the study of comet impacts on the early Earth using the multiscale shock technique [35], the SPRINT metadynamics of Pietrucci and Andreoni [68], simulations with strong electric fields to induce reactivity [79], high-temperature simulations of quantum dot growth [107], and transition state search using chemical dynamics simulations [94]. The ab initio nanoreactor [99] is part of this class of simulations, which is described in detail in this chapter. All of these simulations use various electronic structure methods to explore reactivity, including HF/DFT and semiempirical models. Reactive force fields such as ReaxFF [20, 91] and empirical valence bond [80, 103, 104] have contributed significant insights into reaction mechanisms, but they have not been broadly applied to discover qualitatively new pathways. In particular, the optimal parameters of ReaxFF-type force fields are known to depend on the particular system and reaction conditions, which makes their application to these types of problems very challenging [42]. On the other hand, EVB-type models are not able to discover new chemical structures because the possible molecular topologies are "built in" to the functional form.

The literature clearly shows that the computational discovery of reaction pathways is a field that is rapidly gaining interest. Both the rules-based methods and dynamics-based methods have important roles to play, due to their relative advantages and disadvantages. The rules-based methods are advantageous in their reduced computational cost, as they do not require MD simulations to explore the configuration space; their ability to comprehensively search over the rule set is also an advantage, but this could become a limitation when the candidate pathways being generated are too numerous to search over [27]. Importantly, there is always an inherent risk of missing certain classes of pathways in rules-based methods, because the rules are fundamental assumptions about what kinds of reactivity are possible. Dynamics-based methods, on the other hand, are able to generate mechanisms free of rules. For example, the nanoreactor study found numerous cases of proton relays through water molecules that reduced the activation barrier in many of the discovered reaction pathways, as well as intermolecular reactions that have inspired detailed follow-up studies [62, 64]; we are not aware of such mechanisms being discovered by the rules-based approaches. In the future, we anticipate dynamics-based and rules-based methods will combine forces to further accelerate the understanding of reactivity.

The* ab initio *nanoreactor employs ab initio molecular dynamics (AIMD) to discover the reactivity that is possible in a molecular system. The nanoreactor is based on an AIMD simulation that uses a simple level of theory and basis set, for example, Hartree–Fock (HF)/3-21G, in order to access longer simulation timescales at the expense of less accurate energetics. The simulations are carried out using an efficient

quantum chemistry implementation running on graphics processing hardware [89]. Because the nanoreactor does not assume specific reaction coordinates, the reaction events occur as a natural consequence of the classical trajectory on the potential energy surface, which allows for qualitatively new reaction pathways to be observed. Although the core concept is very simple, the major challenge lies in the high cost of evaluating the potential energy in ab initio MD simulations. For many systems of interest, each AIMD time step represents ~1 fs of physical time or even less, but many reactions of interest occur on nanosecond or longer timescales (10^6 or more time steps). This is cost-prohibitive to perform with AIMD simulations, which generally yield ~1 ps/day in simulation time. Hence, there exists a great need to explore accelerated methods.

The nanoreactor simulation addresses this challenge by introducing additional energy via a time-dependent external potential, reducing the simulation time needed to cross over activation barriers (Fig. 6.6). Importantly, the design of this external potential should not be biased in favor of specific mechanistic hypothesis or reaction coordinates, because the main goal for the simulation is to find the possible reaction pathways. A simple example of such a potential is the spherically symmetric, flat-bottomed potential where the energy is given as

$$E = \sum_{i}^{N_{\text{atoms}}} \frac{m_i k}{2} (r_i - R)^2 \Theta(r_i - R)$$

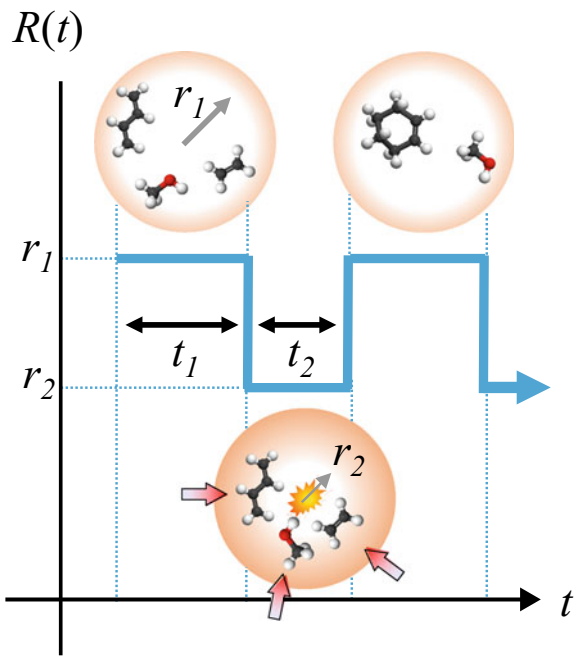


Fig. 6.6 Illustration of ab initio nanoreactor simulation. The blue curve indicates the radius parameter R as a function of time

where m_i is the atomic mass, k the force constant, R the sphere radius that separates the flat-bottom and harmonic regions, and Θ the Heaviside step function. The potential is made time-dependent through the parameters k and R, which oscillate between values k_1, R_1 and k_2, R_2 as a rectangular waveform time intervals of t_1 and t_2, respectively. The simulation uses Born-Oppenheimer molecular dynamics with a small time step of 0.5 fs, chosen to ensure the simulation remains stable when atomic velocities are high. The molecules diffuse freely in the region $r < R_1$ during the time interval t_1, about 1.0 ps or 2000 time steps; the simulation temperature is high enough (1500–3000 K) that intermolecular interactions are rapidly dissociated and the molecules are distributed throughout the spherical region, but not high enough to dissociate the covalent bonds within the short 1.0 ps window. The temperature controlled using a Langevin thermostat with a time constant of 300 time steps (0.15 ps). The value of k_1 is relatively unimportant, as long as molecules are prevented from diffusing far beyond R_1; we typically choose $k_1 = 1.0$ kcal mol^{-1} A^{-2}. After this time interval, the sphere radius is switched to R_2 (we assume $R_2 < R_1$). At this moment, the atoms located in the region $R_2 < r < R_1$ experience a strong inward force proportional to their mass and distance from R_1 as $\mathbf{F} = -mk_2(r - R_2)\hat{\mathbf{r}}$. We call this force the "spherical piston."

This spherical piston will cause all atoms at the same distance from R_2 to accelerate uniformly toward the center. The force is proportional to the atomic mass and is intended to accelerate molecules uniformly without dissociating or severely distorting them. As the molecules enter the $r < R_2$ inner region, the spherical piston no longer applies any forces but the molecules possess a high center-of-mass velocity; the speed of atoms, if converted to random thermal energy, would correspond to temperatures of >10,000 K. Shortly after the switching of the potential (i.e., inward stroke of the piston), the molecules collide near the center of the simulation sphere with sufficient kinetic energy to cross over many activation barriers. During this time, the friction term in the Langevin equation rapidly dissipates the energies in excess of the set temperature, so that newly formed molecules cool to the set temperature instead of recrossing the energy barriers. At the end of the time interval t_2, the sphere radius is reset to R_1, allowing molecules to freely diffuse throughout the large sphere; the cycle then repeats.

Conventional molecular simulations produce statistical samples of physical properties that converge to the thermodynamic distribution given sufficient simulation time. The behavior of the nanoreactor simulations should be considered on a different footing, because the simulations are not at equilibrium and never reach long enough times to sample all of the possible connectivities of atoms in the system. Rather, it is instructive to consider the evolution of chemical species as the simulation evolves toward longer times. Intuitively, the first "products" to appear in the nanoreactor are chemically close to the "starting materials" and differ in only one or two chemical bonds. These early products are now available to react with other species in the simulation, leading to new species that are increasingly chemically different from the starting materials. If the starting materials are highly unsaturated, the products are generally larger in size—for example, a 500 ps simulation that started with 39 C_2H_2 (acetylene) molecules resulted in over half of the atoms belonging

to a single highly connected species, $C_{46}H_{36}$ [99]. Because the translational motion of larger species is slower compared to small species, they are subjected to smaller boundary forces and the rate of reactivity in the simulation is accordingly reduced. In mechanistic studies with known starting materials and products, very long nanoreactor simulations (i.e., >100 ps) are not strictly necessary, because promising reaction intermediates may be identified and used to initiate new simulations.

The nanoreactor discovers a large number of new species and reaction pathways, but the high-energy and nonequilibrium simulation conditions are not directly relevant to most experiments. In order to obtain mechanistic insights that are transferable to ambient conditions, a procedure is used to map the reaction events to minimum energy paths. The two main steps of this procedure—*trajectory analysis and energy refinement*—are described in the following two subsections.

Trajectory analysis and visualization. The cycling of the spherical piston, followed by the cooling of the Langevin thermostat gives rise to intermittent bursts of reactivity that follow the inward stroke (i.e., when the sphere radius is switched to R_2). Because the molecules are free to react according to the topography of the potential energy surface, a wide range of new compounds may be formed that depends on the initial composition of the simulation. An important step in understanding the simulation results is to visualize the simulation trajectory, but it is difficult to visually discern changes in bonding and molecular structure in a high-energy simulation trajectory. To facilitate the understanding of nanoreactor simulations, we developed analysis and visualization tools to highlight new molecules that are formed in the simulation trajectory.

The first step of trajectory analysis is the automatic perception of new molecular species. Our analysis involves constructing a connectivity graph of each simulation frame, where nodes and edges represent atoms and bonds, respectively. For each frame, edges are drawn to connect atom pairs with distances below a threshold calculated from covalent radii as $r_{ij} < 1.4(c_i + c_j)$. The factor of 1.4 accounts for the temporary stretching of bonds beyond their equilibrium length. The connected subgraphs are then determined from the total graph using standard approaches; these are the bonded molecules in a trajectory frame. The most interesting features of nanoreactor simulations are the reactions that lead to new molecules. To detect interesting reaction events and new products, the molecules in each trajectory frame are categorized by comparing their connectivity graphs. The nodes are labeled by the atomic symbol (element), and standard graph isomorphism procedures are used to determine the chemical equivalence of molecular graphs. We implemented this procedure in the Python programming language and used the *networkx* graph library.

One limitation of using purely distance-based criteria is that atom pairs that are momentarily compressed by the piston to near-bonding distances—but do not experience any real chemical bonding—are still treated as "connected" by the trajectory analysis. These transient nonbonding connections are only below the threshold for a very short time and are filtered out in subsequent analysis steps described below. Moreover, the graph isomorphism procedure only distinguishes between structural isomers, but cannot distinguish between geometric isomers (*cis*/*trans*) or stereoisomers; if the application requires distinguishing between more subtle isomerization

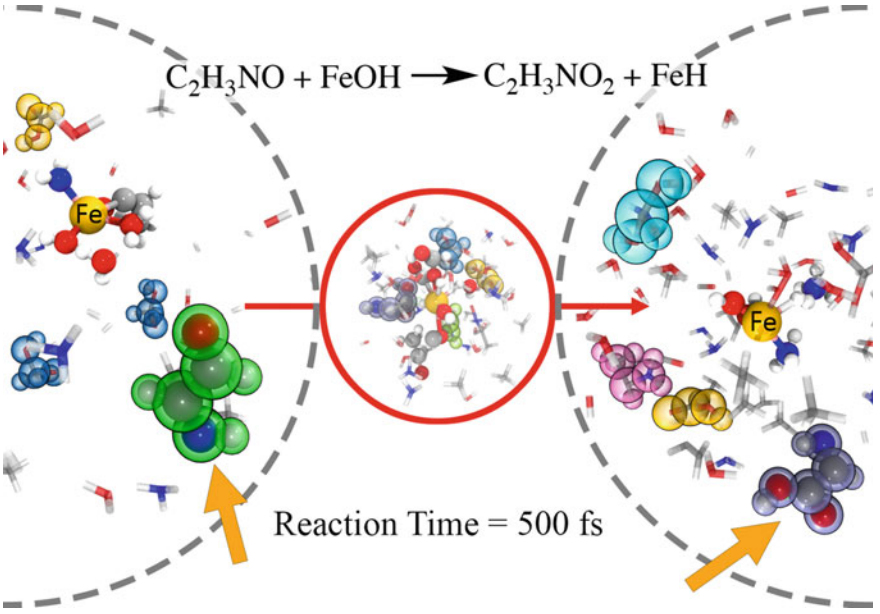


Fig. 6.7 Visualization of a nanoreactor simulation trajectory containing a reaction event where an O atom is transferred from an iron-containing complex (labeled as Fe above) to iminoacetaldehyde (green) to afford 2-iminoacetic acid (blue, lower right). Molecules that differ from the starting materials (H_2O, NH_3, H_2, CH_4, and CO) are highlighted using colored spheres

events, node-based properties such as chirality may be added to the graph isomorphism procedure.

With the graph-based analysis tool, new molecules may be highlighted in distinct colorings as the trajectory is visualized (Fig. 6.7). This is done by writing out a file that contains the connectivity and color data of each atom for each frame of the trajectory; a script in the visualization program (VMD) then updates the colors as the trajectory is played back. Because every chemically distinct species is highlighted in a different color, we can readily see the different kinds of compounds that appear. The analysis capability could be easily used to visually inspect other kinds of reactive simulations, not just the nanoreactor.

The connectivity graph analysis is also the first step toward more detailed analyses of the individual reaction events. Although the simulation is high-energy and nonequilibrium, the concept of stable molecules is fairly well-defined as most molecules remain intact between the periodic compressions of the piston potential. If we define the following binary time series $E_A(t)$ as the *existence* of a molecular graph A in the simulation at time t, the time-frame of the reaction event may be defined as the interval $[t^* - \Delta, t^* + \Delta]$; t^* is the time where $E_A(t)$ switches from "on" to "off" or vice versa, i.e., $E_A(t^*) \neq E_A(t^* + \delta)$, and Δ is a user-defined window size, e.g., 100 frames. In other words, the interval $[t^* - \Delta, t^* + \Delta]$ is a short length of

the trajectory where molecule A appears or disappears. If the existence time series flips several times in rapid succession, we expand the interval to $[t^{*\prime} - \Delta, t^{*\prime\prime} + \Delta]$ where $t^{*\prime} < t^* < t^{*\prime\prime}$ and no flips are observed in the pre-interval $[t^{*\prime} - \Delta, t^{*\prime}]$ and post-interval $[t^{*\prime\prime}, t^{*\prime\prime} + \Delta]$.

Now that the reaction time interval has been identified, it remains to select the atoms that participate in the reaction. This concept is intuitive because not all atoms in the nanoreactor participate in each reaction event; selecting a subset of atoms also enables higher levels of theory to be used than in the nanoreactor simulation itself. Because atoms are conserved in the simulation, the disappearance of A simply means that the nodes of A, denoted as S_A, are now connected in a different way and some other graphs have appeared. The new graphs containing members of S_A that appear at the same moment that A disappears are denoted as B, C, etc. (the reaction may produce more than one molecule). The atoms of the new graphs $S_B \cup S_C \ldots$ may be a superset of S_A, which indicates that A is not the only participating reactant molecule. The reactant molecules are now expanded to all graphs that contain members of $S_B \cup S_C$, and the procedure is iterated back and forth until the reactant and product molecules contain the same set of atoms. As more molecules are included in the reactants and products, the time interval $[t^{*\prime} - \Delta, t^{*\prime\prime} + \Delta]$ may be further extended such that no flips are observed in any of the existence time series for all species in the pre-interval and post-interval. At the conclusion of this procedure, the final set of atoms S_{final} and the time interval $[t^{*\prime} - \Delta, t^{*\prime\prime} + \Delta]$, are extracted from the trajectory and saved to the disk.

The large-amplitude motions in the nanoreactor simulation can render the extraction of reaction events challenging, especially when the existence time series undergoes many rapid flips over long times. This may occur, for example, if a proton is transferred back and forth across a hydrogen bond. One possible solution is to rectify the individual existence time series by applying filters to reduce the noise. We employed a two-state hidden Markov model, which describes the "measured" existence time series as a noisy observation of an underlying Markov process (MP). The MP is described by an intrinsic transition (i.e., flipping) probability between the "on" and "off" states, which we parameterize using a symmetric 2×2 matrix as $\begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$; here $\epsilon = 0.001$, which roughly corresponds with our expected frequency of reactions from the simulation (roughly once per 1000 frames). The output probability measures the likelihood of the observed values given the current value of the MP; this is also parameterized using a symmetric 2×2 matrix as $\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$, where p describes how much the MP is likely to deviate from the observed values. The initial probability of the Markov process is set to a uniform distribution (0.5, 0.5), and the Viterbi algorithm is used to compute the most likely sequence of states of the MP; we use this output as the rectified version of the existence time series. An advantage of this approach is that we gain significant control over the frequency of flips in the time series, but we may miss important reaction events if the rectification is too strong. Another drawback is that because rectification occurs separately for each graph, the rectified existence time series no longer satisfy the conditions that (a) the

nodes of concurrently existing graphs should be nonoverlapping and (b) the union of all nodes in concurrently existing graphs should be equal to the set of all atoms in the simulation. Improving the methods of treating the existence time series is currently a topic we are actively investigating; one possibility is to use ab initio bond orders as a criterion for determining connectivity, as this may produce molecular graphs that are more stable over time.

The outputs of the trajectory analysis consist of small "extracts" of the nanoreactor trajectory, each of which contains an individual reaction event. As described above, each reaction event contains a subset of atoms S_{final} and the sequence of atomic coordinates over the time interval $[t^{*\prime} - \Delta, t^{*\prime\prime} + \Delta]$. Detailed analysis of these reaction events is essential for applying the high-energy and high-temperature nanoreactor simulations to understand reactivity under milder (e.g., standard, 25 °C, 1 atm) conditions. Because the atomic coordinates of the reaction event involve changes in chemical bonding, it is intuitive to search for corresponding reactant and product species that are minima on the potential energy surface, as well as a transition state that connects them through a minimum energy path. A procedure that utilizes automated quantum chemistry calculations to convert the reaction events into minimum energy paths is described in the following section, as well as Figs. 6.8 and 6.9.

Energy refinement of nanoreactor pathways. Many analyses of reaction rates begin with knowledge of the minimum energy path, which connects the reactant, transition state, and product structures via the pathway of steepest descent. The reactant and product are local minima on the potential energy surface whereas the transition state is a saddle point with one imaginary frequency. By contrast, the reaction events extracted from the nanoreactor trajectory contain large-amplitude displacements in the orthogonal degrees of freedom. Thus, the analysis of energetics along the reaction pathway should involve relaxing the orthogonal degrees of freedom to obtain a minimum energy path.

The following procedure is used to obtain minimum energy paths from the extracted reaction events:

(a) The structures along the path are energy minimized using a small interval (e.g., 10 trajectory frames/MD steps; Fig. 6.8, top left).
(b) For each energy-minimized structure, a connectivity graph is created to determine the topology of molecules. A sequence of minimized structures with the same connectivity is called a *segment*. Segments are ordered by their corresponding trajectory frames.
(c) For every pair of segments with different topologies, an *initial pathway* is created that joins the final frame of the earlier segment (the "reactant structure") with the first frame of the later segment (the "product structure"). The main goal is to connect the reactant and product minima with a sequence of closely spaced, essentially continuous structures (Figs. 6.8 and 6.9, top right).

    (i) The sequence of frames resulting from energy minimization of the initial frame is reversed; thus, the first step is the energy-minimized reactant structure, and the last step is on the pathway extracted from the nanoreactor

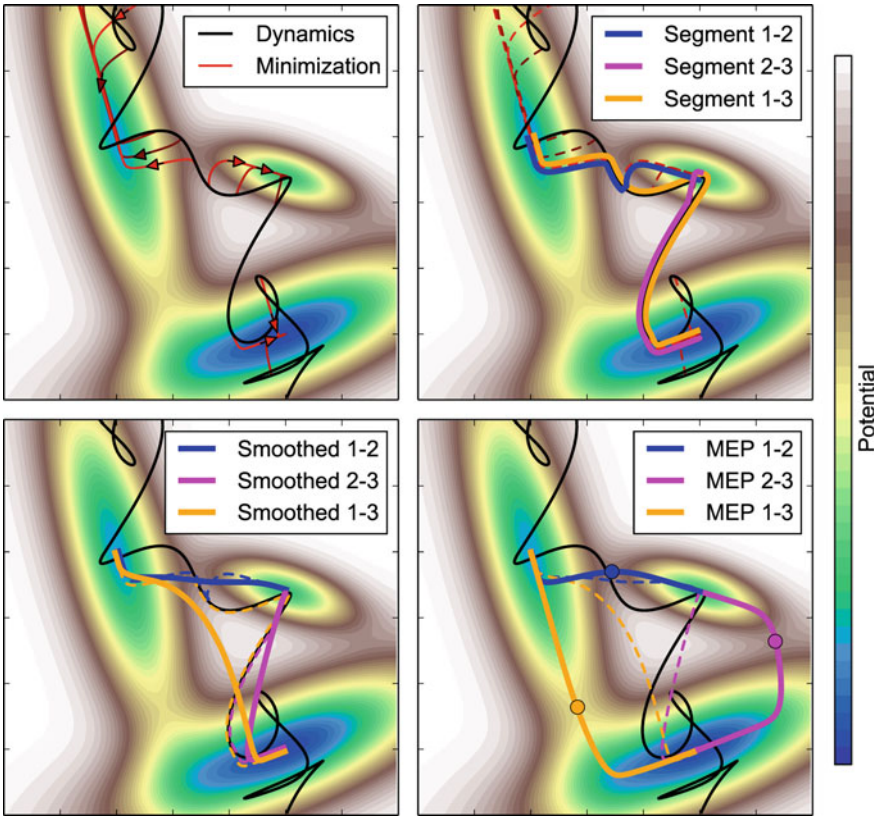


Fig. 6.8 The stages of processing a nanoreactor pathway are illustrated for a toy example with a two-dimensional potential energy surface. *Upper left* the high-energy nanoreactor trajectory (black) passes through several low-energy regions corresponding to distinct chemical species; this trajectory segment is first recognized as a reaction event by the trajectory analysis tool. Energy minimization calculations at evenly spaced frames lead to the local energy minima. *Upper right* initial pathways are created by concatenating structures from energy minimization with trajectory frames to provide continuous (but not smooth) pathways that connect the local minima. *Lower left* an internal-coordinate smoothing procedure produces initial guesses for reaction path optimization. *Lower right* application of standard reaction path optimization methods leads to the final results

MD trajectory. This step assumes that the energy minimization calculation takes sufficiently small steps, such that the sequence of optimization steps is essentially continuous. Although not rigorous, this is often true in practice.

(ii) The MD trajectory frames are concatenated to the result from (i), starting with the final frame of the reactant segment and the first frame of the product segment.

(iii) The sequence of frames resulting from energy minimization of the product frame is concatenated to the result from (ii).

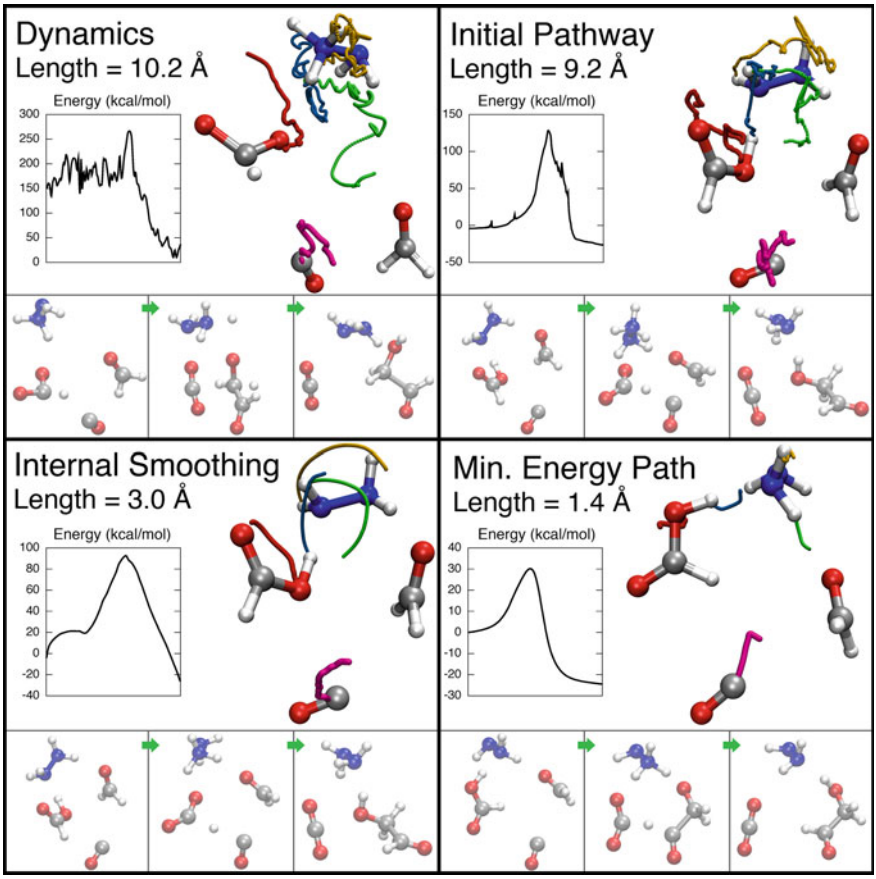


Fig. 6.9 The stages of processing a nanoreactor pathway are shown for a representative example reaction. The positions of five selected atoms are shown using lines over the whole pathway. The inset shows the relative energy over the pathway, and the small panels show representative frames. The reaction involves forming a C–C bond between CO (bottom) and formaldehyde. *Dynamics*: the extracted subset of atoms and frames from the nanoreactor simulation trajectory containing the detected reaction event. *Initial pathway*: the initial and final structures are energy-minimized; the sequences of structures from energy minimization are joined with the MD trajectory to create a continuous path joining the two energy minima. *Internal smoothing*: the high-frequency motions are removed from the initial pathway via a smoothing procedure in the primitive internal coordinates. *Minimum energy path*: the final result obtained after traditional reaction path optimization, transition state search, and intrinsic reaction coordinate calculations are carried out on the "internal smoothing" pathway

(iv) The result is a continuous, but not necessarily smooth sequence of structures connecting the energy-minimized reactant and product.

Segments need not be consecutive in order to be considered as a pathway, and more than one pathway may be created with the same segment as the initial or final point.

(d) The initial pathway is smoothed; this step is needed to ensure good behavior of the reaction path optimization methods. This step may be carried out by smoothing the primitive internal coordinates (i.e., distances, angles, and dihedral angles) using a low-pass filter or window function, then fitting the Cartesian coordinates to maximize agreement with the smoothed internal coordinates (which are not exactly realizable due to redundancies). The resulting smoothed pathway does not contain kinks or high-frequency motions of the initial pathway, but traces out a qualitatively similar path through the configuration space (Figs. 6.8 and 6.9, bottom left).

(e) Reaction path optimization calculations are carried out starting from the smoothed pathway (Figs. 6.8 and 6.9, bottom right).

(i) A small number of equally spaced frames (typically 11–21) is extracted from the smoothed pathway as input to the path optimization.

(ii) The path optimization may be carried out using the string method, nudged elastic band, or other methods within this class. The goal of these methods is to generate an equally spaced chain of structures where the perpendicular component of the gradient is zero. A transition state structure is estimated as the highest-energy structure along the path. Some methods such as climbing image NEB treat the highest-energy structure differently from the others.

(iii) A transition state (TS) optimization is performed to precisely locate the TS structure. This optimization calculation is difficult, and is facilitated by having a good TS estimate from (ii). A calculation of the Hessian may be provided to further ensure convergence. After TS optimization, a second Hessian calculation may be performed to verify the existence of one imaginary mode.

(iv) An intrinsic reaction coordinate (IRC) calculation proceeds from the TS structure in the forward and backward directions of the imaginary mode; this is a steepest descent energy minimization that connects the TS back to the reactant and product. The IRC is the same as the minimum energy path connecting the reactant and product.

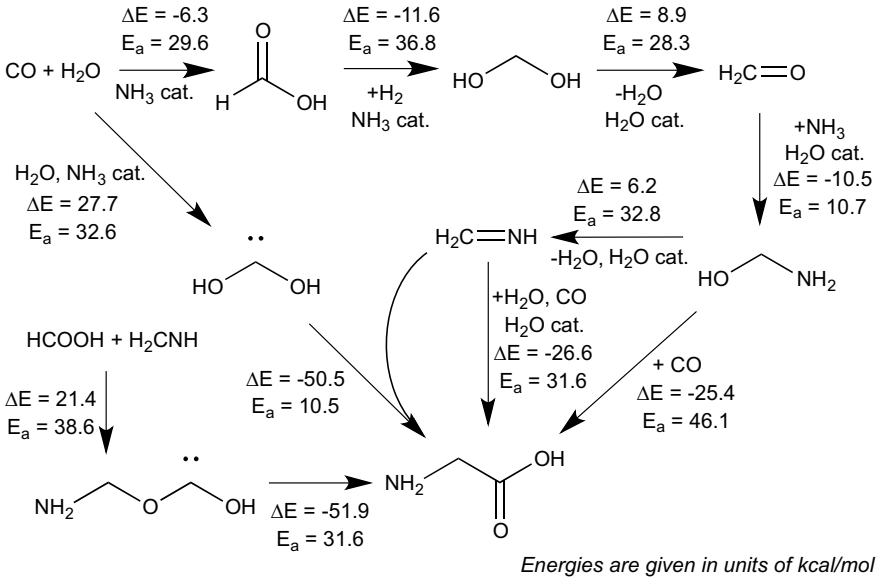


Fig. 6.10 Several reaction pathways leading to glycine (bottom middle) that were discovered in the nanoreactor simulations. The reaction energies and activation energies were obtained following energy refinement of the reaction events, which was carried out at the B3LYP/6-31 + G(*d*, *p*) level of theory

The calculations described above are carried out automatically using a software package written in Python that executes the workflow automatically by calling external quantum chemistry codes. The individual calculations, such as geometry optimization, TS optimization, and IRC calculation are included as standard features in several quantum chemistry packages such as Gaussian, GAMESS, NWChem, and Q-Chem.

As an example application, a series of nanoreactor simulations were carried out on a collection of molecules inspired by the famous Urey-Miller experiment (14 H_2O, 14 CH_4, 14 NH_3, 14 CO, 16 H_2) comprising over 1200 ps of total simulation time, with some individual trajectories over 400 ps in length. The aggregate simulation data contained over 700 reaction pathways and more than 600 distinct products, and included the "discovery" of simple amino acids such as glycine and alanine. In particular, one of the pathways found by the nanoreactor is a previously unreported mechanism for glycine synthesis (Fig. 6.10) that has inspired further theoretical mechanistic studies [61, 63].

To summarize the above, the input to the energy refinement procedure is a reaction event extracted from the nanoreactor trajectory, and the result is one or more intrinsic reaction coordinates (IRC), or minimum energy paths (MEP) that corresponds to elementary reaction steps at 0 K. The MEP connects well-defined critical points on the molecular potential energy surface, and is a useful input to kinetic models for

estimating the rate constants of elementary steps, such as variational transition state theory [6]. Thus, the high-energy reaction events in the nanoreactor can be used to produce elementary reaction steps with quantifiable rates at various experimental conditions.

6.3 Conclusions

Force fields and the nanoreactor are two complementary approaches for understanding reactivity in the condensed phase and/or under extreme conditions, which are some of the most interesting and challenging problems in the chemical sciences. Force fields comprise a broad class of empirical models of the molecular potential energy surface, and include functional forms that assume fixed chemical bonding topologies as well as reactive force fields that can form and break bonds. The nanoreactor, on the other hand, is a new simulation method that augments ab initio molecular dynamics with time-varying external potential to cause a large number of reaction events in a short time, followed by a detailed trajectory analysis and energy refinement procedure that produces minimum energy paths corresponding to the discovered reaction events.

At the intersection of these complementary methods is a modern research field full of possibilities, including the development and improvement of reactive force fields using reaction pathways from the nanoreactor, harnessing semiempirical methods for nanoreactor simulations to discover new mechanisms, and new methods for the detection and analysis of reaction events in simulation trajectories. These developments will help theory and computation to become a more predictive partner of experiment in unraveling complex and elusive reaction pathways.

References

1. Abascal JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 123(23). https://doi.org/10.1063/1.2121687
2. Akin-Ojo O, Song Y, Wang F (2008) Developing ab initio quality force fields from condensed phase quantum-mechanics/molecular-mechanics calculations through the adaptive force matching method. J Chem Phys 129(6). https://doi.org/10.1063/1.2965882
3. Albaugh A, Demerdash O, Head-Gordon T (2015) An efficient and stable hybrid extended Lagrangian/self-consistent field scheme for solving classical mutual induction. J Chem Phys 143(17):174104. https://doi.org/10.1063/1.4933375
4. Allinger NL, Yuh YH, Lii JH (1989) Molecular mechanics—the MM3 force-field for hydrocarbons. 1. J Am Chem Soc 111(23):8551–8566. https://doi.org/10.1021/ja00205a001
5. Applequist J, Carl JR, Fung KK (1972) Atom dipole interaction model for molecular polarizability—application to polyatomic-molecules and determination of atom polarizabilities. J Am Chem Soc 94(9):2952. https://doi.org/10.1021/ja00764a010
6. Bao JL, Truhlar DG (2017) Variational transition state theory: theoretical framework and recent developments. Chem Soc Rev 46(24):7548–7596. https://doi.org/10.1039/C7CS00602K

7. Beauchamp KA, Lin YS, Das R, Pande VS (2012) Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. J Chem Theory Comput 8(4):1409–1414. https://doi.org/10.1021/ct2007814
8. Behler J (2017) First principles neural network potentials for reactive simulations of large molecular and condensed systems. Angew Chem Int Ed 56(42):12828–12840. https://doi.org/10.1002/anie.201703114
9. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. J Phys Chem 91(24):6269–6271. https://doi.org/10.1021/j100308a038
10. Bergeler M, Simm GN, Proppe J, Reiher M (2015) Heuristics-guided exploration of reaction mechanisms. J Chem Theory Comput 11(12):5712–5722. https://doi.org/10.1021/acs.jctc.5b00866
11. Betz RM, Walker RC (2015) Paramfit: automated optimization of force field parameters for molecular dynamics simulations. J Comput Chem 36(2):79–87. https://doi.org/10.1002/jcc.23775
12. Bofill JM (1994) Updated Hessian matrix and the restricted step method for locating transition structures. J Comput Chem 15(1):1–11. https://doi.org/10.1002/jcc.540150102
13. Bolhuis PG, Chandler D, Dellago C, Geissler PL (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annu Rev Phys Chem 53:291–318. https://doi.org/10.1146/annurev.physchem.53.082301.113146
14. Bourasseau E, Haboudou M, Boutin A, Fuchs AH, Ungerer P (2003) New optimization method for intermolecular potentials: optimization of a new anisotropic united atoms potential for olefins: prediction of equilibrium properties. J Chem Phys 118(7):3020–3034. https://doi.org/10.1063/1.1537245
15. Brommer P, Gahler F (2007) Potfit: effective potentials from ab initio data. Modell Simul Mater Sci Eng 15(3):295–304. https://doi.org/10.1088/0965-0393/15/3/008
16. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4(2):187–217. https://doi.org/10.1002/jcc.540040211
17. Caleman C, van Maaren PJ, Hong MY, Hub JS, Costa LT, van der Spoel D (2012) Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. J Chem Theory Comput 8(1):61–74. https://doi.org/10.1021/ct200731v
18. Cao JS, Voth GA (1994) The formulation of quantum-statistical mechanics based on the Feynman path centroid density. 2. Dynamical properties. J Chem Phys 100(7):5106–5117. https://doi.org/10.1063/1.467176
19. Chang YT, Miller WH (1990) An empirical valence bond model for constructing global potential-energy surfaces for chemical-reactions of polyatomic molecular-systems. J Phys Chem 94(15):5884–5888. https://doi.org/10.1021/j100378a052
20. Chenoweth K, van Duin ACT, Goddard WA (2008) ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. J Phys Chem A 112(5):1040–1053. https://doi.org/10.1021/jp709896w
21. Chmiela S, Tkatchenko A, Sauceda HE, Poltavsky I, Schütt KT, Müller K-R (2017) Machine learning of accurate energy-conserving molecular force fields. Sci Adv 3(5)
22. Cisneros GA, Wikfeldt KT, Ojamae L, Lu JB, Xu Y, Torabifard H, Bartok AP, Csanyi G, Molinero V, Paesani F (2016) Modeling molecular interactions in water: from pairwise to many body potential energy functions. Chem Rev 116(13):7501–7528. https://doi.org/10.1021/acs.chemrev.5b00644
23. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. J Am Chem Soc 117(19):5179–5197. https://doi.org/10.1021/ja00124a002
24. Craig IR, Manolopoulos DE (2004) Quantum statistics and classical mechanics: real time correlation functions from ring polymer molecular dynamics. J Chem Phys 121(8):3368–3373. https://doi.org/10.1063/1.1777575

25. Dang LX, Chang TM (1997) Molecular dynamics study of water clusters, liquid, and liquid-vapor interface of water with many-body potentials. J Chem Phys 106(19):8149–8159. https://doi.org/10.1063/1.473820
26. Day TJF, Soudackov AV, Cuma M, Schmitt UW, Voth GA (2002) A second generation multistate empirical valence bond model for proton transport in aqueous systems. J Chem Phys 117(12):5839–5849. https://doi.org/10.1063/1.1497157
27. Dewyer AL, Zimmerman PM (2017) Finding reaction mechanisms, intuitive or otherwise. Org Biomol Chem 15(3):501–504. https://doi.org/10.1039/C6OB02183B
28. Dykstra CE (1993) Electrostatic interaction potentials in molecular-force fields. Chem Rev 93(7):2339–2353. https://doi.org/10.1021/cr00023a001
29. Ercolessi F, Adams JB (1994) Interatomic potentials from first-principles calculations—the force-matching method. Europhys Lett 26(8):583–588. https://doi.org/10.1209/0295-5075/26/8/005
30. Faller R, Schmitz H, Biermann O, Muller-Plathe F (1999) Automatic parameterization of force fields for liquids by simplex optimization. J Comput Chem 20(10):1009–1017. https://doi.org/10.1002/(sici)1096-987x(19990730)20:10%3c1009:aid-jcc3%3e3.0.co;2-c
31. Fatima S, Christopher R (2018) The CHARMM36 force field for lipids can be used with more accurate water models. https://doi.org/10.26434/chemrxiv.6137474.v1
32. Finnis MW, Sinclair JE (1984) A simple empirical N-body potential for transition-metals. Philos Mag Phys Condens Matter Struct Defects Mech Prop 50(1):45–55. https://doi.org/10.1080/01418618408244210
33. Fukui K (1981) The path of chemical-reactions—the IRC approach. Acc Chem Res 14(12):363–368. https://doi.org/10.1021/ar00072a001
34. Gao M, Lyalin A, Maeda S, Taketsugu T (2014) Application of automated reaction path search methods to a systematic search of single-bond activation pathways catalyzed by small metal clusters: a case study on H-H activation by gold. J Chem Theory Comput 10(4):1623–1630. https://doi.org/10.1021/ct500068b
35. Goldman N, Reed EJ, Fried LE, William Kuo IF, Maiti A (2010) Synthesis of glycine-containing complexes in impacts of comets on early Earth. Nat Chem 2(11):949–954. https://doi.org/10.1038/nchem.827
36. Habershon S (2016) Automated prediction of catalytic mechanism and rate law using graph-based reaction path sampling. J Chem Theory Comput 12(4):1786–1798. https://doi.org/10.1021/acs.jctc.6b00005
37. Habershon S, Manolopoulos DE, Markland TE, Miller TF (2013) Ring-polymer molecular dynamics: quantum effects in chemical dynamics from classical trajectories in an extended phase space. In: Johnson MA, Martinez TJ (eds) Annual review of physical chemistry, vol 64, pp 387–413. https://doi.org/10.1146/annurev-physchem-040412-110122
38. Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM (1984) A consistent empirical potential for water-protein interactions. Biopolymers 23(8):1513–1518. https://doi.org/10.1002/bip.360230807
39. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120(20):9665–9678. https://doi.org/10.1063/1.1683075
40. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Struct Funct Bioinform 65(3):712–725. https://doi.org/10.1002/prot.21123
41. Hulsmann M, Kirschner KN, Kramer A, Heinrich DD, Kramer-Fuhrmann O, Reith D (2016) Optimizing molecular models through force-field parameterization via the efficient combination of modular program packages. In: Snurr RQ, Adjiman CS, Kofke DA (eds) Foundations of molecular modeling and simulation. Molecular modeling and simulation-applications and perspectives. Springer, Singapore, pp 53–77. https://doi.org/10.1007/978-981-10-1128-3_4
42. Jensen BD, Bandyopadhyay A, Wise KE, Odegard GM (2012) Parametric study of ReaxFF simulation parameters for molecular dynamics modeling of reactive carbon gases. J Chem Theory Comput 8(9):3003–3008. https://doi.org/10.1021/ct300491d

43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935. https://doi.org/10.1063/1.445869
44. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236. https://doi.org/10.1021/ja9621760
45. Kokkila Schumacher SIL, Hohenstein EG, Parrish RM, Wang LP, Martinez TJ (2015) Tensor hypercontraction second-order Moller-Plesset perturbation theory: grid optimization and reaction energies. J Chem Theory Comput 11(7):3042–3052. https://doi.org/10.1021/acs.jctc.5b00272
46. Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci 99(20):12562–12566. https://doi.org/10.1073/pnas.202427399
47. Lamoureux G, MacKerell AD, Bt Roux (2003) A simple polarizable model of water based on classical Drude oscillators. J Chem Phys 119(10):5185–5197. https://doi.org/10.1063/1.1598191
48. Laury ML, Wang LP, Pande VS, Head-Gordon T, Ponder JW (2015) Revised parameters for the AMOEBA polarizable atomic multipole water model. J Phys Chem B 119(29):9423–9437. https://doi.org/10.1021/jp510896n
49. Li Y, Abberton BC, Kroger M, Liu WK (2013) Challenges in multiscale modeling of polymer dynamics. Polymers 5(2):751–832. https://doi.org/10.3390/polym5020751
50. Li YP, Li JC, Wang F (2013) Liquid-liquid transition in supercooled water suggested by microsecond simulations. Proc Natl Acad Sci U S A 110(30):12209–12212. https://doi.org/10.1073/pnas.1309042110
51. Liang T, Shin YK, Cheng YT, Yilmaz DE, Vishnu KG, Verners O, Zou CY, Phillpot SR, Sinnott SB, van Duin ACT (2013) Reactive potentials for advanced atomistic simulations. In: Clarke DR (ed) Annual review of materials research, vol 43, pp 109–129. https://doi.org/10.1146/annurev-matsci-071312-121610
52. Lifson S, Warshel A (1968) Consistent force field for calculations of conformations vibrational spectra and enthalpies of cycloalkane and N-alkane molecules. J Chem Phys 49(11):5116. https://doi.org/10.1063/1.1670007
53. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins Struct Funct Bioinform 78(8):1950–1958. https://doi.org/10.1002/prot.22711
54. Mackerell AD, Wiorkiewiczkuczera J, Karplus M (1995) An all-atom empirical energy function for the simulation of nucleic-acids. J Am Chem Soc 117(48):11946–11975. https://doi.org/10.1021/ja00153a017
55. Maeda S, Harabuchi Y, Takagi M, Taketsugu T, Morokuma K (2016) Artificial force induced reaction (AFIR) method for exploring quantum chemical potential energy surfaces. Chem Rec 16(5):2232–2248. https://doi.org/10.1002/tcr.201600043
56. Mayo SL, Olafson BD, Goddard WA (1990) DREIDING—a generic force-field for molecular simulations. J Phys Chem 94(26):8897–8909. https://doi.org/10.1021/j100389a010
57. McKiernan KA, Wang L-P, Pande VS (2016) Training and validation of a liquid-crystalline phospholipid bilayer force field. J Chem Theory Comput 12(12):5960–5967. https://doi.org/10.1021/acs.jctc.6b00801
58. Momany FA, Carruthers LM, McGuire RF, Scheraga HA (1974) Intermolecular potentials from crystal data. 3. Determination of empirical potentials and application to packing configurations and lattice energies in crystals of hydrocarbons, carboxylic-acids, amines, and amides. J Phys Chem 78(16):1595–1620. https://doi.org/10.1021/j100609a005
59. Mu XJ, Wang QT, Wang LP, Fried SD, Piquemal JP, Dalby KN, Ren PY (2014) Modeling organochlorine compounds and the sigma-hole effect using a polarizable multipole force field. J Phys Chem B 118(24):6456–6465. https://doi.org/10.1021/jp411671a

60. Mullen RG, Shea J-E, Peters B (2015) Easy transition path sampling methods: flexible-length aimless shooting and permutation shooting. J Chem Theory Comput 11(6):2421–2428. https://doi.org/10.1021/acs.jctc.5b00032
61. Nhlabatsi ZP, Bhasi P, Sitha S (2016) Possible interstellar formation of glycine from the reaction of CH_2[double bond, length as m-dash]NH, CO and H_2O: catalysis by extra water molecules through the hydrogen relay transport. Phys Chem Chem Phys 18(1):375–381. https://doi.org/10.1039/C5CP04987C
62. Nhlabatsi ZP, Bhasi P, Sitha S (2016) Possible interstellar formation of glycine from the reaction of CH_2NH, CO and H_2O: catalysis by extra water molecules through the hydrogen relay transport. Phys Chem Chem Phys 18(1):375–381. https://doi.org/10.1039/C5CP04987C
63. Nhlabatsi ZP, Bhasi P, Sitha S (2016) Possible interstellar formation of glycine through a concerted mechanism: a computational study on the reaction of CH_2[double bond, length as m-dash]NH, CO_2 and H_2. Phys Chem Chem Phys 18(30):20109–20117. https://doi.org/10.1039/C5CP07124K
64. Nhlabatsi ZP, Bhasi P, Sitha S (2016) Possible interstellar formation of glycine through a concerted mechanism: a computational study on the reaction of CH_2NH, CO_2 and H_2. Phys Chem Chem Phys 18(30):20109–20117. https://doi.org/10.1039/C5CP07124K
65. Paul W, Smith GD (2004) Structure and dynamics of amorphous polymers: computer simulations compared to experiment and theory. Rep Prog Phys 67(7):1117–1185. https://doi.org/10.1088/0034-4885/67/7/r03
66. Pendleton IM, Perez-Temprano MH, Sanford MS, Zimmerman PM (2016) Experimental and computational assessment of reactivity and mechanism in C(sp(3))-N bond-forming reductive elimination from palladium(IV). J Am Chem Soc 138(18):6049–6060. https://doi.org/10.1021/jacs.6b02714
67. Peters B, Trout BL (2006) Obtaining reaction coordinates by likelihood maximization. J Chem Phys 125(5):10. https://doi.org/10.1063/1.2234477
68. Pietrucci F, Andreoni W (2011) Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale. Phys Rev Lett 107(8):085504. https://doi.org/10.1103/PhysRevLett.107.085504
69. Ponder JW, Case DA (2003) Force fields for protein simulations. Protein Simul 66:27
70. Ponder JW, Wu CJ, Ren PY, Pande VS, Chodera JD, Schnieders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T (2010) Current status of the AMOEBA polarizable force field. J Phys Chem B 114(8):2549–2564. https://doi.org/10.1021/jp910674d
71. Qi R, Wang LP, Wang QT, Pande VS, Ren PY (2015) United polarizable multipole water model for molecular mechanics simulation. J Chem Phys 143(1):12. https://doi.org/10.1063/1.4923338
72. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM (1992) UFF, A full periodic-table force-field for molecular mechanics and molecular-dynamics simulations. J Am Chem Soc 114(25):10024–10035. https://doi.org/10.1021/ja00051a040
73. Rappe AK, Goddard WA (1991) Charge equilibration for molecular dynamics simulations. J Phys Chem 95(8):3358–3363. https://doi.org/10.1021/j100161a070
74. Ren PY, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107(24):5933–5947. https://doi.org/10.1021/jp027815+
75. Ren PY, Wu CJ, Ponder JW (2011) Polarizable atomic multipole-based molecular mechanics for organic molecules. J Chem Theory Comput 7(10):3143–3161. https://doi.org/10.1021/ct200304d
76. Rick SW, Stuart SJ, Berne BJ (1994) Dynamical fluctuating charge force fields: application to liquid water. J Chem Phys 101(7):6141–6156. https://doi.org/10.1063/1.468398
77. Rosso L, Mináry P, Zhu Z, Tuckerman ME (2002) On the use of the adiabatic molecular dynamics technique in the calculation of free energy profiles. J Chem Phys 116(11):4389–4402. https://doi.org/10.1063/1.1448491

78. Saitta AM, Saija F (2014) Miller experiments in atomistic computer simulations. Proc Natl Acad Sci U S A 111(38):13768–13773. https://doi.org/10.1073/pnas.1402894111
79. Saitta AM, Saija F (2014) Miller experiments in atomistic computer simulations. Proc Natl Acad Sci 111(38):13768–13773. https://doi.org/10.1073/pnas.1402894111
80. Schmitt UW, Voth GA (1998) Multistate empirical valence bond model for proton transport in water. J Phys Chem B 102(29):5547–5551. https://doi.org/10.1021/jp9818131
81. Sheng HW, Kramer MJ, Cadien A, Fujita T, Chen MW (2011) Highly optimized embedded-atom-method potentials for fourteen fcc metals. Phys Rev B 83(13):20. https://doi.org/10.1103/PhysRevB.83.134118
82. Sheppard D, Terrell R, Henkelman G (2008) Optimization methods for finding minimum energy paths. J Chem Phys 128(13). https://doi.org/10.1063/1.2841941
83. Silvestrelli PL, Parrinello M (1999) Water molecule dipole in the gas and in the liquid phase. Phys Rev Lett 82(16):3308–3311. https://doi.org/10.1103/PhysRevLett.82.3308
84. Smith JS, Isayev O, Roitberg AE (2017) ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem Sci 8(4):3192–3203. https://doi.org/10.1039/c6sc05720a
85. Stillinger FH, Weber TA (1985) Computer-simulation of local order in condensed phases of silicon. Phys Rev B 31(8):5262–5271. https://doi.org/10.1103/PhysRevB.31.5262
86. Suleimanov YV, Green WH (2015) Automated discovery of elementary chemical reaction steps using freezing string and Berny optimization methods. J Chem Theory Comput 11(9):4248–4259. https://doi.org/10.1021/acs.jctc.5b00407
87. Sun H (1998) COMPASS: an ab initio force-field optimized for condensed-phase applications—overview with details on alkane and benzene compounds. J Phys Chem B 102(38):7338–7364. https://doi.org/10.1021/jp980939v
88. Tersoff J (1988) Empirical interatomic potential for carbon, with applications to amorphous-carbon. Phys Rev Lett 61(25):2879–2882. https://doi.org/10.1103/PhysRevLett.61.2879
89. Ufimtsev IS, Martinez TJ (2009) Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. J Chem Theory Comput 5(10):2619–2628. https://doi.org/10.1021/ct9003004
90. Van Belle D, Wodak SJ (1995) Extended Lagrangian formalism applied to temperature control and electronic polarization effects in molecular dynamics simulations. Comput Phys Commun 91(1):253–262. https://doi.org/10.1016/0010-4655(95)00051-G
91. van Duin ACT, Dasgupta S, Lorant F, Goddard WA (2001) ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 105(41):9396–9409. https://doi.org/10.1021/jp004368u
92. van Erp TS (2007) Reaction rate calculation by parallel path swapping. Phys Rev Lett 98(26):268301. https://doi.org/10.1103/PhysRevLett.98.268301
93. van Erp TS, Moroni D, Bolhuis PG (2003) A novel path sampling method for the calculation of rate constants. J Chem Phys 118(17):7762–7774. https://doi.org/10.1063/1.1562614
94. Varela JA, Vazquez SA, Martinez-Nunez E (2017) An automated method to find reaction mechanisms and solve the kinetics in organometallic catalysis. Chem Sci 8(5):3843–3851. https://doi.org/10.1039/C7SC00549K
95. Wagner W, Pruss A (2002) The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. J Phys Chem Ref Data 31(2):387–535. https://doi.org/10.1063/1.1461829
96. Wang JM, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25(2):247–260. https://doi.org/10.1016/j.jmgm.2005.12.005
97. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174. https://doi.org/10.1002/jcc.20035
98. Wang L-P, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, Rice JE, Swope WC, Martínez TJ, Pande VS (2017) Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. J Phys Chem B 121(16):4023–4039. https://doi.org/10.1021/acs.jpcb.7b02320

99. Wang L-P, Titov A, McGibbon R, Liu F, Pande VS, Martínez TJ (2014) Discovering chemistry with an ab initio nanoreactor. Nat Chem 6(12):1044–1048. https://doi.org/10.1038/nchem.2099
100. Wang LP, Chen JH, Van Voorhis T (2013) Systematic parametrization of polarizable force fields from quantum chemistry data. J Chem Theory Comput 9(1):452–460. https://doi.org/10.1021/ct300826t
101. Wang LP, Head-Gordon T, Ponder JW, Ren P, Chodera JD, Eastman PK, Martinez TJ, Pande VS (2013) Systematic improvement of a classical molecular model of water. J Phys Chem B 117(34):9956–9972. https://doi.org/10.1021/jp403802c
102. Wang LP, Martinez TJ, Pande VS (2014) Building force fields: an automatic, systematic, and reproducible approach. J Phys Chem Lett 5(11):1885–1891. https://doi.org/10.1021/jz500737m
103. Warshel A (2003) Computer simulations of enzyme catalysis: methods, progress, and insights. Annu Rev Biophys Biomol Struct 32:425–443. https://doi.org/10.1146/annurev.biophys.32.110601.141807
104. Warshel A, Weiss RM (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. J Am Chem Soc 102(20):6218–6226. https://doi.org/10.1021/ja00540a008
105. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) A new force-field for molecular mechanical simulation of nucleic-acids and proteins. J Am Chem Soc 106(3):765–784. https://doi.org/10.1021/ja00315a051
106. Wood MA, Cherukara MJ, Antillon E, Strachan A (2017) Molecular dynamics simulations of shock loading of materials: a review and tutorial. In: Parrill AL, Lipkowitz KB (eds) Reviews in computational chemistry, vol 30, pp 43–92
107. Xie L, Zhao Q, Jensen KF, Kulik HJ (2016) Direct observation of early-stage quantum dot growth mechanisms with high-temperature ab initio molecular dynamics. J Phys Chem C 120(4):2472–2483. https://doi.org/10.1021/acs.jpcc.5b12091
108. Yu H, Hansson T, van Gunsteren WF (2002) Development of a simple, self-consistent polarizable model for liquid water. J Chem Phys 118(1):221–234. https://doi.org/10.1063/1.1523915
109. Zhang H, Yin C, Jiang Y, van der Spoel D (2018a) Force field benchmark of amino acids: I. Hydration and diffusion in different water models. J Chem Inf Model. https://doi.org/10.1021/acs.jcim.8b00026
110. Zhang L, Han J, Wang H, Car R, EW (2018b) Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. Phys Rev Lett 120(14):143001
111. Zheng L, Chen M, Yang W (2008) Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. Proc Natl Acad Sci 105(51):20227–20232. https://doi.org/10.1073/pnas.0810631106
112. Zubarev DY, Rappoport D, Aspuru-Guzik A (2015) Uncertainty of prebiotic scenarios: the case of the non-enzymatic reverse tricarboxylic acid cycle. Sci Rep 5:8009. https://doi.org/10.1038/srep08009

Chapter 7
Application of ReaxFF-Reactive Molecular Dynamics and Continuum Methods in High-Temperature/Pressure Pyrolysis of Fuel Mixtures

Chowdhury Ashraf, Sharmin Shabnam, Yuan Xuan and Adri C. T. van Duin

Abstract Rocket engines, gas turbines, HCCI engines, and other such combustion devices frequently exceed the critical pressure of the fuel or the oxidizer. Modeling of combustion processes at high-pressure operating condition is required to determine the reaction rates based on which chemical kinetic models are developed. The current need is to focus on the transfer from low pressure to high-pressure conditions as this can have a significant effect on the chemistry as well as the reaction rates. The ReaxFF reactive force field method is a computationally feasible method used to study the combustion kinetics of fuels and fuel mixtures at supercritical condition. In this chapter, ReaxFF-MD is used to investigate the effect of a highly reactive fuel on the properties of a less reactive fuel at different levels of concentration, temperature, and density/pressure. The activation energies, based on Arrhenius-type rate laws, are compared with those from Continuum simulations and the limitations of the latter has been elaborated on. The study reveals a pressure/temperature regime and mixing conditions, where simple first-order kinetics-based Arrhenius-type relations cannot be applied. The reason can be attributed to different initial reaction mechanisms and product distributions of the two fuels considered. These results indicate how ReaxFF-based molecular dynamics simulations can provide significant atomistic insights on the combustion properties of fuel mixtures at supercritical conditions, where experiments are difficult to perform.

Keywords Pyrolysis · Arrhenius parameters · Fuel mixture · Toluene · n-Dodecane

Chowdhury Ashraf and Sharmin Shabnam have contributed equally to this manuscript.

C. Ashraf · S. Shabnam · Y. Xuan · A. C. T. van Duin (✉)
Department of Mechanical Engineering, The Pennsylvania State University, State College, PA, USA
e-mail: acv13@psu.edu

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_7

7.1 Introduction

Many combustion devices such as rocket, gas turbines, diesel, and HCCI engines achieve an operating pressure, which is much higher than the critical pressure of either the fuel or the oxidizer. For example, rocket engines using H_2 as fuel can achieve a pressure excess of 100 atm [1], while the critical pressure of the fuel (H_2) and oxidizer (O_2) is 13 atm and 50 atm, respectively [2]. Similarly, pressure inside the combustion chamber of diesel engine increases from ~25 atm during injection to 60 atm [3, 4] after ignition, which is beyond the critical pressure of any of the components of diesel fuels. The physical and the chemical mechanisms of combustion at this pressure/temperature condition is drastically different than its low-pressure counterpart, since molecules are densely packed and may experience van der Waals interactions and caging effects under high pressure [5]. Therefore, a detailed understanding of this complex process is important for the further development of these devices.

When a fuel is sprayed into the combustion chamber, it mixes with a stream of high-pressure fluid, which is primarily the oxidizer. This molecular mixing process has a great influence on autoignition and is a strong precursor of combustion. Additionally, during combustion, the heavier fuel species first undergo chemical decomposition, i.e., pyrolysis and then oxidation. Due to the nonuniform nature of the combustion (both pyrolysis and oxidation) reactions, highly active and fast moving radicals are generated in one location and mix with fuel and oxidizer at a different location inside the combustion chamber and significantly alter the combustion dynamics. Thus, this turbulent mixing of multiple chemical species under high pressure conditions has been an active area of research for the combustion community. However, despite this great interest, the experimental studies [6–12] are limited to only single or binary species mixing instead of multicomponent mixing. Additionally, most studies are based on qualitative visualization with rare quantitative analysis due to the difficulties of performing experiments at supercritical conditions, and therefore cannot be used to validate the proposed kinetic models at high-pressure conditions. Furthermore, these studies except a few [6, 13, 14] have used simple species like N_2, H_2, and O_2 in their turbulent mixing study, which require very simple chemical kinetic models, whereas the real fuels used inside combustion chambers are rather complex ones.

Since the experiments fail to provide sufficient information regarding the complex physical (diffusion, turbulence) and chemical (reaction) coupling of turbulent mixing, computational fluid dynamics (CFD) simulations have emerged as a potential solution to study most of the physical aspects of these problems. Among different CFD techniques available, direct numerical simulation (DNS) has proved to be a powerful tool of combustion and turbulence research [15], since it can resolve any length and timescales of the flow using high order accurate methods. Masi et al. [16] have developed a DNS model to describe the multi- species high-pressure turbulence mixing. Combining this model with the rate of single-step chemical reaction consistent with ignition prediction [17], Jossette Bellan from Jet Propulsion Lab

(JPL) has recently studied both the diffusion [18] and turbulent reaction rate [19] of high pressure mixing. Additionally, Gnanaskandan et al. [5] studied the length of the potential core of round fluid jet entering high-pressure chamber using both DNS and large eddy simulations (LES) methods, and Foster et al. [20] studied the effect of Soret and Dufor cross diffusions in turbulent mixing. The pressure considered in these studies typically ranges from 60 to 100 atm.

All the CFD studies mentioned above mainly focused on the physical aspects of turbulent chemistry; their chemical model was fairly simplistic. For example, Bellan and coworkers [16, 18, 19] only considered single-step chemical reaction for n-heptane oxidation, while Foster et al. [20] considered very simple 7-step 10 species reduced mechanism for a similar problem. However, as mentioned above, a highly coupled relation between chemical and physical properties governs the turbulent mixing phenomenon, which cannot be fully captured by the simple chemical kinetic model.

Chemical kinetics models for various hydrocarbons are well documented in the literature [21]. However, these models are only developed for low-pressure/temperature condition. Additionally, these models were developed only considering temperature dependence on the reaction rates via simple Arrhenius-type rate laws, while neglecting pressure dependence on combustion pathways; which can significantly alter the chemical properties at high pressure. As experiments are difficult to be conducted at supercritical pressure and temperature, we need a computationally feasible method, which can simulate complex combustion reactions without requiring any user input of possible reactions; so that the complete reaction network can be captured. Quantum mechanical (QM)-based ab initio methods are the best choice for accurately predicting the reactions for such systems, however, they have serious system size and simulation time limitation [22]. Since these methods solve the Schrodinger equation to estimate reaction energies and barriers, they can only be applied to small systems typically containing a couple of hundred of atoms for a shorter time scale [22]. Recently, ReaxFF reactive force field method [23] has proven to be a useful alternative of QM-based methods for combustion chemistry simulations [24–31] as it can simulate larger systems for a longer time scale.

The focus of this book chapter is to demonstrate how the ReaxFF reactive force field method can be used as a valuable tool to study combustion kinetics of fuels and fuel mixtures at supercritical condition. The chapter is organized as follows: in Sect. 7.2, we will introduce the ReaxFF method itself while mentioning some of its previous applications in combustion study. In Sect. 7.3, we will elaborate on our systems and simulation setup. Next, some observations on combustions kinetics based on our simulations will be made in Sect. 7.4; and results obtained from ReaxFF simulations will be compared with continuum simulations. Lastly, we will provide some concluding remarks mentioning the pros and cons of ReaxFF simulation for studying similar problems and challenges and opportunities regarding the development of a multiscale ReaxFF/continuum simulation capability for high-temperature/high-density pyrolysis and combustion simulations.

7.2 ReaxFF Background

Since force field based methods are computationally inexpensive and can provide a reasonable agreement with quantum mechanical simulations, they are now being used extensively to get an atomistic insight of complex physical and chemical problems. While potentials such as DREIDING [32], MM [3, 33–35], MM [4, 36] COMPASS [37], etc., have been used to perform atomistic scale simulations of hydrocarbon fuels, mostly to study their thermodynamic properties, due to their rigid connectivity requirement they cannot be used to simulate chemical reactions. However, potentials such as first-[38] and second-generation [39] reactive empirical bond order (REBO), charge-optimized many-body (COMB) potential [40, 41], modified embedded atom method (MEAM) [42] and reactive force field (ReaxFF) [23] can dynamically simulate bond formation and bond breaking and capture chemical reactions. In this chapter, we will keep our discussion limited to the ReaxFF method and its application to high-pressure combustion.

ReaxFF adopts the concept of bond order introduced by Tersoff [43] and Brenner [38] and calculates the bond order of every pair of atoms as a function of their interatomic distances. Based on a certain cutoff value, ReaxFF finds the connectivity between atoms at each step, which enables ReaxFF to simulate dynamic bond breaking and bond formation during simulation. All the connectivity-dependent terms like bond, angle, and torsion energies are calculated based on bond order, thus nonbonded atom pairs do not contribute to these partial energies. Long-range interactions like van der Waals and Coulomb are not connectivity dependent and calculated between every pair of atoms, any excessive short-range nonbonded interactions are avoided by including a shielding term. ReaxFF calculates atomic charges using a geometry-dependent charge calculation scheme and uses electronegativity equalization method (EEM) [44] for this purpose. Additionally, for long-range interactions ReaxFF uses a cutoff distance (which is typically set to 10 Å) to reduce the computational cost. To eliminate any energy discontinuity and reduce the range of the Coulomb interactions, a seventh-order Taper function is employed [23]. Equation (7.1) shows the different energy components of ReaxFF total energy, while Fig. 7.1 demonstrates the complicated internal scheme of ReaxFF. A more detailed description of the ReaxFF energy terms can be found in the previous literature [23, 24].

$$E_{\text{system}} = E_{\text{bound}} + E_{\text{over}} + E_{\text{under}} + E_{\text{lp}} + E_{\text{val}} + E_{\text{tor}} + E_{\text{vdWaals}} + E_{\text{coulomb}} \quad (7.1)$$

where E_{bond}, E_{over}, E_{under}, E_{lp}, E_{val}, E_{tor}, E_{vdWaals}, E_{coulomb} represent bond energy, over-coordination energy penalty, under-coordination stability, lone pair energy, valence angle energy, torsion angle energy, van der Waals energy, and coulomb energy, respectively.

ReaxFF force field parameters are trained against QM-calculations in describing energies and barriers for chemical reactions while solving Newton's equation of motion to generate a dynamic description of complex reactive systems. Thus, ReaxFF-based molecular dynamics (MD) simulations are a number of magnitudes

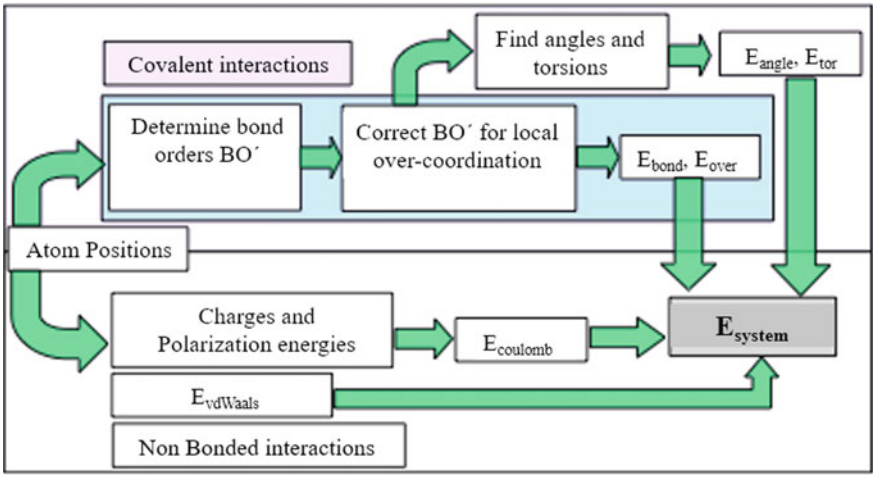


Fig. 7.1 Internal scheme of ReaxFF

faster than QM-based simulations while approaching the accuracy of those methods. Additionally, the inclusion of polarizable charge description and bond-order-dependent 3- and 4-body terms in ReaxFF makes it uniquely applicable for both metallic and covalent systems. Thus, ReaxFF has now become a great computational tool to study many reactive systems; at the same time, it has enabled researchers throughout the world to investigate previously inaccessible computational chemistry problems. Apart from studying combustion problems, ReaxFF method has been extensively used to investigate a wide range of applications in materials [45–50] catalysts [51, 52], and other chemical systems [53–57].

Combustion, being a complex reactive system, requires an atomistic-level understanding of the intricate details regarding the underlying reactions and species to facilitate better engine and fuel design. To this end, the first ReaxFF combustion force field (CHO-2008) was developed by Chenoweth et al. [24] in 2008. Since then, it has been implemented in a wide range of applications in scientific community for studying pyrolysis and oxidation of variety of fuels, for example, JP-10 [25] n-dodecane [58], 1-heptane [30], n-octanol [29], toluene [59], Illinois no. 6 coal [26], 1,6-dicyclopropane-2,4-hexyne [60] and liginin [61], and many more. The recent development of GPU-enabled ReaxFF [62] has made it suitable to simulate large complex systems like coal [63] and lignin [64] pyrolysis. Apart from the hydrocarbon fuel, this description has also been used to investigate a wide range of aspects related to carbon-based materials including the oxidation of graphene [65], structural and chemical properties of graphene oxide [66], chemomechanics of crack propagation in graphene [67] and dynamics of carbon nano-onion formation [68]. For a more detailed review of this force field, we refer to the recent paper by Dontgen et al. [31].

Though ReaxFF CHO-2008 description was very good in describing the initial pyrolysis and oxidation of large hydrocarbons, where the initial combustion process

is mainly dominated by the pyrolysis process, it failed to describe the chemistry of smaller hydrocarbon properly [28]. Additionally, CHO-2008 parameter set was not transferable to study mechanical properties of condensed phase carbons [69]. Furthermore, the O_2 in CHO-2008 description is highly reactive, which can abstract H from hydrocarbon chain at a faster-than-expected rate. This might send us a false alarm of system initiation during the simulation. Srinivasan et al. [69] derived an improved set of ReaxFF carbon parameters for describing carbon-condensed phases and taking those parameters on board, Ashraf et al. [28] have recently published an extended version of ReaxFF combustion force field (CHO-2016), which addresses the limitations of CHO-2008 description while retaining the overall quality of CHO-2008 description for larger hydrocarbon combustion. Thus, a single force field can now be used to study not only both the condensed and the gas phase carbon, but also the combustion of any hydrocarbon irrespective of fuel size or structure. This ensures a greater transferability of ReaxFF force field to study any combustion related problems.

Chemical reactions occurring in combustion are considered 'rare events' in MD simulations as not every collision between molecules lead to a reaction. Additionally, reactive MD simulations are expensive as it needs to calculate the bond order at every time step, thus the simulations are limited to only a couple of nano-seconds or less, whereas the timescales associated with combustion reactions at experimental conditions are micro-second-level phenomenon. MD simulations are thus typically carried out at high temperatures and pressures to accelerate the system dynamics and capture the appropriate chemical reactions. Therefore, ReaxFF-based MD simulations are suitable for studying the complex chemistry at high-pressure condition although sometimes the pressure in ReaxFF simulation may exceed the highest possible pressure that a rocket engine can survive. However, despite this shortcoming, it is arguably the most appropriate currently available method to study the high-pressure combustion chemistry. Almost all of the ReaxFF combustion studies mentioned above were performed at very high pressure to study the kinetics of different fuels. Additionally, Ashraf et al. [27] have recently demonstrated that ReaxFF-MD simulations can also be used to study dynamic properties like ignition front speed at supercritical conditions.

In this chapter, we will demonstrate how we can use ReaxFF-MD simulations to calculate Arrhenius parameters for different hydrocarbons at very high pressure and temperature condition. Since the hydrocarbons do not typically exist in a single-component form, we will also investigate how the blending of a highly reactive hydrocarbon with a less reactive one alters the combustion properties of the mixture. This enabled us to identify pressure/temperature and mixing conditions, where the simple first-order kinetics and Arrhenius-type relation fails to prevail, indicating that more complex relationship is required to calculate mixture activation energies. Since experiments are difficult to conduct and measurements are hard to perform at very high pressure/temperature condition, these information from ReaxFF-MD simulation will certainly open up the possibility to study the mixing effects of multiple hydrocarbons at supercritical condition.

7.3 Simulation Details

Since typical transportation fuels used for combustion are highly complex mixture of various hydrocarbons, for modeling purpose, Kim et al. [70] developed two different surrogates for diesel so that various physical and chemical processes inside of a diesel engine can be replicated. The surrogate fuels are either a mixture of n-dodecane/*iso*-cetane/methylcyclohexane/toluene or a mixture of n-dodecane/*iso*-cetane/decalin/toluene, where n-dodecane is the most reactive one and toluene is the least. As such, for our study, we decided to investigate the effect of n-dodecane addition on toluene pyrolysis using the reactive molecular dynamics (RMD) simulations. We also used two different densities for the overall system, to investigate the effect of density/pressure on the pyrolysis of the mixtures. All of the MD simulations in this section were performed employing the ReaxFF method with a constant number of atoms (N) in a constant volume (V) while keeping the temperature constant (T) using a thermostat, as described by the NVT-MD ensemble. Also, we used the recently developed combustion force field by Ashraf et al. [28] in this study.

7.3.1 Single-Component System

First, to determine the applicability of the force field, toluene and n-dodecane pyrolysis have been investigated independently using homogeneous system. We placed 40 toluene molecules in two cubic periodic unit boxes of dimension 31.20 Å × 31.20 Å × 31.20 Å and 25.00 Å × 25.00 Å × 25.00 Å, where the overall system densities are 0.2 kg/dm^3 and 0.4 kg/dm^3 respectively. Similarly, in case of n-dodecane, we placed 40 molecules in cubic boxes of dimension 38.39 Å × 38.39 Å × 38.39 Å and 30.47 Å × 30.47 Å × 30.47 Å, respectively, to generate systems of similar densities like toluene. Next, each system is energy minimized and equilibrated using NVT simulation for 10 ps at a temperature of 1500 K. Test cases has been run beforehand to ensure that no thermal decomposition occurs at this temperature for both toluene and n-dodecane. After equilibration, 10 different initial configurations of the system were selected to perform a series of NVT-MD simulations. For each case, the system temperature was varied from 2000 to 2600 K at 100 K interval. The average initial pressure of the system for the first 5 ps of the NVT-MD simulations was in the range of 26–75 MPa. Although the use of such high temperatures and pressures is rarely seen in experiments, this is essential in MD simulations to keep the computational time within a reasonable scope. The high temperature and subsequent high pressure results in a larger number of collisions in the system and thereby reduces the reaction time (Fig. 7.2).

During all of the simulations, the time step has been kept as 0.1 fs, which is appropriate for describing hydrocarbon reaction mechanism at high temperatures [24]. Since n-dodecane is highly reactive, a simulation time of 50 ps was enough to calculate the Arrhenius parameters. On the other hand, the toluene simulations

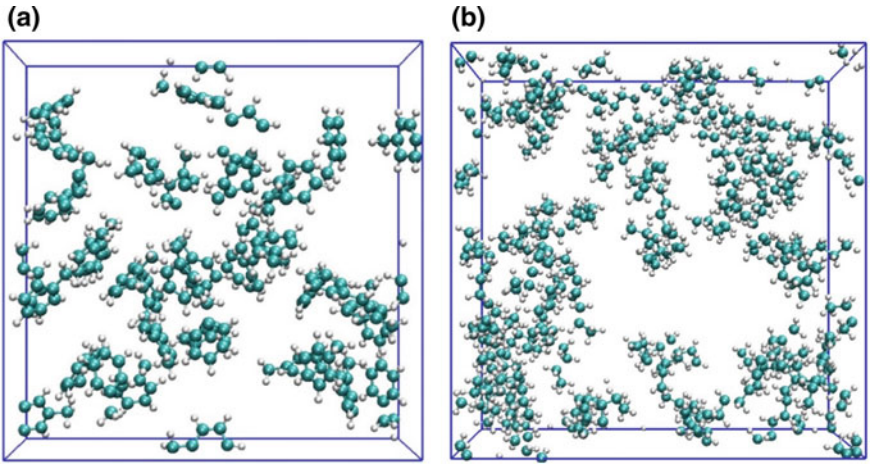


Fig. 7.2 Snapshots of representative single-component system of equilibrated **a** toluene, and **b** n-dodecane with density of 0.2 kg/dm^3. The carbon and hydrogen atoms are displayed in cyan and white, respectively

were run for 200 ps due to the less reactive nature of toluene. The results from the simulations with 10 different starting configurations were then averaged to obtain the reactant decomposition and overall product distribution. This information is then used to get the Arrhenius parameters for individual components.

7.3.2 n-*Dodecane and Toluene Mixture (Multicomponent System)*

Once we validated the quality of the force field against the single-component system, we moved on to explore the effect of n-dodecane presence to the pyrolysis of toluene at high-pressure/temperature condition which is the main focus of this chapter. To do this, we performed a series of NVT-MD simulations of toluene and *n*-dodecane mixtures, where these two hydrocarbons are introduced in different ratios. Table 7.1 represents a summary of all the input configurations for these systems.

All of the systems studied were equilibrated using NVT-MD after placement of the molecules at a temperature of 1500 K for 10 ps. Then, 10 different initial configurations of the molecules were generated and a series of NVT-MD simulation at temperatures 2000, 2100, 2200, 2300, 2400, 2500, 2600 K were carried out for all of them. The total simulation time was 200 ps with the time step being 0.1 fs. We repeated all the simulations changing the system density to 0.4 kg/dm^3 to study and compare the effect of a higher density/pressure of the system.

For all the simulations, whether single-component or multicomponent, the following parameters were the same. To keep the system temperature constant, the

Table 7.1 Initial configurations of different *n*-dodecane and toluene mixtures

n-Dodecane: Toluene	% of n-dodecane in the mixture	Density (kg/dm^3)	Temperature range (interval)	Box dimension
1:40	2.44	0.2	2000 K–2600 K (100 K)	31.75 Å × 31.75 Å × 31.75 Å
5:40	11.11	0.2	2000 K–2600 K (100 K)	33.52 Å × 33.52 Å × 33.52 Å
10:40	20.00	0.2	2000 K–2600 K (100 K)	35.50 Å × 35.50 Å × 35.50 Å
20:40	33.33	0.2	2000 K–2600 K (100 K)	38.91 Å × 38.91 Å × 38.91 Å
40:40	50.00	0.2	2000 K–2600 K (100 K)	44.34 Å × 44.34 Å × 44.34 Å

temperature was controlled by the Berendsen thermostat using a temperature damping constant of 100.0 fs. For molecular recognition, we used a bond-order cut off of 0.3. The choice of bond-order cutoff does not alter the simulation pathway, but it is only used to identify the intermediates and products formed during the MD simulations.

7.3.3 Continuum Simulations

In addition to the ReaxFF-based simulations described above, zero-dimensional (time-evolving), continuum-scale pyrolysis simulations of the same series of test cases (see Table 7.1) are performed. The intent is to highlight the deficiency of Arrhenius-type chemical kinetics in predicting chemical evolutions (e.g., pyrolysis) under high-pressure high temperature conditions, by comparing the results obtained from both approaches. For this purpose, the initial temperatures, density, and mole fractions of n-dodecane and toluene in these continuum simulations are set to be the same as those used in the molecular dynamics simulations as described in Table 7.1, and these simulations are performed with constant volume and constant temperature to best duplicate the same running conditions as for the ReaxFF simulations. In these simulations, a cubic equation of state is used to account for real gas effects [71], and a widely used chemical kinetic mechanism [27] with Arrhenius-type rate coefficients is used. This chemical mechanism contains 179 species and 1895 chemical reactions (forward and backward reactions counted separately), and it has been extensively validated for jet fuel surrogate components under lower pressure and temperature ranges (up to 40 bars and 1250 K) [72].

7.4 Results and Discussion

7.4.1 *Kinetic Analysis of Toluene and n-Dodecane Pyrolysis as Single-Component System*

For single-component systems of toluene and n-dodecane, first-order kinetics has been used to study the thermal decomposition. We used the consumption rate from the ReaxFF-MD simulations in this analysis and the number of reactant molecules has been chosen to represent the reactant concentration. From the NVT-MD simulations at different temperatures, the decomposition of toluene and n-dodecane has been found to change as a function of time and temperature. Using this information, we used integrated first-order rate law to determine the rate constant at each temperature:

$$\ln(N_0) - \ln(N_t) = \mathrm{kt} \tag{7.2}$$

where N_0 is the number of molecules initially in the system and N_t is the number of molecules at any time t. At each temperature, the quantity $\ln(N_0) - \ln(N_t)$ has been plotted against time and the rate constant has been determined from the slope of the linear fitting of the plot. We used these rate constants from different temperatures in the Arrhenius plot of toluene and n-dodecane as shown in Fig. 7.3. The activation energy (E_a) and the pre-exponential factor (A) were calculated with the help of a linear fit of the plot and the Arrhenius equation described as

$$k = A \exp(-E_a/\mathrm{RT}) \tag{7.3}$$

Next, we compared the values of Arrhenius parameters obtained from ReaxFF-MD simulations with their experimental counterparts, which are listed in Table 7.2.

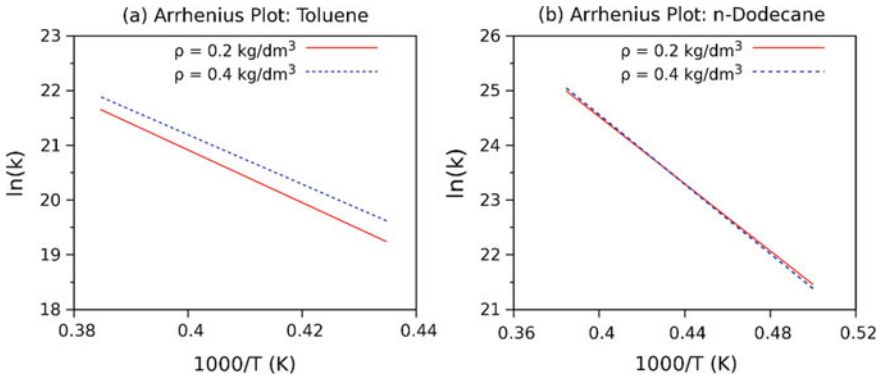


Fig. 7.3 Arrhenius plot for calculated and fitted rate constants of **a** toluene, **b** n-dodecane as single-component systems with density $\rho = 0.2$ and 0.4 kg/dm^3

Table 7.2 Fitted Arrhenius equation and parameters for n-dodecane and toluene

Molecule	ReaxFF			Experiment [73, 74]	
	Density (kg/dm^3)	E_a (kcal/mol)	A (1/s)	E_a (kcal/mol)	A (1/s)
Toluene	0.2	95.71	28.3×10^{16}	88.9–97.0	$0.28 - 1 \times 10^{16}$
	0.4	89.90	11.59×10^{16}		
n-Dodecane	0.2	60.94	0.95×10^{16}	61.32	0.12×10^{16}
	0.4	63.11	1.53×10^{16}		

It is important to note that, the experimental values are obtained at a much lower pressure and temperature condition than ReaxFF simulations. For example, Colket et al. [73] investigated the pyrolysis of toluene at a temperature range of 1200–1850 K with a total pressure of approximately 1 MPa. According to them, the formation of benzyl radical is the most dominant reaction pathway, which is also accompanied by the pathway that leads to phenyl radical formation. The activation energies mentioned in Table 7.2 are for these two key reaction pathways.

Liu et al. [74] investigated the supercritical thermal cracking of n-dodecane at a temperature range 700–800 K and pressure of 3–4 MPa. For both toluene and n-dodecane, the results obtained in this work shows good agreement with those obtained from the experiments. Wang et al. [58] investigated the pyrolysis of n-dodecane using ReaxFF-MD simulations and the force field parameters developed by Chenoweth et al. [24] using a temperature range of 2100–3000 K. The activation energies of n-dodecane derived in the study were 63.68 and 66.14 kcal/mol for densities $\rho = 0.17$ and 0.33 kg/dm^3. The activation energy of n-dodecane derived in this work using the recently developed force field by Ashraf et al. [28] showed a better agreement with the experimental results than those derived by Wang et al. [58] using Chenoweth et al. [24] force field. Though in our simulations pressure and temperature are much higher than their experimental counterparts, the reasonable agreement with experimental values indicates that the simple Arrhenius type relation holds good for single-component system, even at a very high pressure and temperature.

Figure 7.4 shows the evolution of toluene and dodecane molecules as a function of time for both of the densities 0.2 and 0.4 kg/dm^3. At temperatures 2000, 2100, and 2200 K, the number of toluene decomposed was very low, so the simulation results are not used in calculations of the Arrhenius parameters since these would include a significant statistical uncertainty. The figures indicate the reactivity of n-dodecane as opposed to that of toluene. We observed that at 2000 K, n-dodecane began to decompose within 5 ps of the simulation time; at temperatures higher than 2000 K the decomposition began in less than 5 ps. While in the case of toluene, at temperatures 2300, 2400, and 2500 K, the decomposition does not begin before 20 ps. Only at a high temperature of 2600 K, toluene starts to decompose within 5 ps of the simulation time. The trend of decomposition initiation is observed in both 0.2 and 0.4 kg/dm^3 density cases.

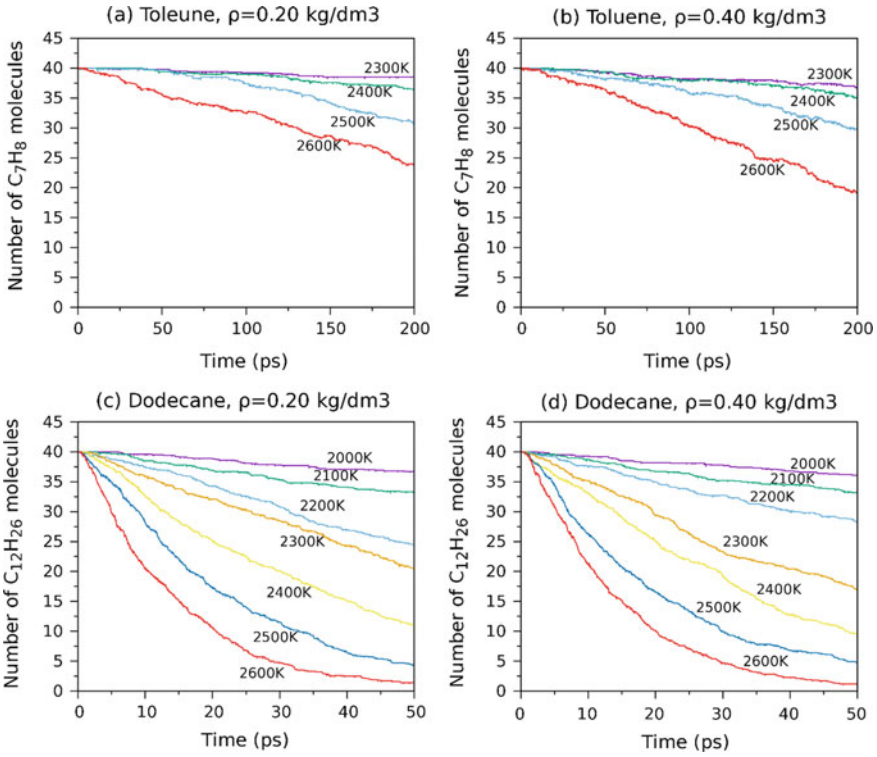
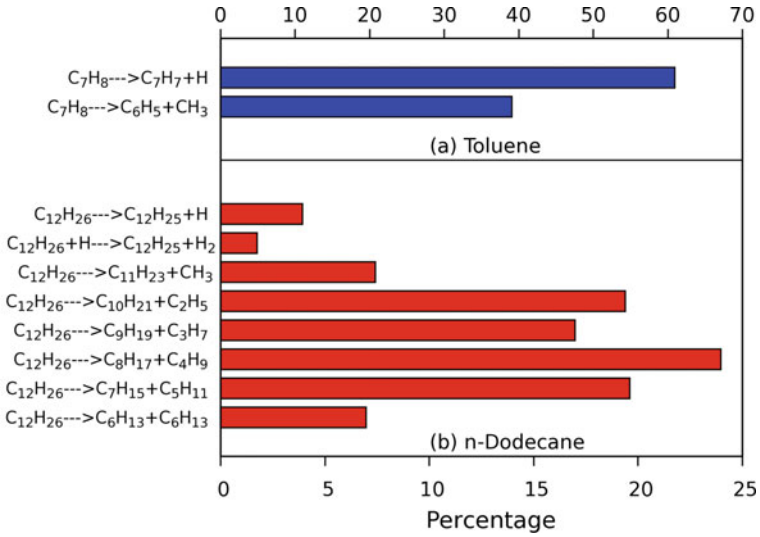


Fig. 7.4 Time evolution of the number of toluene molecules in single-component systems with density **a** $\rho = 0.2$ kg/dm^3 and **b** 0.4 kg/dm^3, and n-dodecane molecules in single-component systems with density **c** $\rho = 0.2$ kg/dm^3 and **d** 0.4 kg/dm^3

As mentioned above, experimentally, the initiation mechanism for toluene pyrolysis is dominated by decomposition of toluene into benzyl (C_7H_7) and hydrogen atom with activation energy 88.9 kcal/mol [71]. The hydrogen atom generated from this toluene decomposition reacts with another toluene molecule to produce another benzyl radical and hydrogen molecule. Another reaction pathway is the formation of phenyl (C_6H_5) and methyl radical (CH_3). A similar initiation mechanism is observed in ReaxFF simulations shown in Scheme 7.1a. Figure 7.5a demonstrates the distribution of toluene and its major decomposition species at 2600 K. The decomposition of toluene starting after 10 ps is on the rise accompanied by a gradual rise of both benzyl and hydrogen molecule.

The pyrolysis of n-dodecane is dominated by radical production in ReaxFF as shown in Fig. 7.5b. This is in agreement with experiment [75]. The initiation reactions principally involve C–C bond fission to produce two alkyl radicals as shown in



Scheme 7.1 Initiation reaction mechanism and their relative percentages occurred in **a** toluene, **b** n-dodecane pyrolysis in a single-component system revealed by ReaxFF–MD simulations at 2600 K

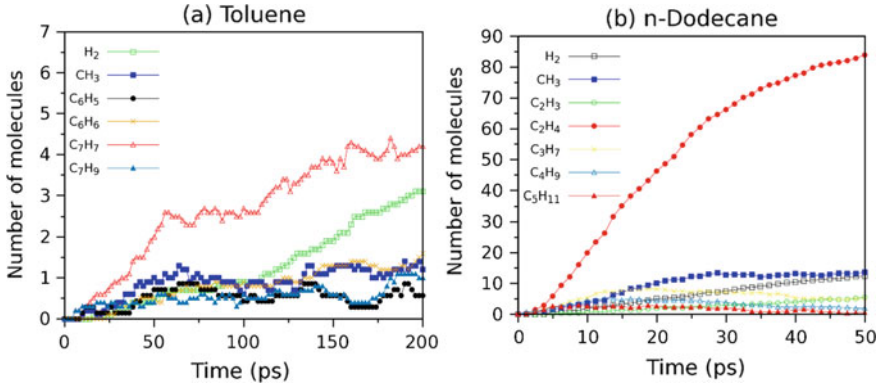


Fig. 7.5 Time evolution of the major species generated in the pyrolysis of **a** toluene, **b** n-dodecane in single-component systems with density $\rho = 0.2$ kg/dm^3 and temperature $T = 2600$ K

Scheme 7.1b. On rare occasion, we observe abstraction of hydrogen from n-dodecane. As the simulation proceeds, larger alkyl radicals undergo further decomposition, which increases the number of smaller radicals such as CH_3, C_2H_5, C_3H_7 and creates hydrogen molecules and stable double-bond-containing molecules like C_2H_4 and C_3H_6.

7.4.2 *Pyrolysis of Toluene and* n-*Dodecane as Multicomponent System*

7.4.2.1 Kinetic Analysis

To investigate the change in activation energy of a less reactive fuel component due to the presence of a more reactive molecule, we analyzed a multicomponent system of n-dodecane and toluene using ReaxFF NVT-MD simulations. The rate constants associated with the decomposition of toluene in the mixture are determined using the first-order kinetics and the number of toluene molecules in (7.3). We derived the Arrhenius parameters in (7.3) from the fitted plot of these rate constants at various temperatures.

Figure 7.6 shows the activation energies (E_a) of toluene in the mixtures as a function of the percentage of n-dodecane introduced in the mixture for both ReaxFF-MD and continuum simulations. The zero-percentage result indicates the pyrolysis of toluene as a single component. As the percentage of n-dodecane increases, the activation energy of toluene decreases rapidly. According to ReaxFF-MD results, the presence of a single n-dodecane molecule during the thermal decomposition of toluene reduces the activation energy from 95.69 to 81.68 kcal/mol (14.64% reduction). If n-dodecane molecule number is increased to 5, the activation energy reduces further to 66.78 kcal/mol. When the percentage of n-dodecane is 33.33 in the mixture (n-dodecane and toluene ratio 20:40) we see that the activation energy of toluene (52.98 kcal/mol) even drops below that of single-component n-dodecane pyrolysis (52.98 kcal/mol) with the same density. The results from continuum simulations show similar behavior. Both set of results show good qualitative agreement in predicting the mixture activation energy, when the n-dodecane concentration is low (less than 20%). This suggests the capability of ReaxFF-MD to predict the activation energy of toluene within this mixing ratio. However, the results show larger deviations when n-dodecane concentration is higher than 20%. Unlike ReaxFF-MD, the activation energy of toluene from continuum simulation in all of the mixtures remains higher than n-dodecane only system. This indicates that with the presence of higher amount of n-dodecane, the first-order kinetics-based Arrhenius method fails to predict the mixture activation energy with sufficient accuracy.

7.4.2.2 Effect of Toluene Presence on the Pyrolysis of *n*-Dodecane

Before investigating what is causing the dramatic decrease of activation energy of toluene at the presence of n-dodecane, we want to check if the presence of toluene also affects the activation energy of n-dodecane. To get an overview, here we have compared the single-component system of n-dodecane and the mixture of 40 toluene and 40 n-dodecane molecules. The rate constants were determined in this case using (7.3) and the time evolution of only n-dodecane molecules in the mixture. We calculated the activation energy similarly using (7.3) and the linear fitting of Arrhenius plot

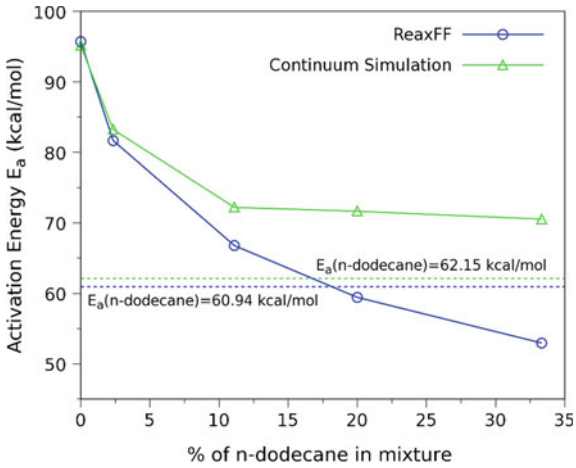


Fig. 7.6 Activation energies of toluene at various compositions of n-dodecane and toluene mixture from ReaxFF-MD simulations and Continuum simulations. The green and blue dashed line represents the activation energy of n-dodecane calculated by ReaxFF and Continuum simulations, respectively

of rate constants at different temperatures. The value of activation energy is found to be 58.56 kcal/mol, which is very close to the activation energy 60.94 kcal/mol found in Sect. 7.4.1. Since the activation energy of n-dodecane in the mixture is close to that obtained from individual n-dodecane system, we can conclude the pyrolysis of n-dodecane is mostly independent of the presence of toluene.

Figure 7.7 shows the time evolution of n-dodecane molecules in both single-component system and a 40:40 mixture of toluene and n-dodecane at 2300 K and at 2600 K. The time-dependent number of n-dodecane molecules in the system is averaged from 10 different NVT-MD simulations of the same system but with different initial configurations. The simulation time of n-dodecane in extended to 100 ps to get a better comparison of n-dodecane decomposition. The figures indicate that there is no significant change in the pyrolysis of n-dodecane even though the two systems are different. We can conclude from these results that the initiation mechanism of n-dodecane pyrolysis is mostly independent and is not affected by the presence of a less reactive fuel like toluene.

7.4.2.3 Effect of n-Dodecane Presence on the Pyrolysis of Toluene

Although the pyrolysis of n-dodecane is mostly unaffected by the presence of toluene in the mixture, this is not the case for toluene itself. Figure 7.8 shows the number of toluene molecules decomposed in the single-component system and the mixtures with different concentrations of n-dodecane for a total simulation time of 200 ps.

A preliminary analysis of the product distribution observed during the decomposition of toluene at temperatures from 2000 to 2600 K shows that as the number of n-dodecane in increased in the mixture, the number of toluene decomposed also increases, even as the number of toluene molecules introduced initially is constant (40). This indicates the existence of several simultaneous processes during the ther-

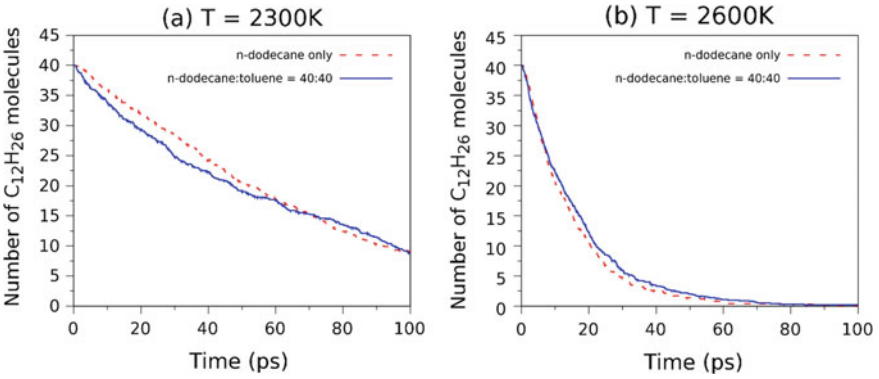


Fig. 7.7 Time evolution the numbers of n-dodecane molecules at temperatures **a** $T = 2300$ K and **b** $T = 2600$ K

mal decomposition of toluene. At low temperatures, since the decomposition of toluene in single-component system is very low, the role of n-dodecane plays a much more significant role in further decomposition of toluene.

The Arrhenius plots for the pyrolysis simulations of toluene and n-dodecane mixtures with increasing concentration of n-dodecane is shown in Fig. 7.9. According to this, at high temperature (2500, 2600 K) the rate constant has been mostly unaffected by the inclusion of n-dodecane in the mixture. However, at low temperatures (2000, 2100, 2200, 2300, 2400 K) the rate constants gradually increase with increasing number of n-dodecane in the mixture. The result is an overall upward shift of the Arrhenius plot at lower temperatures. Additionally, with the increase of n-dodecane molecules in the system, lower temperatures become accessible to calculate the rate constants. For example, in toluene-only case, enough toluene was not decomposed to calculate a rate constant at 2200 K, however, with the presence of n-dodecane, we were able to calculate rate constant even at 2000 K. Thus, the decrease in the slope of the plot which gives the value of—E_a/R implies lower activation energies for increasing n-dodecane number in the mixture.

Though further investigation is required, the different mechanism of toluene and n-dodecane pyrolysis as single-component systems might explain the underlying mechanism responsible for lowering the activation energies of the mixture. Figure 7.4 shows that, at lower temperature (2000 K), roughly 15% of n-dodecane molecules decompose within the 50 ps simulation time of n-dodecane only system, while it reaches to over 90% at higher temperature (2600 K). The decomposition behavior of n-dodecane is observed in the mixtures too; the process is accompanied by the production of a large radical pool. The main species in the radical pool consists of a combination of different alkyl radicals ranging from C_1–C_{12} initially and smaller radicals such as methyl (CH_3), ethyl (C_2H_5), propyl (C_3H_7) and in the later part of the simulation, hydrogen (H) radical etc. At low temperatures, the intramolecular reaction between the small free radicals and toluene molecules cause further decomposition of toluene into benzyl radical (C_7H_7) and H. For example, the reaction

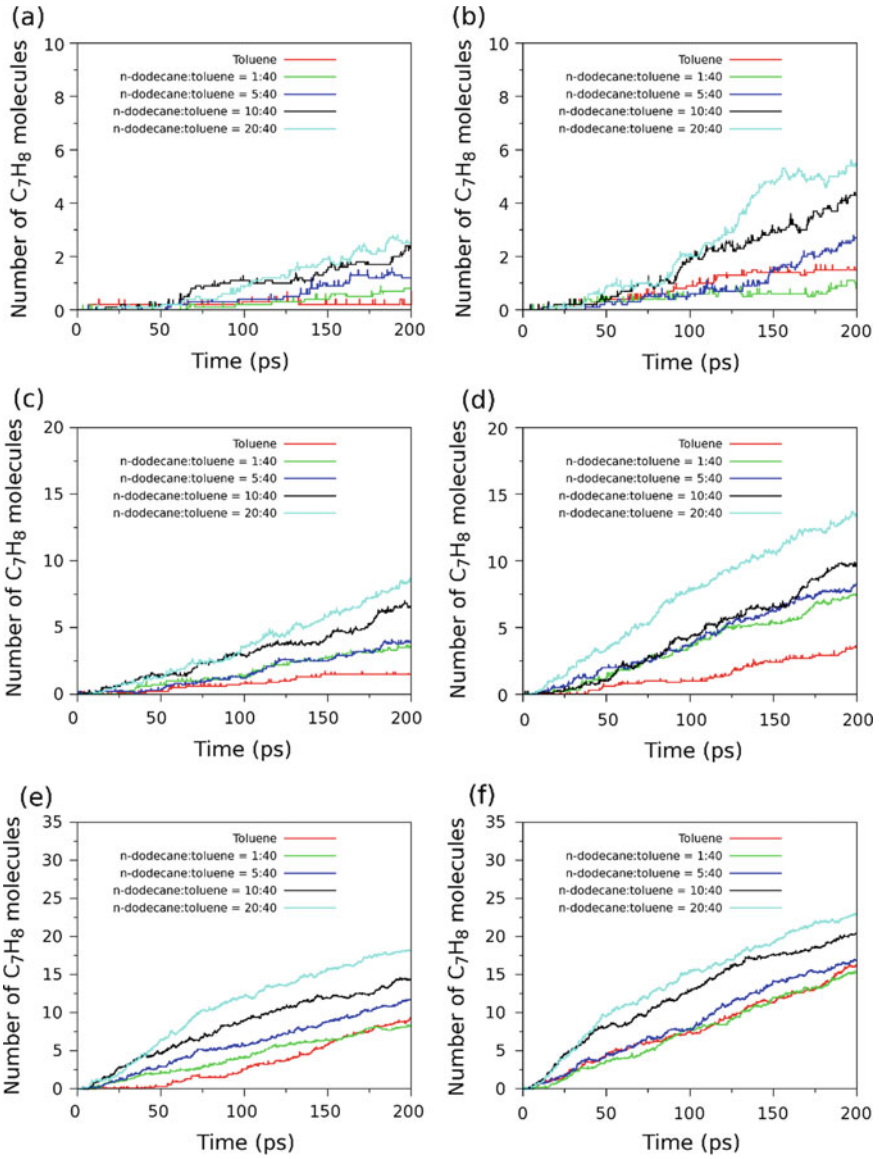


Fig. 7.8 Time evolution the numbers of toluene molecules for single-component system and in mixtures with various concentrations of n-dodecane at temperatures **a** $T = 2100$ K, **b** $T = 2200$ K, **c** $T = 2300$ K, **d** $T = 2400$ K, **e** $T = 2500$ K, and **f** $T = 2600$ K

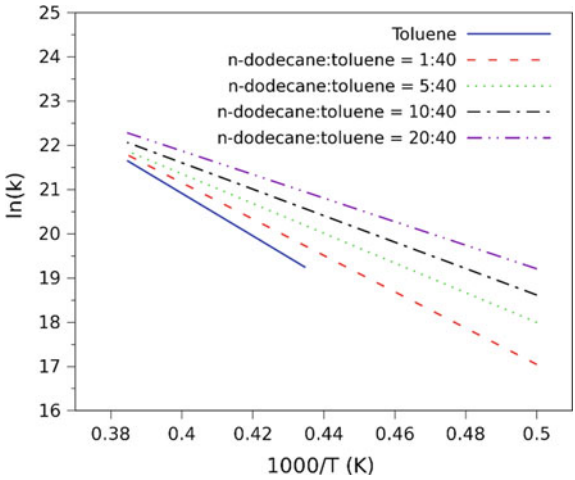


Fig. 7.9 Arrhenius plot with the fitted natural logarithm of rate constant versus the inverse of temperature for toluene as single-component system and in mixtures with various concentrations of n-dodecane at densities of $\rho = 0.2$ kg/dm^3

between methyl (CH_3) and toluene (C_7H_8) to generate methane (CH_4) and benzyl has an energy barrier of only 11.1 kcal/mol [70]. With the presence of n-dodecane, such reactions with low barriers cause further decomposition of toluene in a shorter time.

7.4.2.4 Density/Pressure Effect on the Pyrolysis of Toluene and n-Dodecane Mixture

To investigate the density/pressure effect on the pyrolysis of toluene and n-dodecane mixture we performed additional simulations for the systems with a high density of 0.4 kg/dm^3. The rate constants and the Arrhenius parameters are calculated based on toluene using the same method in Sect. 7.4.1. Figure 7.10a shows the comparison of the activation energies from ReaxFF simulations for density 0.2 and 0.4 kg/dm^3 with increasing percentage of n-dodecane in the mixture. Overall, the results show that the activation energy for a system density of 0.4 kg/dm^3 is less than that for a system density of 0.2 kg/dm^3 when the percentage of n-dodecane is low in the mixture (2.34 and 11.11%). For 10 and 20 molecules of n-dodecane in the mixture, the density/pressure effect is insignificant and the activation energies for both cases are almost similar. For the high density case, we observe that the presence of only 11.11% n-dodecane in the mixture results in a surprisingly lower activation energy of toluene than that of n-dodecane as a single-component system. Figure 7.10b shows results for similar cases obtained through continuum-scale pyrolysis simulations. At higher density, the activation energy of toluene decreases sharply from 95.17 to just 62.79 kcal/mol (34%) with the addition of only one n-dodecane molecule in the toluene-only system. After that, the activation energy keeps on increasing as the percentage of n-dodecane increases. Similar to ReaxFF-Md results, when n-dodecane number in the mixture is 10 and 20, the results are very similar for the

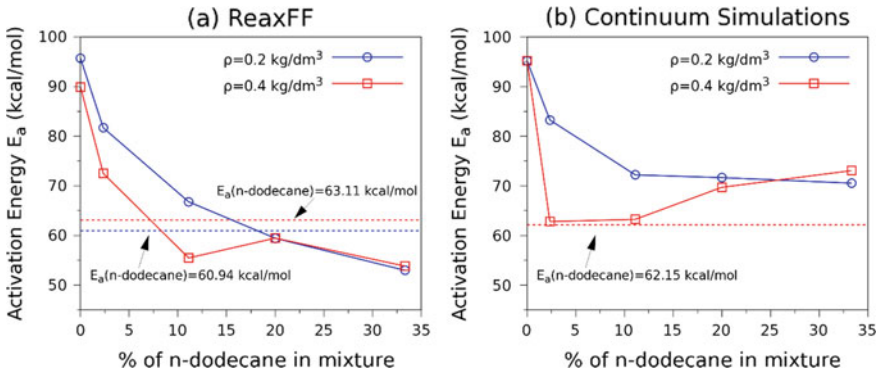


Fig. 7.10 Activation energies of toluene at various compositions of n-dodecane and toluene mixture from **a** ReaxFF-MD simulations. **b** Continuum simulations with two different densities. The blue and red dashed line represents the activation energy of only n-dodecane in a system density of 0.2 kg/dm^3 and 0.4 kg/dm^3, respectively

two density cases. These results are in qualitative agreement with the ReaxFF-MD results in the sense that at higher concentrations of n-dodecane in the mixture, the toluene activation energy does not show any dependency on either the amount of n-dodecane in the mixture or the overall system density.

These ReaxFF-MD results indicate that apparent first-order kinetics is no longer suitable for calculation of activation energy of toluene in the mixture at high density/high pressure system. We can also conclude that when the ratio of n-dodecane and toluene exceeds a certain threshold (in our case more than 5:40), the method has limitations in accurately capturing the effect of the n-dodecane in the activation energy of toluene. It is evident from the results that the decomposition of toluene follows a more complex mechanism in the mixture in high density/pressure than in low density/pressure. Therefore, in order to determine the activation energy of a less reactive component such as toluene in a mixture with a highly reactive molecule such as n-dodecane, we need to look beyond the simple first-order kinetics and the Arrhenius equation.

A complete investigation of this surprising behavior of toluene decomposition with the presence of n-dodecane is beyond the scope of this study, as the purpose of this book chapter is to demonstrate the capability of ReaxFF-MD simulation for high pressure/temperature combustion. However, we would like to propose a hypothesis for this interesting finding, which requires more rigorous studies to confirm. According to our hypothesis, the main reason for low activation energy of toluene in the presence of n-dodecane is mostly related to the very different initiation mechanism of toluene and n-dodecane decomposition, when they are studied as a single-component system. As mentioned several times in this work, n-dodecane decomposition leads to a large alkyl radical pool build up, and at high density, these radicals can easily find a toluene molecule to decompose as they are diffusion limited at that density. Since n-dodecane can generate radicals even at lower temperatures, this leads to higher

toluene decomposition, which significantly effects toluene activation energy. That is why, a small amount of n-dodecane can significantly reduce toluene activation energy when simple first-order kinetics and pressure-independent Arrhenius equation is used at high density (0.4 kg/dm^3). However, at a relatively lower density such as 0.2 kg/dm^3 and lower number of n-dodecane in the system, active alkyl radicals has more space to move around and it takes longer time for them to find a toluene molecule to collide with as the alkyl pool build up is not that significant. However, even at low density, simple Arrhenius equation is not applicable if the initial number of n-dodecane is high as it significantly increases the number of alkyl radicals in the system. Since we used first-order kinetics to calculate the rate constant, while the toluene decomposition is mostly the secondary reaction in the system initiated by active radicals, our analysis no longer holds at this situation.

Additionally, we hypothesize that, at higher temperatures, toluene decomposition shows more temperature dependence than density/pressure dependence. Also, at high temperature, the presence of n-dodecane does not greatly affect the rate constant of toluene which is supported in Fig. 7.9 too. At higher temperatures, probably toluene in mixture decomposes in a similar manner as the toluene-only system as those high energy decomposition routes become accessible. As these routes no longer require an effective collision, toluene mainly decomposes in those routes.

As mentioned above, our hypothesis requires further investigation to confirm, however, with the aid of ReaxFF-MD simulation we were able to gain some useful information of high pressure system, which was never explored before.

7.5 Conclusions

In this work, we applied ReaxFF method to study the pyrolysis of toluene and n-dodecane as both single-component systems and mixtures subjected to high temperature and pressure conditions. A series of NVT-MD simulations for a temperature range of 2000–2600 K has been performed to investigate the decomposition of the reactant molecules and to calculate the temperature-dependent rate constants. Two different densities, 0.2 and 0.4 kg/dm^3 of the system were used to investigate the effect of a higher density/pressure on the Arrhenius parameters. First-order kinetics was used to calculate the Arrhenius parameters based on the toluene decomposition and corresponding rate constants calculated at different temperatures. In case of mixtures of toluene and n-dodecane, zero-dimensional (time-evolving) continuum-scale pyrolysis simulations for the same tests cases have been performed and the results were then compared to those obtained from ReaxFF-MD.

For individual systems of toluene and n-dodecane, the activation energy and the pre-exponential factor calculated in both high- and low-density cases show good agreement with previous experimental results. For mixtures having relatively lower density (0.2 kg/dm^3) of the system, we observed from both of the methods that the overall effect of n-dodecane is to reduce the activation energy of toluene. The two methods showed satisfactory correlation when the n-dodecane concentration

is below 20%. Overall, for all the cases the two methods showed good qualitative agreement. In addition, we observed that as the percentage of n-dodecane in the mixture is increased beyond a threshold value (in our case >11.11%), the ReaxFF-MD simulations predict an activation energy even lower than that of n-dodecane in single-component system. This suggests that the pyrolysis of toluene follows a much more complex mechanism in the presence of a greater concentration of n-dodecane. In this case, the simple Arrhenius relation based on first-order kinetics does not hold anymore to accurately predict the activation energy of toluene. A preliminary investigation of the time evolution of major species generated by n-dodecane showed that the pyrolysis of n-dodecane is radical dominated. The large alkyl radical pool generated by the thermal decomposition of n-dodecane are highly reactive and they cause further dissociation of toluene. This played a critical role in low temperatures especially, since at low temperatures the toluene decomposition by itself is very limited.

In the higher density case (0.4 kg/dm^3) even small amounts of n-dodecane had much more significant impact on the decomposition of toluene and its activation energy. This was observed in both ReaxFF-MD and continuum simulation results. Beyond a certain mixture ratio, the activation energy did not show any dependence on the number of n-dodecane in the mixture or the overall system density. We hypothesize that a different initiation mechanism is at work in these cases, which needs further studies to be confirmed.

These results demonstrate that although ReaxFF-MD is useful in determining the Arrhenius parameters for single-component systems, in case of mixtures the same methodology only applies in certain cases of densities and mixture ratios. This is also confirmed using the results of continuum-scale pyrolysis simulations. When the overall system density/pressure is high or the concentration of a highly reactive molecule in the mixture is increased, the simple first-order kinetics-based Arrhenius equation does not apply. We need to look further into the complex reaction pathways to determine the effective mixture activation energies in these cases.

Acknowledgements CA, SS and ACTvD acknowledge funding from AFOSR FA9550-17-1-0173.

References

1. Sutton GP (2006) History of liquid propellant rocket engines. AIAA
2. Poling BE, Prausnitz JM, O'connell JP et al (2001) The properties of gases and liquids, vol 5. Mcgraw-hill, New York
3. Bianchi GM, Pelloni P, Corcione FE, Allocca L, Luppino F (2001) Modeling atomization of high-pressure diesel sprays. J Eng Gas Turbines Power 123(2):419–427
4. A validated numerical simulation of diesel injector flow using a VOF method. http://papers.sae.org/2000-01-2932/. Accessed 22 May 2017
5. Gnanaskandan A, Bellan J (2017) Numerical simulation of jet injection and species mixing under high-pressure conditions. J Phys: Conf Ser 821(1):012020
6. Falgout Z, Rahm M, Wang Z, Linne M (2015) Evidence for supercritical mixing layers in the ECN spray A. Proc Combust Inst 35(2):1579–1586

7. Mayer W, Telaar J, Branam R, Schneider G, Hussong J (2003) Raman measurements of cryogenic injection at supercritical pressure. Heat Mass Transf 39(8–9):709–719
8. Oschwald M, Schik A, Klar M, Mayer W (1999) Investigation of coaxial LN2/GH2-injection at supercritical pressure by spontaneous Raman scattering. 35th Joint propulsion conference and exhibit, p 2887
9. Oschwald M, Schik A (1999) Supercritical nitrogen free jet investigated by spontaneous Raman scattering. Exp Fluids 27(6):497–506
10. Roy A, Joly C, Segal C (2013) Disintegrating supercritical jets in a subcritical environment. J Fluid Mech 717:193–202
11. Roy A, Segal C (2010) Experimental study of fluid jet mixing at supercritical conditions. J Propuls Power 26(6):1205–1211
12. Segal C, Polikhov SA (2008) Subcritical to supercritical mixing. Phys Fluids 20(5):052101
13. Falgout Z, Rahm M, Sedarsky D, Linne M (2016) Gas/fuel jet interfaces under high pressures and temperatures. Fuel 168:14–21
14. Manin J, Pickett LM, Crua C (2015) Microscopic observation of miscible mixing in sprays at elevated temperatures and pressures. ILASS meeting, Rayleigh, NC
15. Moin P, Mahesh K, Direct numerical simulation: a tool in turbulence research. http://Dx.doi.org/10.1146/annurev.fluid.30.1.539. http://www.annualreviews.org/doi/10.1146/annurev.fluid.30.1.539. Accessed 23 May 2017
16. Masi E, Bellan J, Harstad KG, Okong'o NA (2013) Multi-species turbulent mixing under supercritical-pressure conditions: modelling, direct numerical simulation and analysis revealing species spinodal decomposition. J Fluid Mech 721:578–626
17. Borghesi G, Bellan J (2015) Irreversible entropy production rate in high-pressure turbulent reactive flows. Proc Combust Inst 35(2):1537–1547
18. Bellan J (2017) Direct numerical simulation of a high-pressure turbulent reacting temporal mixing layer. Combust Flame 176:245–262
19. Bellan J (2017) Evaluation of mixture-fraction-based turbulent-reaction-rate model assumptions for high-pressure reactive flows. Combust Flame 179:253–266
20. Foster J, Miller RS (2010) Fundamentals of high pressure combustion. na
21. Detailed chemical kinetic models for the combustion of hydrocarbon fuels. ResearchGate. https://www.researchgate.net/publication/221949556_Detailed_chemical_kinetic_models_for_the_combustion_of_hydrocarbon_fuels. Accessed 23 May 2017
22. Wang L-P, Titov A, McGibbon R, Liu F, Pande VS, Martínez TJ (2014) Discovering chemistry with an ab initio nanoreactor. Nat Chem 6(12):1044–1048
23. van Duin ACT, Dasgupta S, Lorant F, Goddard WA (2001) ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 105(41):9396–9409
24. Chenoweth K, van Duin ACT, Goddard WA (2008) ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. J Phys Chem A 112(5):1040–1053
25. Chenoweth K, van Duin ACT, Dasgupta S, Goddard WA III (2009) Initiation mechanisms and kinetics of pyrolysis and combustion of JP-10 hydrocarbon jet fuel. J Phys Chem A 113(9):1740–1746
26. Castro-Marcano F, Kamat AM, Russo Jr MF, van Duin ACT, Mathews JP (2012) Combustion of an illinois no. 6 coal char simulated using an atomistic char representation and the ReaxFF reactive force field. Combust Flame159(3):1272–1285
27. Ashraf C, Jain A, Xuan Y, van Duin AC (2017) ReaxFF based molecular dynamics simulations of ignition front propagation in hydrocarbon/oxygen mixtures under high temperature and pressure conditions. Phys Chem Chem Phys 19(7):5004–5017
28. Ashraf C, van Duin AC (2017) Extension of the ReaxFF combustion force field towards syngas combustion and initial oxidation kinetics. J Phys Chem A
29. Bharti A, Banerjee T (2016) Reactive force field simulation studies on the combustion behavior of n-octanol. Fuel Process Technol 152:132–139

30. Castro-Marcano F, van Duin ACT (2013) Comparison of thermal and catalytic cracking of 1-heptene from ReaxFF reactive molecular dynamics simulations. Combust Flame 160(4):766–775
31. Döntgen M, Przybylski-Freund M-D, Kröger LC, Kopp WA, Ismail AE, Leonhard K (2015) Automated discovery of reaction pathways, rate constants, and transition states using reactive molecular dynamics simulations. J Chem Theory Comput 11(6):2517–2524
32. Mayo SL, Olafson BD, Goddard WA (1990) Dreiding: a generic force field for molecular simulations. J Phys Chem 94(26):8897–8909
33. Allinger NL, Yuh YH, Lii JH (1989) Molecular mechanics. The MM3 force field for hydrocarbons. 1. J Am Chem Soc 111(23):8551–8566
34. Lii JH, Allinger NL (1989a) Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. J Am Chem Soc 111(23):8566–8575
35. Lii JH, Allinger NL (1989b) Molecular mechanics. The MM3 force field for hydrocarbons. 3. The Van Der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. J Am Chem Soc 111(23):8576–8582
36. Allinger NL, Chen K, Lii J-H (1996) An improved force field (MM4) for saturated hydrocarbons. J Comput Chem 17(5–6):642–668
37. Sun H (1998) Compass: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. J Phys Chem B 102(38):7338–7364
38. Brenner DW (1990) Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. Phys Rev B 42(15):9458–9471
39. Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys: Condens Matter 14(4):783
40. Yu J, Sinnott SB, Phillpot SR (2007) Charge Optimized Many-body Potential For The $\mathrm{Si}/{\mathrm{SiO}}_{2}$ System. Phys. Rev. B 75(8):085311
41. Shan T-R, Devine BD, Hawkins JM, Asthagiri A, Phillpot SR, Sinnott SB (2010) Second-generation Charge-optimized Many-body Potential For $\text{Si}/{\text{SiO}}_{2}$ And Amorphous Silica. Phys Rev B 82(23):235302
42. Nouranian S, Tschopp MA, Gwaltney SR, Baskes MI, Horstemeyer MF (2014) An interatomic potential for saturated hydrocarbons based on the modified embedded-atom method. Phys Chem Chem Phys 16(13):6233
43. Tersoff J (1988) Empirical interatomic potential for carbon, with applications to amorphous carbon. Phys Rev Lett 61(25):2879–2882
44. Mortier WJ, Ghosh SK, Shankar S (1986) Electronegativity-equalization method for the calculation of atomic charges in molecules. J Am Chem Soc 108(15):4315–4320
45. LaBrosse MR, Johnson JK, van Duin ACT (2010) Development of a transferable reactive force field for cobalt. J Phys Chem A 114(18):5855–5861
46. Nomura K, Kalia RK, Nakano A, Vashishta P, van Duin ACT, Goddard WA (2007) Dynamic transition in the structure of an energetic crystal during chemical reactions at shock front prior to detonation. Phys Rev Lett 99(14):148303
47. Ostadhossein A, Cubuk ED, Tritsaris GA, Kaxiras E, Zhang S, van Duin AC (2015) Stress effects on the initial lithiation of crystalline silicon nanowires: reactive molecular dynamics simulations using ReaxFF. Phys Chem Chem Phys 17(5):3832–3840
48. Strachan A, van Duin ACT, Chakraborty D, Dasgupta S, Goddard WA (2003) Shock waves in high-energy materials: the initial chemical events in nitramine RDX. Phys Rev Lett 91(9):098301
49. Wood MA, Cherukara MJ, Kober EM, Strachan A (2015) Ultrafast chemistry under nonequilibrium conditions and the shock to deflagration transition at the nanoscale. J Phys Chem C 119(38):22008–22015
50. Verners O, van Duin ACT (2015) Comparative molecular dynamics study of Fcc–Ni Nanoplate stress corrosion in water. Surf Sci 633:94–101
51. Zou C, Van Duin A (2012) Investigation of complex iron surface catalytic chemistry using the ReaxFF reactive force field method. JOM 64(12):1426–1437

52. Shin YK, Kwak H, Vasenkov AV, Sengupta D, van Duin ACT (2015) Development of a ReaxFF reactive force field for Fe/Cr/O/S and application to oxidation of butane over a pyrite-covered Cr2O3 catalyst. ACS Catal 5(12):7226–7236
53. Hong S, van Duin ACT (2015) Molecular dynamics simulations of the oxidation of aluminum nanoparticles using the ReaxFF reactive force field. J Phys Chem C 119(31):17876–17886
54. Verlackt CCW, Neyts EC, Jacob T, Fantauzzi D, Golkaram M, Shin Y-K et al (2015) Atomic-scale insight into the interactions between hydroxyl radicals and DNA in solution using the ReaxFF reactive force field. New J Phys 17(10):103005
55. Yeon J, van Duin ACT (2016) ReaxFF molecular dynamics simulations of hydroxylation kinetics for amorphous and nano-silica structure, and its relations with atomic strain energy. J Phys Chem C 120(1):305–317
56. Yue D-C, Ma T-B, Hu Y-Z, Yeon J, van Duin AC, Wang H et al (2015) Tribochemical mechanism of amorphous silica asperities in aqueous environment: a reactive molecular dynamics study. Langmuir 31(4):1429–1436
57. Zou C, Shin YK, van Duin ACT, Fang H, Liu Z-K (2015) Molecular dynamics simulations of the effects of vacancies on nickel self-diffusion, oxygen diffusion and oxidation initiation in nickel, using the ReaxFF reactive force field. Acta Mater 83:102–112
58. Wang Q-D, Wang J-B, Li J-Q, Tan N-X, Li X-Y (2011) Reactive molecular dynamics simulation and chemical kinetic modeling of pyrolysis and combustion of n-dodecane. Combust Flame 158(2):217–226
59. Cheng X-M, Wang Q-D, Li J-Q, Wang J-B, Li X-Y (2012) ReaxFF molecular dynamics simulations of oxidation of toluene at high temperatures. J Phys Chem A 116(40):9811–9818
60. Liu L, Bai C, Sun H, Goddard WA (2011) Mechanism and kinetics for the initial steps of pyrolysis and combustion of 1,6-dicyclopropane-2,4-hexyne from ReaxFF reactive dynamics. J Phys Chem A 115(19):4941–4950
61. Beste A (2014) ReaxFF study of the oxidation of lignin model compounds for the most common linkages in softwood in view of carbon fiber production. J Phys Chem A 118(5):803–814
62. Kylasa SB, Aktulga HM, Grama AY (2014) PuReMD-GPU: a reactive molecular dynamics simulation package for GPUs. J Comput Phys 272:343–359
63. Zheng M, Li X, Liu J, Wang Z, Gong X, Guo L et al (2014) Pyrolysis of liulin coal simulated by GPU-based ReaxFF MD with cheminformatics analysis. Energy Fuels 28(1):522–534
64. Zhang T, Li X, Qiao X, Zheng M, Guo L, Song W et al (2016) Initial mechanisms for an overall behavior of lignin pyrolysis through large-scale ReaxFF molecular dynamics simulations. Energy Fuels 30(4):3140–3150
65. Goverapet Srinivasan S, van Duin ACT (2011) Molecular-dynamics-based study of the collisions of hyperthermal atomic oxygen with graphene using the ReaxFF reactive force field. J Phys Chem A 115(46):13269–13280
66. Bagri A, Mattevi C, Acik M, Chabal YJ, Chhowalla M, Shenoy VB (2010) Structural evolution during the reduction of chemically derived graphene oxide. Nat Chem 2(7):581–587
67. Huang X, Yang H, van Duin ACT, Hsia KJ, Zhang S (2012) Chemomechanics control of tearing paths in graphene. Phys Rev B 85(19):195453
68. Ganesh P, Kent PRC, Mochalin V (2011) Formation, characterization, and dynamics of onion-like carbon structures for electrical energy storage from nanodiamonds using reactive force fields. J Appl Phys 110(7):073506
69. Srinivasan SG, van Duin ACT, Ganesh P (2015) Development of a ReaxFF potential for carbon condensed phases and its application to the thermal fragmentation of a large fullerene. J Phys Chem A 119(4):571–580
70. Kim D, Martz J, Violi A (2014) A surrogate for emulating the physical and chemical properties of conventional jet fuel. Combust Flame 161(6):1489–1498
71. Oefelein JC (2006) Mixing and combustion of cryogenic oxygen-hydrogen shear-coaxial jet flames at supercritical pressure. Combust Sci Technol 178(1–3):229–252
72. The Force: CaltechMech. http://www.theforce.caltech.edu/CaltechMech/. Accessed 9 June 2017

73. Colket MB, Seery DJ (1994) Reaction mechanisms for toluene pyrolysis. Symposium (International) on combustion, vol 25. Elsevier, pp 883–891
74. Liu G, Han Y, Wang L, Zhang X, Mi Z (2008) Supercritical thermal cracking of n-dodecane in presence of several initiative additives: products distribution and kinetics. Energy Fuels 22(6):3960–3969
75. Herbinet O, Marquaire P-M, Battin-Leclerc F, Fournet R (2007) Thermal decomposition of n-dodecane: experiments and kinetic modeling. J Anal Appl Pyrolysis 78(2):419–429

Chapter 8
Shock-Induced Chemistry: Molecular Dynamics and Coarse Grain Modeling

Md Mahbubul Islam, Mathew Cherukara, Edwin Antillon and Alejandro Strachan

Abstract The fast loading rates associated with shockwaves in solids make molecular dynamics (MD) a particularly well-suited tool for their study. This chapter focuses on recent methods to study shock-induced chemistry using all-atom reactive MD and coarse-grained simulations and their application. We describe insight on the formation of hot spots formed following the shock-induced collapse of pores and their transition to a deflagration wave in high energy density materials obtained from large-scale MD simulations using the reactive force field ReaxFF. Experimental validation of such simulations is critical to assess the predictive capabilities of these methods to describe new materials and show how to extract observables from the simulations that can be directly contrasted with experiments. Such direct comparisons are not just critical for validation but also contribute to the interpretation of the experimental results. We also describe coarse-grained simulations to study the possibility and effectiveness of shock-induced, endothermic, volume-collapsing reactions; these simulations quantify how the various characteristics of the chemical reactions attenuate the propagating shockwave and provide key information to experimentalists designing and synthesizing such materials.

8.1 Introduction

Shock or dynamical loading of a material causes a sudden increase in pressure and temperature which, in turn, triggers a wide range of processes through which the shocked material relaxes after the insult. The response is often a combination of pro-

M. M. Islam · A. Strachan (✉)
School of Materials Engineering and Birck Nanotechnology Center, Purdue University, West Lafayette, IN 47907, USA
e-mail: strachan@purdue.edu

M. Cherukara
Center for Nanoscale Materials, Argonne National Laboratory, Argonne, IL 60439, USA

E. Antillon
UES Inc., Dayton, Ohio 45433, USA

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_8

cesses including plastic deformation [1–3] to relax the uniaxial compression caused by the shock, stress-induced phase transformations [4, 5], and chemical reactions [6, 7]. The fast strain rates involved (approximately 10^9 1/s) and the extreme pressures and temperatures achieved make shocks an attractive means to study materials behavior not accessible otherwise and to study materials at extreme conditions. For example, some crystal structures of certain minerals like silica form only at extreme conditions but remain in a metastable state after unloading [8]; consequently, their presence provides unique information about planetary systems. In addition, the fast loading results in nonequilibrium states and can result in processes not observed under equilibrium of slow-loading conditions. For example, Ravelo and Levitas found "virtual melting" as a new stress relaxation mechanism [9].

In this Chapter, we focus on recent molecular-level simulations of shock-induced chemical reactions in two classes of materials that, at first sight, appear to be complete opposites but, upon further analysis, show remarkable similarities. Section 8.3 discusses reactive atomistic simulations of high energy density (HE) materials, that react exothermically leading to gaseous products which, under appropriate conditions, can turn a shock into a detonation. Section 8.4 discusses coarse-grained simulations to explore materials that can weaken shockwaves via endothermic, volume-collapsing reactions. In both cases coupling the shock excitation to the degrees of freedom (DoFs) capable of causing chemical reactions is a complex process and does not happen instantaneously. Actually, in some cases chemical reactions occur when the system has not fully relaxed and equilibrated locally after the passage of the shockwave. Thus, of the focus areas of this chapter is how the nonequilibrium states right behind the shock front and the kinetics associated with the chemical reactions affect materials response. Before discussing the various applications we provide, in Sect. 8.2, a brief introduction of the simulation methods utilized and the details of the simulations presented in Sects. 8.3 and 4. A recent review of molecular simulations of shock processes [10] provides additional details on simulation techniques, applications to nonreactive systems and tutorials to perform atomistic shock simulations online using nanoHUB [11] cloud computing.

8.2 Atomistic and Coarse Grain Simulations of Shocks

Molecular dynamics simulations describe the temporal evolution of a group of atoms by solving classical equations of motion, given in Hamilton's form by

$$\dot{r}_i = v_i$$
$$\dot{v}_i = \frac{f_i}{m_i}$$

where $\{r_i\}$, $\{v_i\}$, and $\{f_i\}$ are the set of all atomic positions, velocities, and forces and $\{m_i\}$ atomic masses. In the absence of external fields, the forces originate from

atomic interactions and are obtained as the negative gradient of the total potential energy with respect to atomic positions.

Given initial conditions (atomic positions and velocities), the simulation predicts, in a deterministic manner, the temporal evolution of the system. When a group of atoms is evolved in time with realistic interactions, that include anharmonicities, the equations of motion take the system toward thermodynamic equilibrium; absent external perturbation the system will satisfy Maxwell–Boltzmann statistics. It is important to emphasize that while thermodynamic equilibrium is achieved under the appropriate conditions, it is not assumed in the simulation and nonequilibrium processes are captured explicitly.

8.2.1 Nonequilibrium Simulations of Shock Loading

Nonequilibrium shock simulations are often set up as an impact simulation between the previously equilibrated target and piston systems, both described in atomistic detail. The initial condition for such simulations is obtained by adding the desired impact velocity to the atoms in the target over the thermal velocities. It is common practice to zero the c.m. velocity of the entire system to avoid an overall translation of the systems within the simulation cell. Almost invariably, periodic boundary conditions are imposed in the cross-sectional directions; this mimics the response of a section of the material of interest away from the lateral free surfaces (far enough that the waves caused by the lateral expansion do not reach the section of interest during the simulated time).

With such initial conditions, the dynamical evolution of the system is described with adiabatic MD simulations (constant energy, number of atoms and volume or NVE ensemble). The passage of the shock leads to both the compression and heating of the shocked material and, as discussed above, can result in a plethora of materials responses to the extreme loads. Shock fronts are locally very sharp (a few atomic distances) and, thus, materials experience fast deformation and heating rates. These fast rates, in turn, lead to nonequilibrium thermodynamics states that do not satisfy equipartition of energy. As will be discussed below, this lack of equilibrium can affect processes triggered at or near the shock front.

8.2.2 Simulating Shock States Using Equilibrium MD

The main advantage of nonequilibrium shock simulations is that the shock front is described explicitly and the material experiences the ultrafast loading rates characteristic of shock loading, including the lack of local thermal equilibrium. An important drawback of such simulations is that the timescales achievable are approximately limited by the time it takes the shock to travel through the sample since a rarefaction wave moving back into the sample will be generated when the shock meets a free

surface. Consider a shock wave propagating through a sample of length L and a speed u_p; for a typical velocity of 5 km/s or 5 nm/ps, a simulation cell 100 nm is length will be traversed in only 20 ps. That is, the regions near the impact would have been in the shock state for only 20 ps when the rarefaction wave starts moving in. Extending the simulation a factor of n, requires increasing the simulation cell length by the same factor, resulting in an n^2 increase in computational cost (assuming a linear increase in cost with system size). In order to address this issue and enable longer simulation times for shocked systems, two methods have been proposed to simulate the state after the passage of a shock by modifying the equations of motion of the systems to compress and heat up the to the state corresponding to the shock state. Reed et al. proposed the MSST method [12] and Maillet, Ravelo, and collaborators proposed the Hugoniostat [13] described next.

For a steady-state shock with particle velocity u_p and shock velocity u_s, mass, energy and momentum conservation across the shock front result in the Rankine–Hugoniot jump conditions. These relate the unshocked state (density, energy, and pressure) with the shocked one. The Hugoniostat method uses a thermostat and a barostat to take the system from its initial state to the desired shock state for a given pressure. We note that this method can only be applied for single-wave situations, see [13, 14]. Multiple wave structures like in the case of plasticity following elastic loading should be simulated using different initial conditions for each wave.

8.2.3 Reactive MD Simulations

Shock-induced chemical reactions are important in applications ranging from explosives [15] to materials that can undergo volume-collapsing reactions that can be used for shock attenuation [16]. The development of reactive potentials over the last decades enabled the simulation of shock-induced chemistry. Here, we use ReaxFF [17, 18] to predict the decomposition and reaction of high energy density materials, Sects. 8.3 and 8.4. Building on earlier work [19], ReaxFF uses the concept of partial bond order between atoms to describe covalent interactions. These partial bond orders are many-body functions of the atomic positions and capture the character of the bonds (sigma, pi, double pi) and, importantly for MD simulations vary smoothly and approach zero as a bond is broken. All covalent terms, bond stretch, angles, and torsions depend on these bond orders which are also used to penalize over and under coordination. A second key element of ReaxFF is that electrostatic interactions are calculated using environment-dependent partial atomic charges obtained using electronegativity equalization method EEM [20].

8.2.4 Dynamics with Implicit Degrees of Freedom

While MD simulations are a powerful tool to explore dynamical loading, it is also computationally intensive and coarse grain models capable of capturing longer time and larger spatial scales are desirable for many applications. Here is use a particle-based coarse grain or mesoscale description, where particles describe group of atoms. Over the last decade or so, Strachan and collaborators have been developing the dynamics with implicit degrees of freedom (DID) family of methods to couple particle dynamics via MD with an implicit description of additional degrees of freedom. These implicit degrees of freedom can be atoms internal to the mesoparticles [21, 22], valence electrons [23] or an external electrochemical potential [24, 25].

In Sect. 8.4, we use DID to explore the possibility of volume-collapsing chemical reactions to weaken shockwaves with the objective of developing materials for protection against high-velocity impact or blasts. Particles represent a single molecule or small group of molecules that can undergo a stress-induced chemical reaction. In order to describe the average effect of the degrees of freedom internal to the particles, we include two internal variables to the mesoparticles: (i) an average particle radius to describe the volume collapsing chemistry and (ii) the average temperature of the remaining internal degrees of freedom. The dynamics of the average molecular radii is described via the Hamiltonian of the system. As described in Sect. 8.5, we add a potential energy term that captures how the energy of the molecules depends on their radius; a two well potential is used to describe the stress-induced phase transformation. The dynamics of the particles and their average radii is governed by the following Hamiltonian [26, 27]:

$$H = \sum_{i<j} \phi_{\text{inter}}\left(\vec{r}_j - \vec{r}_i - \sigma_j - \sigma_i\right) + \sum_i \phi_{\text{intra}}(\sigma_i) + \sum_i \frac{p_i^2}{2m_i} + \sum_i \frac{\pi_i^2}{2m_i^*}$$

where σ_i and π_i are the average molecular radius and momentum associated with particle i. Interactions between particles are described with a two-body potential, ϕ_{inter}, that depends on the distance between the particle surfaces. The dynamics of the particle's average radii is governed by an intramolecular potential, ϕ_{intra}, the details of which will be discussed in Sect. 8.4.

The internal degrees of freedom not described by the radii, are described in an average manner and their state is governed by the internal temperature of each mesoparticle, that evolves in time using DID equations of motion by exchanging energy with the local particles [27]:

$$\dot{T}_i^{\text{int}} = \nu_{\text{meso}} \frac{T_i^{\text{meso}} - T_i^{\text{int}}}{m_i C_i^{\text{int}} \langle \omega_{\text{inter}}^2 \rangle \Theta_0} \left| F_i^{\text{inter}} \right|^2 + \nu_{\text{rad}} \frac{T_i^{\text{rad}} - T_i^{\text{int}}}{m_i^* C_i^{\text{int}} \langle \omega_{\text{rad}}^2 \rangle \Theta_0} \left| F_i^{\text{rad}} \right|^2$$

where C_i^{int} is the specific heat associated with the internal DoFs, ν_{meso} and ν_{rad} describe the strength of the internal to intermolecular coupling and the internal to

radial coupling, respectively. Θo is a reference temperature, and the ratio $\frac{|F_i^{\text{rad}}|^2}{m_i^* \langle \omega_{\text{rad}}^2 \rangle}$ provides a natural timescale for the corresponding interaction.

Note that energy flow between internal DoF and the particles is determined by the relative local temperatures. If the local particle temperature around particle i is larger than its internal temperature, the particles are slowed down and the extra energy is captured by the internal DoF. Lin et al. [23] provides a detailed description of DID including its foundation on statistical mechanics. As described there, key features of the method include: (i) total energy and linear momentum are conserved; (ii) Galilean invariance; (iii) correct description of an isolated particle moving in free flight.

8.3 Shock-Induced Chemistry in High Energy Density Materials

The ultrafast loading rates associated with shock propagation in materials can reveal aspects of material behavior and properties that are inaccessible otherwise. Such nonequilibrium states include metastable phases [28], defect nucleation and multiplication [29] and of particular relevance to this section, nonequilibrium chemical reaction pathways [30]. In general, the process of shock loading a material, triggers a series of material responses that seek to minimize the potential energy in the system and alleviate the extreme shock-induced conditions of pressure and temperature. Such material responses include plastic deformation [1], phase transitions [31], and conformational changes in the case of organic crystals [32]. In each of these, the material response weakens the shock, as energy from the compressive wave is diverted to plastically deform the material, drive the material over a phase change or change the molecular conformation. Another mechanism triggered by shocks, that is the subject of this section, is chemical reactions. However, unlike the other mechanisms discussed above, not always does this result in a weakening of the shock. In HE materials, shock loading can trigger a series of net-exothermic reactions that form gaseous products. The exothermicity and volume expansion strengthens the shock, accelerates the chemistry and leads to a detonation if the chemical wave catches up with the shock wave. In a detonation, the wave propagation velocity (chemical and mechanical waves overlap) depends only on chemical kinetics and is consequently, independent of the initial piston velocity.

Experimental characterization of all shock-related phenomena is extremely difficult due to the short spatio-temporal length scales involved (from nm to microns and from ps to ns), but the problem is exacerbated when chemical reactions that are induced or assisted by the shock need to be taken into account. Recent progress on ultrafast spectroscopy coupled to laser shocks [33–37] is providing an unprecedented picture into chemistry at extreme conditions. Despite the impressive results, these experiments are not without limitations. For example, it is difficult to perform peak assignment in spectroscopic studies at extreme conditions. Thus, we believe only the combination of experiments and atomistic simulations will provide a definite descrip-

tion of the detailed chemistry of HE materials at extreme conditions. Interestingly, computational capabilities and models have matured to the point where it is possible to simulate large enough system sizes with accurate interatomic potentials, that can capture complex chemical reactions under nonequilibrium conditions [38]. In this section, we discuss shock-induced chemistry of poly vinyl nitrate, a high-energy amorphous polymer, and compare the predictions with ultrafast spectroscopy. Following the discussion of decomposition of a homogeneous, amorphous polymer we discuss recent results on the coupling of a shock wave with microstructural defects, the formation of hotspots and their criticality.

8.3.1 Atomistic Modeling of Shocks in HE Materials

Modeling a complete detonation from the initial chemical events to a detonation wave with all atom simulations remains beyond the scope of current computational capability. However, reactive MD simulations have provided unique insight into several aspects of the shock to detonation cycle, including the initial chemical reactions, and the role of preexisting material defects in the creation of hotspots.

Limited by the computational resources of the time, early MD work that studied reactive chemistry was based on the so-called AB system, a simplified representation of explosives, where diatomic AB molecules transform into A_2 and B_2 [39]. More recently, large-scale atomic simulations (up to ~36 million atoms) using a similar monoatomic model for nitrocubane, showed for the first time how hotspots can spontaneously form in a homogeneously heated sample, and how "critical" hotspots transition into a spherical detonation [40]. When a hotspot is formed, a competition arises between heat conduction away from the hotspot and accelerated kinetics due to the elevated temperatures. Smaller hotspots that have a larger surface area to volume quench, while larger hotspots can become self-sustaining giving rise to a deflagration wave that could eventually transition to a detonation. In their study, Hu et al. [40] found that for nitrocubane hotspots that reach a critical radius of ~1.5 nm continue to grow with a radial velocity that becomes supersonic. Figure 8.1 shows snapshots from the simulation, where only N_2 molecules are shown. A multitude of hotspots form spontaneously during the initial stages (Fig. 8.1a), most of which quench due to thermal conduction away from the hotspot, while a few larger ones become self-sustaining, eventually reaching supersonic velocities (Fig. 8.1d).

The presence of material defects have long been thought to play an important role in the initiation of chemistry and subsequent detonation, and the presence or absence of defects can have strong repercussions on the sensitivity of an explosive. For instance, the shock strengths required for the detonation of single-crystal explosives are significantly higher than shocks required to detonate granular powders [41]. Defects in the material such as voids, interfaces, or dislocation localize the energy from the shock through material jetting and shear into hotspots [42]. As with thermal hotspots, higher temperatures at hotspots lead to faster chemistry, which, in turn, increases the likelihood of a self-sustaining chemical wave [43].

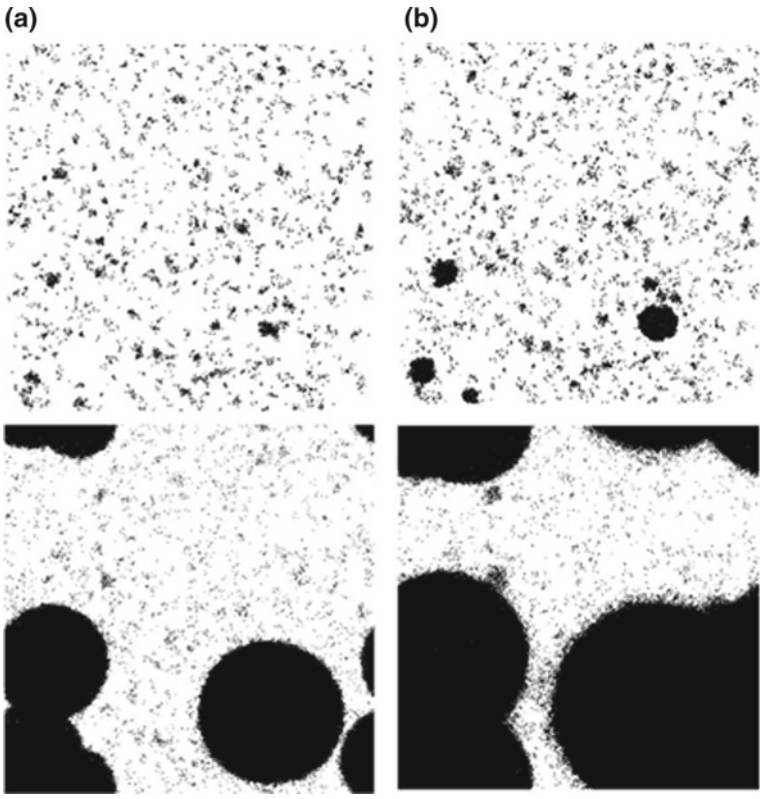


Fig. 8.1 Reaction progression during thermal cookoff of nitrocubane initially at 1160 K leading to the formation of N_2. Only N_2 molecules are shown for clarity. Times are **a** 5.0 ps, **b** 7.3 ps, **c** 11.4 ps, and **d** 13.2 ps. Reproduced with permission from [40]. Copyright (2011) American Chemical Society

Early MD work by Holian at Los Alamos and others, using a nonreactive LJ potential [44], showed significant temperature increase at the far wall of a collapsing void. The authors attributed this localized heating to recompression of rarified ejecta at the far wall of the pore, observing that the temperature increase in the collapsed pore scales with void size, but only up to a point [44]. A few years later, Herring et al. used the AB model to perform one of the first atomistic simulations of a shock to detonation transition [45]. Figure 8.2 shows the evolution of a shocked sample of the model material that initially contained a void of radius 5 nm. Atoms are colored by the nature of their bond, reactant AB molecules are in blue, products A_2 and B_2 are in green, while red atoms denote free radicals, i.e., isolated A or B atoms. Figure 8.2a shows a snapshot at an early stage, soon after the collapse of the void, where a deflagration wave starting from the collapsed void can be seen on the far left (region in green). Subsequently, more hotspots are seen to develop between the deflagration wave from the collapsed pore and the shock front (Fig. 8.2b). Finally,

Fig. 8.2 MD snapshots of a shock to detonation transition in a model AB explosive with a 5 nm radius void shocked at $U_p = 4$ km/s. Shock travels from left to right. Blue atoms denote unreacted material, green atoms represent product molecules while red atoms are unbonded monoatomic species. Snapshots show **a** Collapse of a void and the formation of a deflagration wave, **b** formation of hotspots between the deflagration wave and the shock front that ultimately give rise to a detonation (**c**). Reproduced with permission from [45]. Copyright (2010), American Physical Society

Fig. 8.2c shows the formation of a detonation wave that is separated from the original deflagration wave.

8.3.2 Shock-Decomposition of Poly Vinyl Nitrate (PVN)—MD Versus Experiments

In this section, we discuss the direct comparison of shock-induced chemistry between MD simulations and experiments. As discussed above, the microstructure in plastic-bonded HE systems plays a key role in localizing the energy of the shockwave in hotspots and in initiation. Such processes invariably complicate comparisons between experiments and MD simulations. We believe that homogeneous materials, including liquid and amorphous materials [46–48] are better choices for such direct comparisons.

PVN is, a homogeneous amorphous energetic polymer, used as a binder material. McGrane et al. [33] used ultrafast IR spectroscopy on laser-shocked PVN over a range of shock pressures and observed chemical reactions in 100s of picosecond timescales when shocked above a threshold of 18 GPa. The spectra reveal chemical reaction initiation via the disappearance of the NO_2 group stretching frequency; however, the limited spectral range of the experiments did not enable the characterization of detailed chemistry. A definitive understanding of the shock-induced chemistry of PVN requires a synergistic combination of experimental and computational studies, where experiments validate simulations and simulations help interpret experimental findings. We carried out shock simulations of PVN to establish a one-to-one comparison of the mechanical and chemical response of the material with the laser shock experiments; these results appeared in [49].

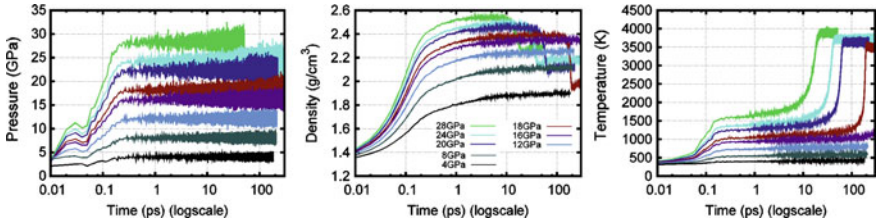


Fig. 8.3 Evolution of pressure, density, and temperature during shock simulations at various pressures. The sudden decrease in density and increase in temperature is indicative of the rapid chemical reactions

8.3.2.1 Shock Loading of PVN

A bulk amorphous PVN geometry at a density of 1.33 gm/cm^3 and molecular weight of 17806 gm/mol was used for the study. The shock simulations were performed utilizing Hugoniostat [13]-based MD simulation method. During shock loading, the uniaxial compression heats up the system to the final desired shock states. The ReaxFF-2014 force field [50] was used for all the simulations. The evolution of the thermodynamic properties such as pressure, density, and temperature was recorded from the shock simulations, and their temporal evolution is shown in Fig. 8.3. The applied rapid strain rate is representative to the shock process. The figures exhibit that the thermodynamic quantities were evolved to their final shocked states within around 1 ps. The applied compressive loading increases the system temperature, and for strong shock pressures, the system undergoes a rapid increase in volume and temperature followed by an induction period. The induction period is contingent on the applied shock and decreases with the increasing pressure. At 18 GPa shock pressure, the simulation-predicted induction time (~180 ps) is in good agreement with McGrane et al. [33] data. The rapid decrease in the density, that is, the increase in the volume is due to the exothermic reactions generated gaseous species.

The pressure–density and temperature–density data were derived from the simulations and shown in Fig. 8.4a, b. Since explosive materials produce gases upon exothermic reaction, that is, at a given pressure–temperature state, the products are less dense than the reactants, as such the products Hugoniot lies above the unreacted Hugoniot in the P, ρ space. The unreacted Hugoniot dictates the state from which the rapid chemical reactions start. It can be seen that both P-ρ and T-ρ plots have two regimes: unreactive Hugoniot for a relatively weak loading and reactive Hugoniot due to the exothermic volume-expanding reactions. The ReaxFF simulations nicely capture the shock pressure at which transition occurs and is in excellent agreement with McGrane et al. [33] reported data. The exothermic chemistry caused the system temperature to increase up to approximately 2000 K.

The Hugoniot relations [14] were used to calculate shock and particle velocity values. The data is presented in Fig. 8.4c along with the unreactive experimental [51] results. The reaction products Hugoniot regime exhibit a sudden increase in the shock velocities. The trend of a rapid increase of shock velocities at the onset

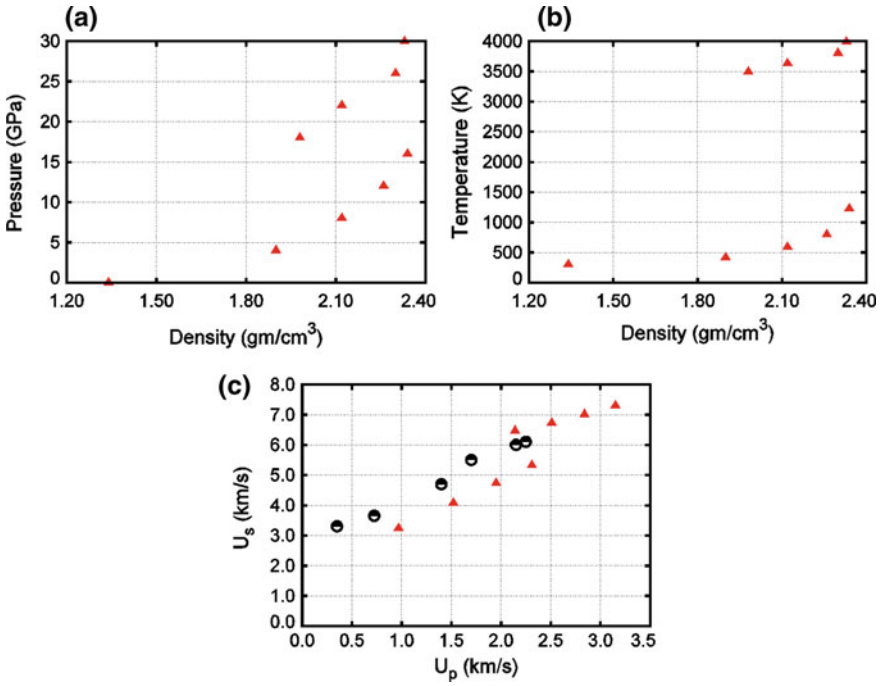


Fig. 8.4 Plot of unreacted and reaction products Hugoniots in **a** P-ρ, **b** T-ρ space, the product Hugoniots lies above the unreacted Hugoniots, **c** shock versus particle velocity data and comparison with the experimental results [51] (black half-filled circles)

of rapid reactions has also been observed in nitromethane shock experiment [52]. ReaxFF-2014 slightly underestimates the experimental unreactive u_s–u_p data. The simulation exhibit a transition to rapid exothermic chemistry at u_p= 2.2 km/s and u_s= 6.5 km/s. The sound speed (value of u_s at $u_p = 0$) computed from the simulations is 1.75 ± 0.05. This value is relatively smaller compared with the experimental data [51] of 3.2 ± 0.3 km/s.

8.3.2.2 Vibrational Analysis and Comparison with Shock Spectroscopy

In order to validate the ReaxFF simulations of shock-induced chemistry in PVN against the ultrafast spectroscopic experiments in a direct manner, we compute the time evolution of the vibrational density of states from the Fourier transform of atomistic velocities in the reactive MD simulations. The signature peaks of the various chemical groups and their evolution during the shock simulations provide key information about the reaction mechanisms and the associated time scales. The predicted time-resolved spectra at a shock pressure of 18 GPa is shown in Fig. 8.5. The NO_2-stretching frequency can be considered as an indicator of the rapid chemical

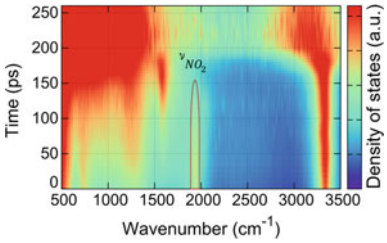


Fig. 8.5 Predicted time-resolved full spectra at 18 GPa shock simulation. The disappearance of NO_2 stretching frequency at ~150 ps is indicative to the onset of rapid chemical reactions

decomposition of the PVN. The figure demonstrates the disappearance of nitro group stretching mode (at ~1900 cm^{-1}) at a time-scale of around 150 ps; this marks the beginning of rapid exothermic chemistry, see Fig. 8.3. This predicted time scale is in excellent agreement with the McGrane et al. [33] reported experimental spectra of shocked PVN at 18 GPa. Furthermore, the calculated spectra yield important information about the evolution of the various intermediates. For example, a peak at 1600 cm^{-1} starts to develop in the spectra at about 50 ps and completely disappears at ~200 ps. The spectra calculated for the individual intermediates and the species analysis during the shock simulations show that the characteristic frequency corresponds to the NO. Thus, spectra capture the entire NO evolution history. Overall, the results demonstrate that the simulation-predicted spectra can help to unravel the experimental spectra to illustrate various reaction initiation mechanisms and related timescales.

8.3.3 Shock to Deflagration Transition and the Role of Hot Spots

While the studies described in Sect. 8.3.1 in model systems provided a wealth of information into the formation and criticality of dynamical hotspots, their simple nature can hide key aspects of the processes that operate in real HE materials where complex inter- and intramolecular processes are known to play key roles in the initiation of chemistry. As described earlier in the chapter, ReaxFF simulations can provide an accurate, all-atom description of the mechano-chemistry in a variety of nitramines [7, 53]. The prohibitive factor until recently has been the vastly greater computational cost ($10\times - 100\times$) as compared to the reduced models. In recent years, however, several research groups have performed multimillion atom ReaxFF simulations on state-of-the-art computational clusters to study hotspot formation resulting from material inhomogeneities such as voids [54] and the interfaces of polymer-bonded explosives [55].

More recently, using large-scale ReaxFF simulations, Wood et al. [30] provided the first atomic picture of a shock to deflagration transition in RDX using a realistic potential. They observed a crescent-shaped hotspot following pore collapse, that eventually gives rise to a deflagration wave. The authors observed three distinct

stages in the formation of the deflagration wave from the hotspot; in the first stage (up to ~10 ps from the collapse of the pore), the initial crescent-shaped hotspot that is formed grows until it is ~5 nm in width. Immediately following impact of ejecta with the far wall of the pore, the molecular center of mass temperature (T_{com}) and the molecular vibrational temperature or internal temperature (T_{vib}) of the impacted molecules differ greatly (T_{com} ~ 4000 K and T_{vib} ~ 1500 K at t_0). During the first stage of reaction, these temperatures equilibrate as energy is transferred from the molecular COM degrees of freedom to internal degrees of freedom. Surprisingly, within this short period, a few product molecules are also seen to form (Fig. 8.6).

During the second stage of reaction (~10–25 ps), the initial, crescent-shaped hotspot grows at a rapid rate, faster into the amorphous ejecta within the pore as compared to the crystalline material ahead of the pore. During this stage, the temperature within the growing hotspot reaches a steady-state value of ~4000 K and two distinct reaction fronts are seen, each with a surprisingly narrow width of ~5 nm. The third and final stage as described by the authors involves the steady-state growth of these reaction fronts at ~250 m/s.

A further surprise from their study was that thermal hotspots that they engineered to have the same temperature distribution, pressure, and morphology as the dynamically created hotspots showed much slower kinetics than the dynamic hotspot, suggesting that nonequilibrium effects arising from the mechanical impingement and shear of molecules during pore collapse, plays a vital role in accelerating the kinetics.

8.4 Shock Energy Absorption via Volume-Reducing Chemical Reactions

Section 8.3 focuses on HE materials where a shock can unleash exothermic, volume expanding reactions that can cause a detonation. In this section, we focus on exploring the possibility of using endothermic, volume collapsing chemistry to weaken a shockwave. The motivation of this effort is to contribute to the design of shock wave energy dissipation (SWED) materials that can be used for protection such as in traumatic brain injury [56] from impact and blast [57–61]. Our goal is to characterize how the details of the chemistry (enthalpy of formation, amount of volume reduction and kinetics) affect the ability of the material to weaken a shockwave. Since we are interested in general aspects of the problem and not in the detailed chemistry, we developed a coarse-grained model to describe SWED materials and characterize what features of its chemical reactions have the largest effort on shock attenuation.

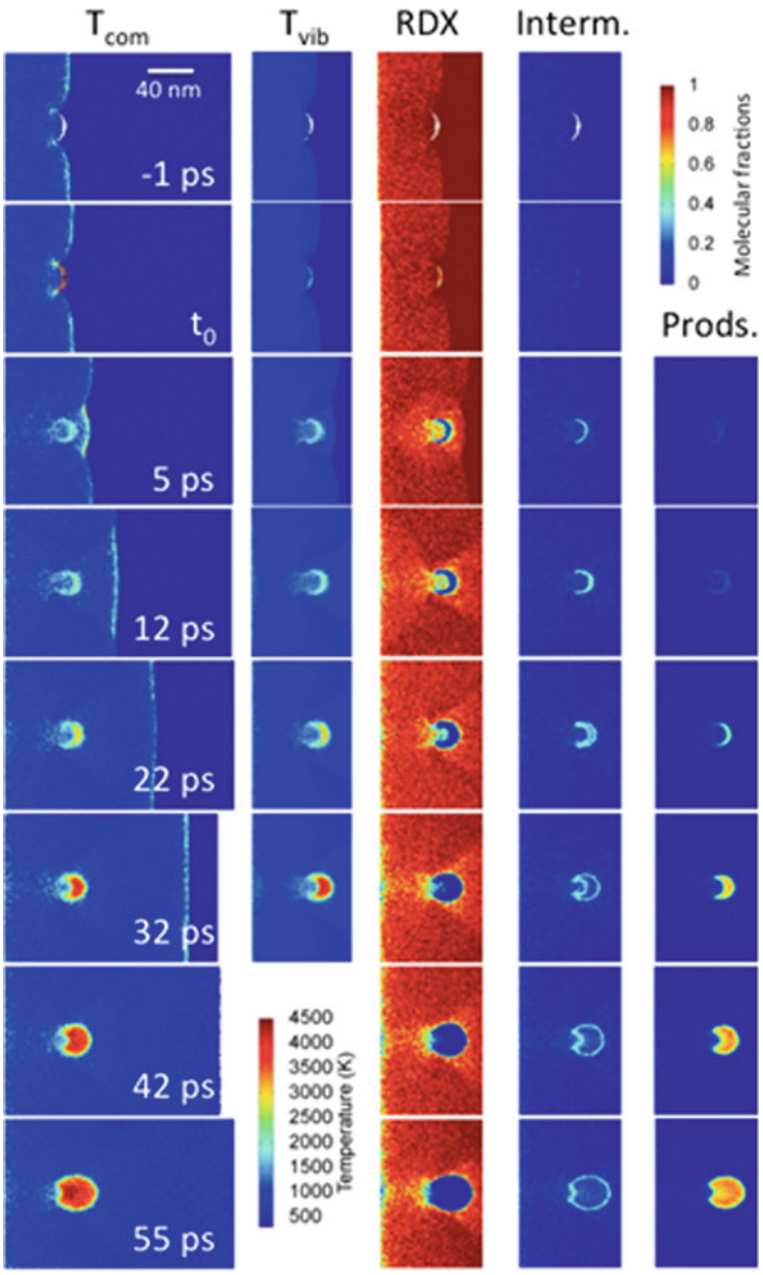


Fig. 8.6 Center of mass molecular temperatures (T_{com}), molecular vibrational temperatures (T_{vib}) and fractions of unreacted RDX, fractions of intermediate species, and fractions of final products at different stages. Reproduced with permission from [27]. Copyright (2015) American Chemical Society

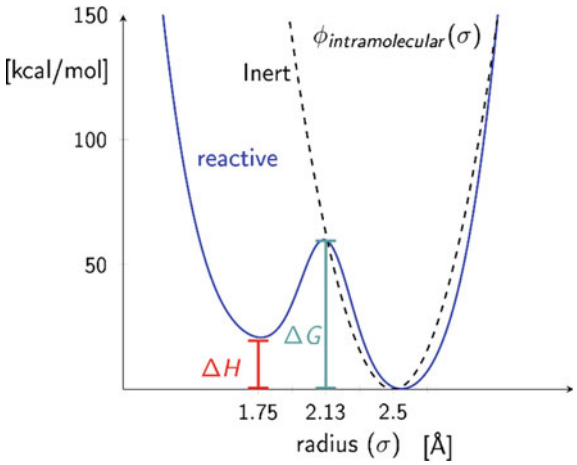


Fig. 8.7 Intramolecular potential for an inert and reactive materials as a function of the coarse-grained particle radius. For the reactive function, the initial state has a radius of size of 2.5 A while the final states are equal to 1.75 A; this corresponds to a 35% volume collapse. The final state can have a variable endothermic state ($\Delta H > 0$) an activation barrier $\Delta G > \Delta H$. The inert case is described by a harmonic potential with a stiffness that yields material properties close to molecular crystals such as HMX and Anthracene

8.4.1 A Family of SWED Materials

As described in Sect. 8.2, ChemDID provides a coarse-grained description of intramolecular chemistry via a potential energy function that controls the dynamics of the average radius of the particles. The potential energy shown in Fig. 8.7, was designed to describe endothermic, volume reducing reactions. This model entertains a specific type of chemistry proceeding via an order parameter (particle size) to transition from a low-energy high-volume state to a high-energy low-volume state. The key characteristics of the chemistry are: (i) endothermicity that we will vary between 0 and 30 kcal/mol, (ii) the volume reduction that varies from 10 to 65%, and (iii) the activation barrier from 0 to 100 kcal/mol.

The interaction between mesoparticles is described using the Morse potential (two exponentials) with distances measured from the surfaces of the particles, i.e., taking into account their radii, see [27]. The parameters of the two-body potential were chosen to result in density and stiffness comparable to HMX-molecule. The initial condition for the simulations consist of a target made of an FCC crystal with a lattice parameter $a = 10.1$ A and a molecular mass of $m = 296.1$ g/mol, which yields a density of 2.0 g/cm^3; this is close to the density of a molecular crystals such as HMX-molecule 1.9 g/cm^3. The system consists of 320,000 molecules with 200 lattice units in the z-directions and 20 lattice units along the x- and y-directions. Periodic boundary conditions are imposed along the x- and y-directions, while the z-direction has free surfaces. In order to follow the precise energy exchange between

the intra- and intermolecular DoF, we use an NVE ensemble where the system has been previously thermalized at 300 K and zero pressure.

8.4.2 Shock Loading and Energy Dissipation

The sample is impacted with a thin (one lattice constant thick), rigid and infinitely massive piston traveling at a constant speed in the z-direction. The infinitely massive piston does not slow down due to the interactions with the target, nor rarefaction waves are generated from its free surface. Thus, this setup generates a sustained shock where the piston does not slow down but it moves at a constant speed.

While in an inert material, plastic deformation or phase transformations are the only mechanisms that allow for stress relaxation, in a SWED material volume-reducing chemical reactions can significantly reduce pressure build up and attenuate the leading shock wave. The material under stress will attempt to deform or rearrange in order to alleviate its local stress build up. The material that is in its immediate proximity will quickly approach the same velocity and consequently this region experiences the largest strains rates where chemical reactions start to nucleate. As the piston is driven into the sample, a two-wave structure emerges separating a chemical region from the plastic region. Above a critical strain rate (HCL discussed below), a chemical region close to the piston nucleates and grows at a speed U_c, followed by a plastic regime growing at speed U_s; an elastic precursor to the plastic wave is present but it is negligible and very close in speed to the plastic wave, and therefore will be ignored for the rest of the discussion. Figure 8.8a shows the velocity profiles along the z-direction at various times after the piston impacts the target and Fig. 8.8b shows the pressure–volume (Hugoniot) curve representing the equation of state for the two regimes: one where both chemistry and plasticity nucleate (two-wave region), and one where only plasticity occurs (inert). The transition between the plastic regime and the chemical regime is denoted as the Hugoniot Chemical Limit (HCL). At extreme strains, the velocities of the chemical wave and the plastic wave will converge to the same value, and a single wave structure emerges, this is the overdriven regime.

The interplay between the various parameters playing a role in the kinetics of the model is discussed next.

In order to understand the roles that endothermicity, volume change, and the activation barrier have in the energy dissipation phenomena. We quantify the local pressure in both the inert and reactive parts of the Hugoniot in Fig. 8.9. Taking as a base case with an activation barrier $\Delta G = 30$ kcal/mol, volume change $= 35\%$ and endothermicity $\Delta H = 0$ kcal/mol. We see that increasing the activation barrier from $\Delta G = 30$ kcal/mol (red points) to 60 kcal/mol (blue points) will postpone the critical transition points (HCL) from about 7 to 17 Gpa, where a two-wave domain will persist at larger impact speeds until reaching the overdriven regime. Increasing the change in the volume collapse in the reactive material shifts the Hugoniot toward the left indicating a larger volume collapse. The pink points show the Hugoniot curve for

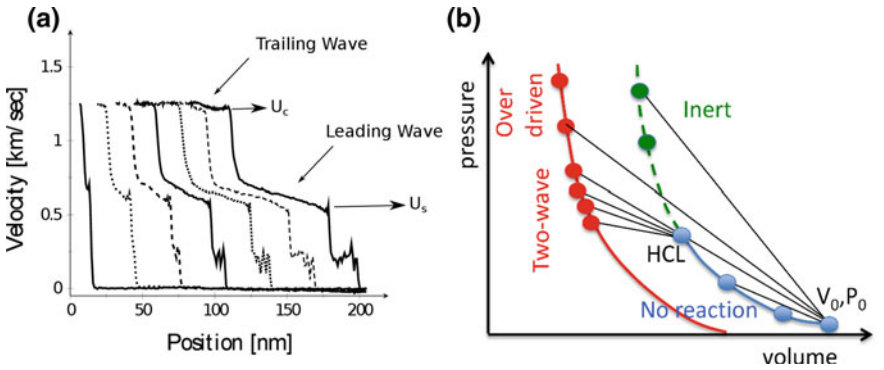


Fig. 8.8 **a** Progression of velocity profiles (in steps of 10 ps) showing the evolution of a leading wave traveling at speed U_s and a trailing wave (chemical wave) traveling at speed U_c. **b** A Hugoniot equation of state for reactive and inert materials. The transition from the inert-to-reactive is known as the Hugoniot Chemical Limit

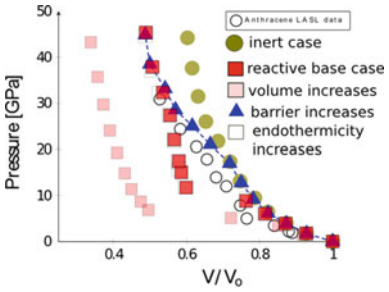


Fig. 8.9 **a** Hugoniot for various model materials for a range of piston speeds between $U_p = 0.25$ and 3.0 km/s (in steps of 0.25 km/s). The increase of volume collapse plays a major role in reducing the pressure until the Hugoniot state reaches the overdriven regime. In this regime, endothermicity only has a minor effect compared to a material with only volume-reducing properties. Larger activation barrier postpones the nucleation of chemical nucleation until larger impact speeds are accessible; For reference, experimental data on Anthracene is also shown

a volume change of 65% illustrating significant lower pressures in both the reactive and inert parts of the Hugoniot compared to the base case. The effect of endotrmicity was found to be null except for the overdriven regime. The empty symbols show a reactive material with endothermicity of $\Delta H = 30$ kcal/mol, which are consistent with the base case almost everywhere; in the overdriven regime, the pressures are lower but not significantly smaller.

In order to be have ideal shockwave absorbing properties, we argue that the HCL pressure needs to be as low as possible. This is because the plastic wave speed will travel at a fixed value in this two-wave regime independent of the impact speed of the piston and consequently its pressure will also be constant. Our model shows that the endothermicity plays a very weak role on the HCL, while the activation barrier and volume reduction both play a role in the HCL. Volume collapse lower than 20%

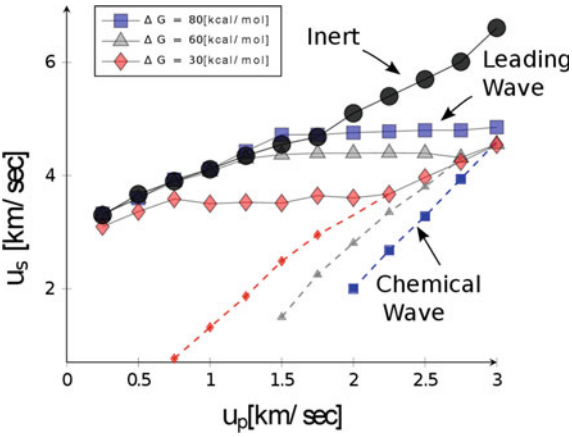


Fig. 8.10 Velocity of the shocked material (U_s) versus piston impact speed (U_p) for an inert material and three reactive materials with activation barriers between 30 and 80 kcal/mol. The chemical wave speed (U_c) is also shown in the dashed lines

are not expected at the molecular level, however, metal–organic frameworks (MOFs) can show larger volume collapse by breaking the ligands making up the structure leading to volume changes in the order of 60–80% [62].

The pressure in the shocked region can be related to the velocity of the shock wave by applying the Hugoniot–Rankine conditions.

$$P_s = \rho u_s \dot{x}_s$$

where ρ is the density of the unshocked material, u_s is the velocity of the leading shockwave, and $\dot{x}_s$ is the particle velocity. This last part can be shown to be a fraction of the velocity of the leading shockwave [27]. Hence, the pressure in the shocked region is a direct function of the shockwave velocity. Therefore, it is important to understand how chemical reaction can affect the shockwave velocity. At lower activation barriers, chemistry starts to nucleate earlier as a function of piston impact speed. Figure 8.10 shows the shock velocity of the leading wave (solid lines) and the chemical wave speed (dashed lines) versus the applied piston speed (U_p).

Above a critical impact speed, the HCL limits are reached, and a reactive region is able to grow at a steady rate. Chemistry nucleation helps to absorb the shock energy and helps to maintain the leading shockwave at a relatively constant speed. On the other hand, the inert case can only relieve stress through plastic deformations, and the shockwave velocity bends down slightly with piston speed until reaches the overdriven regime relatively early at about $U_p = 1.75$ km/s. The reactive model materials can postpone the overdriven regimes for the reactive materials after $U_p =$ 2.25, 2.75, and 3.0 km/s for $\Delta G = 30$, 60, and 80 kcal/mol, respectively. This shows that while rapid chemistry nucleation rate damps the shockwave energy the most, it ultimately arrives at the overdriven regime earlier.

In conclusion, endothermic, volume-reducing chemistry can weaken shockwaves [26, 27, 62, 63]. The role of endothermicity was found to be minimal while the volume-reducing contribution affects the damping the most. This study focuses on

sustained shockwaves, where the mass of the piston is assumed to be infinitely massive, and thus corresponds the extreme values expected during the collision with a finite piston. Once the local pressure comes down below the HCL, chemical reactions are not expected to nucleate and grow at larger scales. However, for the sizes considered in this study, the sustained wave simulations represent a realistic scenario of what might occur during the chemical reactions induced during shock loading conditions.

8.5 Conclusion and Outlook

This chapter focused on an atomic level and coarse grain studies of shock-induced chemistry. Initiating chemistry requires for the energy in the initial impact to be transferred into bond vibrations with characteristic lengths of Angstroms and vibrational periods on tens of femtoseconds (~10^{-14} s). In most solid HE materials, microstructural features like voids, grain boundaries, and cracks help localize the energy of the shock spatially into hotspots. At the same time, energy is localized in real space it needs to delocalize (or equilibrate) in the frequency domain. This is because, the macroscopic shock couples more strongly to long-frequency long-wavelength modes and inter- and intra-vibrational relaxation processes are responsible to locally equilibrate the system and transfer energy to high-frequency modes, this is called up-pumping [36]. A combination of large-scale atomistic simulations and recent experiments is providing unprecedented detail into these processes.

We described simulations playing various roles. On the one hand, detailed simulations about dynamical hotspots in RDX provide new insight into the reactivity of nanoscale hotspots. Specifically, we find that dynamically created hotspots are more reactive than counterparts with identical size and thermodynamic conditions by created under equilibrium conditions. On the other hand, coarse grain simulations can be used to help the design of new materials with the desired response of shock loading by relating how the characteristics of the chemical reactions affect shock propagation. Finally, first principles simulations have the potential to help the design of new materials with a reduced set of experiments, but the accuracy of its predictions must be rigorously quantified. We showed the emerging possibility of direct validation of predicted shock-induced chemistry against experiments. Ongoing efforts to achieve ultrafast, broadband spectroscopy in shock experiments will enable a better comparison and provide definite information about the detailed chemistry of HE materials under shock loading.

References

1. Holian BL, Lomdahl PS (1998) Plasticity induced by shock waves in nonequilibrium molecular-dynamics simulations. Sci 280:2085–2088
2. Bringa EM, Caro A, Wang Y et al (2005) Ultrahigh strength in nanocrystalline materials under shock loading. Sci 309:1838–1841
3. Chen MW, McCauley JW, Dandekar DP, Bourne NK (2006) Dynamic plasticity and failure of high-purity alumina under shock loading. Nat Mater 5:614
4. Duvall GE, Graham RA (1977) Phase transitions under shock-wave loading. Rev Mod Phys 49:523
5. Kadau K, Germann TC, Lomdahl PS, Holian BL (2002) Microscopic view of structural phase transitions induced by shock waves. Sci 296:1681–1684
6. Yang Y, Wang S, Sun Z, Dlott DD (2004) Propagation of shock-induced chemistry in nanoenergetic materials: the first micrometer. J Appl Phys 95:3667–3676
7. Strachan A, van Duin ACT, Chakraborty D et al (2003) Shock waves in high-energy materials: the initial chemical events in Nitramine RDX. Phys Rev Lett 91:098301. https://doi.org/10.1103/PhysRevLett.91.098301
8. Hirai H, Kondo K (1991) Modified phases of diamond formed under shock compression and rapid quenching. Sci 253:772–774
9. Levitas VI, Ravelo R (2012) Virtual melting as a new mechanism of stress relaxation under high strain rate loading. Proc Natl Acad Sci 109:13204–13207
10. Wood MA, Cherukara MJ, Antillon E, Strachan A (2017) Molecular dynamics simulations of shock loading of materials: a review and tutorial. Rev Comput Chem 43–92
11. Strachan A, Klimeck G, Lundstrom M (2010) Cyber-enabled simulations in nanoscale science and engineering. Comput Sci Eng 12:12–17
12. Reed EJ, Fried LE, Joannopoulos JD (2003) A method for tractable dynamical studies of single and double shock compression. Phys Rev Lett 90:235503. https://doi.org/10.1103/PhysRevLett.90.235503
13. Maillet J-B, Mareschal M, Soulard L et al (2000) Uniaxial hugoniostat: a method for atomistic simulations of shocked materials. Phys Rev E 63:016121. https://doi.org/10.1103/PhysRevE.63.016121
14. Ravelo R, Holian BL, Germann TC, Lomdahl PS (2004) Constant-stress hugoniostat method for following the dynamical evolution of shocked matter. Phys Rev B 70:014103. https://doi.org/10.1103/PhysRevB.70.014103
15. Bdzil JB, Stewart DS (2007) The dynamics of detonation in explosive systems. Annu Rev Fluid Mech 39:263–292. https://doi.org/10.1146/annurev.fluid.38.050304.092049
16. Yang K, Lee J, Sottos NR, Moore JS (2015) Shock-induced ordering in a nano-segregated network-forming ionic liquid. J Am Chem Soc 137:16000–16003
17. Van Duin AC, Dasgupta S, Lorant F, Goddard WA (2001) ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 105:9396–9409
18. Senftle TP, Hong S, Islam MM et al (2016) The ReaxFF reactive force-field: development, applications and future directions. Npj Comput Mater 2:15011. https://doi.org/10.1038/npjcompumats.2015.11
19. Tersoff J (1988) New empirical approach for the structure and energy of covalent systems. Phys Rev B 37:6991
20. Mortier WJ, Ghosh SK, Shankar S (1986) Electronegativity-equalization method for the calculation of atomic charges in molecules. J Am Chem Soc 108:4315–4320. https://doi.org/10.1021/ja00275a013
21. Strachan A, Holian BL (2005) Energy exchange between mesoparticles and their internal degrees of freedom. Phys Rev Lett 94:014301
22. Lynch K, Thompson A, Strachan A (2008) Coarse grain modeling of spall failure in molecular crystals: role of intra-molecular degrees of freedom. Model Simul Mater Sci Eng 17:015007
23. Lin K-H, Holian BL, Germann TC, Strachan A (2014) Mesodynamics with implicit degrees of freedom. J Chem Phys 141:064107. https://doi.org/10.1063/1.4891308

24. Onofrio N, Guzman D, Strachan A (2015) Atomic origin of ultrafast resistance switching in nanoscale electrometallization cells. Nat Mater 14:440–446. https://doi.org/10.1038/nmat4221
25. Onofrio N, Strachan A (2015) Voltage equilibration for reactive atomistic simulations of electrochemical processes. J Chem Phys 143:054109. https://doi.org/10.1063/1.4927562
26. Antillon E, Banlusan K, Strachan A (2014) Coarse grain model for coupled thermo-mechano-chemical processes and its application to pressure-induced endothermic chemical reactions. Model Simul Mater Sci Eng 22:025027
27. Antillon E, Strachan A (2015) Mesoscale simulations of shockwave energy dissipation via chemical reactions. J Chem Phys 142:084108
28. Millot M, Dubrovinskaia N a, Černok A, et al (2015) Shock compression of stishovite and melting of silica at planetary interior conditions. Sci 347:418–420
29. Erhart P, Bringa EM, Kumar M, Albe K (2005) Atomistic mechanism of shock-induced void collapse in nanoporous metals. Phys Rev B 72:052104
30. Wood MA, Cherukara MJ, Kober EM, Strachan A (2015) Ultrafast chemistry under nonequilibrium conditions and the shock to deflagration transition at the nanoscale. J Phys Chem C 119:22008–22015. https://doi.org/10.1021/acs.jpcc.5b05362
31. Knudson MD, Desjarlais MP, Dolan DH (2008) Shock-wave exploration of the high-pressure phases of carbon. Sci 322:1822–1825
32. Dreger ZA, Gruzdkov YA, Gupta YM, Dick JJ (2002) Shock wave induced decomposition chemistry of pentaerythritol tetranitrate single crystals: time-resolved emission spectroscopy. J Phys Chem B 106:247–256
33. McGrane SD, Moore DS, Funk DJ (2004) Shock induced reaction observed via ultrafast infrared absorption in poly(vinyl nitrate) films. J Phys Chem A 108:9342–9347. https://doi.org/10.1021/jp048464x
34. Dlott DD (2011) New developments in the physical chemistry of shock compression. Annu Rev Phys Chem 62:575–597. https://doi.org/10.1146/annurev.physchem.012809.103514
35. Dlott DD (1990) Theory of ultrahot molecular solids: vibrational cooling and shock-induced multiphonon up pumping in crystalline naphthalene. J Chem Phys 93:1695–1709. https://doi.org/10.1063/1.459097
36. Dlott DD (1990) Shocked molecular solids: vibrational up pumping, defect hot spot formation, and the onset of chemistry. J Chem Phys 92:3798–3812. https://doi.org/10.1063/1.457838
37. Bassette WP, Dlott DD (2016) High dynamic range emission measurements of shocked energetic materials: Octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine (HMX). J Appl Phys 119:225103. https://doi.org/10.1063/1.4953353
38. Shan T-R, Thompson AP (2014) Micron-scale reactive atomistic simulations of void collapse and hotspot growth in pentaerythritol tetranitrate. In: 15th Int. Detonation Symp. Sandia National Laboratories (SNL-NM), Albuquerque, NM (United States), Albuquerque, NM (United States), p SAND2015–1243C
39. Brenner DW, Robertson DH, Elert ML, White CT (1993) Detonations at nanometer resolution using molecular dynamics. Phys Rev Lett 70:2174
40. Hu Y, Brenner DW, Shi Y (2011) Detonation initiation from spontaneous hotspots formed during cook-off observed in molecular dynamics simulations. J Phys Chem C 115:2416–2422
41. Marsh SP (1980) LASL shock hugoniot data. University of California Press
42. Cherukara MJ, Germann TC, Kober EM, Strachan A (2014) Shock loading of granular Ni/Al composites. Part 1: mechanics of loading. J Phys Chem C 118:26377–26386
43. Cherukara MJ, Germann TC, Kober EM, Strachan A (2016) Shock loading of granular Ni/Al composites. Part 2: shock-induced chemistry. J Phys Chem C 120:6804–6813
44. Holian BL, Germann TC, Maillet J-B, White CT (2002) Atomistic mechanism for hot spot initiation. Phys Rev Lett 89:285501
45. Herring SD, Germann TC, Grønbech-Jensen N (2010) Effects of void size, density, and arrangement on deflagration and detonation sensitivity of a reactive empirical bond order high explosive. Phys Rev B 82:214108
46. Campbell AW, Davis WC, Travis JR (1961) Phys Fluids 4(4):498

47. Bowden FP, Yoffe AD (1952) Initiation and growth of explosion in liquids and solids. CUP Archive
48. Field JE (1992) Hot spot ignition mechanisms for explosives. Acc Chem Res 25:489–496
49. Islam MM, Strachan A (2017) Decomposition and reaction of polyvinyl nitrate under shock and thermal loading: a ReaxFF reactive molecular dynamics study. J Phys Chem C. https://doi.org/10.1021/acs.jpcc.7b06154
50. Wood MA, van Duin AC, Strachan A (2014) Coupled thermal and electromagnetic induced decomposition in the molecular explosive αHMX: a reactive molecular dynamics study. J Phys Chem A 118:885–895
51. Moore DS, McGrane SD, Funk DJ (2004) Ultrafast spectroscopic investigation of shock compressed energetic polymer films. AIP Conf Proc 706:1285–1288. https://doi.org/10.1063/1.1780473
52. Brown KE, McGrane SD, Bolme CA, Moore DS (2014) Ultrafast chemical reactions in shocked nitromethane probed with dynamic ellipsometry and transient absorption spectroscopy. J Phys Chem A 118:2559–2567. https://doi.org/10.1021/jp4125793
53. Strachan A, Kober EM, van Duin AC et al (2005) Thermal decomposition of RDX from reactive molecular dynamics. J Chem Phys 122:054502
54. Shan T-R, Thompson AP (2014) Shock-induced hotspot formation and chemical reaction initiation in PETN containing a spherical void. In: J Phys Conf Ser, IOP Publishing, p 172009
55. An Q, Zybin SV, Goddard WA III et al (2011) Elucidation of the dynamics for hot-spot initiation at nonuniform interfaces of highly shocked materials. Phys Rev B 84:220101
56. Kulkarni SG, Gao X-L, Horner SE et al (2013) Ballistic helmets—their design, materials, and performance against traumatic brain injury. Compos Struct 101:313–331
57. Morinière FD, Alderliesten RC, Benedictus R (2014) Modelling of impact damage and dynamics in fibre-metal laminates—a review. Int J Impact Eng 67:27–38
58. Sadighi M, Alderliesten RC, Benedictus R (2012) Impact resistance of fiber-metal laminates: a review. Int J Impact Eng 49:77–90
59. López-Puente J, Arias A, Zaera R, Navarro C (2005) The effect of the thickness of the adhesive layer on the ballistic limit of ceramic/metal armours: an experimental and numerical study. Int J Impact Eng 32:321–336
60. Crupi V, Epasto G, Guglielmino E (2012) Collapse modes in aluminium honeycomb sandwich panels under bending and impact loading. Int J Impact Eng 43:6–15
61. Grujicic M, Pandurangan B, Bell WC et al (2011) Molecular-level simulations of shock generation and propagation in polyurea. Mater Sci Eng A 528:3799–3808
62. Banlusan K, Strachan A (2016) Shockwave energy dissipation in metal-organic framework MOF-5. J Phys Chem C 120:12463–12471. https://doi.org/10.1021/acs.jpcc.6b02283
63. Banlusan K, Antillon E, Strachan A (2015) Mechanisms of plastic deformation of metal-organic framework-5. J Phys Chem C 119:25845–25852

Chapter 9
Data-Driven Methods for Building Reduced Kinetic Monte Carlo Models of Complex Chemistry from Molecular Dynamics Simulations

Qian Yang, Carlos A. Sing-Long, Enze Chen and Evan J. Reed

Abstract Complex chemical processes such as those found in combustion, the decomposition of energetic materials, and the chemistry of planetary interiors, are typically studied at the atomistic level using molecular dynamics (MD) simulations. A nascent but growing trend in many areas of science and technology is to consider a data-driven approach to studying complex processes, and molecular dynamics simulations, especially at high temperatures and pressures, are a prime example of an area ripe for disruption with this approach. MD simulations are expensive, but each simulation generates a wealth of data. In this chapter, we discuss a statistical learning framework for extracting information about the underlying chemical reactions observed in MD data, and using it to build a fast kinetic Monte Carlo (KMC) model of the corresponding chemical reaction network. We will show our KMC models can not only extrapolate the behavior of the chemical system by as much as an order of magnitude in time but can also be used to study the dynamics of entirely different chemical trajectories. We will also discuss a new and efficient data-driven algorithm for reducing our learned KMC models using L1-regularization. This allows us to reduce complex chemical reaction networks consisting of thousands of reactions in a matter of minutes.

9.1 Introduction

Complex chemical processes such as those found in combustion, the decomposition of energetic materials and the chemistry of planetary interiors, are typically studied at the atomistic level using molecular dynamics (MD) simulations. This poses two

Q. Yang (✉)
University of Connecticut, Storrs, CT 06269, USA
e-mail: qyang@uconn.edu

C. A. Sing-Long
Pontificia Universidad Catolica de Chile, Santiago, Chile

E. Chen · E. J. Reed
Stanford University, Stanford, CA 94305, USA

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_9

distinct challenges. First, the length and timescales currently available to MD are generally still too small for meaningful comparison with many experimental measurements. With current algorithms and computational power, molecular dynamics simulations generally require weeks to run on high-performance parallel machines for system sizes of only a few thousand atoms and timescales of nanoseconds. Second, even though we want to run bigger and longer MD simulations, the ones we can currently run are already too large and complex to easily interpret. Often, MD simulations are run and some aggregate property of the system is computed; there are few existing frameworks for understanding the detailed reactions that are driving the aggregate property.

To address the first problem, researchers have worked over the past 20 years to develop a variety of methods for accelerating atomistic simulation, resulting in many innovative approaches such as hyperdynamics [24], parallel replica methods [25], temperature-accelerated MD [20], and enhanced sampling [2]. Kinetic Monte Carlo (KMC) methods have also become an important way to speed up atomistic simulation. In KMC, the system is described by a set of possible states, and rates are given for the transitions between each pair of states, allowing the algorithm to traverse through them over time [26]. A key limitation of KMC is that the set of states and transition rates must be known a priori. Various approaches to get around this problem include adaptive KMC [28] and coupled KMC/MD [10, 23].

To address the second problem, researchers in the combustion, biochemistry, and catalysis communities, among others, have worked to develop methods for model reduction of chemical reaction networks. These approaches include the lumping of species and reactions, graph theoretic approaches, and quasi-steady-state approximations [17], as well as sensitivity matrices [12, 22], integer programming [1, 3, 7, 14], and genetic algorithms [19]. A common challenge with these methods is their dependence on a complex set of parameters, making them difficult to efficiently generalize across systems. Many are also computationally expensive, making them undesirable for studying complex systems with hundreds to tens of thousands of reactions.

These two problems often collide when researchers attempt to build multiscale models [18]. On the one hand, it is currently extremely difficult to simulate physical phenomena such as fracture using atomistic methods due to the computational expense of running simulations at the relevant time and length scales. On the other hand, there exist mesoscale models that can predict these properties of the system with the input of a simple model that captures the important chemistry of the system. This means that researchers often use a combination of physical intuition and data from MD simulations to build a reduced model of the complex chemistry in the system, which they then plug into the mesoscale model. Although this approach is extremely useful and often the only practical way to drive computational predictions of many materials properties, it is ultimately an unsatisfactory process. It is time consuming, limited by physical intuition, and impossible to generalize across different systems. Thus this area is ripe for innovation.

With the increasing availability of large-scale computation, a nascent but growing trend in many areas of science and technology is to consider a *data-driven*

approach. While data from experiments have always played an important part in the development and parameterization of physical models from the beginning of scientific history, what has changed in recent years is the vast *amount* of data that scientists are now able to work with. Whether through high-throughput computation, high-throughput experimentation, or data mining of existing literature, researchers are increasingly working to generate and collect large quantities of data. The availability of this data enables a paradigm shift in the way we think about building useful models of physical phenomena.

Atomistic simulation is a prime example of an area ripe for disruption with the data-driven approach. Although MD simulations are expensive, each simulation generates detailed data consisting of the atomic positions of every atom in a finely resolved time series. It is then natural to wonder whether this wealth of data can be used to do more than simply describe the particular system being studied. In this chapter, we discuss a framework for extracting information about the fundamental underlying elementary reactions observed from MD data, and using it to build a kinetic Monte Carlo model of the corresponding chemical reaction network [30]. We will show that our KMC models can not only extrapolate the behavior of the chemical system by as much as an order of magnitude in time but can also be used to study the dynamics of entirely different chemical trajectories. Then, we will discuss a new and efficient data-driven algorithm for reducing our learned KMC models using L1-regularization [29, 30].

9.2 Data-Driven Approach

In a typical data-driven learning problem, there is an *outcome* that we wish to predict from a set of *features* that describe an object. This is done by using *data* consisting of many observed objects encoded as features and their corresponding outcomes to learn a model for the relationship between features and outcomes.

The data that underlies our framework is provided by molecular dynamics simulations. The features that we use are a set of elementary reactions, and the outcome is the probability distribution of concentrations of each molecule over time. We use the chemical master equation [9], which can be simulated exactly using KMC, to model the relationship between the features and the outcome. To fit the model to the data, we learn the statistically optimal reaction rate constant for each elementary reaction such that the KMC model of the resulting chemical master equation best corresponds to the observed MD data. We discuss more carefully in later sections how our framework defines the statistically optimal reaction rate constant, as well as our metric for how well a given KMC model corresponds to the chemical system as described by molecular dynamics.

We note that the more features that are used, the more complex the model. A key aspect of the data-driven learning problem is choosing a model with just the right amount of complexity: not so few that it is unable to capture important patterns in the data (underfitting), but not so much that it captures noise from the dataset

as false patterns (overfitting), causing the model to perform poorly on new data. In statistics and machine learning, this is known as the bias–variance tradeoff [8]. In our framework, the complexity of the model is given by the number of elementary reactions that we use to describe the MD data.

We extract elementary reactions from the molecular dynamics dataset by using bond length and duration criteria. The bond length criteria is chosen to be fixed based on radial distribution functions. The bond duration criteria, on the other hand, are a parameter that plays an important role in determining the complexity of the model. If the bond duration τ is chosen to be too small, then many reactions will be identified, creating a complex model that is likely to be overfit to noise in the data. In this case, the noise corresponds to atomic vibrations. If τ is chosen to be too large, then too few reactions may be identified, creating a model that may not be able to express the full complexity of the system, underfitting the data. Thus, it is important to choose τ to optimize the bias–variance tradeoff.

One aspect of our learning problem that is a bit different from the traditional setting is that both the features and the outcomes have been projected into a different space from the original data. This is because the bond durations we choose also determine the molecule concentrations over time in the molecular dynamics simulation, so what we are looking for is a self-consistent model. We will show that these statistically learned models can successfully extrapolate in both time and chemical space.

9.2.1 Stochastic Model

We use the chemical master equation (CME) as the underlying stochastic model for describing the complex chemistry in our system. The chemical master equation is a system of ordinary differential equations (ODEs) that gives the probability $\mathbb{P}(x, t)$ of a system being in a particular state $X(t) = x$ in molecular concentration space at time t. The chemical master equation can be simulated exactly using Gillespie stochastic simulation (GSS), which is equivalent to KMC and only requires knowing the set of reactions and their corresponding rate constants. The model assumes that the system has constant volume and is well stirred, so that the spatial position of the molecules does not appreciably affect the rate of reaction between them [9].

We first perform molecular dynamics simulations under high temperature and pressure conditions. In our work, we use the ReaxFF potential [11] to simulate a computational cell of 216 methane molecules for approximately half a nanosecond. This allowed us to observe more reactions in a single MD trajectory within a reasonable computational time. Note that our algorithm does not depend on the particular potential being used. It is possible to do the same analysis with a different potential, for example, using ab initio molecular dynamics.

Our simulations are run using LAMMPS [15] at a temperature of 3300 K and a pressure of 40.53 GPa. The bond length criteria that we use to determine whether two atoms are bonded were determined using radial distribution functions taken from previous work [16]. We say that two atoms are bonded if they are within a bond

length distance of each other for at least a bond duration of τ timesteps. Similarly, we say that two atoms that were previously bonded are unbonded only if they remain beyond a bond length distance from each other for more than τ timesteps. We choose τ to minimize the error between the stochastic model and the MD data, as described below. In our optimal model, we find approximately 2000 distinct reactions and 500 distinct molecules.

We treat the reactions as if they are elementary. That is, we assume they have a single energy barrier with a corresponding rate constant k. We associate with every reaction j a propensity function $a_j(x)$, such that $a_j(x)dt$ gives the probability of that reaction occurring in the time interval $[t, t + dt)$ given the vector of molecular concentrations $X(t) = x$ at time t. It can be shown that this propensity function is proportional to the number of possible combinations of the reactant molecules at time t [4, 5]. The constant of proportionality (e.g., the reaction rate constant) k_j and the molecular concentrations $X(t)$ give rise to the propensity functions in the following way:

- Unimolecular reactions $X_m \rightarrow$ products: $a_j(X) = k_j X_m$
- Bimolecular reactions $2X_m \rightarrow$ products: $a_j(X) = k_j X_m(X_m - 1)$
- Bimolecular reactions $X_m + X_{m'} \rightarrow$ products: $a_j(X) = k_j X_m X_{m'}$.

If necessary, this can be extended naturally to elementary reactions involving more than two reactants (which are rare). The chemical master equation is fully determined by the set of all propensity functions that correspond to reactions in the system. The Gillespie stochastic simulation algorithm can then be used to simulate the evolution over time of the molecular concentrations in the chemical system [4, 5]. At each iteration, the algorithm first chooses at random the next reaction to occur based on the relative probabilities of each reaction given by their propensity functions and then computes the time until the next reaction. This algorithm is equivalent to KMC. Each time a reaction occurs, the concentrations of the reactants and products change, leading to a transition between states defined by the molecular concentration vector.

9.2.2 Maximum Likelihood Estimation

While the propensity functions $a_j(x)$ are defined with respect to infinitesimal time intervals $[t, t + \mathrm{d}t]$, molecular dynamics are numerical simulations and have an integration timestep of Δt. Therefore, we cannot estimate the k_j for the chemical master equation directly from molecular dynamics data. However, we can try to get around this by making the assumption that Δt is small enough such that the $a_j(X(t'))$ is approximately constant throughout the time interval $t' \in [t, t + \Delta t)$ for all reactions. This is known as the tau-leaping approximation [6]. Intuitively, this means that the molecular concentration $X(t')$ does not change appreciably over this interval, which means that only a few reactions have occurred, and that the molecular concentrations of all reactant molecules are large enough to make any resulting changes in $a_j(X(t'))$ relatively small for all reactions j. Although this is a strong assumption,

we will find that it is sufficient to give us reasonable models. This assumption allows us to model the number of times $n_j(t, t+\Delta t)$ that each reaction j occurs in the time interval $[t, t+\Delta t)$ as instances of a conditionally independent Poisson random variable $\mathcal{P}_j|X(t)$ with mean and variance given by $\lambda_j = a_j(X(t))\tau$.

At each timestep, projecting the molecular dynamics simulations via bond length and duration criteria gives rise to a set of molecular concentrations $X(t)$ and the number of times $n_j(t, t+\Delta t)$ that reaction j has occurred. Each $n_j(t+\Delta t)$ is an observation of $\mathcal{P}_j|X(t)$, which are all conditionally independent but not identically distributed Poisson random variables. Then, we can use maximum likelihood estimation [27] to estimate k_j from our observations. The likelihood of observing a particular sequence of $n_j(t, t+\Delta t)$ can be given by the equivalent expression

$$\mathcal{L} = \prod_t \mathbf{Pr}(\mathcal{P}_j = n_j(t, t+\Delta t)|X(t))$$

The log likelihood is then given by

$$\ell = \log(\mathcal{L}) = \sum_t \log \mathbf{Pr}(\mathcal{P}_j = n_j(t, t+\Delta t)|X(t))$$

Since the Poisson distribution has a probability density function given by

$$\mathbf{Pr}(\mathcal{P}_j = n_j) = \frac{\lambda_j^{n_j} e^{-\lambda_j}}{n_j!}$$

maximum likelihood estimation over the $n_j(t, t+\Delta t)$ gives a simple estimate for the reaction rates

$$k_j^* = \frac{1}{\Delta t} \frac{\sum_{t=1}^N n_j(t, t+\Delta t)}{\sum_{t=1}^N h_j(X(t))}$$

where $h_j(x) = a_j(x)/k_j$ corresponds to the propensity function without the constant factor. For example, for bimolecular reactions $X_m + X_{m'} \rightarrow$ products, we have $h_j(x) = X_m X_{m'}$.

Note that each reaction rate is estimated independently. We can see from Fig. 9.1 that there is good agreement between the molecular concentration trajectory given by the Gillespie stochastic simulations (colored lines) and molecular dynamics (black lines). Note that in this figure, we only plot the three molecules with the largest concentration in the system; there are hundreds of other distinct species in the system at any given time.

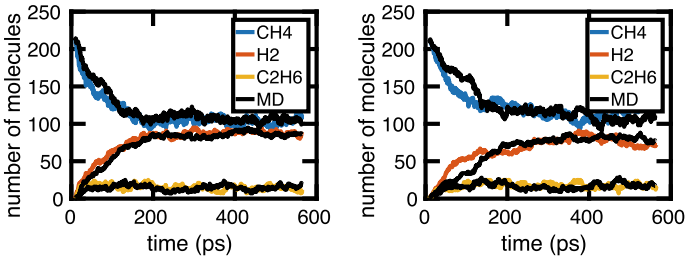


Fig. 9.1 Two examples of Gillespie stochastic simulations of a learned model of the system from two independent molecular dynamics simulations that they were trained from. We see that both achieve reasonable agreement, especially in comparison to fluctuations between independent MD simulations of the same system. Reproduced from [30] with permission from the Royal Society of Chemistry

9.2.3 Model Selection via Bond Duration

As described above, we carefully choose bond duration τ to optimize the complexity of our model. As τ increases, fewer bond formation and bond breaking events are detected. This also leads to fewer distinct reactions being detected, as depicted in Fig. 9.2a.

We can see from Fig. 9.2b that our method of extracting reactions from the MD data results in mostly unimolecular and bimolecular reactions, and a few trimolecular reactions. There are a negligible amount of reactions that are more complex than trimolecular, which is reasonable at these high temperatures and pressures. This suggests that our assumption of elementary reactions with one energy barrier is not implausible.

The complexity of the stochastic model describing the molecular dynamics data decreases with increasing bond duration τ. If we choose a model that is too complex, we might erroneously identify atomic vibrations as elementary reactions. This leads to a high model error. If we choose a model that is too simple, we might miss true elementary reactions. This again leads to an increase in model error. The dotted line near $\tau = 0.1$ ps in Fig. 9.2c indicates the optimal choice of τ for our model of the methane molecular dynamics simulation.

9.2.4 Error Metric

In an ideal setting, many MD simulations of the same chemical system would be available, from which a probability distribution of the molecular concentrations of molecules over time can be constructed, which can then be compared to the probability distribution given by the learned stochastic model. However, for many practical problems, molecular dynamics is too expensive to generate a statistically signifi-

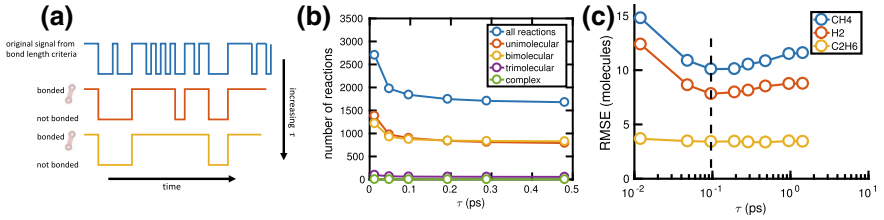


Fig. 9.2 **a** Schematic of how varying the bond duration τ changes the identification of bonded atoms. **b** The number of elementary reactions extracted from the MD data decreases with increasing bond duration τ. **c** Bond durations that are too short erroneously capture atomic vibrations as elementary reactions; bond durations that are too long miss true elementary reactions. The dotted line near $\tau = 0.1$ ps indicates the optimal choice of τ for our model of the methane system. Reproduced from [30] with permission from the Royal Society of Chemistry

cant number of independent simulations. We have thus designed an error metric that assumes a single MD simulation corresponds to the mean overall independent MD simulations of a given system. This is usually not a bad assumption, especially if we are interested in equilibrium properties of the system. Then, we use the root mean square error between the MD simulation and the mean of S stochastic simulations to characterize the error between MD and our stochastic model. We find the error for each molecular species, m, separately:

$$\mathrm{Err}_m = \sqrt{\frac{\sum_{t=1}^{N} \left(X_{\mathrm{MD}}[t] - \mathbb{E}_S\{X_{\mathrm{GSSA}}[t]\}\right)^2}{N}}$$

In the above expression, N corresponds to the total number of sampled timesteps of the molecular dynamics simulation. Gillespie stochastic simulation gives a continuous time model, so we sample it at the same times as the MD. We choose S to be large enough that the error converges. Note that this error should always be compared to the amount of inherent fluctuation in a single MD trajectory for the corresponding molecular species.

9.2.5 Extrapolation

The stochastic models that we learn from molecular dynamics data using the framework above can be shown to extrapolate in both time and chemical space. These extrapolations provide substantive test cases for our model, revealing that the set of learned elementary reactions and rate constants are truly capturing underlying information about the chemistry of these systems.

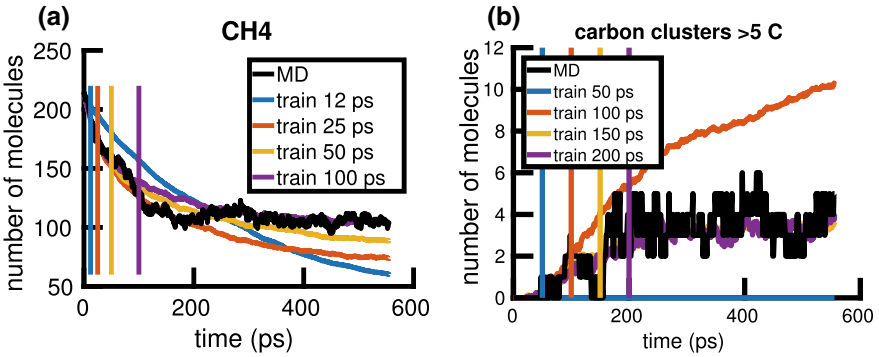


Fig. 9.3 Extrapolation in time. The colored lines correspond to Gillespie simulations averaged over 20 independent runs. The black lines correspond to molecular dynamics. Reproduced from [30] with permission from the Royal Society of Chemistry

9.2.5.1 Time

One potential application for learning stochastic models from MD is to then use them to extrapolate forward in time. We find that it is indeed possible for reactions observed early in our molecular dynamics simulations to model the largest population molecules for 10 times longer or more. For example for CH_4, we can train on only 25 picoseconds (ps) and extrapolate out to 200 ps. Note that training on only 12 ps, however, was not sufficient. For larger molecule species that first occur in the system after about 50 ps, we find that training data from times out to 150 ps are needed to capture their presence and growth in the chemical system over the full half nanosecond. These results are shown in Fig. 9.3.

9.2.5.2 Chemical Space

Our stochastic model also provides a new way to study-related systems in different regions of molecular concentration space. We demonstrate this by building a new stochastic model trained from molecular dynamics simulation of a different system, and show that the resulting model is able to successfully predict the molecular dynamics trajectory of our original methane system. We first simulated 64 molecules of isobutane C_4H_{10}, under the same pressure and temperature conditions as methane. Next, we trained a stochastic model of the isobutane system using the framework described above. Then, we used Gillespie stochastic simulation to simulate this stochastic model but initialized with 216 molecules of CH_4. Given that this new stochastic model was not trained on any of the methane data, this is a true test case. The comparison between simulation using the new stochastic model and the original molecular dynamics is shown in Fig. 9.4a.

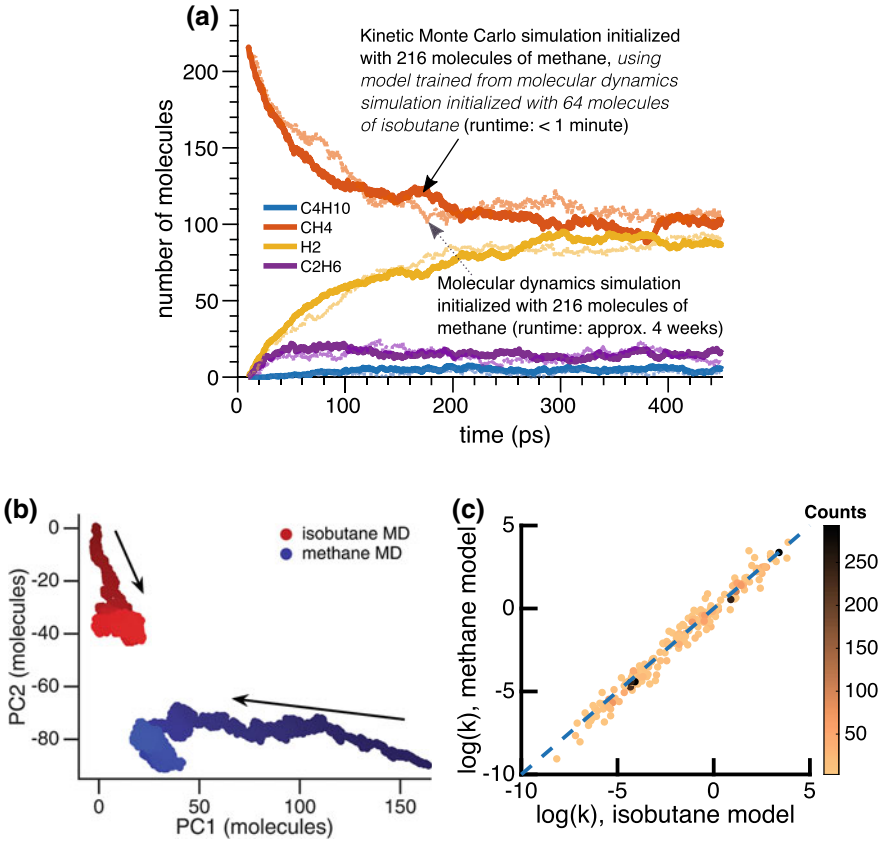


Fig. 9.4 **a** Extrapolation from a stochastic model trained on an isobutane system to a methane system. There is good agreement between the two simulations. **b** To visualize how the isobutane and methane systems traverse completely different regions of chemical space, we project their molecular concentration trajectories onto the first and second principal components of the methane trajectory. The arrows and shading indicate the direction of increasing time. **c** Comparison of the estimated reaction rate constants k_j for reactions that appear in both stochastic models trained from molecular dynamics simulations of the two different chemical systems. Darker colors correspond to reactions that occur more frequently and thus are expected to have better estimates for k_j. Note that none of the k_j differ by more than two orders of magnitude between the stochastic models, even though the spread of the rate constants is more than twelve orders of magnitude

Why do we believe that this extrapolation should work? If we compare their molecular concentration trajectories, we see that they traverse completely different regions of chemical concentration space. We can visualize this by projecting these trajectories into two dimensions. Consider for convenience projecting onto to the first and second principal components of the methane dataset. We can see in Fig. 9.4b that their projections do not intersect. However, our stochastic model is not described in terms of molecular concentrations, but rather in terms of elementary reactions. When we consider the set of reactions that are observed in the methane and isobutane datasets,

we can see that a large portion of the reactions extracted from the methane system are also extracted from the isobutane system, and furthermore that the reaction rate constants for these reactions generally differ by no more than one or two orders of magnitude. This explains the ability of our models to extrapolate—they are capturing something about the underlying chemistry in the system, rather than only information specific to the molecular concentration space that they have traversed, which is true of raw phase space data from molecular dynamics. Note that while each molecular dynamics simulation required several weeks on high-performance parallel machines, the Gillespie simulations completed in less than a minute on a standard laptop computer, and were easily and instantaneously initialized to a different system.

9.3 Model Reduction

As noted above, even for a relatively simple hydrocarbon system initialized with 216 molecules of methane run for half a nanosecond, approximately 2000 reactions are extracted from a single molecular dynamics simulation. In order to understand the key reactions that drive the underlying dynamics of the most important molecules in the system, we would like to find a reduced subset of reactions that can model the system with a minimal loss in accuracy for the important molecules.

Model reduction of complex chemistry with thousands of reactions is a difficult computational problem. This is in part because the concentration changes nonlinearly with respect to the current concentration. To see this, consider a toy problem involving three reactions and three molecular species:

$$\begin{aligned} A + B &\rightarrow C && \text{with rate constant } k_1 \\ A &\rightarrow 2C && \text{with rate constant } k_2 \\ 2C &\rightarrow A && \text{with rate constant } k_3 \end{aligned}$$

Then the corresponding reaction rate equations are given by

$$\begin{aligned} \frac{\mathrm{d}[A]}{\mathrm{d}t} &= -k_1[A][B] - k_2[A] + k_3[C]\,([C] - 1) \\ \frac{\mathrm{d}[B]}{\mathrm{d}t} &= -k_1[A][B] \\ \frac{\mathrm{d}[C]}{\mathrm{d}t} &= k_1[A][B] + 2k_2[A] - 2k_3[C]\,([C] - 1) \end{aligned}$$

Note that the bimolecular reactions cause the reaction rate equations to be a nonlinear dynamical system, which greatly increases the computational complexity of many model reduction algorithms. This is also a feature of the stochastic model based on the chemical master equation that we described above.

An important observation from the reaction rate equations, however, is that its right-hand side is linear in the set of reaction rate constants k_j. We have developed [29, 30] a computationally efficient method for model reduction of chemical reaction networks that takes advantage of this observation, combined with a data-driven approach to capture the nonlinearity of the system. Our method is subject to only one parameter that determines the tradeoff between the amount of reduction in the model and its corresponding loss in accuracy. We also note that it can be easily generalized to other nonlinear dynamical systems outside of chemical reaction networks.

Here, we will describe our model reduction method applied to the stochastic model based on the chemical master equation described above. At each timestep, the molecular concentrations change according to the set of reactions that have taken place. For each reaction j, let the column vector R_j represent the stoichiometric coefficients of the reactant and product molecules. For example, for the toy problem above, the corresponding R_j would be

$$R_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad R_2 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \quad R_3 = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$$

Then, the change in concentration from one timestep to the next in our stochastic model is described by

$$X(t + \Delta t) - X(t) = \sum_{j=1}^{r} R_j \mathcal{P}_j(a_j(X(t))\Delta t)$$

where the $\mathcal{P}_j$'s are all independent Poisson random variables that describe the number of times reaction j will occur, and r is the total number of distinct reactions.

We can see that the first and second moments of the change in concentration are given by

$$\mathbb{E}\{X(t + \Delta t) - X(t)\} = \sum_{j=1}^{r} R_j a_j(X(t))\tau = \sum_{j=1}^{r} R_j k_j h_j(X(t))\tau$$

$$\mathbf{cov}\{X(t + \Delta t) - X(t)\} = R\Lambda(X(t))R^T$$

where $\Lambda(X(t))$ is a diagonal matrix whose jth entry is $a_j(X(t))\Delta t$. Consider for example the simple chemical system described above. Then, the expected value of the changes in concentration of each molecule is given by

$$\begin{aligned}
&\frac{1}{\Delta t}\mathbb{E}\begin{bmatrix} A(t+\Delta t)-A(t) \\ B(t+\Delta t)-B(t) \\ C(t+\Delta t)-C(t) \end{bmatrix} \\
&\quad = k_1\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} A(t)B(t) + k_2\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} A(t) + k_3\begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} C(t)\left(C(t)-1\right) \\
&\quad = \underbrace{\begin{bmatrix} -1 & -1 & 1 \\ -1 & 0 & 0 \\ 1 & 2 & -2 \end{bmatrix}}_{\mathbf{R}} \underbrace{\begin{bmatrix} A(t)B(t) & 0 & 0 \\ 0 & A(t) & 0 \\ 0 & 0 & C(t)\left(C(t)-1\right) \end{bmatrix}}_{\mathbf{D}_t} \underbrace{\begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix}}_{\mathbf{k}}
\end{aligned}$$

We can write the above compactly as

$$\mu_{t+\Delta t}|X(t) = \mathbf{R}\mathbf{D}_t\mathbf{k}$$

and a similar expression linear in the vector of rate constants $\mathbf{k}$ can be obtained for the covariance of the changes in concentration:

$$\Sigma_{t+\Delta t}|X(t) = \mathbf{H}\mathbf{D}_t\mathbf{k}$$

Note that both the first and second moments of the change in concentration from any given point in molecular concentration space $X(t) = x$ can be expressed linearly in the vector of reaction rate constants $\mathbf{k}$. We thus formulate our model reduction problem in a moment-matching framework. We seek the smallest subset of reactions that is able to maximally match the first and second moments of the changes in concentration of the full reaction network at any given point in molecular concentration space.

9.3.1 L_1 Regularization

By sampling the molecular dynamics trajectory at different timesteps, we generate samples of relevant points in molecular concentration space. We can write one linear system for each sampled point for the first and second moments (expected value and covariance) of the change in concentration at the next timestep. We can then stack all of them into one large linear system as follows:

$$\frac{1}{\tau}\begin{bmatrix} \mu_2|_{X_1} \\ \Sigma_2|_{X_1} \\ \cdots \\ \mu_{n+1}|_{X_n} \\ \Sigma_{n+1}|_{X_n} \end{bmatrix} = \begin{bmatrix} RD_1 \\ HD_t \\ \cdots \\ RD_n \\ HD_n \end{bmatrix} \mathbf{k}$$
$$b = A\mathbf{k}$$

This large stacked linear system captures the nonlinearity of the chemical reaction network by sampling points in molecular concentration space traversed by the nonlinear dynamical system.

We would like to choose a subset of reactions for the reduced model by setting as many of the reaction rate constants k_j to zero as possible. Note that we first scale the problem using our estimated k_j^*'s from the full stochastic model described above, so that the vector $\mathbf{k} = \mathbf{1}$ is an exact solution of $A\mathbf{k} = b$. Then to find the reduced model, we formulate the nonnegative LASSO (least absolute selection and shrinkage operator) problem [13, 21]:

$$\begin{aligned} \min_{\mathbf{k}} \quad & \|A\mathbf{k} - b\|_2 \\ \text{s.t.} \quad & \mathbf{k} \geq 0, \quad \|\mathbf{k}\|_1 \leq \lambda \end{aligned}$$

By expressing our stochastic model in terms of a linear system of equations, our model reduction problem can be expressed with a simple quadratic objective. This makes solving the problem very efficient numerically, in contrast to many other model reduction methods where the optimization objective is expressed in terms of the error in the molecular concentrations, which are solutions of a nonlinear dynamical system.

Note that we have two constraints in our optimization problem. The first constraint is a physical one—no reaction rate can be negative. The second constraint is intended to bias the solution towards one where as many of the k_j's are zero as possible, i.e., to find a *sparse* $\mathbf{k}$. Note that if we wanted to exactly constrain the solution to have no more than λ zero entries, the problem would be NP hard. Thus, we use instead L_1-regularization, which allows us to solve the problem for thousands of reactions in a matter of minutes, scaling polynomially in the total number of reactions r.

To see why L_1-regularization promotes sparsity, consider the optimization problem

$$\begin{aligned} \min_{x} \quad & \|a^T x - b\|_2 \\ \text{s.t.} \quad & \|x\|_p \leq \lambda \end{aligned}$$

For different choices of norm p, the solution will have different characteristics. In Fig. 9.5, we illustrate a comparison between the commonly used L_2-regularization and the sparsity-promoting L_1-regularization. The dotted lines in each case enclose all solutions that satisfy the corresponding regularization constraint. Due to the shape of each region, solutions in the case of L_1-regularization tend to occur exactly on

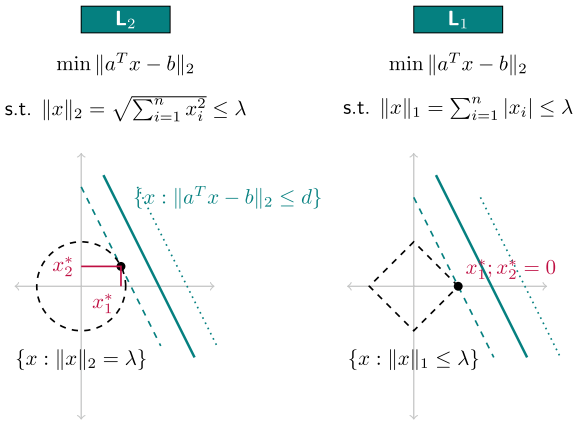


Fig. 9.5 Regularization using the L_2 and L_1 norms. The L_1 norm tends to find solutions with many zero entries. Reproduced from [30] with permission from the Royal Society of Chemistry

one or several coordinate axes, corresponding to zeros in the solution. In contrast, solutions in the case of L_2-regularization seldom occur exactly on coordinate axes.

Our model reduction method uses the single parameter λ to adjust between the accuracy of the reduced chemical reaction network and the number of reactions in the reduced model. This allows us to achieve a high degree of granularity in finding reduced models of different sizes and enables the method to be uniform across different chemical reaction networks, since there are no other system-specific parameters to adjust. In practice, we may add an upper bound on the rate constants $\mathbf{k} \leq 1/\epsilon$, where $\epsilon \leq 1$ is usually on the order of 0.01, to require that the reaction rates in the reduced model are not too much larger than estimated physical rates, although we have found the reduced models to usually be quite similar whether or not we include this upper bound.

9.3.2 Results

In Fig. 9.6, we show the results of applying our model reduction method to the methane system. Each colored line corresponds to Gillespie stochastic simulations of a reduced model, for different values of λ and thus different size chemical reaction networks. The full stochastic model consists of 2000 reactions. Using only about 7.1% of those reactions, we can model the number of CH_4 molecules over time with an error of less than 10%, compared to a baseline fluctuation of about 7% in CH_4 concentrations between independent molecular dynamics simulations. This corresponds to an increase in error of just 3% after removing 93% of reactions from the network. We note however that more reactions are needed to simultaneously model the number of molecules of H_2 and C_2H_6 over time with similar accuracy.

Since L_1-regularization does not require that the k_j's in the solution to be either 0 or 1, the reduced model will allow the k_j's to be somewhat different from that

estimated from the data. However, we would like to ensure that the reduced k_j's are still physically meaningful. It would be unreasonable, for example, if a reaction in the reduced model has a reaction rate that is a significant order of magnitude different from physically reasonable rates. Since we do not actually know the true physical rates for most reactions, we consider instead the estimated rates obtained from our molecular dynamics data. By comparing them with the reduced solution, we find that the structure of our problem is such that most of the reduced k_j's are either 0, or no more than two orders of magnitude different from the estimated k_j's. See Fig. 9.7.

Finally, we compare our L_1-regularization based model reduction method and a more naive method such as simply removing the most infrequent reactions from the system, which we will call count-based reduction. The differences are illustrated in Fig. 9.8. In this example, we use the first 200 ps of a molecular dynamics simulation of the methane system to train a stochastic model consisting of 629 reactions. We then reduce this model using both L_1 regularization and count-based reduction. We use these reduced models to extrapolate the system out to 500 ps and compare their accuracy. At first glance, it seems that the system can be reduced to about only

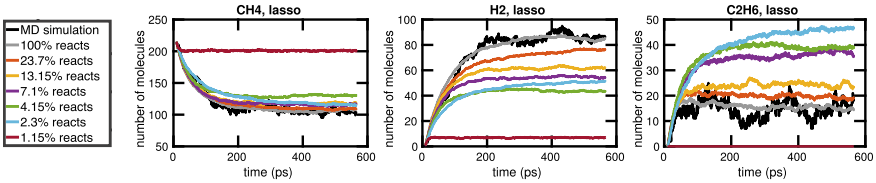


Fig. 9.6 Gillespie stochastic simulations using reduced models determined by L_1-regularization. Reproduced from [30] with permission from the Royal Society of Chemistry

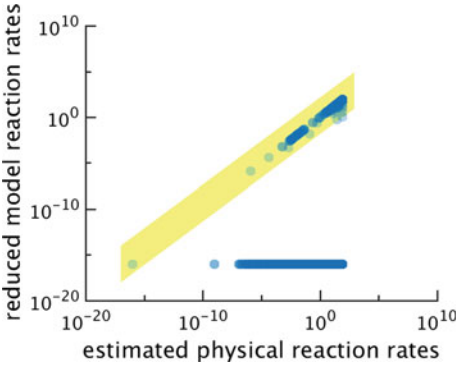


Fig. 9.7 The reduced model reduction rates compared with the estimated physical reaction rates for one choice of λ, here $\lambda = 125$ where 7.1% of reactions are retained. All reaction rates falling within the region highlighted in yellow are no more than two orders of magnitude different from their estimated physical rate in the reduced model determined by L_1-regularization. Note that almost all reduced reaction rates either fall within the yellow region or are approximately numerically zero. This phase transition in the reduced rates is typical over different choices of λ

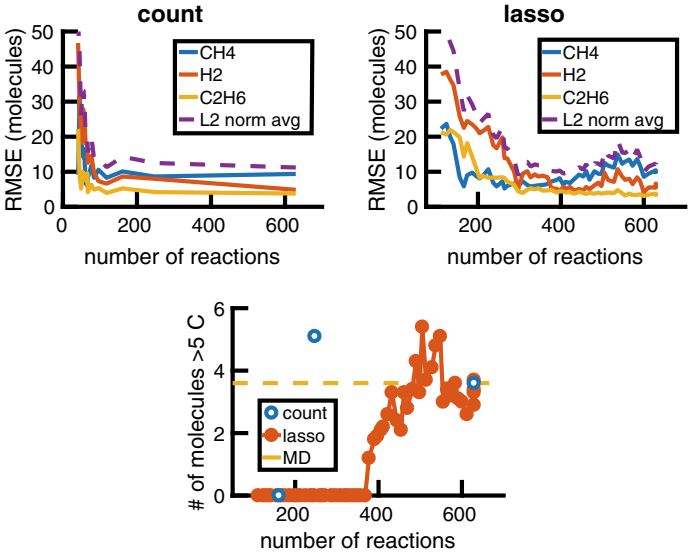


Fig. 9.8 L_1-regularization based model reduction (upper right, lasso) appears to require more reactions to model the largest population molecules for the same amount of error than count-based reduction (upper left, count). However, count-based reduction has in comparison very little granularity in finding reduced models of different sizes and performs poorly on smaller population molecules of importance, such as the growth of carbon clusters in the methane system (bottom). Reproduced from [30] with permission from the Royal Society of Chemistry

100 reactions before incurring significant growth in error for the largest population molecules in the system, CH_4, H_2, and C_2H_6. In contrast, the optimal reduced model found using the L_1-regularization method seems to require approximately 350–450 reactions. However, in our methane system, one property of interest is the growth of carbon clusters in the system. In correspondence with the size of the computational cell and the total simulation time, we characterize any molecules with more than 5 carbons as carbon clusters. Note that the count-based reduction method is completely unable to find a reduced reaction network that can correctly capture the average number of carbon clusters at the end of 500 ps. After the full stochastic model, the next largest reduced model consists of only 247 reactions and overestimates the number of carbon clusters. The L_1-regularization method, on the other hand, can achieve fine granularity in the size of the reduced reaction networks that it finds. In particular, a reduced model with approximately 450 reactions, which is also among the best reduced models for capturing the large population molecules, also captures the average number of carbon clusters after 500 ps.

9.4 Conclusions

In conclusion, we discover that MD simulations contain a lot of useful underlying information that can be exploited using statistical learning to build models of complex chemistry. We use a data-driven method to build stochastic models from slow MD simulations that can be rapidly simulated using KMC. These models have the potential to extrapolate more than an order of magnitude in time, and furthermore can be used to simulate systems in different regions of chemical space. We also develop a new data-driven method for model reduction of chemical reaction networks. We show that by using data to sample the nonlinearities in the system and using L1-regularization to relax the sparsity constraint, we can express model reduction very simply as a quadratic program with simple linear and bound constraints. This allows us to reduce complex chemical reaction networks consisting of thousands of reactions in a matter of minutes.

One can easily imagine a future in which MD simulations used for research on many different chemical systems are routinely archived and analyzed in order to add to and modify an existing repository of elementary chemical reactions and reaction rates. This repository would form a chemical genome that can then be used to quickly simulate all kinds of new chemical systems.

References

1. Androulakis IP (2000) Kinetic mechanism reduction based on an integer programming approach. AIChE J 46(2):361–371
2. Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta 1850(5):872–877
3. Bhattacharjee B, Schwer DA, Barton PI, Green WH (2003) Optimally-reduced kinetic models: reaction elimination in large-scale kinetic mechanisms. Combust Flame 135(3):191–208
4. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J Comput Phys 22:403–434
5. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 93555(1):2340–2361
6. Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys 115(4):1716–1733
7. Hannemann-Tamás R, Gábor A, Szederkényi G, Hangos KM (2013) Model complexity reduction of chemical reaction networks using mixed-integer quadratic programming. Comput Math Appl 65(10):1575–1595
8. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer, New York
9. Higham DJ (2008) Modeling and simulating chemical reactions. SIAM Rev 50(2):347–368
10. Kabbe G, Wehmeyer C, Sebastiani D (2014) A coupled molecular dynamics/kinetic Monte Carlo approach for protonation dynamics in extended systems. J Chem Theory Comput 10(10):4221–4228
11. Mattsson TR, Lane JMD, Cochrane KR, Desjarlais MP, Thompson AP, Pierce F, Grest GS (2010) First-principles and classical molecular dynamics simulation of shocked polymers. Phys Rev B 81:054103

12. Meskine H, Matera S, Scheffler M, Reuter K, Metiu H (2009) Examination of the concept of degree of rate control by first-principles kinetic Monte Carlo simulations. Surf Sci 603(10–12):1724–1730
13. Osborne MR, Presnell B, Turlach BA (2000) On the LASSO and its Dual. J Comput Graph Stat 9(2):319–337
14. Petzold L, Zhu W (1999) Model reduction for chemical kinetics: an optimization approach. AIChE J 45(4):869–886
15. Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117(1):1–19
16. Qi T, Reed EJ (2012) Simulations of shocked methane including self-consistent semiclassical quantum nuclear effects. J Phys Chem A 116:10451–10459
17. Radulescu O, Gorban AN, Zinovyev A, Noel V (2012) Reduction of dynamical biochemical reactions networks in computational biology. Front Genet 3(JUL):1–17
18. Rice BM (2012) Multiscale modeling of energetic materials: easy to say, harder to do. AIP Conf Proc 1426(2012):1241–1246
19. Sikalo N, Hasemann O, Schulz C, Kempf A, Wlokas I (2014) A genetic algorithm-based method for the automatic reduction of reaction mechanisms. Int J Chem Kinet 46(1):41–59
20. Sørenson MR, Voter AF (2000) Temperature-accelerated dynamics for simulation of infrequent events. J Chem Phys 9599:2014
21. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc 58(1):267–288
22. Turanyi T (1990) Reduction of large reaction mechanisms. New J Chem 14(11):795–803
23. Violi A, Sarofim AF, Voth GA (2004) Kinetic Monte Carlo-molecular dynamics approach to model soot inception. Combust Sci Technol 176(5–6):991–1005
24. Voter AF (1997) Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett 78:3908–3911
25. Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57:R13985–R13988
26. Voter AF (2007) Introduction to the kinetic Monte Carlo method. In: Radiation effects in solids, chap 1, pp 1–23. Springer, Dordrecht
27. Wasserman LA (2005) All of Statistics: a concise course in statistical inference. Springer, New York
28. Xu L, Henkelman G (2008) Adaptive kinetic Monte Carlo for first-principles accelerated dynamics. J Chem Phys, 129(11)
29. Yang Q, Sing-Long CA, Reed EJ (2016) L1 regularization-based model reduction of complex chemistry molecular dynamics for statistical learning of kinetic monte carlo models. MRS Adv 1(24):1767–1772
30. Yang Q, Sing-Long CA, Reed EJ (2017) Learning reduced kinetic monte carlo models of complex chemistry from molecular dynamics. Chem Sci 8:5781–5796

Chapter 10
Toward a Predictive Hierarchical Multiscale Modeling Approach for Energetic Materials

Brian C. Barnes, John K. Brennan, Edward F. C. Byrd, Sergei Izvekov, James P. Larentzos and Betsy M. Rice

Abstract This chapter describes efforts to enable multiscale modeling of energetic material response to insult through a concurrent hierarchical multiscale framework. As a demonstration, a quantum-derived, particle-based coarse-grain model of an energetic material is used to provide part of the constitutive response in a finite element multiphysics simulation. Bottom-up coarse-grain models of hexahydro-1,3,5-trinitro-*s*-triazine (RDX) and the methods used to perform reactive simulations at the microscale will be described. Simulations demonstrating microstructure-dependent initiation are also presented. Research opportunities addressing the remaining challenges related to detonation are discussed.

Keywords Hierarchical multiscale modeling · Energetic materials · Dissipative particle dynamics · Bottom-up coarse-grain models · Condensed phase chemistry · Quantum mechanics

10.1 Introduction

A grand challenge for modeling and simulation (M&S) of the properties and response of energetic materials (EM) is virtual assessment of EM performance in munitions, providing a substantial time and monetary savings in the development of materials for next-generation weapons. Current M&S capabilities have numerous shortfalls that do not yet allow accurate, predictive in silico assessment, or even reliable suc-

B. C. Barnes · J. K. Brennan · E. F. C. Byrd · S. Izvekov · J. P. Larentzos · B. M. Rice (✉)
US Army Research Laboratory, Aberdeen, USA
e-mail: betsy.rice.civ@mail.mil

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28,
https://doi.org/10.1007/978-3-030-05600-1_10

cess in virtual design or screening of new EM. These include a lack of micro- and mesoscale modeling capabilities necessary to represent salient physical and chemical features of the materials, deficiencies in multidisciplinary linkages of the relevant scales, and the existence of high levels of empiricism in continuum simulations. Furthermore, the inaccuracy and uncertainty in the descriptions of energetic material response at the microscale and beyond are problematic, especially in the treatment of chemical reactivity. In light of these shortfalls, vigorous research efforts have been pursued by us and others [1–8] to develop simulation capabilities to adequately capture microstructural dependencies on macroscopic events—a simulation challenge for many materials [9–16] and inherent to the initiation of EMs [17].

Our efforts in this area focus on the development of a multiscale M&S framework to predict EM response when subjected to thermal or mechanical insults. While particle-based and continuum level simulation methodologies and models have been extensively studied and advanced [18, 19], the coupling of grain-scale, nonequilibrium microstructural changes between the microscale and macroscale is not commonly realized due in part to methodological challenges [20] in multiscale modeling, as well as computational limitations of mesoscale modeling. As an illustration, consider the case of plate impact on either a single-crystal or formulated energetic material. Classical molecular simulations of one-dimensional (1D) impact under boundary conditions of uniaxial strain typically involve 10^4–10^7 particles (reaching up to micron length scales in one dimension) for durations reaching hundreds of picoseconds. However, experimental setups may be measured in centimeters or larger, in all dimensions, while observations occur over microseconds. Direct molecular simulation of these macroscale systems is simply unfeasible. Current engineering or continuum models used by most in the EM community are parameterized against macroscale experimental data for particular materials, where any atomistic or mesoscale mechanisms affecting material response are implicitly present, as opposed to naturally emerging via models that represent microstructural features as statistical distributions or material history variables. Reaction models used at the continuum level, whether based on matching a specific set of experiments or analysis of atomistic simulations, rarely incorporate explicit dependencies of the reaction chemistry on evolving microstructure, particularly under extreme temperatures and pressures. Furthermore, the role of initial reaction chemistry in changing mechanical constitutive behavior for the interactions of materials at the grain level is not typically utilized in continuum models.

For these reasons, some of our recent efforts have been focused on advancing the nascent microscale models/methods and scale-bridging approaches required for the multiscale M&S framework. Our overall long-term goals have been to develop the appropriate models and approaches that will overcome these gaps, leading to a predictive capability to simulate the coupled thermophysical, chemical and mechanical material processes that affect material response. Within this chapter, we will describe our efforts to properly depict microstructural features in continuum simulations of EM via a multiscale hierarchical approach that bridges higher resolution modeling and the continuum.

We will first describe the higher resolution model and method development at the coarse-grain (CG) level that enables particle-based reactive simulations at the microscale. Particle-based microscale simulation methods utilizing CG models currently offer a promising route for extending atomistic modeling toward the microscale with a significant gain in computational efficiency. CG models, generated by grouping a set of smaller entities (e.g., atoms or molecules) into a single larger entity, are built in a *bottom-up* fashion, such that they incorporate the key underlying physics from the higher resolution scale. During this coarse-graining process, the reduction in molecular degrees-of-freedom (d.o.f.) provides a gain in computational efficiency; however, the loss of information must be adequately recovered through the CG methodology. Furthermore, at the atomistic scale, the formation and breaking of chemical bonds is treated explicitly, while at the microscale, the CG models and methods must collectively capture and recover the relevant chemical features lost during coarse-graining.

We will then discuss our efforts to understand condensed phase chemistry under extreme temperature and pressure, and provide perspectives regarding realistic chemistry that is to be included into a CG model that treats chemical reactivity. This will be followed by an overview of multiscale approaches for continuum simulations that make use of information from particle-based simulations, including our chosen approach, a concurrent hierarchical multiscale simulation method (HMS) that couples continuum and particle-based CG simulations. Finally, we will describe our scale-bridging and algorithmic research efforts within HMS, including the use of machine learning to increase computational efficiency, and provide demonstrations of the HMS approach using our CG models and methods. The chapter will conclude with a discussion of new research opportunities and future directions.

10.2 Coarse-Grain Models

Simulation of the complete range of EM responses, including the competing mechanisms of energy flow, mass flow, and chemical reactivity, requires modeling at length and timescales that are far beyond those amenable to atomistic-scale approaches. Quantum chemical approaches based on ab initio, density functional theory (DFT) or semiempirical calculations can provide detailed information about chemical reactions and transition state structures, but tend to be limited in the number of atoms that can be treated; thus, they are unable to capture the full extent of heterogeneity present in real microstructured materials. Classical reactive potentials that describe the energy landscape and barriers between initial material, the reaction products, and the relevant transition states are available [21], for example, ReaxFF [22, 23] is a commonly employed atomistic force field for modeling EMs. All-atom reactive molecular dynamics (MD) simulations of EMs have been applied to examine the initiation and growth of hot spots created near a single microstructural heterogeneity [24], when the material is subjected to thermal and mechanical loading. However, these simulations of an isolated, nanometer-sized heterogeneity still required

petascale computational resources. As such, the computational expense quickly becomes impractical, requiring years of wall clock time when attempting to simulate realistic samples containing a collection of microscale heterogeneities. These well-recognized limitations of modeling at the atomistic scale provided the motivation for the work described here: the development of microscale models and methods to bridge these spatial and temporal modeling regimes while ensuring multiscale consistency. Until recently, the requisite microscale computational capabilities were either nonexistent or lacking. In this section, we will describe our efforts in developing tools for building CG models, while our efforts in developing the necessary CG methodologies will be described in the following section.

Discrete particle-like descriptions for computationally feasible modeling of EM at the microscale can be obtained through two distinct approaches. One is derived from macroscale data and governing equations (discretization of continuum models), which has demonstrated success in going beyond simple homogeneous systems, but still remains within an essentially macroscopic phenomenological framework, thereby severely limiting its predictive capability [25]. The second approach, which we will discuss in detail, is bottom-up particle-based coarse-graining, in which groups of atoms are mapped into a statistically equivalent ensemble of structure-less CG particles interacting via CG force fields derived from microscopic information [26–29].

In this approach, the atomistic coordinate space is reduced to a smaller space of CG coordinates commonly associated with center-of-mass (CoM) coordinates of molecular clusters representing the CG particles, while the CG dynamics is considered to be Newtonian. Considerable efforts have been expended in the search for CG conservative force fields that lead to a correct representation of equilibrium atomistic statistics and, hence, correct thermodynamics of the CG ensemble [30]. Among the growing efforts to develop conservative force fields for CG models, of particular interest for the work presented here were efforts to generalize CG models for fluids with complicated EOS. These efforts resulted in the development of many-body interaction models [31–33]. Conservative forces in such models can be viewed as effective (mean) forces, where these forces arise due to changes in a many-body potential of mean force (PMF). Approximations to the CG PMF energy surface can be obtained from microscopic data with a number of bottom-up methods such as structure inversion [34], force-matching [35–41], or entropy-matching approaches [42].

One efficient technique for producing well-performing bottom-up approximations to the CG PMF is the multiscale coarse-graining (MS-CG) method [35–38, 43–50]. The MS-CG method has been described elsewhere [36, 40, 41, 51] as a parameter-free force-matching approach that yields an optimal pairwise decomposition of CG conservative forces, and therefore of the associated CG PMF. Consequently, the MS-CG approach is our method of choice for parameterization of the CG models for EM. The MS-CG method can be naturally extended to include a dependence on thermodynamic state variables such as density or temperature, which further enhances the transferability of the CG models [39, 41, 52]. This is important for simulations of shocked EM, where an exceptionally broad spectrum of thermodynamic conditions ranging from ambient to extreme temperatures and pressures may be sampled.

For realistic molecular environments found in condensed-matter systems such as EM, however, the CG conservative models typically fail to yield correct time correlations of the CG dynamical variables leading to accelerated dynamics and affecting the corresponding transport properties. This failure has been attributed to the absence of interparticle friction and thermal noise in the CG description, both of which lead to energy dissipation due to the coupling to the atomistic intraparticle (irrelevant) dynamics with the surroundings. Consequently, bottom-up modeling of both thermodynamic and transport properties require thermodynamically consistent modeling of both the CG conservative (PMF) and nonconservative (dissipative) interactions.

In developing a bottom-up description of the CG nonconservative interactions, we have capitalized on recent advances in formulating the CG dynamics from first principles [51, 53–55]. In these works, the CG equations of motion in the form of generalized Langevin equations (GLE) have been derived from the microscopic Newtonian equations by means of the Mori-Zwanzig formalism. The Mori-Zwanzig formalism leads to the thermodynamically consistent decomposition of the microscopic forces into CG conservative and nonconservative contributions. Within the pairwise and Markovian limits, which are always valid for systems interacting with nonlinear potentials at sufficiently aggressive coarse-graining, the GLE dynamics acquires a classical Galilean invariant form of the Dissipative Particle Dynamics (DPD) [56, 57] equations with the conservative forces precisely prescribed by the MS-CG approach. Initially, the DPD equations were introduced phenomenologically to describe the hydrodynamics of simple liquids using a particle-based approach. The derivation of the DPD equations from first principles in this fashion provides one with a recipe to parameterize the CG models in a fully bottom-up fashion and has led to new systematic approaches to extract the distance-dependent radial and shear dissipative and random forces directly from the atomistic data [51, 58–60]. Our recently proposed multiscale methodology [51, 60] to extract nonconservative interactions from atomistic interactions and dynamics data complements the MS-CG approach to provide a robust framework for bottom-up parameterization of both the conservative and nonconservative forces used in the DPD methodology.

In the following sections, we briefly review the basics of the approaches for first principles parameterization of the conservative and nonconservative force fields for use in the DPD equations of motion and then discuss their applications to two well-studied energetic materials, hexahydro-1,3,5-trinitro-*s*-triazine (RDX, $C_3H_6N_6O_6$) and nitromethane (NM, CH_3NO_2).

10.2.1 Conservative DPD Force Fields

10.2.1.1 The MS-CG Method

The MS-CG method is described [35–37, 40, 47, 51] as a force-matching-based approach for constructing a least-squares optimal pairwise decomposition of a (conservative) force field $\left(F_I^C\right)$ and corresponding many-body CG PMF $\left(W^{\mathrm{PMF}}\right)$. In the

MS-CG approach, the intrinsic many-body force F_I^C is approximated by a pairwise and central force field $F_I^{C,2b}$. The latter force field is determined using a database of the microscopic forces F_I associated with the CG coordinates R_I by minimizing the merit function

$$\chi^2(\alpha) = \left\langle \frac{1}{N}\sum_{I=1}^{N}\left|F_I - F_I^{C,2b}(\alpha)\right|^2 + \left|3VP^{\text{atm}} - 3Nk_BT^{\text{atm}} - \sum_{I=1}^{N} F_I^{C,2b}(\alpha)\cdot R_I\right|^2 \right\rangle \quad (10.1)$$

with respect to spline parameters α used to represent each pair term in the $F_I^{C,2b}$. In the atomistic system, the CG coordinates R_I are located at the CoM of an atomic group mapped into a CG particle. The optional pressure constraint introduced by the second term in (10.1), where T^{atm} and P^{atm} are, respectively, the atomistic temperature and pressure, leads to a CG model suitable for *NPT* simulations [36, 39, 45].

One approach to account for, on average, the many-body interactions within the two-body representation is to use pair terms $f^C(R_{IJ}, \rho)$ in $F_I^{C,2b}(\alpha)$ that are functions of a particle number density distribution, ρ, and where R_{IJ} is the interparticle distance [39, 41, 52]. This density dependency is not merely an abstract construction, but realistically describes the EOS-dependent interactions at the CG scale. As matter of fact, the force-matching expressed by (10.1) yields a different force $f^C(R, \rho, T)$ at each thermodynamic state point (e.g., density ρ and temperature T), when applied to the atomistic system at different state points. In general, both the density ρ and temperature T dependency are needed to ensure a correct EOS for the CG system.

For a homogeneous, single component system, the particle density could be defined based on the global density $\rho \equiv \langle\rho\rangle = N/V$, while alternatively, the notion of a local particle density $\rho(R_I)$ associated with the Ith particle could also be implemented. In the latter approach, the PMF remains a regular function of particle position, and therefore, this approach avoids the issues that otherwise plague global density-dependent interactions [61]. A local particle density definition is more appropriate for systems whose local structure may dramatically vary, such as a microstructured material under thermal or mechanical loading. A consistent approach for defining $\rho(R_I)$ from a discrete set of masses is to use weight functions centered on particles $\rho(R_I) = \sum_{J\neq I}\omega_\rho(R_{IJ})$, where the density weight function $\omega_\rho(R_{IJ})$ has compact support [52]. Neglecting the explicit temperature dependency of $f^C(R_{IJ}, \rho, T)$, an assumption that often is justified, the density dependency can be introduced numerically via interpolation of $f^C(R_{IJ}, \rho)$ to a set of forces $\left\{f^C(R_{IJ}, \rho^s)\right\}$ calculated for a reference system at a set of specific thermodynamic densities $\{\rho^s\}$. Although linear interpolation is considered to be adequate [39, 41] in most DPD applications, it leads to unacceptable energy conservation in the energy-conserving DPD method (DPD-E), described hereafter. A computationally efficient density dependency that results in an exactly integrable force field (and hence better energy conservation) can be written as [52, 62]

$$f^C(R_{IJ}, \rho_I, \rho_J) = f_0^C(R_{IJ}) + \frac{1}{2}\omega_D(R_{IJ})[A(\rho_I) + A(\rho_J)] \quad (10.2)$$

where $f_0^C(R) = f^C\left(R, \rho^0\right)$ is the MS-CG pair term for the density ρ^0 at ambient conditions, $\omega_D(R)$ is a preselected distance weighting function, and $A(\rho)$ is chosen to generate the desired EOS. Integrating the density-dependent MS-CG force field with pair terms given by (10.2) results in the pairwise decomposition of the CG PMF into a sum of potentials $w^{\mathrm{PMF}}(R_{IJ}, \rho)$.

10.2.1.2 Application to RDX and NM: Parameterization of Conservative MS-CG Models

The MS-CG methodology has been applied to derive accurate and transferrable CG conservative force fields for RDX [41, 52, 63] and NM [39, 51, 60]. The mapping of the atomistic RDX and NM structure into a one-bead CG representation is displayed in Fig. 10.1a. The performance of the MS-CG approach for aggressive coarse-graining where clusters of many molecules mapped into a CG particle was explored for liquid NM and proved to be successful [60]. Specifically, a hierarchy of MS-CG models starting with one NM molecule per Voronoi cell (Fig. 10.1a) and up to 64 NM molecules per cell (Fig. 10.1b) was able to accurately describe the structure and density of the atomistic liquid NM under constant *NPT* conditions.

Calculated MS-CG potential terms $w^{\mathrm{PMF}}(R, \rho)$ for different pressures (densities) for the density-dependent MS-CG models of RDX and NM are shown in Fig. 10.2 [39, 41, 52]. The density dependency for the NM model was introduced through linear interpolation of MS-CG potentials for different densities. For RDX, the model with local density dependency using the representation in (10.2) (referred to as the

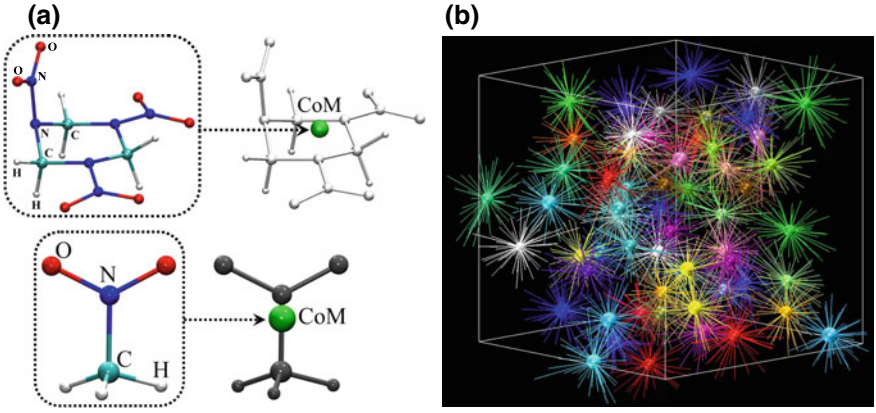


Fig. 10.1 Panel **a**: Mapping of RDX (top) and NM (bottom) into a one-site representation. Panel **b**: Visualization of partitioning of liquid NM (3840 molecules) into centroidal Voronoi cells with 64 molecules per cell. The large balls represent CoMs (centroids) of the Voronoi cells, while the lines shown connect the Voronoi cell centroids to the molecular CoM that are associated with the cell. Colors are used to visually delineate individual clusters (reprinted from Izvekov, S.; Rice, B. M., *J. Chem. Phys.* **2014,** *140* (10), 104104, with the permission of AIP Publishing)

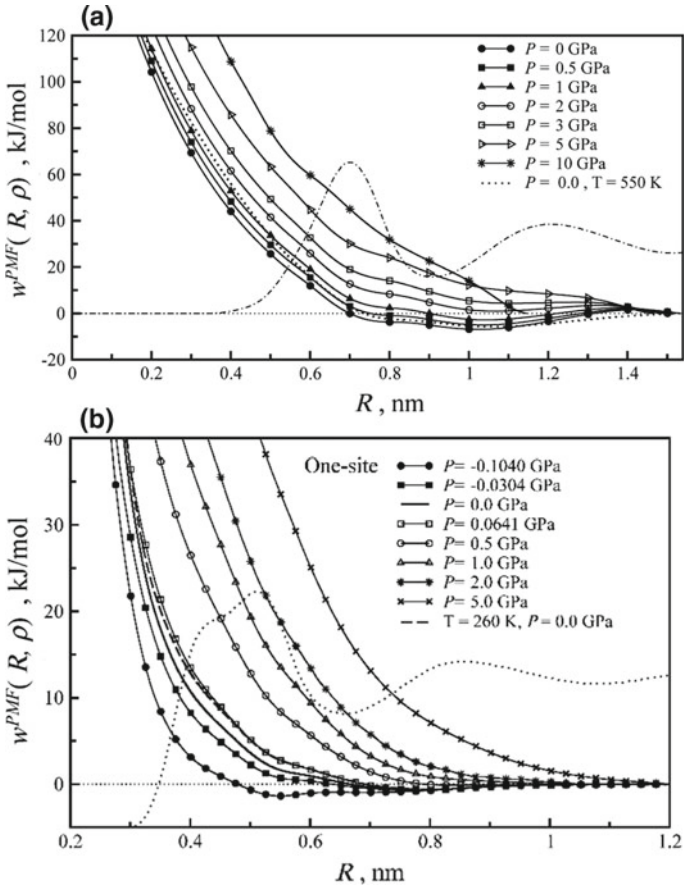


Fig. 10.2 One-site density-dependent MS-CG models of RDX (Panel **a**) and NM (Panel **b**). The scaled CoM RDFs are shown by the dot-dashed and dotted lines (Reprinted from (a) Izvekov, S.; Chung, P. W.; Rice, B. M., *J. Chem. Phys.* **2011,** *135* (4), 044112 and (b) Izvekov, S.; Chung, P. W.; Rice, B. M., *J. Chem. Phys.* **2010,** *133* (6), 064109, with the permission of AIP Publishing.)

MS-CG-D(E) model) was introduced to achieve the energy conservation required for DPD-E simulations of RDX [52].

Due to the systematic representation of microscopic interactions, the MS-CG conservative models describe the equilibrium properties of these explosives reasonably well under different thermodynamic conditions and in different phases. In particular, the MS-CG model of RDX well describes various material properties, including the structure of the ideal RDX crystal [52]. A summary of key thermodynamic and mechanical properties for crystalline RDX from (i) experiment, (ii) atomistic simulations using the reference atomistic force field of Smith and Bharadwaj (the SB model) [64], and (iii) MD simulations of the MS-CG model at ambient pressure is given in Table 10.1.

Table 10.1 Lattice constants a, b, c, density ρ, volumetric thermal expansion coefficient α_V, isothermal bulk modulus B_T, and atmospheric melting temperature T_{melt} of crystalline RDX, from experiment, the atomistic reference and the MS-CG model

Property	Experiment	Atomistic	MS-CG[a]
a (Å)	13.18[b]	13.45[c]	13.37
b (Å)	11.57[b]	11.53[c]	11.58
c (Å)	10.71[b]	10.53[c]	10.59
ρ (kg/m^3)	1806[b]	1805[c]	1801
α_V (10^{-5} 1/K)	19.34[d]	16.38[c], 10.94[e]	8.32
B_T (GPa)	12.1[f], 11.99[g]	13.0[c]	13.17
T_{melt} (K)	478.6[h]	488.5[i]	468.8

[a]Moore et al. [52]
[b]Choi and Prince [65] 300 K
[c]Munday et al. [66] 300 K
[d]Cady [67]
[e]Podeszwza et al. [68]
[f]Olinger et al. [69] 293 K
[g]Haycraft [70] 295 K
[h]Hall [71]
[i]Sellers et al. [72]

The elastic stiffness constants C_{ij} are compared in Table 10.2 to the results for the reference atomistic model, density functional theory (DFT) calculations, and to experiment. The table also reports Cauchy pressures for an MS-CG model crystal, which are measures of the deviation from purely two-body interactions. A similar or better level of agreement is achieved for MS-CG models of NM [39].

Introducing density dependency into the MS-CG potentials leads to significantly improved mechanical properties of the materials under elevated pressures. This is particularly challenging due to the well-known problem of representability, i.e., at a given state point, no single pair potential may exist that can capture all the properties of a given material [61]. The exceptional transferability of the MS-CG models make them suitable for application to systems under thermal and mechanical loading [75]. The level of accuracy and transferability achieved with the MS-CG approach would be difficult to reproduce using the conventional top-down approaches [29, 76, 77].

10.2.2 Nonconservative DPD Force Fields

10.2.2.1 Multiscale Coarse-Graining of Nonconservative Interactions

A characteristic feature of the DPD methodology is the dissipative force that acts between particles, which provides a means of depicting the atomistic model dynamics with a CG model. We have refined the standard dissipative contributions and incorpo-

Table 10.2 Stiffness constants C_{ij} for the MS-CG model of crystalline RDX compared to experimental, ab initio, and empirical atomistic model data

C_{ij} (GPa)	Experiment[d,e]	ab initio[c]	Atomistic[a,b]	MS-CG[a]
C_{11}	36.67[d], 25.02[e]	29.96	30.64[a], 25.00[b]	21.03
C_{12}	1.38, 8.21	7.48	12.68, 10.60	10.06
C_{13}	1.67, 5.81	4.52	7.92, 7.60	11.49
C_{22}	25.67, 19.60	25.51	29.41, 23.80	20.98
C_{23}	9.17, 5.90	5.28	10.64, 8.80	11.50
C_{33}	21.64, 17.93	23.61	30.23, 23.40	23.53
C_{44}	11.99, 5.17	5.34	4.16, 3.10	6.11
C_{55}	2.72, 4.07	4.83	6.54, 5.20	6.10
C_{66}	7.68, 6.91	8.59	10.03, 7.70	5.23
C_{13}–C_{14}	N/A	N/A	N/A	5.38
C_{12}–C_{66}	N/A	N/A	N/A	4.83

[a]Moore et al. [52] 0 K
[b]Munday et al. [66] 300 K
[c]Taylor [73] 0 K, DFT-D3
[d]Haycraft [70] 295 K
[e]Haussuhl [74] 293 K

rated them into the DPD-RX framework, including multidirectional dissipative interactions [78, 79] that are both parallel and perpendicular to the interparticle separation axis. Directional dissipative force contributions attempt to capture the CG d.o.f. that contribute to the molecular shape or polarity. The multiscale bottom-up approach to derive the radial (parallel) $\left[\gamma^{||}(R)\right]$ and shear (perpendicular) $\left[\gamma^{\perp}(R)\right]$ friction functions dependent on the interparticle separation R is proposed and described in detail elsewhere [60]. The friction functions $\gamma^{||}(R)$ and $\gamma^{\perp}(R)$ describe the amplitudes of the nonconservative force components that are parallel and perpendicular to the interparticle separation, respectively. The approach exploits the statistical independence of the random forces and the initial particle velocities, which is a generic property of the GLE dynamics, and hence the DPD method. This property allows a unique relationship between the friction functions, and both the three-body velocity–velocity correlation functions $c_{VV}^{\alpha\beta}$, $\alpha, \beta \in \{||, \perp\}$ and the two-body correlation functions $\left[c_{\Delta FV}^{\alpha}\right]$ of residual force, $\Delta F_I = F_I - F_I^C$, with velocity. By introducing the mesh $\{R_l, l = 1, \ldots, N_{\text{bin}}\}$ of interparticle distances and considering the correlation functions at a sufficiently large moment of time T_M, the orthogonality relation leads to the following linear system of equations

$$\sum_{\tilde{l}=1}^{N_{\text{bin}}} \left[c_{VV}^{\alpha||}(T_M, R_l, R_{\tilde{l}})\gamma^{||}(R_{\tilde{l}}) + c_{VV}^{\alpha\perp}(T_M, R_l, R_{\tilde{l}})\gamma^{\perp}(R_{\tilde{l}})\right] = -c_{\Delta FV}^{\alpha}(T_M, R_l), \alpha = ||, \perp \quad (10.3)$$

Equation (10.3) allows the friction functions $\gamma^{||}(R)$, $\gamma^{\perp}(R)$ to be uniquely determined on the distance mesh. The friction functions are thermodynamically consistent with the MS-CG conservative force used to compute the correlation functions $c^{\alpha}_{\Delta FV}$. Following the standard DPD formalism, the random forces for both the parallel and perpendicular contributions are modeled a priori with fluctuation–dissipation theorem-compliant white noise (uncorrelated in time). The appropriate fluctuation–dissipation relations for both the parallel and perpendicular contributions can be found elsewhere [75].

10.2.2.2 Application to RDX and NM: Parameterization of Nonconservative Forces

The accurate description of transport phenomena, such as diffusion and viscosity, with the DPD approach, requires the parameterization of radial, $\gamma^{||}(R)$, and shear, $\gamma^{\perp}(R)$, friction functions. The application of the multiscale methodology presented in Sect. 10.2.2.1 for RDX [63] and NM [51, 60] leads to accurate $\gamma^{||}(R)$ and $\gamma^{\perp}(R)$ friction functions that are thermodynamically consistent with the conservative interactions (Fig. 10.3). The radial friction function $\gamma^{||}(R)$ for NM was found to exhibit a typical form at fine coarse-graining [51, 60]. In contrast to this and to bottom-up nonconservative DPD forces of real molecular liquids discussed in the literature, the dissipative dynamics of RDX is dominated by the shear friction $\gamma^{\perp}(R)$, and hence cannot be accurately simulated using standard DPD, which accounts for only the radial friction $\gamma^{||}(R)$. The analysis of DPD simulations using MS-CG forces for RDX highlights the importance of the perpendicular friction on the short-time dynamics and transport properties. Evidently, the way in which the dissipation in molten RDX is partitioned (with the shear component being dominant) is a result of concerted translational and intramolecular conformational dynamics of the RDX molecules. Therefore, the RDX dissipation dynamics is distinctly different from those observed in liquid NM and other molecular liquids [41, 51], which are well described by nonconservative DPD forces with dominant radial friction. We note in the DPD methods described below, both shear and radial friction terms have been included.

10.2.3 Outlook

Our future efforts are directed toward formulating novel extensions of the MS-CG method to create a hierarchy of high-fidelity bottom-up CG models for explosive formulations. We intend to create CG particle interactions (potentials) that are dependent upon other aspects of the particle's local environment beyond local density, such as the three-body local arrangement or more complex order parameters. We also intend to incorporate a wide range of multibody interactions as well as dipolar and higher order electronic polarizability interactions that are pertinent to EM composites. For

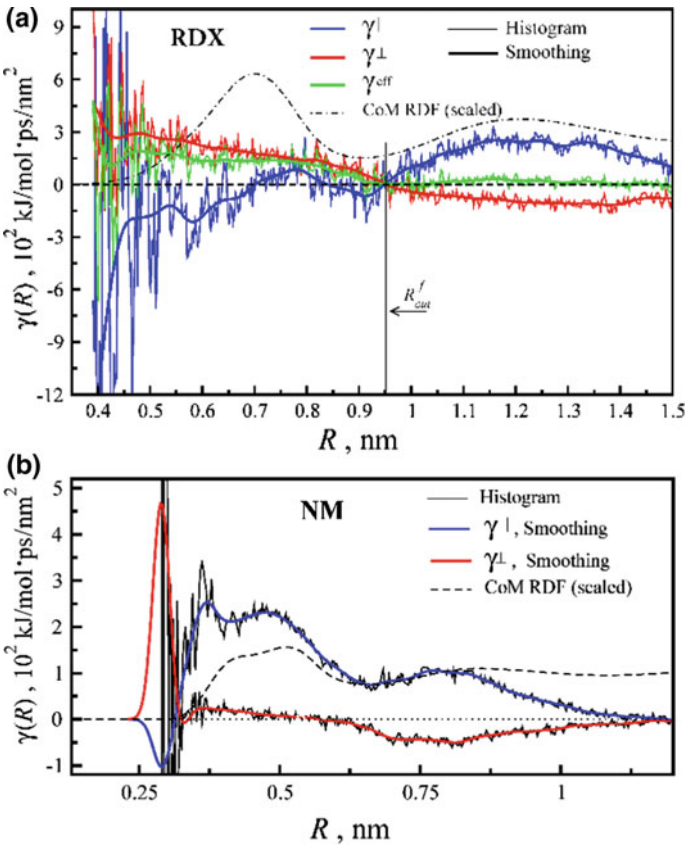


Fig. 10.3 Histogram data (thin noisy) and Gaussian filter smoothing [51, 63] (thick) of friction functions $\gamma^{||}(R)$, $\gamma^{\perp}(R)$ for the one-site MS-CG models of RDX (Panel **a**) and NM (Panel **b**) obtained by solving (10.3). The scaled CoM radial distribution functions (RDFs) are shown by dot-dashed lines. For the MS-CG-D(E) model for RDX, the vertical line and arrow show the choice of cutoff distance $\left(R^{f}_{\text{cut}}\right)$ for the frictions, and $\gamma^{\text{eff}}(R)$ is the effective friction [34]. (From Izvekov, S.; Rice, B. M., *Phys. Chem. Chem. Phys.* **2015**, *17* (16), 10795–804.)

example, we believe that noncentral three-body interactions might play an important role in the microstructure evolution of EMs under shock, particularly in the molecular level plastic response, while the dipolar and higher order electronic polarizability interactions have longer range effects that will influence the structural response at longer scales. In order to have thermodynamically consistent capabilities to model the microscale, however, coarse-grained models cannot be generated separately from the development of the methods that use them, since features lost during the required process of coarse-graining will not be adequately recovered unless reintroduced in some fashion, namely the simulation method. Thus, we now turn our attention to the DPD-based methodologies used to simulate solid EM under a range of threats.

10.3 Coarse-Grain Methods

Computationally reasonable particle-based simulations of material behavior governed by micro- and mesoscale structural heterogeneities require the development of CG models such as those just described. However, careful consideration must be taken for the choice of not only a CG *model*, but as critically, for the choice of an appropriate CG *methodology*, especially since the use of their results in higher scale simulations will strongly influence the outcome. To determine properties other than static (equilibrium) properties, the CG-molecular dynamics (CG-MD) approach is not adequate due to the well-known speedup of the dynamics of CG models compared to their atomistic model counterparts, which is a direct consequence of losing d.o.f. during coarse-graining [80, 81]. Moreover, at the atomistic scale, the formation and breaking of chemical bonds is treated explicitly and is conceptually intuitive, but at the microscale, the CG models and methods must collectively capture and recover the relevant physics and chemistry lost during coarse-graining.

The development of CG modeling methods is an active field, where most of the attention has been given to treating the static behavior of *soft matter* (see e.g., [82, 83]). These systems are amenable to coarse-graining because some atomic motions (e.g., hydrogen vibrations) contribute minimally to the backbone behavior and resulting equilibrated microstructure. Further, the weakly repulsive interaction potentials for such systems allow on the order of 10^{-13} s time steps to be used, providing access to phenomena that occur on microsecond timescales. However, only a relatively limited number of CG model studies have attempted to reproduce nonequilibrium behavior [84–92], in part due to the challenge of mapping the CG and atomistic model dynamics. Thus, we have made significant investments in CG methods development to describe nonequilibrium behavior.

An overview of our efforts described here has entailed creating computational capabilities targeting the microscale that allow for physically realistic simulations of the thermomechanical response of EM composites with microstructure. Particle-based microscale simulation methods utilizing CG models currently offer a promising route into and beyond the microscale, the critical time and spatial regime that cannot be accessed with atomistic MD. We have chosen microscale methods that are built upon the DPD method [56, 57], a technique initially developed for the simulation of soft matter, but recently applied to condensed phase matter [75, 92]. Our microscale approach, which incorporates the salient physics, couples CG models with reactive microscale methods, where both are described within this chapter. These efforts are designed to provide predictive capabilities that are amenable to direct coupling with continuum level models in a multiscale modeling framework, or to provide crucial information for development of higher fidelity continuum material models. The direct coupling efforts are described in Sect. 10.5.

10.3.1 CG Method: Variants of DPD

10.3.1.1 General Description

The DPD method is now a well-established CG particle simulation method that has evolved substantially since its inception in 1992 [56, 57]. Advances in both method and model development now allow DPD simulation of a wide range of material classes from soft matter, such as polymers and biomolecules, to condensed matter, such as metals and crystals [52, 89, 93]. DPD is well grounded in statistical mechanics and stochastic dynamics, allowing for a physics-based interpretation of the parameters and their determination from higher resolution models. Advances in the method continually arise from work in various material communities, further extending its potential applicability and utility.

The original formulation of the DPD method conserves total momentum only and thus is limited to modeling isothermal processes. For the purpose of simulating EM composite response, the energy-conserving DPD method (DPD-E) [92, 94–96] is particularly critical since it enables nonequilibrium simulation scenarios and thermally variant conditions. DPD-E uniquely treats the CG d.o.f. through both the dissipative forces and a particle internal energy term. The particle internal energy term plays two roles within the DPD-E method. First, it provides a numerical means of ensuring energy conservation during the simulation. Moreover, the particle internal energy term provides an additional mechanism to recover the coarse-grain d.o.f., which is essential for accurately reproducing the atomistic model behavior [75].

Building upon the ideas of Maillet and co-workers [86, 91], we recently developed a general DPD framework that incorporates chemical reactivity (DPD-RX) [75]. Originally constructed for DPD-E, the DPD-RX method can be formulated upon either this variant or the constant-enthalpy DPD variant (DPD-H) [92], notated as DPD-RX-E and DPD-RX-H, respectively. In either variant, a reaction progress variable is assigned to each particle that monitors the time evolution of an *extent-of-reaction* associated with each of the prescribed reactions that may occur within each particle. As such, the DPD-RX approach does not necessitate a reactive potential that involves explicit bond breaking and bond forming. The chemical reactivity can be modeled using complex or reduced reaction mechanisms and allows for both unimolecular and multimolecular collision reactions to be simulated via both direct and indirect approaches. Aside from including the extent-of-reaction and introducing an additional term to the particle internal energy (u_{chem}), the DPD-E and DPD-H formalisms do not change. In practice, for every time step, DPD-RX dynamics are separated into three elementary, physical processes: (1) *inert dynamics*—execution of the DPD-E (or DPD-H) equations of motion; (2) *CG-reactor chemistry*—extent-of-reaction change within each CG particle based upon the prescribed chemistry; and (3) *reaction energy update*—partitioning of the chemical energy release or gain by updating the CG particle internal energy, during which the total energy of the given CG particle does not change. As the chemistry of a CG particle changes, so

does its interaction potential. The interaction potential changes in such a way that it captures both heat exchange and pressure–volume work due to chemical reactivity.

10.3.1.2 Equations of Motion

In the DPD-E method, at any time t, a particle I is specified by its mass m_I, position $\mathbf{r}_I$, momentum $\mathbf{p}_I$, and internal energy u_I. The particle internal energy accounts for the energy absorbed or released by the d.o.f of the higher resolution model that are unresolved as a result of coarse-graining; u_I is coupled to the internal temperature θ_I through a CG equation of state (CG-EOS) typically defined as, $u_I = u_I(\theta_I)$. The variation of the internal energy $\mathrm{d}u_I$ can be written as the sum of contributions that correspond to the mechanical work done on the system, $\mathrm{d}u_I^{\mathrm{mech}}$, and the heat conduction between particles, $\mathrm{d}u_I^{\mathrm{cond}}$, i.e., $\mathrm{d}u_I = \mathrm{d}u_I^{\mathrm{mech}} + \mathrm{d}u_I^{\mathrm{cond}}$. In DPD-E, two types of temperatures are defined, a kinetic temperature, T_{kin}, which is associated with the external d.o.f., and the internal temperature, θ_I, which is associated with the internal d.o.f. At equilibrium conditions, these two temperatures will be statistically equivalent, but not necessarily under nonequilibrium conditions.

Extending the DPD-E method to the reactive case requires the variance of the particle internal energy $\mathrm{d}u_I$ to include an additional contribution that corresponds to the energy associated with changes in chemistry, $\mathrm{d}u_I^{\mathrm{chem}}$. The total variance is given by $\mathrm{d}u_I = \mathrm{d}u_I^{\mathrm{mech}} + \mathrm{d}u_I^{\mathrm{cond}} + \mathrm{d}u_I^{\mathrm{chem}}$, where the total energy of the system is assumed to remain constant during changes in $\mathrm{d}u_I^{\mathrm{chem}}$. The set of equations of motion for DPD-E is accompanied by an update of u_I^{chem} specified as

$$\mathrm{d}u_I^{\mathrm{chem}} = -\mathrm{d}u_I^{\mathrm{CG}} \quad (I = 1, \ldots, N) \tag{10.4}$$

where u_I^{CG} is the CG interaction energy of particle I. Equation (10.4) follows from the requirement of total system energy conservation, such that the total energy of each particle is assumed to be conserved upon any concentration change due to the reaction. Analogous to the nonreactive DPD-E approaches, conservation of the total momentum and the total system energy, $E = U^{\mathrm{CG}} + K + \sum_I u_I$, is satisfied, where $K = \sum_I \frac{\mathbf{p}_I \cdot \mathbf{p}_I}{2m_I}$ is the total kinetic energy and U^{CG} is the total CG interaction energy. For the DPD-RX approach presented here, any change in the CG particle chemistry is also reflected through a species-dependent CG-EOS, $u_I = u_I\left(\theta_I, N_\xi\right)$, where N_ξ is the number of molecular species ξ in the product gas mixture. The choice of $u_I\left(\theta_I, N_\xi\right)$ is a modeling decision, where a possible formulation is to consider the isolated molecule contributions based upon the molar heat of formation, $\Delta H_{f,\xi}(T_r)$, and the isobaric heat capacity of each molecular species $C^0_{P,\xi}(\theta_I)$, taken at some reference state [75]. The development of alternative formulations is an ongoing pursuit by our group.

10.3.1.3 Reaction Model

The DPD-RX framework presented here builds upon the treatment of the CG particles seemingly as interacting continuous-stirred tank reactors [97] or *CG-reactors*. The CG-reactor depicts temporal changes in the species type and concentration of the molecules representing the CG particle, where these changes in chemistry are governed by the prescribed set of reaction mechanisms and kinetics, termed the *reaction model*. For further illustration of the DPD-RX methodology, consider a specific reaction model, namely, the thermal decomposition of crystalline RDX into a mixture of product gases. From global reaction rate models of high-temperature combustion of nitramines, a reduced reaction mechanism was determined, where the resulting RDX decomposition pathway is a four-step reaction rate model [75] consisting of a unimolecular, irreversible reaction:

$$\mathrm{RDX} \rightarrow 3\mathrm{HCN} + \frac{3}{2}(\mathrm{NO_2} + \mathrm{NO} + \mathrm{H_2O}) \tag{10.5}$$

and three bimolecular, irreversible reactions:

$$\begin{aligned} \mathrm{HCN} + \mathrm{NO_2} &\rightarrow \mathrm{NO} + \frac{1}{2}(\mathrm{N_2} + \mathrm{H_2}) + \mathrm{CO} \\ \mathrm{HCN} + \mathrm{NO} &\rightarrow \mathrm{CO} + \mathrm{N_2} + \frac{1}{2}\mathrm{H_2} \\ \mathrm{NO} + \mathrm{CO} &\rightarrow \frac{1}{2}\mathrm{N_2} + \mathrm{CO_2} \end{aligned} \tag{10.6}$$

Each reaction rate constant is modeled by a standard temperature-dependent Arrhenius expression, where the temperature used is a local weight-average internal temperature of CG particle I, defined as

$$\bar{\theta}_I^{-1} = \frac{\sum_{J=1} \omega_{\mathrm{Lucy}}(r_{IJ})\, \theta_J^{-1}}{\sum_{J=1} \omega_{\mathrm{Lucy}}(r_{IJ})} \tag{10.7}$$

where J includes the neighboring particles of I and itself, and $\omega_{\mathrm{Lucy}}(r_{IJ})$ is the Lucy function [98]. A local-average particle internal temperature used in the reaction rate expressions implicitly depicts multimolecular conditions and the local environment of a chemically reacting mixture. Nevertheless, the forms of $\overline{\theta_I}$ and ω_{Lucy} are modeling choices, where alternative forms are possible. Further discussion of the choice of the reaction model specifically for EM simulations is given below in Sect. 10.4.

10.3.1.4 Particle Model

In the DPD-RX method, the particle interaction potential is not a reactive type potential that mimics chemistry through bond breaking and subsequent formation of tran-

sition structures and reaction products (e.g., [23, 99]). Rather, the evolving chemistry modeled by the CG-reactor depiction is directly coupled to the interaction potential of the CG particle. In the application presented here, the chemical character of any particle can vary between two end states, unreacted RDX and a final product gas mixture, with many continuous chemical states in between. An RDX molecule is represented by the isotropic one-site CG model (CG-RDX) [52] obtained by the MS-CG method described above, while the product gas mixture is modeled using the exponential-six interaction potential [100]. A notable feature of this model is that the expansion of hot product gases is inherently captured through the chemistry-dependent particle interactions via changes in the particle excluded volumes.

10.3.1.5 Practical Aspects

Beyond attempting to reproduce the correct physics, the development and implementation of the CG framework requires several practical considerations. A plethora of numerical integration schemes have been extensively explored (e.g., [85, 92, 101–106]), allowing for stable, accurate, and efficient simulations. Recently our group has adopted an efficient integration scheme for the DPD variants based on the Shardlow-splitting algorithm (SSA) [92, 104, 107]. Compared to the traditional DPD integrators, the SSA allows for larger time steps, on the order of 10^3, making simulations of CG models of EM composites possible for the DPD-E and DPD-H variants. Advancements to the DPD methods and algorithms are continually incorporated within the highly-scalable open-source LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) simulation package [108] to provide a long-term, sustainable modeling framework that can readily leverage high-performance computing resources. Such practical considerations are critical for enabling simulations with 10^7–10^9 particles at the relevant time and length scales for EM composite models.

10.3.2 Sample Applications

The DPD variants can be applied to gain a fundamental understanding of the energetic material response to shock. Consider the shock profiles of the CG-RDX model with those of a nonreactive atomistic model [64] at conditions under which reactions are not expected to occur (see Fig. 10.4). Clearly, the MD simulation of the CG-RDX model significantly overpredicts the thermal response, due to improperly accounting for the energy and momentum exchange. This leads to a kinetic temperature increase that is significantly higher than the atomistic model temperature at and behind the shock front. In the DPD-E simulation, the transfer of mechanical energy from the plate impact into the CG d.o.f. has appropriately occurred via the heat and momentum exchange between the particle internal energies. In the CG-MD simulation, the particles behave effectively as hard spheres, while the DPD-E simulation allows for

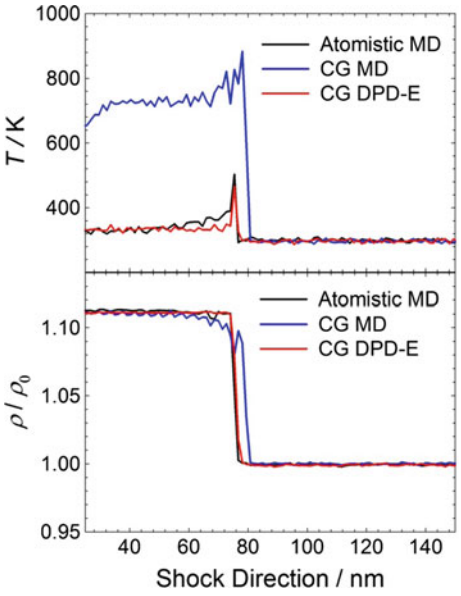


Fig. 10.4 Comparison of the nonequilibrium, nonreactive response of pure crystalline RDX: (atomistic MD) fully atomistic model using MD; (CG-MD) one-site CG-RDX model using MD; and (CG DPD-E) due to mechanical shock generated by plate impact at $u_p = 0.5\,\mathrm{km/s}$. Kinetic temperature and density profiles are snapshots taken 15.0 ps after impact (Adapted with permission from Brennan et al., *J. Phys. Chem. Lett.*, **2014**, *5* (12), pp 2144–2149. Copyright 2014 American Chemical Society.)

some of the momentum exchange to be absorbed into the particle internal energies via the CG-EOS.

Next, consider applications of the DPD-RX-E method to examine the effects of microstructural heterogeneities on the material response to mechanical shock. Variations in the local geometry and density may lead to significantly different hotspot formation, which dictates the macroscopic material response. In the examples below, several microstructures are generated to examine the role that defects and complex grain boundaries play in the material response. In the first scenario, 10-nm spherical inclusions are randomly distributed throughout the 2.5 μm sample. The sample is slammed into a reflective wall in the −z-direction, generating a shock compression wave that propagates through the sample in the +z-direction. As the shock front passes through the material, the spherical inclusions collapse, transferring mechanical energy to heat the particles and initiating chemical reactions. The particle internal temperature and reaction progress (defined as the fraction of RDX that is present in the particle) is shown in Fig. 10.5, where after approximately 0.5 ns, the shock front reaches the end of the 2.5 μm sample. In cases with sufficiently large shock velocities and/or defect sizes, the spherical inclusions are observed to combine and coalesce into larger reaction zones.

To examine the effects of grain boundaries on material response, a polycrystalline RDX microstructure geometry is constructed using a Voronoi tessellation method, where the polyhedra are treated as individual grains (Fig. 10.6a). The particle-based model of each grain is subsequently grown as a perfect crystal that is free of any intragranular defects. The resulting structure is a large polycrystalline network of randomly oriented HCP-like ordered grains with complex, planar interfaces and inter-

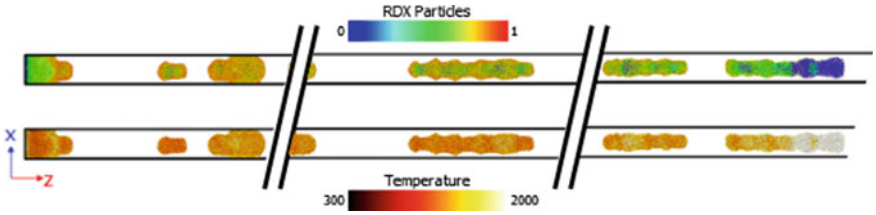


Fig. 10.5 Snapshots along the sample for the reaction progress (top) and particle internal temperature (bottom) at various positions along the z-axis of a 2.5 μm shocked RDX sample. The initial, unshocked sample included a random distribution of 10-nm spherical inclusions that collapse, creating localized hot spots that initiate chemistry. The unreacted material in the sample is not shown for visual clarity and to depict the surface area of the reaction zones

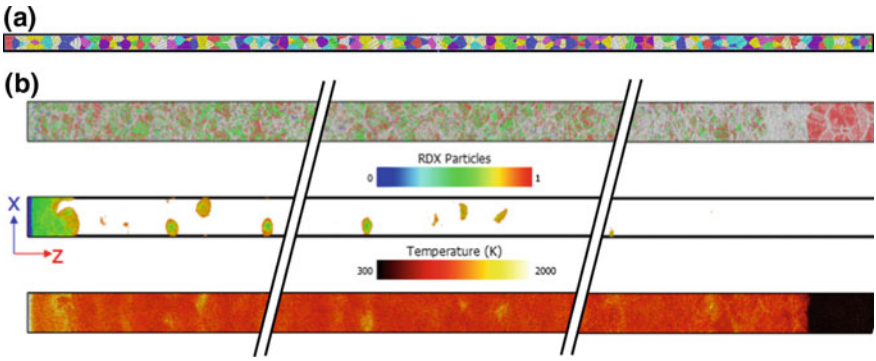


Fig. 10.6 **a** The initial, unshocked polycrystalline RDX sample. **b** Snapshots along the sample for the common neighbor analysis (top), reaction progress (middle), and internal temperature (bottom) at various positions along the z-axis of the 2.5 μm sample. For the reaction progress snapshots, the unreacted material in the sample is not shown for visual clarity and to depict the surface area of the reaction zones. The common neighbor analysis uses the following coloring scheme: HCP = pink, FCC = Green, BCC = Blue, ICO = Yellow, and White = Other

granular voids. As the material undergoes shock compression, a common neighbor analysis is performed to highlight the initial, unshocked HCP-like grains and the resulting microstructure that evolves due to shock (see Fig. 10.6b). These particle-based methods are able to capture the microstructural heterogeneities and short-time physics that evolve immediately after the shock wave passes through the sample, enabling the exploration and understanding of the fundamental mechanisms that influence material performance. In this particular microstructure, energy localizes at the inter-granular voids, creating localized hot spots where chemical reactions begin to evolve. The behavior of the void collapse differs from the previous microstructure due to the varying volume sizes, shapes, and locations.

10.3.3 Outlook

For the goal of understanding EM composite performance, we developed a CG method framework that allows us to begin exploring the role of explicit microstructural heterogeneities in these materials. While numerous atomic-scale reactive potentials are available within the literature (e.g., [23, 99, 109]) that also allow the exploration of the effects of microstructure, they are limited to studies of single, isolated defects due to computational cost [110]. CG approaches such as those presented here allow exploration into previously unattainable temporal scales. Given that the time step in a particle simulation is dictated by the highest frequency modes, which are coarsened away in a CG model, time steps that are thousands of times larger may be taken [92, 104]. Application of CG models and methods that treat chemical reactivity extend the length and timescales well beyond those currently realizable in atomistic simulations, enabling exploration of microstructure-dependent material systems that was not previously possible.

While the DPD-RX framework enables simulations at vastly different length and timescales as compared to atomistic-scale simulations, the computational speedup comes at the cost of losing atomistic detail. By design, the CG model behavior cannot replicate the total fidelity of the atomistic model behavior. Rather, the intent is for the CG model simulation to produce results of sufficient accuracy across a broad range of material properties and behavior for the problem of interest, but for considerably less computational expense. In other words, efforts are made to strike a balance between maintaining the d.o.f. that reproduce the key atomistic model behavior, while decreasing the overall computational cost. To this end, we continue to optimize the recovery of the salient d.o.f. through improvements and refinements of the CG model development and the DPD-RX framework. Several possible refinements and future research directions for the DPD-RX framework are highlighted next.

The particle interaction model for the product gas mixture (pgm) is a first-generation model, where an assortment of modifications for extending the transferability of the model to describe a wider range of chemical states typically present under the extreme conditions of shock and thermal loading is needed. In determining the interaction potential for the pgm model, significant effort is required to adequately sample the reacting environment in a statistically reliable manner, which is dependent upon both the density and species concentrations. Simulation data from finer-scale models, as well as an analytical EOS for the exponential-6 potential may reduce some of the burdens. Furthermore, to minimize the limitations of the single-site exponential-6 interaction potential currently being used for the pgm model, the development of a density-dependent model is under consideration.

Some further refinements that would make the method framework more generally applicable are possible. For example, the DPD-RX framework can be extended to permit intra- as well as interparticle reactions, i.e., reactions influenced by the composition and temperature of either the particles themselves solely, or also the surrounding particles, respectively. The intraparticle approach implicitly accounts for mass diffusion of the reacting species, which under certain circumstances (e.g.,

short timescales) may be sufficient to accurately represent the salient behavior, but in other scenarios, mass diffusion may play a key role. The interparticle approach more directly mimics multimolecule reactions, where explicit treatment of mass diffusion can occur via species transfer between CG particles. Further work is underway to extend the DPD-RX framework to allow other variations of reactions, including non-bond breaking reactions such as molecular conformational transitions [111].

Finally, specific to understanding the reactive behavior of EMs under extreme conditions, improvements to the reaction model itself are needed. Currently, the four-step reduced kinetics model for RDX decomposition described above is used, where this reaction model exhibits density-dependence consistent with the condensed phase. Moreover, the DPD-RX framework ensures that the chemical energy content from the reaction model, starting with pure RDX and ending with product gases, is thermodynamically consistent through the use of standard state data taken from either thermochemical data tables or ab initio calculations. Simulation studies using the current reaction model are expected to provide reliable qualitative trends, as a means of mapping the relative roles of various types of microstructure heterogeneity in EM composites. Nonetheless, much is unknown about the condensed phase chemistry of EM, so uncertainty remains regarding the accuracy of our particular reduced reaction model. Hence, more investigation and development is needed to provide the most accurate depiction of the condensed phase chemistry. The DPD-RX framework is sufficiently general such that it allows any (practical) number and type of reactions and the associated species to be implemented, including pressure-dependent reaction models. Therefore, if new insight gained from higher resolution simulation techniques, such as those described in the next section, or novel experiments lead to the development of more accurate reaction models, then these models can be readily implemented.

10.4 Condensed Phase Chemistry Under Extreme Conditions

10.4.1 Development of Reaction Models for DPD-RX

A description of the chemical reactivity is needed to complete our approach to model the microscale response. As indicated above, this description must be amenable for use at the CG scale within the DPD-RX methodology. In this CG representation, the description of the interparticle interactions is dependent on the CG material state whether unreacted, partially reacted, or fully reacted. As described earlier, the current DPD-RX methodology assumes each CG particle behaves as a well-stirred reactor [75], whose chemical composition is governed during the simulation by a corresponding extent-of-reaction variable for each reaction in a prescribed chemical kinetics mechanism. In its current form, the DPD-RX method requires a description of the chemical composition of the CG-reactor during the decomposition of the

solid EM to gaseous products, which in turn is dependent on the chemical kinetics mechanism.

Derivations of full chemical kinetics mechanisms based on elementary reactions [112] and subsequent reduction to computationally tractable reduced-order chemical kinetics models (ROM) are conceptually straightforward and have been very successful when used in numerous applications including combustion modeling [113]. The DPD-RX methodology demonstrated above for shock loading of the CG-RDX model uses such a reduced four-step chemical kinetics model [75], which was based on gas-phase reaction mechanisms for nitramine combustion [114] combined with a single-step rate of RDX decomposition above the melting temperature [115]. However, while this chemical kinetics model qualitatively exhibits the expected behavior, it may not properly represent the actual condensed phase reactivity in extreme nonequilibrium conditions, such as those associated with high-pressure, high-temperature shock states (>10 GPa and >1000 K). The material environment in these highly dense conditions (>2.0 g/cm^3) allows for different chemistries that would not be accessible at less extreme conditions. Thus, because a reaction model is required in the DPD-RX framework, it is necessary to obtain an accurate microscale description of the extent of chemical reaction of an energetic material subjected to insult, whether a detonation or a sub-detonative response (often even more challenging to model) is the final result. Some of the challenges that exist for transitioning the condensed phase chemistry behavior into the DPD-RX framework will be considered at the end of this section. However, prior to that, we will describe the challenges in first determining just the reaction mechanism for chemistry occurring under these extreme conditions, where consideration for determining the associated kinetics are left for discussion elsewhere.

10.4.2 Scope of the Problem and Challenges

As chemical reactions are inherently atomistic processes, the proper simulation methodology to obtain rate information is through atomistic approaches, preferably using accurate quantum mechanical (QM) methods. QM methods applied to the systems at hand, are typically limited to simulations of a few picoseconds using <10^5 atoms. Numerous classical and quantum-mechanically based MD simulations of thermally and mechanically initiated energetic materials have been published [116–153], some of which attempt to describe reaction rates and mechanisms of the material in extreme conditions. However, as nicely detailed in a review by Manaa and Fried [117], determination of accurate rate information from atomistic simulations of energetic materials in highly nonequilibrium high-density states is a daunting undertaking. The most obvious challenges pertain to the complexity of the event, spawning several questions that must be considered: Is it possible to unravel the complex dependence of concurrent reactive processes occurring in a heterogeneously dense environment under extremely nonequilibrium conditions into a series of individual reaction steps? Does the simulation adequately sample the phase space for a suf-

ficiently long time to observe all important reactions? Is all relevant phase space represented in the simulation? Is the level of theory adequate to properly describe the material state under these extreme conditions? Do computational tricks imposed for simulation efficiency [18] introduce artifacts that might influence outcomes?

Provided these questions can be adequately addressed, there are additional concerns as to whether chemical species can be properly identified for monitoring throughout a simulation, the most direct manner in which to determine the extent of reaction. Using empirical classical reactive models, assignment of bonds can be readily accomplished, thus allowing for monitoring species evolution throughout a trajectory. However, there is always the question whether the material states, which might be far from those of the training set used to parameterize the empirical model, are properly described. Thus, a more predictive, less empirical quantum-based level of theory is appropriate for use in these simulations. This leads to questions about the adequacy of the QM method. At what conditions do approximations in the chosen QM theory [116] break down? Furthermore, within a QM representation, there is no unique way to define electron localization and therefore, no unique way of defining whether atom pairs are bonded, or whether species are radicals or ions. For example, species may be identified based on bond distance and lifetime criteria, a reasonable, but nonunique scheme to determine chemical moieties [117]. However, different sets of bond distance and lifetime criteria could result in different sets of observed chemical moieties. Without a unique means of identifying species, individual reaction steps, the key to the classical chemical kinetics mechanisms, cannot unambiguously be determined and monitored to measure lifetimes.

Perhaps the single most difficult challenge is accurately simulating the overall event. Current computational capabilities allow for certain QM methods (e.g., DFT) to simulate systems under extreme conditions, thus reducing reliance on empirical models but at an increased computational expense. It would be desirable to use highly accurate QM methods, such as the "gold standard" of QM (coupled cluster with singles, double, and perturbative triples excitations [CCSD(T)]) [116]; however, this method is extremely computational intensive and is limited to approximately 20–40 atoms for a single time step. Outside of this consideration of accuracy versus computation time, the simulations suffer from several other deficiencies that might introduce error into the results. First, the systems being simulated are highly idealized, and a realistic material environment is inadequately described. The computational requirements for QM simulations of this type preclude inclusion of material heterogeneities in the simulation cell and do not allow for simulations of processes that go beyond a few picoseconds (assuming system sizes no larger than 10^5 atoms). For some simulations, selection of appropriate initial conditions is tricky and fraught with opportunity to introduce bias. For example, in quantum molecular dynamics (QMD) simulations that target a specific thermodynamic state at an extreme temperature and/or pressure (such as Wu et al. [154] and Rice and Byrd [155]), equilibrating the system might result in chemical reactions occurring before the desired thermodynamic state is reached. This, in turn, could influence subsequent chemistry once the targeted thermodynamic state is reached. Thus, it must be considered whether the initial state and subsequent equilibration protocol had biased the resulting chemistry as the desired

thermodynamic state was approached. Rice and Byrd attempted to address these issues by performing two isothermal–isobaric QMD simulations of formic acid at extreme conditions in which chemistry was observed, and for which the initial conditions were dramatically different [155]. In these two simulations, densities and potential energies converged to the same values on the time scale of the simulations. This information was used to predict a shock Hugoniot point using assumed final thermodynamic states. However, due to the limited simulation time, it could not be determined whether full chemical equilibrium was reached, thus potentially influencing the predicted shock Hugoniot point. Furthermore, in simulations of this kind, the equations of motion are coupled to a thermostat and/or barostat, which could influence chemical reactions as energy is adjusted to achieve the desired temperature and pressure. For example, the dissipation rate of thermal energy originating from exothermic reactions will depend on a thermostat's damping parameter.

Simulations other than those that target-specific thermodynamic states, such as the aforementioned examples [154, 155], are also susceptible to potential errors by virtue of their simulation protocol or process. For example, in some simulations to explore onset of thermal decomposition [117, 126, 130, 131, 140, 142, 151, 156], a system is first optimized or equilibrated to a state in which reaction does not occur and is then heated, leading to the question of how the heating rate is influencing the chemistry, and whether the heating rate is realistic. For shock simulations, two approaches are used to explore shock-initiated chemistry: direct mechanical shock simulation or the multiscale shock technique (MSST) [157, 158]. For the former approach, of which there are multiple schemes available for mechanically generating a shocked sample, the system is often overdriven in order to observe chemical reaction within a computationally feasible time frame. The question is raised as to whether the chemistry resulting from the overdriven shock (leading to a higher degree of material compression) is relevant to the chemistry associated with steady-state detonation or that initiated by a weaker shock. On the other hand, while several studies [118, 119, 121–124, 132, 133, 135, 136, 145] have been used to explore the chemistry resulting from a shock using the MSST approach (which allows for smaller simulation sizes, and thus longer times), the inherent assumptions within MSST regarding stress gradients and thermal gradients limit its accuracy in describing material states immediately behind the shock front. Thus, it is unlikely that direct mechanical shock and MSST simulations would yield the same initial chemistries directly behind the shock front. This, in turn, could influence subsequent chemistry across the reaction zone. An example limitation of MSST is that it does not explicitly model a wave traveling through the sample. As such, some aspects of anisotropic material response cannot be captured, as microstructural features (e.g., voids, grain boundaries) will not experience a directional shock wave moving through them. Another advantage of MSST over direct shock simulation is that it allows for the convenient inclusion of quantum nuclear effects [159, 160], which may decrease the shock strength necessary to observe the onset of reactions. For both MSST and direct mechanical shock simulations, it is possible that the chemistry is biased due to the initial conditions, equilibration protocol, and simulation process.

Furthermore, these nonequilibrium simulations are performed within the constraints of periodic boundary conditions, which can be problematic if the material's thermodynamic properties or local structure becomes significantly inhomogeneous within the simulation cell (such as conditions under mechanical loading). These local variances (which may further increase with reaction during the simulation) could interact with the periodic images and adversely influence the outcome. Some of these errors are introduced because of system size and time limitations; they can be somewhat mitigated by using empirical or semiempirical approaches, such as reactive force fields [22, 23] or tight-binding DFT (DFTB) [161] to perform substantially larger and longer simulations [117]. However, the accuracy of these methods cannot be assumed for conditions beyond those used for parameterization.

Even when using DFT, the accuracy of the methods at extreme conditions is a limiting factor. As opposed to gas-phase processes, the highly accurate multi-reference QM approaches used in the evaluation of elementary reaction rates cannot be applied to the condensed phase, due to computational costs. (For a further discussion of quantum mechanical methods, see Taylor and Rice [116]). Additionally, at extreme conditions, electronic excited states might play a role in chemistry; these could not be treated by single-reference DFT. Furthermore, pseudopotentials used for computational efficiency might introduce errors for highly compressed material. However, as DFT is the most reasonably accurate ab initio approach available to predict reactions in a shocked condensed phase system at this time [116, 117], we consider it the best approach to model chemistry of an energetic material under extreme conditions, and thus, DFT is being used in our attempts to determine a condensed phase reaction model for use in DPD-RX.

10.4.3 Some Illustrations of the Challenges

An illustration of some of the difficulties associated with DFT simulations described above is evident in a heroic study by Wu and co-workers [154], in which they attempted to examine the chemistry of solid PETN compressed to the estimated Chapman–Jouguet (CJ) density of 2.38 g/cm^3 and heated to temperatures of 3,000 and 4,200 K (the estimated CJ temperature) using DFT isothermal–isochoric (NVT) simulations. By "painstakingly" tracking reactions during the simulations, Wu et al. found, for example, "over 3,000 unique reactions, 78% of which only occur once in the simulation" in a periodic simulation cell containing only four molecules [154]. With computational platforms and algorithms substantially improved since the 2009 Wu et al. study [154] that allow for system sizes and simulation times of 10^5 atoms and picosecond timescales, respectively, using DFT, the complexity and number of the reactions discovered are expected to increase. Wu et al. were able to determine that material conversion under these conditions involved catalysis by water and its decomposition products, a distinctly different process than those assumed in more traditional proposed decomposition mechanisms [154]. The catalytic process is possible due to the extreme conditions, in which water rapidly dissociates and provides

a continual source of OH and H. By virtue of this ease of dissociation, Wu et al. suggest that "bonds containing hydrogen are inherently nonmolecular, and thus the CJ state should not be treated as a mixture of conventional molecules" [154].

Likewise, our own forays into large-scale QM studies of materials under shock conditions have shown us that unusual chemistries can occur. For example, we performed QMD NVT simulations of formic acid at shock conditions [155], a detonation product for which there was conflicting experimental information regarding its reactivity at high shock pressures, and is a standard species considered in thermochemical code calculations. QMD NVT simulations at a state point well above the purported transition pressure for reaction showed chemical reactivity, determined from time traces of all original covalent bonds in parent formic acid, the nearest-neighbor hydrogen atom from each of the oxygen atoms in the parent molecule, as well as nearest-neighbor distances between heavy atoms in adjacent molecules. The simulations showed that hydrogens exhibited extensive mobility, migrating back and forth among species, resulting in hydrogen exchange reactions to reform formic acid or forming protonated formic acid or formate moieties. Also observed were long-lived extended networks composed of fragments and atoms from various parent molecules, which might be the early stages of polymerization under these conditions.

Other quantum-based simulation studies of materials under high temperature and pressure states also have shown similar mobility of hydrogen atoms [133–135, 162, 163], as well as charge transfer [118, 136, 164]. Similarly, we have observed "nonmolecular" hydrogen behind the shock front in overdriven shock simulations of PETN using large-scale, DFT, Born–Oppenheimer MD simulations. For example, we followed a hydrogen atom propelled forward toward the shock front, weaving through free space among a tangled mass of atoms so densely packed that chemical speciation would require a herculean effort, and would rely on a certain level of empiricism. This atom quickly migrated forward ahead of the mass flow, before it was subsequently captured by a different moiety closer to the shock front, thus arresting its free motion. In this shock simulation using QM forces, it appears to be impractical (or impossible) to define the chemistry in terms of unimolecular or bimolecular mechanisms.

While the material state in the overdriven shock simulations of PETN is at higher compression and temperature than the Wu et al. simulation of PETN at a single thermodynamic state corresponding to the experimental CJ condition [154], it is notable that a few features are similar, such as the presence of nonmolecular hydrogens. It is unfortunate that current state of the art precludes QMD simulation for sufficiently long times to simulate a steady-state detonation of an EM; thus, the material state in the reaction zone of a steady-state detonation remains to be discovered. Until novel algorithms and computational resources allowing for realistic atomistic simulations of shock initiation leading to steady-state detonation are available or advanced experimental methods are developed to interrogate the reaction zone, that question will remain unanswered. These and other studies of chemistry of materials under extreme conditions clearly support the conclusion of Wu et al. that "the traditional

approach based on molecular reactions that is commonly applied in gas-phase combustion chemistry is no longer adequate for describing chemical reactions under these extreme conditions [154]."

10.4.4 Outlook

Due to the lack of understanding of the material state and the complex chemical conversion occurring behind the shock front of a reacting EM, we argue that the most immediate need at present is to perform an extensive series of quantum-based (either DFT or DFTB) MD simulations of an EM subjected to shocks of different strengths to monitor emergent material response, specifically identifying features behind the shock front in which molecularity is retained or lost, and where chemical speciation can or cannot be determined. A lofty goal would be to compare the heterogeneous material state within a simulated detonation reaction zone against that proposed in the Nonequilibrium Zeldovich–von Neumann–Doring (NEZND) model of detonation for condensed phase explosives [165]. Pursuit of this goal is facilitated by the emergence of exascale computing [166–168], novel computational methods that will allow for QM simulations with larger system sizes and longer times (e.g., [169]), and data mining approaches [170–174] to cull crucial information directly from these atomistic simulations leading to the much-needed understanding.

Clearly, both the time and financial commitment and resources required to achieve an understanding of condensed phase chemistry is daunting. Moreover, the path forward is speculative, where we have described a possible approach for determining only the reaction mechanisms. Even more daunting is the task of determining the reaction kinetics associated with the reaction pathways. Still further, there are considerations for transitioning this information into generating a reaction model that is of practical use in the DPD-RX framework. For example, a complete reaction mechanism is required that encapsulates the reactive behavior from the initial unreacted EM through to the final product gas species. Thus, research is needed to formulate ROM for use in the DPD-RX framework that captures the important features of the detailed chemical kinetics mechanisms obtained from the atomistic simulations. Still other open-ended questions remain. If radicals, polyradicals [175], or transient states are part of the reaction mechanism, can the required input for the CG-EOS for each species be readily determined (i.e., a reference state heat of formation and temperature-dependent constant-pressure heat capacity)? Will these transient states be so short-lived that they won't play a role within the time scale of the DPD-RX time step (~5–20 fs)? If they are sufficiently long-lived, can CG models be readily developed for these species? If we develop a highly-detailed reaction mechanism, will these details be significant on the time scale of the DPD-RX simulation?

These questions provide ample opportunities for research investigations designed to afford a multiscale description of condensed phase chemical reactivity of EM at extreme conditions. Our expectation is that the ROMs will be progressively enhanced as the condensed phase reaction modeling efforts evolve. These reaction model devel-

opments will be complemented by any necessary coinciding adaptions of DPD-RX. This complementary approach provides a convenient and natural framework to incorporate the complexities arising from the coupling of microstructure with a quantum mechanical understanding of chemistry.

10.5 Hierarchical Multiscale Simulation: Reaching the Experimental Scale

Prediction of the macroscale response of energetic material, including all aspects of chemical reactivity, provides a significant motivation for continued research in multiscale modeling. Continuum "burn" (chemistry) models may often reproduce shock-to-detonation transitions when parameterized using sufficient experimental data, or when using chemical equilibrium approaches for ideal explosives, but subdetonative reactivity and deflagration-to-detonation remain difficult to predict for many systems. As chemical reactivity may be influenced by many factors, such as material EOS or microstructure, our research plan involves systematically increasing the complexity of our multiscale simulations. In this section, we will discuss: (a) common multiscale approaches, (b) properties of constitutive models for energetic materials that are currently addressed by multiscale simulation, (c) our current hierarchical multiscale approach and implementation, (d) a demonstration of our concurrent hierarchical approach, and (e) future research opportunities.

10.5.1 Multiscale Approaches

Information from particle-based simulations, such as those using the DPD-RX method, can be utilized in continuum simulations through a variety of modern multiscale approaches. We will briefly survey common multiscale approaches, generally following the taxonomy described by Tadmor and Miller [20]. In describing multiscale approaches, we will typically be describing a continuum, or macroscale, simulation that leverages information from high-fidelity, smaller length-scale calculations. In the literature, these may also be referred to as upper and lower scales, or alternatively as coarse and fine scales. We will avoid referring to the fine scale as the "microscale" in order to avoid confusion of the term with micron length scales or microsecond timescales, which nevertheless may be present in particular multiscale approaches.

10.5.1.1 Sequential Upscaling

Fast-running constitutive material models for continuum simulations may be directly parameterized from the results of fine scale simulations, much like how CG models may be constructed from the results of atomistic simulations. Examples of this include construction of an EOS for a pure material from rigorous first principles calculations [176], calculation of elastic constants for use in a continuum model [177], or the fitting of a chemistry model to a numerical description of molecular simulation results [175]. We will not discuss the sequential approach in detail; it is historically the most common approach for transferring information between scales. When carefully applied by a subject matter expert, it may provide good results for the selected problem. It is often the case that a particular upscaled material property is understood to only be applicable within a constrained range of states, i.e., the transferability problem (particularly important in force field development [178]).

10.5.1.2 Concurrent

Concurrent approaches involve performing simulations with both macroscale and fine scale methods at run time [20]. This allows for a higher fidelity description of phenomena of interest than sequential upscaling, but with increased computational expense. Fracture, for example, is difficult to realistically model at the continuum scale, but one approach is the embedding of an atomistic domain in a continuum simulation [179]. A particular property's response may be dependent upon a large number of variables, making generation of an upscaled model in advance prohibitive due to the "curse of dimensionality [14, 180]." Microstructural effects that give rise to emergent behavior may be difficult to "build in" to a predictive constitutive model [11]. In each of those cases, a concurrent multiscale approach able to directly query the results of a high-fidelity model may provide an accurate solution. Note, the nomenclature for describing concurrent simulation is highly inconsistent across the open literature. We follow Tadmor and Miller's [20] usage and further categorize concurrent multiscale simulation as "partitioned-domain" or "hierarchical".

Partitioned-Domain

Partitioned-domain approaches involve a decomposition of the simulation into multiple spatial domains, which may be overlapping. An inexpensive, less computationally costly model is used in the larger domain(s), and an expensive, more computationally costly model is used in the smaller region of critical interest [181]. Examples of this include a QM/MM simulation for modeling the active site of proteins [182], or an embedded atomistic domain in a continuum finite element mesh. Challenges for partitioned-domain approaches involve describing the "handshaking" region between different methods/models, and the accessible simulation timescales being limited by the computationally most expensive per-time-step domain (typically

also the domain requiring the smallest time step). While quite successful in solving some problems, partitioned-domain approaches may not be applicable to problems where the domain requiring high-fidelity information is itself macroscale in size; the high-fidelity, expensive model could consume nearly the entire extent of the system being simulated. Another limitation is when long timescales need to be accessed. For example, some partitioned-domain methods are limited by the timescale of the step size in the expensive simulation (e.g., continuum cycles operating at MD time steps, roughly ~1 fs), although recent research has attempted to address this issue [179].

Hierarchical

Hierarchical approaches often involve the interleaving of macroscale and fine scale simulations [13, 183–185]. A macroscale simulation, such as a continuum finite element simulation, may require information about its elements from the fine scale at each time step [186]. This information may be provided from constrained fine scale simulations run in-between continuum time steps. This information exchange between scales characterizes a hierarchical approach. An example of this from the field of computational mechanics is the FE^2 simulation, where two finite element simulations are performed, with the fine scale being a much more highly resolved "representative volume element" (RVE) [187]. Intensive properties of the RVE may be used at the macroscale in a technique called computational homogenization [188]. More general mathematical frameworks for hierarchical simulation include the "equation-free" approach and heterogeneous multiscale method (HMM) [184, 189, 190]. In HMM calculations, generalized macroscale governing equations that rely upon a set of variables may have missing variables provided by constrained fine scale calculations. The constraints and simulations necessary at the fine scale are application-specific, and left for the subject matter expert to determine [189]. Using an HMM approach, one may calculate many different material properties across the entire macroscale domain using high-fidelity fine scale simulations. An example we will describe later in this section is calculating EOS response from particle-based simulations.

10.5.2 Constitutive Material Models

We briefly describe two common components of a constitutive model for energetic materials, which may be described by a multiscale model. Other possible components of a material model may include (but are not limited to) models for failure, elasticity, plasticity, yield strength, hardening, as well as properties such as thermal conductivity, melt curves, heat capacity, or viscosity. In a shock-to-detonation transition, brittle failure or thermal conductivity may not be a concern for modeling; the material will detonate on a short timescale and those properties may not affect

the transition. However, every continuum simulation for energetic material response requires an EOS, even if the material does not undergo reaction. Therefore we start with a discussion of the material EOS, and then discuss chemistry, since energetic materials will eventually leverage their energy-releasing chemistry during physically relevant simulations.

10.5.2.1 Equation of State

First and foremost for our application is the material EOS. This is the relationship between material energy, density, temperature, and pressure. A simple example is the ideal gas law for a gaseous EOS. Pressure is often calculated through an analytical EOS such as the Mie-Grüneisen equation, Jones–Wilkins–Lee EOS, or a number of other forms [191]. More complicated recent examples include the use of the Peng–Robinson EOS for multiphase mixtures [192], or construction of a free energy-based EOS for solid RDX [176]. In many material models, the temperature may not actually be calculated and is not present in the continuum governing equations. The temperature may be calculated if a cold curve, or the static lattice energy as a function of volume, is available. Tabulated forms such as Livermore EOS tables or SESAME tables are also used in modern hydrocodes, which may include data for all relevant state variables (including temperature). An accurate EOS is critical for calculation of peak shock pressure and related state variables in a continuum simulation, or for stress–strain response at a variety of strain rates. Calculation of realistic stress–strain curves requires both accurate elasticity and plasticity models before the onset of failure. Calculation of the EOS is particularly challenging for energetic materials, as chemical composition will change over the course of a reactive event, such as a detonation. An ideal steady-state detonation described by Chapman–Jouguet theory is at the state point where the Rayleigh line of the unreacted products is tangent to the shock Hugoniot for the product gas [193]. Calculation of those states, and therefore the detonation properties of an energetic material, requires an accurate EOS for unreacted material and any species in the reaction mechanism leading to the final products.

10.5.2.2 Chemistry

While continuum simulations of metals, ceramics, or some composites (e.g., fiber composites at low strain rate) may not need to consider chemical reactivity, energetic materials release stored energy through chemical reactions. When modeling explosives, phenomenological models such as the Tarver "ignition and growth" reactive flow model [165, 194] are often used to describe detonation kinetics. In that model, progress from reactants to products depends upon material density, pressure, and amount previously reacted. The reactants and products are described by separate EOS models. Complex hydrocodes also have support for advanced chemistry models that may explicitly account for many chemical compounds and complicated chemi-

cal kinetics, including chemical equilibrium detonation models [195]. In the absence of robust experimental data, it may be difficult to parameterize any type of chemistry model, motivating the need for multiscale modeling. Further, phenomenological models such as the ignition and growth model are less useful for some problem domains; for example, a model that describes shock-to-detonation may not provide predictive physical insight regarding the effects of low-velocity sub-detonative impacts on an energetic material. Therefore, phenomenological models are developed to model particular classes of problems. Significant progress has been made in developing models for thermal cook-off [8, 196] or reactions due to low-speed impact [197]. In principle, a high-fidelity multiscale model can be used to describe the response to loading at a variety of conditions, including the key chemical processes in a material [171, 198]. Advances in the DPD-RX methodology (Sect. 10.3) provide a path forward for high-fidelity modeling at micron length scales that may address conditions ranging from nonreactive, low-speed impact to shock-to-detonation transitions.

10.5.3 Hierarchical Multiscale Simulation

10.5.3.1 Definition of Scales

In this section, we will describe our particular approach for bridging to continuum, which we refer to as "hierarchical multiscale simulation" (HMS). The macroscale in our work is a continuum, Lagrangian finite element simulation in ALE3D [171, 199]. The fine scale in our work generates results from the set of DPD variants implemented in LAMMPS and the CG-RDX model described above. The constitutive material model in ALE3D receives its EOS information from these fine scale simulations. We will also describe results where chemistry is provided from fine scale simulations. This combination of macroscale and fine scale defines a concurrent, hierarchical approach using the Heterogeneous Multiscale Method. A software framework, referred to as the HMS framework, handles communication and transformation of information between the macroscale and fine scale [200, 201]. The HMS framework also handles computational aspects of the multiscale implementation, such as queuing of the fine scale calculations and storage of the fine scale results. Figure 10.7 provides a conceptual representation of the connection between scales, with the HMS framework handling communication between the macroscale and fine scale, and requests being dispatched from the macroscale on a per-element basis.

10.5.3.2 Machine Learning

In some cases, direct evaluation of the fine scale for every element at every continuum time step (or "cycle") may be prohibitively expensive for simulations with millions of finite elements, over many hundreds of continuum cycles. Consider the case of a 1

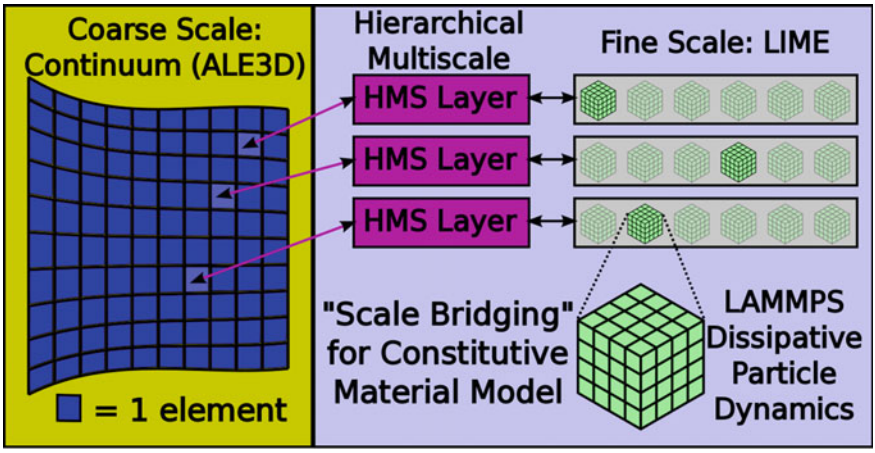


Fig. 10.7 Conceptual representation of the connection between the macroscale, the HMS framework, and the fine scale

million element simulation run for 7000 cycles, with a fine scale that requires 3 min and 64 cores for solution (for one element, at one cycle). On a 100,000 core machine, completion of that simulation would take over 25 years. In order to make our HMS approach computationally tractable, the HMS framework implements a feature to leverage previously computed fine scale results. The central idea is that a set of such previously computed fine scale results may be used to approximate the result for a new property calculation, if the new fine scale state is easily interpolated from previously computed states. Interpolation may be many orders of magnitude faster than an actual fine scale evaluation. Publications variously describe this general approach as surrogate modeling, adaptive sampling, or machine learning [180, 200, 202–209]. When used with a database managed by the HMS framework that is expanding over time (i.e., over the course of the macroscale simulation), where additional information in the database improves interpolation efficiency and accuracy, we consider this to be regression via supervised machine learning. HMS applications requiring millions of elements and hundreds of cycles become computationally tractable with an efficient machine learning implementation. It is critical that the development of this data-driven model is done in close coordination with a subject matter expert for the underlying physics-driven model, otherwise unphysical results may go unnoticed. In short, theory and data should work hand in hand [172].

10.5.3.3 Speculative Evaluations

The use of machine learning to make problems tractable introduces a new challenge for HPC environments, where the number of cores available is not a severe constraint. As the efficiency of the machine learning algorithm (i.e., its interpolation rate) may

vary from cycle to cycle, the number of required fine scale evaluations will fluctuate. If the number of cores requested for the job exceeds the number of cores necessary for fine scale evaluations, then some remaining cores may be idle. This inefficiency may be overcome through the launching of "speculative" evaluations of the fine scale model on those otherwise idle cores. While not directly used in constitutive model evaluations for the current cycle, these speculative evaluations may be added to the HMS framework database of fine scale results. Additional data in that database will improve the efficiency and accuracy of future machine-learned model evaluations, therefore potentially decreasing wall clock run time and increasing the overall fidelity of the simulation. This is particularly important, as wall clock run time may be determined by the slowest element; improvement of the fine scale database to a point where no fine scale evaluations are needed for a particular cycle results in impressive wall clock speedups. Speculative HMS may also leverage batch schedulers that allow changes in the number of cores available for jobs that are currently running; additional speculative evaluations could be initiated if there are idle nodes on a supercomputer, or the number of speculative evaluations was lowered for the case in which a higher priority job is needed to acquire additional batch nodes. The inherently parallel nature of fine scale evaluations—each fine scale simulation is independent of others—makes speculative HMS easily amenable to full utilization of petascale and potentially even exascale HPC resources, given a sufficiently complicated problem and expensive fine scale evaluations.

10.5.3.4 The HMS Framework

Key to the effective execution of this HMS strategy is the software framework that interfaces the macroscale and fine scale [200, 201]. Typically, simulation software is not written with concurrent, hierarchical multiscale coupling in mind. The software is most often run in a standalone manner using well-documented material models, such as analytical forms available in hydrocodes or empirical potential forms built-in to MD programs. Therefore, communicating constraints and results between those programs is a software engineering task. This framework is also responsible for the machine learning (or surrogate model) and scheduling of fine scale evaluations, which are computational science problems. While not adding significant computational overhead, the HMS framework is responsible for handling the parallel execution of the updates to the fine scale database (evaluations necessary for the macroscale model), as well as the updates to the machine-learned model. It should be sufficiently general that if the HMS requirements change, e.g., additional macroscale variables are requested for evaluation in a more complicated fine scale model, then large parts of the software framework do not need to be rewritten. Similarly, it should be flexible enough to allow for easy changes to the machine learning method or its hyperparameters and to accommodate restarts to simulations that are terminated due to hardware error or queue run time limitations. In general, the design and implementation of the HMS framework is an effort that is distinct from, but as important as, the design and implementation of the fine scale model.

10.5.4 HMS with LIME as the Fine Scale

In this final section, we describe HMS results using a fine scale model for RDX. The fine scale solver we use is the LAMMPS Integrated Materials Engine (LIME) [210], an automation and analysis tool written in Python, and developed specifically for our HMS effort. LIME instantiates, executes, and returns results from DPD simulations using our CG model for RDX. In an HMM review article, E. and co-workers describe the design of the fine scale solver as "often the most difficult step, and is subject to continuous improvement [189]." Fine scale solvers must reliably return an accurate result to the HMS framework without human intervention in a minimal amount of time, over any range of inputs that may be encountered during the macroscale simulation. Further, execution of many fine scale solvers must effectively scale—task parallelism leads to efficient leveraging of supercomputing resources [186]. If the fine scale solver is extremely costly for reading or writing many files or large files to a shared file system, i.e., input and output (I/O) intensive, then simultaneous execution of thousands of fine scale solvers may degrade HMS performance. This may also occur for MPI initialization of programs [211]. Further, jobs unrelated to the multiscale simulation may also be degraded. As such careful design of the fine scale solver is necessary.

10.5.4.1 Nonreactive Case

The nonreactive case of HMS using LIME utilizes a CG model for RDX and DPD-E. In this simulation, the macroscale requires EOS information from the fine scale. In particular, it needs temperature and pressure (T, P) as a function of energy and density (E, ρ). The pressure desired is the hydrostatic component of the stress tensor. The hydrostatic nature of the response and the two input state variables provide sufficient constraints to allow for the construction of the fine scale problem. The temperature and pressure returned may be considered to be equilibrium values. This allows for decoupling of the macroscale and fine scale in both length (the fine scale simulation will have a smaller volume than the macroscale element) and time (equilibrium simulations at the fine scale may have a smaller simulation time than the duration of macroscale time steps). The fine scale solver must instantiate a DPD simulation with the same energy and density as the macroscale inputs. It must then equilibrate the simulation cell. Although equilibration may be assumed for sufficiently long equilibration periods, properly detecting equilibration is nontrivial and extremely useful for performing fine scale simulations in minimal amounts of wall clock time. After the cell is equilibrated, the solver must collect enough data to produce accurate ensemble averages for temperature and pressure, and return those averages to the HMS framework. Additional data, such as the standard error and variance of the ensemble averages, may also be returned. Those values may be used for uncertainty quantification at the macroscale. When information from microscale/mesoscale simulations is used to determine material response in continuum simulations, the ideal

continuum model would reproduce the exact result of the fine scale model, if the two models/methods were used to simulate the same physical system over the same amount of time. In general, correspondence of observable properties between scales is a key component of validating a multiscale model.

In Fig. 10.8, we demonstrate this correspondence of results for the *P*-*V* and *T*-*V* planes of the Hugoniot with our CG-RDX model. The sets compared are data provided by LIME and used in an ALE3D plate impact simulation, and data from longer duration results purely from DPD-E simulation in LAMMPS. The T-V plane is particularly sensitive to error. For example, if the simulation cell does not have approximately equal pressure in each cardinal direction when generating data to be used in ALE3D, then the LAMMPS and ALE3D results will diverge by several Kelvin. LIME carefully monitors the normal directions of the pressure tensor in order to ensure they are roughly equal, therefore that error is not present in Fig. 10.8. Additionally, a very small difference is observed when a yield stress (strength model) is present in the ALE3D simulation. Disabling this restores near exact agreement between the sets.

Because temperature determines chemistry in our fine scale model and is the more sensitive property for bridging scales, it is prudent to investigate the accuracy of temperature calculations in further detail. Figure 10.9 demonstrates the standard error of the mean, or the sampling accuracy, of temperature for our fine scale model using LIME. The axes in Fig. 10.9 reflect the inputs to LIME: finite element energy

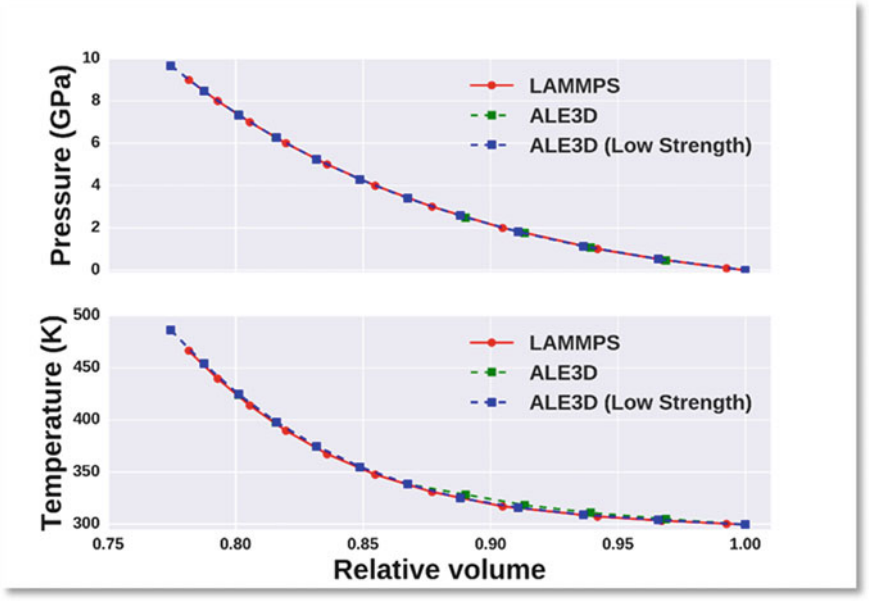


Fig. 10.8 *P*-*V* and *T*-*V* planes of the Hugoniot for particle-based simulations (LAMMPS) and continuum simulations (ALE3D) driven by LIME EOS tables

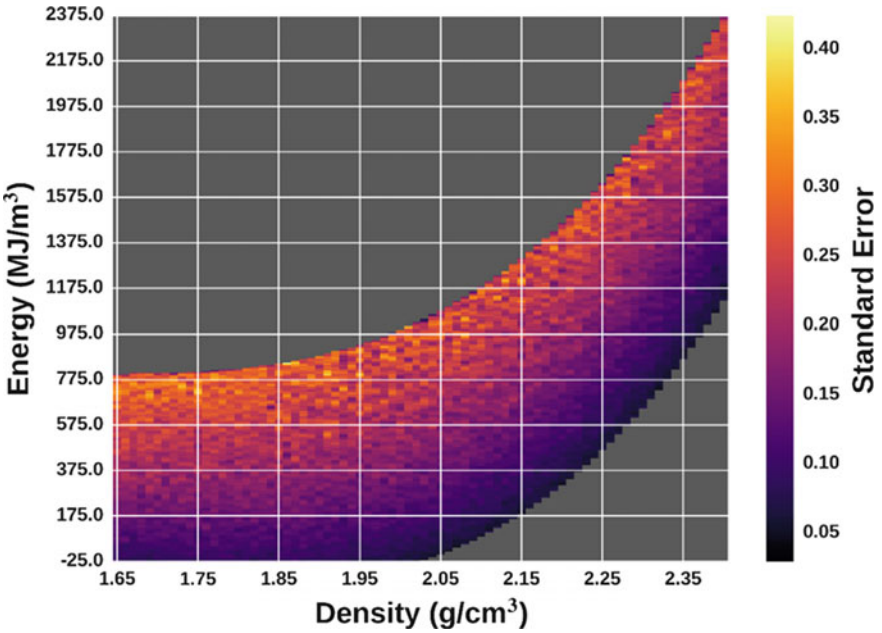


Fig. 10.9 Standard error of the mean for temperature over a range of LIME EOS calculations

and density. In general, LIME provides results with 0.5 K or better accuracy in temperature for a given state point. As would be expected, systems that are higher in energy (hotter) have larger error, since we are demonstrating absolute error and not relative error.

Understanding the efficiency of fine scale simulations across a wide variety of state points is important, in that it determines both a limiting factor for wall clock time of the macroscale simulation, and may reveal a challenge for automation of fine scale calculations. In Fig. 10.10 we examine the total number of DPD-E time steps (equilibration and production, combined) performed during state point evaluations across the wide-ranging surface of energy and density. While the vast majority of state point evaluations using LIME require fewer than ~9000 total DPD-E time steps, there is a significant cluster in the 1.85–2.15 g/cm^3 range at energies corresponding to temperatures of over 500 K, where many more time steps are needed in order to converge LIME. If it would be common for simulations to explore that part of state space, then the algorithms and heuristics in LIME could be further tuned in order to accelerate convergence for those state points.

Putting it all together for the nonreactive case, we demonstrate a Lagrangian finite element Taylor anvil impact simulation using ALE3D as the macroscale and LIME as the fine scale for the material EOS. In Figs. 10.11 and 10.12, snapshots from 2D axisymmetric impact simulations are shown with elements colored by pressure, demonstrating pressure waves traveling through the material and deformation at

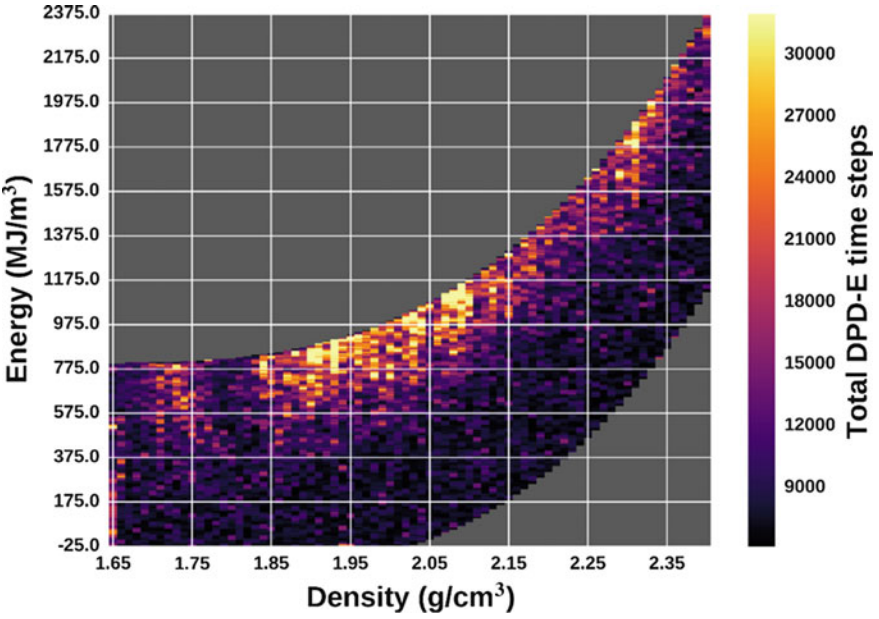


Fig. 10.10 Total number of DPD-E time steps necessary for LIME convergence over a range of state point calculations

material edges that are impacting the hard wall. The simulation in Fig. 10.11 used LIME for EOS calculations in each of the 1600 elements. These calculations were performed "on the fly", meaning that LIME was called during the HMS simulation, for each element, at each cycle. The simulation in Fig. 10.12 also used LIME as the fine scale, but used a machine learning method for statistical regression in the HMS framework in order to provide estimates of LIME response within a controllable error tolerance. The regression method of choice for our HMS framework is kriging, also known as Gaussian process regression, and was previously demonstrated in a study examining a two-scale model for crystal plasticity [200]. This allowed the simulation in Fig. 10.12 to use one million elements for its mesh (a 625× increase in mesh size), without needing to explicitly call LIME for each of those elements at every cycle. Comparing the figures, for this case, it is evident that higher mesh resolutions are critical for resolving physical phenomena in continuum simulations, where the separation of elastic and plastic waves is visible in Fig. 10.12, while not visible in Fig. 10.11.

10.5.4.2 Reactive Case

The second example of HMS using LIME is that of a plate impact test using the CG-RDX model and DPD with reactions (DPD-RX) method described above in

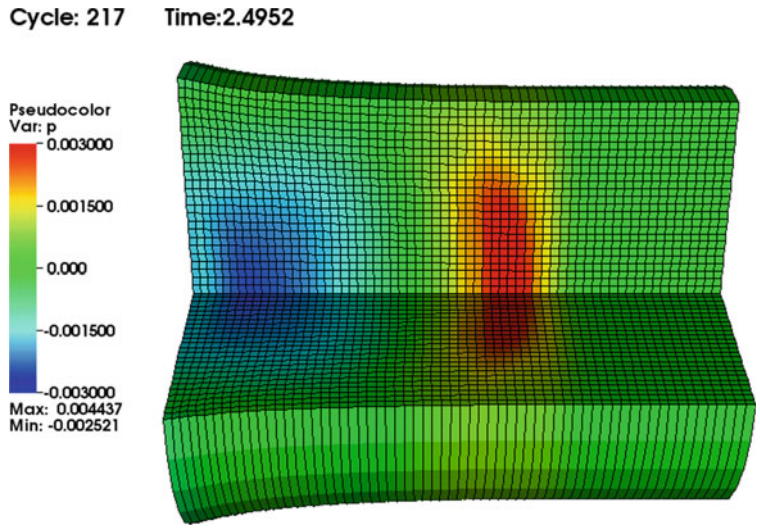


Fig. 10.11 Taylor impact test simulation, colored by pressure, with 1600 elements. EOS response for each element was calculated with LIME "on the fly" every cycle. Time units are microseconds and pressure units are megabar

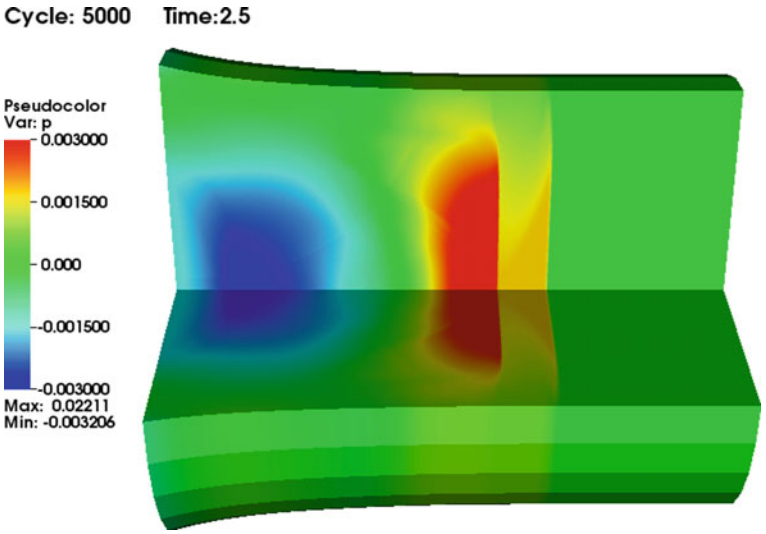


Fig. 10.12 Taylor impact test simulation, colored by pressure, with one million elements. EOS response for each element was calculated using a combination of LIME "on the fly" and machine-learned responses. Time units are microseconds and pressure units are megabar

Sects. 10.2 and 10.3, respectively. In this simulation, the macroscale requires not only EOS information from the fine scale, but also chemical species information, an additional nine variables given our reaction model described in Sect. 10.3. However, the time evolution of chemical kinetics under shock infers that the fine scale may no longer be assumed to be in an equilibrium state. Therefore, a similar instantiation and equilibration scheme may be followed in LIME (adjusted for input chemical species from the macroscale), but production data is collected in "lock step" with the macroscale; in particular, the fine scale solver covers the same production simulation time as the macroscale time step. HMS plate impact simulations, such as the snapshot in Fig. 10.13, demonstrate the consumption of RDX and temperature rises similar in magnitude to those observed in DPD-RX simulation, but systematically lower. This is believed to be in part due to the effects of computational homogenization when a wave has only partially passed through a many-micron wide finite element. Resolving differences between reactions at the macroscale and fine scale and solving the "scale-bridging in time" problem for the nonequilibrium case are active areas of research for us.

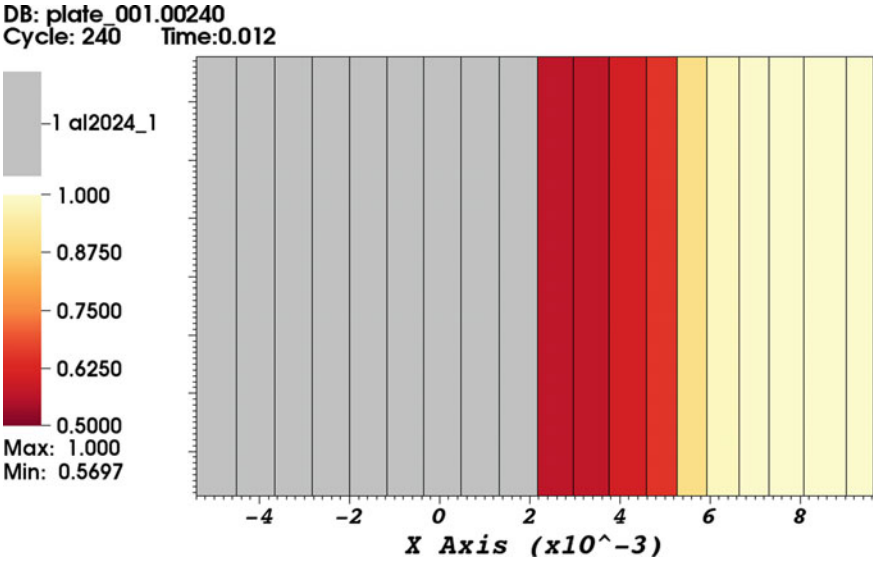


Fig. 10.13 Plate impact simulation, colored by amount of RDX consumed (red scheme, 1.0 = pure RDX, 0.0 = all RDX consumed) and material type (gray, aluminum). This demonstrates direct use of chemistry from fine scale simulations during a continuum shock impact simulation. Only elements near where chemistry is occurring are displayed. Time units are microseconds and length units are centimeters

10.5.5 Outlook

Going forward, many challenges need to be overcome for our HMS approach applied to energetic materials. Simulations to date have relied on an idealized description of RDX as a homogenous, defect-free single-crystal material. However, as described at the beginning of this chapter, it is known that for real energetic materials, initiation thresholds are determined by a number of factors such as polycrystalline grain size, shear, porosity, amount and types of binder or plasticizer present, or more broadly, any inhomogeneities present [1, 2, 6, 117, 194, 212–220]. Incorporating those effects into an HMS model will require continued research and development of the fine scale particle-based models, and then leveraging of those models in an HMS scheme that utilizes additional history variables to describe each of those components of the material microstructure. Capturing all of those effects in a single microscale RVE may be difficult or impossible, requiring use of several statistical volume elements to capture a variety of effects and then return weighted results to the macroscale [221–224]. Including over a dozen history variables for descriptions of microstructure and chemistry in an automated fine scale calculation also leads to new challenges for machine learning in a high-dimensional space, the so-called "curse of dimensionality [14, 180, 225–229]." Regardless of the multiscale approach chosen, currently there is no immediately apparent approach to model the complex reactive response of energetic materials across all possible insults. However, with further developments, HMS provides a viable and promising path forward for accurate, high fidelity simulations in the future.

10.6 Concluding Remarks

Advances in multiscale modeling and simulation methodologies are beginning to make possible virtual design and performance assessment of EM, before actual production, enabling modeling and analysis of complex phenomena across multiple time and length scales. We have described our own developmental efforts to predict the dynamic behavior of EM at all relevant length scales that will lead to understanding of the various dynamic processes, properties, and mechanisms for energetic material response under the full range of conditions. Microstructure plays a dominant role in the macroscale response of the EM, however, computational capabilities have been lacking at the microscale. We described our efforts to fill this computational capability gap through the development of CG models and methods that can simulate microstructure evolution (including effects of chemistry) in response to various stimuli. Upon upscaling of these microscale tools into the macroscale models, the effects that microstructural heterogeneities impose on macroscopic events can be captured. We described multiscale modeling methodologies capable of coupling behaviors from the fine scale to the macroscale, including our choice of a concurrent,

hierarchical multiscale approach, and the computational simulation framework into which simulations at the various scales are effectively integrated.

Relative to atomistic and continuum approaches used to simulate EMs, our microscale and hierarchical multiscale efforts are in the nascent stage of development. Currently, we have demonstrated proof of concept calculations and provided paths forward for advancing the methodologies. We are proceeding in a straightforward methodical manner, and while much more work is needed, our efforts continue to evolve, where progress continues to be promising. Although the effects of the loss of fidelity due to coarse-graining have yet to be rigorously quantified for our models, estimates of the computational gains are encouraging. If we consider MD using the ReaxFF force field as the most viable model comparison for the detailed simulation of microstructure evolution with chemistry, estimates of the computational costs savings of our microscale approaches are on the order of 10^4 in speedup. Such gains allow us to move beyond the simulation of isolated ideal microstructure and begin simulating microstructure typically found in actual EM composites.

We described deficiencies of various components of our efforts, with an emphasis on the reaction models. We believe that future work should focus on the development of reduced-order models that properly depict chemistry of condensed phase materials under extreme temperatures and pressures. Key development needed for our CG models is a more accurate modeling of the plastic response and improvement in the transferability of the product gas mixture model. A next stage development for our CG methods will focus on a more complete representation of species mass diffusion. For the multiscale hierarchical tools, further work is needed on novel statistical sampling of microstructural features at the fine scale, and obtaining an understanding of error propagation and uncertainty quantification across the scales. This includes uncertainties on dynamic yield behavior, fracture behavior, mechanical/chemical physics at extreme states, plastic deformation, anisotropic crystal properties, material slip, and contact behavior—all of which affect the behavior of an EM.

We hope that the description of the research challenges will inspire further development of innovative models and methods, leading to a robust predictive multiscale modeling and simulation capability that will describe EM response under the full range of conditions. Our tools are built around a general computational framework, so that they can be extended to other material systems with moderate modification. Therefore, while we hope our efforts will advance EM design, we also hope that they will be used to study materials beyond EM.

References

1. Baer MR (2002) Modeling heterogeneous energetic materials at the mesoscale. Thermochim Acta 384:351–367
2. Baer MR (2007) Mesoscale modeling of shocks in heterogeneous reactive materials. In: Horie Y (ed) Shock wave science and technology reference library, vol 2. Springer, pp 321–356

3. Barton NR, Winter NW, Reaugh JE (2009) Defect evolution and pore collapse in crystalline energetic materials. Model Simul Mater Sci Eng 17(3):035003. https://doi.org/10.1088/0965-0393/17/3/035003
4. Jackson TL, Hooks DE, Buckmaster J (2011) Modeling the microstructure of energetic materials with realistic constituent morphology. Propel Explos Pyrotech 36(3):252–258. https://doi.org/10.1002/prep.201000096
5. Akiki M, Menon S (2015) A model for hot spot formation in shocked energetic materials. Combust Flame 162(5):1759–1771. https://doi.org/10.1016/j.combustflame.2014.11.037
6. Austin RA, Barton NR, Reaugh JE, Fried LE (2015) Direct numerical simulation of shear localization and decomposition reactions in shock-loaded HMX crystal. J Appl Phys 117(18):185902. https://doi.org/10.1063/1.4918538
7. Rai NK, Udaykumar HS (2015) Mesoscale simulation of reactive pressed energetic materials under shock loading. J Appl Phys 118(24):245905. https://doi.org/10.1063/1.4938581
8. Nichols AL, Becker R, Howard WM, Wemhoff A, Elert M, Furnish MD, Anderson WW, Proud WG, Butler WT (2009) Toward improved fidelity of thermal explosion simulations, AIP Conf Proc 1195:229. https://doi.org/10.1063/1.3295110
9. Geers MGD, Kouznetsova VG, Brekelmans WAM (2010) Multi-scale computational homogenization: trends and challenges. J Comput Appl Math 234(7):2175–2182. https://doi.org/10.1016/j.cam.2009.08.077
10. Ghosh S (2015) Foundational aspects of a multi-scale modeling framework for composite materials. Integrating Mater Manuf Innov 4(1). https://doi.org/10.1186/s40192-015-0036-x
11. Geers MGD, Yvonnet J (2016) Multiscale modeling of microstructure–property relations. MRS Bull 41(08):610–616. https://doi.org/10.1557/mrs.2016.165
12. Yip S, Short MP (2013) Multiscale materials modelling at the mesoscale. Nat Mater 12(9):774–777. https://doi.org/10.1038/nmat3746
13. Elliott JA (2013) Novel approaches to multiscale modelling in materials science. Int Mater Rev 56(4):207–225. https://doi.org/10.1179/1743280410y.0000000002
14. Unger JF, Eckardt S (2011) Multiscale modeling of concrete. Arch Comput Methods Eng 18(3):341–393. https://doi.org/10.1007/s11831-011-9063-8
15. McDowell DL (2010) A perspective on trends in multiscale plasticity. Int J Plast 26(9):1280–1309. https://doi.org/10.1016/j.ijplas.2010.02.008
16. Kouznetsova V, Brekelmans WAM, Baaijens PT (2001) An approach to micro-macro modeling of heterogeneous materials. Comput Mech 27:37–48
17. Tappan A (2013) There's plenty of room in the middle—microenergetics, the mesoscale, and interfaces. Propel Explos Pyrotech 38:475
18. Frenkel D, Smit B (2002) Understanding molecular simulation, 2nd edn. Academic Press, San Diego, CA
19. Belytschko T, Liu WK, Moran B, Elkhodary KI (2014) Nonlinear finite elements for continua and structures. Wiley Ltd., The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom
20. Tadmor EB, Miller RE (2011) Modeling materials: continuum, atomistics, and multiscale techniques. Cambridge University Press, University Press, Cambridge
21. Farah K, Muller-Plathe F, Bohm MC (2012) Classical reactive molecular dynamics implementations: state of the art. ChemPhysChem 13(5):1127–1151. https://doi.org/10.1002/cphc.201100681
22. Chenoweth K, van Duin AC, Goddard III WA (2008) ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. J Phys Chem A 112:1040–1053
23. van Duin AC, Dasgupta S, Lorant F, Goddard WA (2001) ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 105:9396–9409
24. Shan T-R, Wixom RR, Thompson AP (2016) Extended asymmetric hot region formation due to shockwave interactions following void collapse in shocked high explosive. Phys Rev B 94(5):054308. https://doi.org/10.1103/physrevb.94.054308
25. Monaghan JJ (1992) Smoothed particle hydrodynamics. Annual Rev Astronomy Astrophys 30:543–574

26. Müller-Plathe F (2002) Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back. ChemPhysChem 3:754–769
27. Tozzini V (2005) Coarse-grained models for proteins. Curr Opin Struct Biol 15(2):144–150
28. Peter C, Kremer K (2009) Multiscale simulation of soft matter systems—from the atomistic to the coarse-grained level and back. Soft Matter 5(22):4357. https://doi.org/10.1039/b912027k
29. Noid WG (2013) Perspective: coarse-grained models for biomolecular systems. J Chem Phys 139(9):090901. https://doi.org/10.1063/1.4818908
30. Reith D, Putz M, Müller-Plathe F (2003) Deriving effective mesoscale potentials from atomistic simulations. J Comput Chem 24:1624–1636
31. Pagonabarraga I, Hagen MHJ, Frenkel D (1998) Self-consistent dissipative particle dynamics algorithm. Europhys Lett 42:377–382
32. Pagonabarraga I, Frenkel D (2001) Dissipative particle dynamics for interacting systems. J Chem Phys 115(11):5015. https://doi.org/10.1063/1.1396848
33. Trofimov SY, Nies ELF, Michels MAJ (2002) Thermodynamic consistency in dissipative particle dynamics simulations of strongly nonideal liquids and liquid mixtures. J Chem Phys 117(20):9383. https://doi.org/10.1063/1.1515774
34. Schommers W (1973) A pair potential for liquid rubidium from the pair correlation function. Phys Lett A 43(2):157
35. Izvekov S, Voth GA (2005) A multiscale coarse-graining method for biomolecular systems. J Phys Chem B 109:2469–2473
36. Izvekov S, Voth GA (2005) Multiscale coarse graining of liquid-state systems. J Chem Phys 123(13):134105. https://doi.org/10.1063/1.2038787
37. Noid WG, Chu JW, Ayton GS, Krishna V, Izvekov S, Voth GA, Das A, Andersen HC (2008) The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. J Chem Phys 128(24):244114. https://doi.org/10.1063/1.2938860
38. Noid WG, Liu P, Wang Y, Chu JW, Ayton GS, Izvekov S, Andersen HC, Voth GA (2008) The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. J Chem Phys 128(24):244115. https://doi.org/10.1063/1.2938857
39. Izvekov S, Chung PW, Rice BM (2010) The multiscale coarse-graining method: assessing its accuracy and introducing density dependent coarse-grain potentials. J Chem Phys 133(6):064109. https://doi.org/10.1063/1.3464776
40. Lu L, Izvekov S, Das A, Andersen HC, Voth GA (2010) Efficient, regularized, and scalable algorithms for multiscale coarse-graining. J Chem Theory Comput 6:954–965
41. Izvekov S, Chung PW, Rice BM (2011) Particle-based multiscale coarse graining with density-dependent potentials: application to molecular crystals (hexahydro-1,3,5-trinitro-s-triazine). J Chem Phys 135(4):044112. https://doi.org/10.1063/1.3607603
42. Shell MS (2008) The relative entropy is fundamental to multiscale and inverse thermodynamic problems. J Chem Phys 129(14):144108. https://doi.org/10.1063/1.2992060
43. Das A, Andersen HC (2009) The multiscale coarse-graining method. III. A test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation. J Chem Phys 131(3):034102. https://doi.org/10.1063/1.3173812
44. Krishna V, Noid WG, Voth GA (2009) The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. J Chem Phys 131(2):024103. https://doi.org/10.1063/1.3167797
45. Das A, Andersen HC (2010) The multiscale coarse-graining method. V. Isothermal-isobaric ensemble. J Chem Phys 132(16):164106. https://doi.org/10.1063/1.3394862
46. Larini L, Lu L, Voth GA (2010) The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. J Chem Phys 132(16):164107. https://doi.org/10.1063/1.3394863
47. Lu L, Voth GA (2011) The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials. J Chem Phys 134(22):224107. https://doi.org/10.1063/1.3599049
48. Das A, Andersen HC (2012) The multiscale coarse-graining method. VIII. Multiresolution hierarchical basis functions and basis function selection in the construction of coarse-grained force fields. J Chem Phys 136(19):194113. https://doi.org/10.1063/1.4705384

49. Das A, Andersen HC (2012) The multiscale coarse-graining method. IX. A general method for construction of three body coarse-grained force fields. J Chem Phys 136(19):194114. https://doi.org/10.1063/1.4705417
50. Das A, Lu L, Andersen HC, Voth GA (2012) The multiscale coarse-graining method. X. Improved algorithms for constructing coarse-grained potentials for molecular systems. J Chem Phys 136(19):194115. https://doi.org/10.1063/1.4705420
51. Izvekov S (2017) Mori-Zwanzig theory for dissipative forces in coarse-grained dynamics in the Markov limit. Phys. Rev. E 95(1):013303. https://doi.org/10.1103/physreve.95.013303
52. Moore JD, Barnes BC, Izvekov S, Lísal M, Sellers MS, Taylor DE, Brennan JK (2016) A coarse-grain force field for RDX: density dependent and energy conserving. J Chem Phys 144(10):104501. https://doi.org/10.1063/1.4942520
53. Kinjo T, Hyodo SA (2007) Equation of motion for coarse-grained simulation based on microscopic description. Phys Rev E Stat Nonlin Soft Matter Phys 75(5 Pt 1):051109. https://doi.org/10.1103/physreve.75.051109
54. Darve E, Solomon J, Kia A (2009) Computing generalized Langevin equations and generalized Fokker-Planck equations. Proc Natl Acad Sci USA 106:10884–10889
55. Hijon CEP, Vanden-Eijnden E, Delgado-Buscalioni R (2010) Mori-Zwanzig formalism as a practical computational tool. Faraday Discuss 144:301–322
56. Hoogerbrugge PJ, Koelman JMVA (1992) Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhys Lett (EPL) 19(3):155–160
57. Koelman JMVA, Hoogerbrugge PJ (1993) Dynamic simulations of hard-sphere suspensions under steady shear. Europhys Lett 21:363–368
58. Hijon C, Serrano M, Espanol P (2006) Markovian approximation in a coarse-grained description of atomic systems. J Chem Phys 125(20):204101. https://doi.org/10.1063/1.2390701
59. Trement S, Schnell B, Petitjean L, Couty M, Rousseau B (2014) Conservative and dissipative force field for simulation of coarse-grained alkane molecules: a bottom-up approach. J Chem Phys 140(13):134113. https://doi.org/10.1063/1.4870394
60. Izvekov S, Rice BM (2014) Multi-scale coarse-graining of non-conservative interactions in molecular liquids. J Chem Phys 140(10):104104. https://doi.org/10.1063/1.4866142
61. Louis AA (2002) Beware of density dependent pair potentials. J Phys Condens Matter 14:9187–9206
62. Warren PB (2003) Vapor-liquid coexistence in many-body dissipative particle dynamics. Phys Rev E Stat Nonlin Soft Matter Phys 68(6 Pt 2):066702. https://doi.org/10.1103/physreve.68.066702
63. Izvekov S, Rice BM (2015) On the importance of shear dissipative forces in coarse-grained dynamics of molecular liquids. Phys Chem Chem Phys 17(16):10795–10804. https://doi.org/10.1039/c4cp06116k
64. Smith GD, Bharadwaj RK (1999) Quantum chemistry based force field for simulations of HMX. J Phys Chem B 103:3570–3575
65. Choi CS, Prince E (1972) The crystal structure of cyelotrimethylene-trinitramine. Acta Crystallogr B 28:2857–2862
66. Munday LB, Chung PW, Rice BM, Solares SD (2011) Simulations of high-pressure phases in RDX. J Phys Chem B 115(15):4378–4386. https://doi.org/10.1021/jp112042a
67. Cady HH (1972) Coefficient of thermal expansion of pentaerythritol tetranitrate and hexahydro-l,3,5-trinitro-s-triazine (RDX). J Chem Eng Data 17:369–371
68. Podeszwa R, Rice BM, Szalewicz K (2009) Crystal structure prediction for cyclotrimethylene trinitramine (RDX) from first principles. Phys Chem Chem Phys 11(26):5512–5518. https://doi.org/10.1039/b902015b
69. Olinger BRB, Cady HH (1978) The linear and volume compression of B-HMX and RDX to 9 Gpa (90 Kilobar). In: Symposium International Sur Le Comportement Des Milieux Denses Sous Hautes Pressions Dynamiques, Commmissariat à l'Energie Atomique Centre d'Etudes de Vaujours Paris. France, Paris, France, pp 3–8
70. Haycraft JJ, Stevens LL, Eckhardt CJ (2006) The elastic constants and related properties of the energetic material cyclotrimethylene trinitramine (RDX) determined by Brillouin scattering. J Chem Phys 124(2):024712. https://doi.org/10.1063/1.2141958

71. Hall PG (1971) Thermal decomposition and phase transitions in solid nitramines. Trans Faraday Soc 67(578):556
72. Sellers MS, Lísal M, Brennan JK (2016) Free-energy calculations using classical molecular simulation: application to the determination of the melting point and chemical potential of a flexible RDX model. Phys Chem Chem Phys 18(11):7841–7850. https://doi.org/10.1039/c5cp06164d
73. Taylor DE (2014) Pressure dependent elastic constants of alpha and gamma cyclotrimethylene trinitramine: A quantum mechanical study. J Appl Phys 116(5):053513. https://doi.org/10.1063/1.4891999
74. Haussuhl S (2001) Elastic and thermoelastic properties of selected organic crystals: acenaphthene, trans-azobenzene, benzophenone, tolane, trans-stilbene, dibenzyl, diphenyl sulfone, 2,2'-biphenol, urea, melamine, hexogen, succinimide, pentaerythritol, urotropine, malonic acid, dimethyl malonic acid, maleic acid, hippuric acid, aluminium acetylacetonate, iron acetylacetonate, and tetraphenyl silicon. Zeitschrift für Kristallographie – Crystal Mater 216(6):339–353
75. Brennan JK, Lísal M, Moore JD, Izvekov S, Schweigert IV, Larentzos JP (2014) Coarse-grain model simulations of nonequilibrium dynamics in heterogeneous materials. J Phys Chem Lett 5(12):2144–2149. https://doi.org/10.1021/jz500756s
76. Chennamsetty N, Bock H, Lísal M, Brennan JK (2011) An introduction to coarse-graining approaches: linking atomistic and mesoscales. In: Adjiman C, Galindo A (eds) Process systems engineering, vol 6. Molecular systems engineering. WILEY-VCH Verlag GmbH & Co., KGaA, Weinheim
77. Shell MS (2016) Coarse-graining with the relative entropy. In: Rice SA, Dinner AR (eds) Advances in chemical physics, vol 161. Advances in chemical physics, pp 395–441
78. Groot RD, Stoyanov SD (2008) Mesoscopic model for colloidal particles, powders, and granular solids. Phys Rev E Stat, Nonlin Soft Matter Phys 78(5 Pt 1):051403. https://doi.org/10.1103/physreve.78.051403
79. Junghans C, Praprotnik M, Kremer K (2008) Transport properties controlled by a thermostat: an extended dissipative particle dynamics thermostat. Soft Matter 4(1):156–161. https://doi.org/10.1039/b713568h
80. Nielsen SO, Lopez CF, Srinivas G, Klein ML (2004) Coarse grain models and the computer simulation of soft materials. J Phys Condens Matter 16(15):R481–R512. https://doi.org/10.1088/0953-8984/16/15/r03
81. Depa P, Chen C, Maranas JK (2011) Why are coarse-grained force fields too fast? A look at dynamics of four coarse-grained polymers. J Chem Phys 134(1):014903. https://doi.org/10.1063/1.3513365
82. Brennan JK, Lísal M (2009) CECAM workshop: 'Dissipative particle dynamics: addressing deficiencies and establishing new frontiers' (16–18 July 2008, Lausanne, Switzerland). Mol Simul 35(9):766–769. https://doi.org/10.1080/08927020902902783
83. Venturoli M, Maddalenasperotto M, Kranenburg M, Smit B (2006) Mesoscopic models of biological membranes. Phys Rep 437(1–2):1–54. https://doi.org/10.1016/j.physrep.2006.07.006
84. Strachan A, Holian BL (2005) Energy exchange between mesoparticles and their internal degrees of freedom. Phys Rev Lett 94(1):014301. https://doi.org/10.1103/physrevlett.94.014301
85. Stoltz G (2006) A reduced model for shock and detonation waves. I. The Inert Case. Europhys Lett (EPL) 76(5):849–855. https://doi.org/10.1209/epl/i2006-10350-y
86. Maillet JB, Soulard L, Stoltz G (2007) A reduced model for shock and detonation waves. II. The Reactive Case. Europhys Lett (EPL) 78(6):68001. https://doi.org/10.1209/0295-5075/78/68001
87. Lynch K, Thompson A, Strachan A (2009) Coarse grain modeling of spall failure in molecular crystals: role of intra-molecular degrees of freedom. Modell Simul Mater Sci Eng 17(1):015007. https://doi.org/10.1088/0965-0393/17/1/015007

88. Zhou Y, Strachan A (2009) Thermal conduction in molecular materials using coarse grain dynamics: role of mass diffusion and quantum corrections for molecular dynamics simulations. J Chem Phys 131(23):234113. https://doi.org/10.1063/1.3272028
89. Brennan JK, Lísal M (2010) Coarse-grain models for metals: constant-pressure dissipative dynamics simulations. In: Proceedings of the 14th International Detonation Symposium. Office of Naval Research, ONR-351-10-185:1451-1459
90. Moore JD, Izvekov S, Lísal M, Brennan JK (2012) Particle based multiscale modeling of the dynamic response of RDX. AIP Conf Proc 1426:1237–1240. https://doi.org/10.1063/1.3686504
91. Maillet JB, Bourasseau E, Desbiens N, Vallverdu G, Stoltz G (2011) Mesoscopic simulations of shock-to-detonation transition in reactive liquid high explosive. EPL (Europhys Lett) 96(6):68007. https://doi.org/10.1209/0295-5075/96/68007
92. Lísal M, Brennan JK, Bonet Avalos J (2011) Dissipative particle dynamics at isothermal, isobaric, isoenergetic, and isoenthalpic conditions using Shardlow-like splitting algorithms. J Chem Phys 135(20):204105. https://doi.org/10.1063/1.3660209
93. Kroonblawd MP, Sewell TD, Maillet JB (2016) Characteristics of energy exchange between inter- and intramolecular degrees of freedom in crystalline 1,3,5-triamino-2,4,6-trinitrobenzene (TATB) with implications for coarse-grained simulations of shock waves in polyatomic molecular crystals. J Chem Phys 144(6):064501. https://doi.org/10.1063/1.4941332
94. Bonet Avalos J, Mackie AD (1997) Dissipative particle dynamics with energy conservation. Europhys Lett 40(2):141–146
95. Español P (1997) Dissipative particle dynamics with energy conservation. Europhys Lett 40(6):631–636
96. Mackie AD, Bonet Avalos J, Navas V (1999) Dissipative particle dynamics with energy conservation: modelling of heat flow. Phys Chem Chem Phys 1:2039–2049
97. Fogler HS (1992) Elements of Chemical Reaction Engineering. Prentice Hall, Englewood Cliffs, NJ
98. Lucy LB (1977) Numerical approach to testing of fission hypothesis. Astron J 82(12):1013–1024
99. Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB (2002) A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys Condensed Matter 14:783–802
100. Buckingham RA (1938) The classical equation of state of gaseous helium, neon and argon. Proc R Soc Lond A 168:264–283
101. Nikunen P, Karttunen M, Vattulainen I (2003) How would you integrate the equations of motion in dissipative particle dynamics simulations? Comput Phys Commun 153(3):407–423. https://doi.org/10.1016/s0010-4655(03)00202-9
102. Chaudhri A, Lukes JR (2010) Velocity and stress autocorrelation decay in isothermal dissipative particle dynamics. Phys Rev E Stat, Nonlin Soft Matter Phys 81(2 Pt 2):026707. https://doi.org/10.1103/physreve.81.026707
103. Vattulainen I, Karttunen M, Besold G, Polson JM (2002) Integration schemes for dissipative particle dynamics simulations: from softly interacting systems towards hybrid models. J Chem Phys 116(10):3967–3979. https://doi.org/10.1063/1.1450554
104. Larentzos JP, Brennan JK, Moore JD, Lísal M, Mattson WD (2014) Parallel implementation of isothermal and isoenergetic dissipative particle dynamics using shardlow-like splitting algorithms. Comput Phys Commun 185(7):1987–1998. https://doi.org/10.1016/j.cpc.2014.03.029
105. Homman AA, Maillet JB, Roussel J, Stoltz G (2016) New parallelizable schemes for integrating the dissipative particle dynamics with energy conservation. J Chem Phys 144(2):024112. https://doi.org/10.1063/1.4937797
106. Stoltz G (2017) Stable schemes for dissipative particle dynamics with conserved energy. J Comput Phys 340:451–469. https://doi.org/10.1016/j.jcp.2017.03.059

107. Shardlow T (2003) Splitting for dissipative particle dynamics. SIAM J Sci Comput 24(4):1267–1282
108. Plimpton SJ (1995) Fast parallel algorithms for short-range molecular dynamics. J Comput Phys 117:1–19
109. Senftle TP, Hong S, Islam MM, Kylasa SB, Zheng Y, Shin YK, Junkermeier C, Engel-Herbert R, Janik MJ, Aktulga HM, Verstraelen T, Grama A, van Duin ACT (2016) The ReaxFF reactive force-field: development, applications and future directions. npj Comput Mater 2:15011. https://doi.org/10.1038/npjcompumats.2015.11
110. Shan TR, van Duin AC, Thompson AP (2014) Development of a ReaxFF reactive force field for ammonium nitrate and application to shock compression and thermal decomposition. J Phys Chem A 118(8):1469–1478. https://doi.org/10.1021/jp408397n
111. Mathew N, Picu RC (2011) Molecular conformational stability in cyclotrimethylene trinitramine crystals. J Chem Phys 135(2):024510. https://doi.org/10.1063/1.3609769
112. Broadbelt LJ, Pfaendtner J (2005) Lexicography of kinetic modeling of complex reaction networks. AlChE J 51(8):2112–2121. https://doi.org/10.1002/aic.10599
113. Klippenstein SJ (2017) From theoretical reaction dynamics to chemical modeling of combustion. Proc Combust Inst 36(1):77–111. https://doi.org/10.1016/j.proci.2016.07.100
114. Yetter Rad FL, Allen MT, Gatto JL (1995) Development of gas-phase reaction mechanisms for nitramine combustion. J Propul Power 11(4):683–697
115. Kumbhakarna N, Thynell ST, Chowdhury A, Lin P (2011) Analysis of RDX-TAGzT pseudo-propellant combustion with detailed chemical kinetics. Combust Theor Model 15(6):933–956. https://doi.org/10.1080/13647830.2011.591503
116. Taylor DE, Rice BM (2014) Quantum-informed multiscale M&S for energetic materials. In: Sabin JR (ed) Advances in quantum chemistry: energetic materials, vol 69. Academic Press. Cambridge, MA, pp 171–204
117. Manaa MR, Fried LE (2014) The reactivity of energetic materials under high pressure and temperature. In: Sabin JR (ed) Advances in quantum chemistry: energetic materials, vol 69. Academic Press, Cambridge, MA, pp 221–252. https://doi.org/10.1016/b978-0-12-800345-9.00006-4
118. Reed EJ, Riad Manaa M, Fried LE, Glaesemann KR, Joannopoulos JD (2007) A transient semimetallic layer in detonating nitromethane. Nat Phys 4(1):72–76. https://doi.org/10.1038/nphys806
119. Manaa MR, Reed EJ, Fried LE, Goldman N (2009) Nitrogen-rich heterocycles as reactivity retardants in shocked insensitive explosives. J Am Chem Soc 131:5483–5487
120. Manaa MR, Fried LE, Melius CF, Elstner M, Frauenheim T (2002) Decomposition of HMX at extreme conditions: a molecular dynamics simulation. J Phys Chem A 106:9024–9029
121. Ge NN, Wei YK, Ji GF, Chen XR, Zhao F, Wei DQ (2012) Initial decomposition of the condensed-phase beta-HMX under shock waves: molecular dynamics simulations. J Phys Chem B 116(46):13696–13704. https://doi.org/10.1021/jp309120t
122. Zhu W, Huang H, Huang H, Xiao H (2012) Initial chemical events in shocked octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine: a new initiation decomposition mechanism. J Chem Phys 136(4):044516. https://doi.org/10.1063/1.3679384
123. Ge NN, Wei YK, Zhao F, Chen XR, Ji GF (2014) Pressure-induced metallization of condensed phase beta-HMX under shock loadings via molecular dynamics simulations in conjunction with multi-scale shock technique. J Mol Model 20(7):2350. https://doi.org/10.1007/s00894-014-2350-1
124. Reed EJ, Rodriguez AW, Manaa MR, Fried LE, Tarver CM (2012) Ultrafast detonation of hydrazoic acid (HN3). Phys Rev Lett 109(3):038301. https://doi.org/10.1103/physrevlett.109.038301
125. Manaa MR, Fried LE, Reed EJ (2003) Explosive chemistry: Simulating the chemistry of energetic materials at extreme conditions. J Comput Aided Mater Des 10(2):75–97. https://doi.org/10.1023/b:jcad.0000036812.64349.15
126. An Q, Liu W-G, Goddard WA, Cheng T, Zybin SV, Xiao H (2014) Initial steps of thermal decomposition of dihydroxylammonium 5,5′-bistetrazole-1,1′-diolate crystals from quantum mechanics. J Phys Chem C 118(46):27175–27181. https://doi.org/10.1021/jp509582x

127. Wu Q, Xiang D, Xiong G, Zhu W, Xiao H (2016) Coupling of temperature with pressure induced initial decomposition mechanisms of two molecular crystals: an ab initio molecular dynamics study. J Chem Sci 128(5):695–705. https://doi.org/10.1007/s12039-016-1068-2
128. Wu Q, Zhu W, Xiao H (2016) Cooperative effects of different temperatures and pressures on the initial and subsequent decomposition reactions of the nitrogen-rich energetic crystal 3,3′-dinitroamino-4,4′-azoxyfurazan. Phys Chem Chem Phys 18(10):7093–7099. https://doi.org/10.1039/c6cp00096g
129. Wu Q, Chen H, Xiong G, Zhu W, Xiao H (2015) Decomposition of a 1,3,5-triamino-2,4,6-trinitrobenzene crystal at decomposition temperature coupled with different pressures: an ab initio molecular dynamics study. J Phys Chem C 119(29):16500–16506. https://doi.org/10.1021/acs.jpcc.5b05041
130. Ye C-C, An Q, Cheng T, Zybin S, Naserifar S, Ju X-H, Goddard III WA (2015) Reaction mechanism from quantum molecular dynamics for the initial thermal decomposition of 2,4,6-triamino-1,3,5-triazine-1,3,5-trioxide (MTO) and 2,4,6-trinitro-1,3,5-triazine-1,3,5-trioxide (MTO3 N), promising green energetic materials. J Mater Chem A 3(22):12044–12050. https://doi.org/10.1039/c5ta02486b
131. Ye C-C, An Q, Goddard III WA, Cheng T, Liu W-G, Zybin SV, Ju X-H (2015) Initial decomposition reaction of di-tetrazine-tetroxide (DTTO) from quantum molecular dynamics: implications for a promising energetic material. J Mater Chem A 3(5):1972–1978. https://doi.org/10.1039/c4ta05676k
132. Ge NN, Wei YK, Song ZF, Chen XR, Ji GF, Zhao F, Wei DQ (2014) Anisotropic responses and initial decomposition of condensed-phase beta-HMX under shock loadings via molecular dynamics simulations in conjunction with multiscale shock technique. J Phys Chem B 118(29):8691–8699. https://doi.org/10.1021/jp502432g
133. He ZH, Chen J, Wu Q, Ji GF (2016) Special catalytic effects of intermediate-water for rapid shock initiation of beta-HMX. RSC Adv 6(95):93103–93110. https://doi.org/10.1039/c6ra21384g
134. Wu Q, Xiong G, Zhu W, Xiao H (2015) How does low temperature coupled with different pressures affect initiation mechanisms and subsequent decompositions in nitramine explosive HMX? Phys Chem Chem Phys 17(35):22823–22831. https://doi.org/10.1039/c5cp03257a
135. Xue X, Wen Y, Zhang C (2016) Early decay mechanism of shocked ε-CL-20: a molecular dynamics simulation study. J Phys Chem C 120(38):21169–21177. https://doi.org/10.1021/acs.jpcc.6b05228
136. He ZH, Chen J, Ji GF, Liu LM, Zhu WJ, Wu Q (2015) Dynamic responses and initial decomposition under shock loading: A DFTB calculation combined with MSST method for beta-HMX with molecular vacancy. J Phys Chem B 119(33):10673–10681. https://doi.org/10.1021/acs.jpcb.5b05081
137. Joshi KL, Chaudhuri S (2015) Reactive simulation of the chemistry behind the condensed-phase ignition of RDX from hot spots. Phys Chem Chem Phys 17(28):18790–18801. https://doi.org/10.1039/c5cp00950b
138. Furman D, Kosloff R, Zeiri Y (2016) Effects of nanoscale heterogeneities on the reactivity of shocked erythritol tetranitrate. J Phys Chem C 120(50):28886–28893. https://doi.org/10.1021/acs.jpcc.6b11543
139. Zhou T, Lou J, Zhang Y, Song H, Huang F (2016) Hot spot formation and chemical reaction initiation in shocked HMX crystals with nanovoids: a large-scale reactive molecular dynamics study. Phys Chem Chem Phys 18(26):17627–17645. https://doi.org/10.1039/c6cp02015a
140. Wen Y, Xue X, Long X, Zhang C (2016) Cluster evolution at early stages of 1,3,5-triamino-2,4,6-trinitrobenzene under various heating conditions: a molecular reactive force field study. J Phys Chem A 120(22):3929–3937. https://doi.org/10.1021/acs.jpca.6b03795
141. Yu Y, Chen S, Li X, Zhu J, Liang H, Zhang X, Shu Q (2016) Molecular dynamics simulations for 5,5′-bistetrazole-1,1′-diolate (TKX-50) and its PBXs. RSC Adv 6(24):20034–20041. https://doi.org/10.1039/c5ra27912g
142. Wood MA, Strachan A (2016) Nonequilibrium reaction kinetics in molecular solids. J Phys Chem C 120(1):542–552. https://doi.org/10.1021/acs.jpcc.5b09820

143. Guo D, Zybin SV, An Q, Goddard III WA, Huang F (2016) Prediction of the Chapman-Jouguet chemical equilibrium state in a detonation wave from first principles based reactive molecular dynamics. Phys Chem Chem Phys 18(3):2015–2022. https://doi.org/10.1039/c5cp04516a
144. Wood MA, Cherukara MJ, Kober EM, Strachan A (2015) Ultrafast chemistry under nonequilibrium conditions and the shock to deflagration transition at the nanoscale. J Phys Chem C 119(38):22008–22015. https://doi.org/10.1021/acs.jpcc.5b05362
145. Xue X, Wen Y, Long X, Li J, Zhang C (2015) Influence of dislocations on the shock sensitivity of RDX: molecular dynamics simulations by reactive force field. J Phys Chem C 119(24):13735–13742. https://doi.org/10.1021/acs.jpcc.5b03298
146. Wen Y, Zhang C, Xue X, Long X (2015) Cluster evolution during the early stages of heating explosives and its relationship to sensitivity: a comparative study of TATB, beta-HMX and PETN by molecular reactive force field simulations. Phys Chem Chem Phys 17(18):12013–12022. https://doi.org/10.1039/c5cp00006h
147. Zhou TT, Lou JF, Song HJ, Huang FL (2015) Anisotropic shock sensitivity in a single crystal delta-cyclotetramethylene tetranitramine: a reactive molecular dynamics study. Phys Chem Chem Phys 17(12):7924–7935. https://doi.org/10.1039/c4cp05575f
148. Guo D, An Q, Zybin SV, Goddard III WA, Huang F, Tang B (2015) The co-crystal of TNT/CL-20 leads to decreased sensitivity toward thermal decomposition from first principles based reactive molecular dynamics. J Mater Chem A 3(10):5409–5419. https://doi.org/10.1039/c4ta06858k
149. Guo D, An Q, Goddard WA, Zybin SV, Huang F (2014) Compressive shear reactive molecular dynamics studies indicating that cocrystals of TNT/CL-20 decrease sensitivity. J Phys Chem C 118(51):30202–30208. https://doi.org/10.1021/jp5093527
150. Li Y, Kalia RK, Nakano A, K-I Nomura, Vashishta P (2014) Multistage reaction pathways in detonating high explosives. Appl Phys Lett 105(20):204103. https://doi.org/10.1063/1.4902128
151. Zhou T, Liu L, Goddard WA 3rd, Zybin SV, Huang F (2014) ReaxFF reactive molecular dynamics on silicon pentaerythritol tetranitrate crystal validates the mechanism for the colossal sensitivity. Phys Chem Chem Phys 16(43):23779–23791. https://doi.org/10.1039/c4cp03781b
152. An Q, Goddard WA, Zybin SV, Luo S-N (2014) Inhibition of hotspot formation in polymer bonded explosives using an interface matching low density polymer coating at the polymer-explosive interface. J Phys Chem C 118(34):19918–19928. https://doi.org/10.1021/jp506501r
153. Zhou T, Song H, Liu Y, Huang F (2014) Shock initiated thermal and chemical responses of HMX crystal from ReaxFF molecular dynamics simulation. Phys Chem Chem Phys 16(27):13914–13931. https://doi.org/10.1039/c4cp00890a
154. Wu CJ, Fried LE, Yang LH, Goldman N, Bastea S (2009) Catalytic behaviour of dense hot water. Nat Chem 1(1):57–62. https://doi.org/10.1038/nchem.130
155. Rice BM, Byrd EF (2016) Theoretical study of shocked formic acid: Born-Oppenheimer MD calculations of the shock Hugoniot and early-stage chemistry. J Phys Chem B 120(8):1711–1719. https://doi.org/10.1021/acs.jpcb.5b08845
156. van Duin AC, Zeiri Y, Dubnikova F, Kosloff R, Goddard III WA (2005) Atomistic-scale simulations of the initial chemical events in the thermal initiation of triacetonetriperoxide. J Am Chem Soc 127:11053–11062
157. Reed EJ, Fried LE, Joannopoulos JD (2003) A method for tractable dynamical studies of single and double shock compression. Phys Rev Lett 90(23):235503. https://doi.org/10.1103/physrevlett.90.235503
158. Reed EJF, Laurence E, Manaa, MR., Joannopoulos JD (2005) A multi-scale approach to molecular dynamics simulations of shock waves. In: Manaa MR (ed) Chemistry at extreme conditions. Elsevier B.V., Amsterdam, Neth, pp 297–325. https://doi.org/10.1016/b978-044451766-1/50010-x
159. Goldman N, Reed EJ, Fried LE (2009) Quantum mechanical corrections to simulated shock Hugoniot temperatures. J Chem Phys 131(20):204103. https://doi.org/10.1063/1.3262710

160. Qi T, Reed EJ (2012) Simulations of shocked methane including self-consistent semiclassical quantum nuclear effects. J Phys Chem A 116(42):10451–10459. https://doi.org/10.1021/jp308068c
161. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys Rev B 58:7260
162. Maillet JB, Bourasseau E (2009) ab initio simulations of thermodynamic and chemical properties of detonation product mixtures. J Chem Phys 131(8):084107. https://doi.org/10.1063/1.3179671
163. Cawkwell MJ, Niklasson AM, Dattelbaum DM (2015) Extended Lagrangian Born-Oppenheimer molecular dynamics simulations of the shock-induced chemistry of phenylacetylene. J Chem Phys 142(6):064512. https://doi.org/10.1063/1.4907909
164. Reed EJ (2012) Electron-ion coupling in shocked energetic materials. J Phys Chem C 116(3):2205–2211. https://doi.org/10.1021/jp206769c
165. Tarver CM, Forbes JW, Urtiew PA (2007) Nonequilibrium Zeldovich-von Neumann-Doring theory and reactive flow modeling of detonation. Russian J Phys Chem B 1(1):39–45. https://doi.org/10.1134/s1990793107010058
166. Reed DA, Dongarra J (2015) Exascale computing and big data. Commun ACM 58(7):56–68. https://doi.org/10.1145/2699414
167. Dongarra J, Beckman P, Moore T, Aerts P, Aloisio G, Andre JC, Barkai D, Berthou JY, Boku T, Braunschweig B, Cappello F, Chapman B, Xuebin C, Choudhary A, Dosanjh S, Dunning T, Fiore S, Geist A, Gropp B, Harrison R, Hereld M, Heroux M, Hoisie A, Hotta K, Zhong J, Ishikawa Y, Johnson F, Kale S, Kenway R, Keyes D, Kramer B, Labarta J, Lichnewsky A, Lippert T, Lucas B, Maccabe B, Matsuoka S, Messina P, Michielse P, Mohr B, Mueller MS, Nagel WE, Nakashima H, Papka ME, Reed D, Sato M, Seidel E, Shalf J, Skinner D, Snir M, Sterling T, Stevens R, Streitz F, Sugar B, Sumimoto S, Tang W, Taylor J, Thakur R, Trefethen A, Valero M, van der Steen A, Vetter J, Williams P, Wisniewski R, Yelick K (2011) The international exascale software project roadmap. Int J High Perform Comput Appl 25(1):3–60. https://doi.org/10.1177/1094342010391989
168. Geist A, Reed DA (2017) A survey of high-performance computing scaling challenges. Int J High Perf Comput Appl 31(1):104–113. https://doi.org/10.1177/1094342015597083
169. Li Z, Kermode JR, De Vita A (2015) Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. Phys Rev Lett 114(9):096405. https://doi.org/10.1103/physrevlett.114.096405
170. Kalidindi SR, De Graef M (2015) Materials data science: current status and future outlook. Annual Rev Mater Res 45(1):171–193. https://doi.org/10.1146/annurev-matsci-070214-020844
171. Matouš K, Geers MGD, Kouznetsova VG, Gillman A (2017) A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials. J Comput Phys 330:192–220. https://doi.org/10.1016/j.jcp.2016.10.070
172. Coveney PV, Dougherty ER, Highfield RR (2016) Big data need big theory too. Philos Trans A Math Phys Eng Sci 374(2080). https://doi.org/10.1098/rsta.2016.0153
173. Pham TL, Kino H, Terakura K, Miyake T, Dam HC (2016) Novel mixture model for the representation of potential energy surfaces. J Chem Phys 145(15):154103. https://doi.org/10.1063/1.4964318
174. Geiger P, Dellago C (2013) Neural networks for local structure detection in polymorphic systems. J Chem Phys 139(16):164105. https://doi.org/10.1063/1.4825111
175. Lee K, Joshi K, Chaudhuri S, Stewart DS (2016) Mirrored continuum and molecular scale simulations of the ignition of high-pressure phases of RDX. J Chem Phys 144(18):184111. https://doi.org/10.1063/1.4948548
176. Cawkwell MJ, Luscher DJ, Addessio FL, Ramos KJ (2016) Equations of state for the α and γ polymorphs of cyclotrimethylene trinitramine. J Appl Phys 119(18):185106. https://doi.org/10.1063/1.4948673

177. Sewell TD, Bennett CM (2000) Monte Carlo calculations of the elastic moduli and pressure-volume-temperature equation of state for hexahydro-1,3,5-trinitro-1,3,5-triazine. J Appl Phys 88(1):88
178. Larentzos JP, Rice BM (2017) Transferable reactive force fields: extensions of ReaxFF-lg to nitromethane. J Phys Chem A 121(9):2001–2013. https://doi.org/10.1021/acs.jpca.6b11761
179. Park HS, Karpov EG, Liu† WK, Klein PA (2005) The bridging scale for two-dimensional atomistic/continuum coupling. Philos Mag 85(1):79–113. https://doi.org/10.1080/14786430412331300163
180. Roehm D, Pavel RS, Barros K, Rouet-Leduc B, McPherson AL, Germann TC, Junghans C (2015) Distributed database kriging for adaptive sampling. Comput Phys Commun 192:138–147. https://doi.org/10.1016/j.cpc.2015.03.006
181. Miller RE, Tadmor EB (2009) A unified framework and performance benchmark of fourteen multiscale atomistic/continuum coupling methods. Model Simul Mater Sci Eng 17(5):053001. https://doi.org/10.1088/0965-0393/17/5/053001
182. Stjernschantz E, Marelius J, Medina C, Jacobsson M, Vermeulen NPE, Oostenbrink C (2006) Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. J Chem Inf Model 46:1972–1983
183. Tang Y-H, Kudo S, Bian X, Li Z, Karniadakis GE (2015) Multiscale universal interface: a concurrent framework for coupling heterogeneous solvers. J Comput Phys 297:13–31. https://doi.org/10.1016/j.jcp.2015.05.004
184. Kevrekidis IG, Samaey G (2009) Equation-free multiscale computation: algorithms and applications. Annual Rev Phys Chem 60:321–344. https://doi.org/10.1146/annurev.physchem.59.032607.093610
185. Bunder JE, Roberts AJ, Kevrekidis IG (2017) Good coupling for the multiscale patch scheme on systems with microscale heterogeneity. J Comput Phys 337:154–174. https://doi.org/10.1016/j.jcp.2017.02.004
186. Barton NR, Bernier JV, Knap J, Sunwoo AJ, Cerreta EK, Turner TJ (2011) A call to arms for task parallelism in multi-scale materials modeling. Int J Numer Meth Engng 86(6):744–764. https://doi.org/10.1002/nme.3071
187. Kouznetsova V, Geers MGD, Brekelmans WAM (2002) Multi-scale constitutive modelling of heterogeneous materials with a gradient-enhanced computational homogenization scheme. Int J Numer Meth Eng 54(8):1235–1260. https://doi.org/10.1002/nme.541
188. Özdemir I, Brekelmans WAM, Geers MGD (2008) Computational homogenization for the thermo-mechanical analysis of heterogeneous solids. Comput Methods Appl Mech Engrg 198(3–4):602–613. https://doi.org/10.1016/j.cma.2008.09.008
189. Weinan E, Engquist B, Li X, Ren W, Vanden-Eijnden E (2007) Heterogeneous multiscale methods: a review. Commun Comput Phys 2(3):367–450
190. Abdulle A, Weinan E, Engquist B, Vanden-Eijnden E (2012) The heterogeneous multiscale method. Acta Numer 21:1–87. https://doi.org/10.1017/s0962492912000025
191. Wescott BL, Stewart DS, Davis WC (2005) Equation of state and reaction rate for condensed-phase explosives. J Appl Phys 98(5):053514. https://doi.org/10.1063/1.2035310
192. Myint PC, McClelland MA, Nichols AL (2016) Application of the Peng–Robinson equation of state to energetic materials RDX and TNT: pure components, liquid mixtures, and solid mixtures. Industr Eng Chem Res 55(7):2252–2266. https://doi.org/10.1021/acs.iecr.5b04808
193. Cooper PW (1998) Introduction to detonation physics. In: Zukas JA (ed) Explosive effects and applications
194. Lee EL, Tarver CM (1980) Phenomenological model of shock initiation in heterogeneous explosives. Phys Fluids 23(12):2362. https://doi.org/10.1063/1.862940
195. Bastea S, Fried LE (2012) Chemical equilibrium detonation. In: Zhang F (ed) Shock wave science and technology reference library, vol 6. https://doi.org/10.1007/978-3-642-22967-1
196. Henson BF (2002) Ignition chemistry in HMX from thermal explosion to detonation, 620:1069–1072. https://doi.org/10.1063/1.1483723

197. Reaugh JE (2011) HERMES: a model to describe deformation, burning, explosion, and detonation. Technical Report LLNL-TR-516119, Lawrence Livermore National Laboratory
198. Rice BM (2017) A perspective on modeling the multiscale response of energetic materials. AIP Conf Proc 1793:020003. https://doi.org/10.1063/1.4971458
199. Nichols III AL (2007) ALE-3D user's manual. Technical report UCRL-MA-152204, Lawrence Livermore National Laboratory
200. Knap J, Barton NR, Hornung RD, Arsenlis A, Becker R, Jefferson DR (2008) Adaptive sampling in hierarchical simulation. Int J Numer Meth Eng 76(4):572–600. https://doi.org/10.1002/nme.2339
201. Knap J, Spear C, Leiter K, Becker R, Powell D (2016) A computational framework for scale-bridging in multi-scale simulations. Int J Numer Meth Engng 108:1649–1666. https://doi.org/10.1002/nme.5270
202. Schmidt MG, Ismail AE, Sauer RA (2015) A continuum mechanical surrogate model for atomic beam structures. J Multiscale Comp Engrg 13(5):413–442
203. Wirtz D, Karajan N, Haasdonk B (2015) Surrogate modeling of multiscale models using kernel methods. Int J Numer Meth Eng 101(1):1–28. https://doi.org/10.1002/nme.4767
204. Balachandran PV, Xue D, Theiler J, Hogden J, Lookman T (2016) Adaptive strategies for materials design using uncertainties. Sci Rep 6:19660. https://doi.org/10.1038/srep19660
205. Rouet-Leduc B, Barros K, Cieren E, Elango V, Junghans C, Lookman T, Mohd-Yusof J, Pavel RS, Rivera AY, Roehm D, McPherson AL, Germann TC (2014) Spatial adaptive sampling in multiscale simulation. Comput Phys Commun 185(7):1857–1864. https://doi.org/10.1016/j.cpc.2014.03.011
206. Behler J (2016) Perspective: machine learning potentials for atomistic simulations. J Chem Phys 145(17):170901. https://doi.org/10.1063/1.4966192
207. Ling J, Jones R, Templeton J (2016) Machine learning strategies for systems with invariance properties. J Comput Phys 318:22–35. https://doi.org/10.1016/j.jcp.2016.05.003
208. Pilania G, Gubernatis JE, Lookman T (2017) Multi-fidelity machine learning models for accurate bandgap predictions of solids. Comp Mater Sci 129:156–163. https://doi.org/10.1016/j.commatsci.2016.12.004
209. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA (2015) Big data meets quantum chemistry approximations: the delta-machine learning approach. J Chem Theory Comput 11(5):2087–2096. https://doi.org/10.1021/acs.jctc.5b00099
210. Barnes BC, Leiter KW, Becker R, Knap J, Brennan JK (2017) LAMMPS integrated materials engine (LIME) for efficient automation of particle-based simulations: application to equation of state generation. Modell Simul Mater Sci Eng 25(5):055006. https://doi.org/10.1088/1361-651x/aa6e36
211. Feng Y (2017) Python-mpi-bcast. https://github.com/rainwoodman/python-mpi-bcast. Accessed 22 Mar 2017
212. Fried LE (2007) The reactivity of energetic materials at extreme conditions. In: Lipkowitz KBC, Cundari TR (ed) Reviews in computational chemistry, vol 25. Wiley Inc., Hoboken, NJ, pp 159–189. https://doi.org/10.1002/9780470189078.ch4
213. Tarver CM (2006) Detonation reaction zones in condensed explosives. AIP Conf Proc 845:1026–1029. https://doi.org/10.1063/1.2263497
214. Johnson JN, Tang PK, Forest CA (1985) Shock-wave initiation of heterogeneous reactive solids. J Appl Phys 57(9):4323. https://doi.org/10.1063/1.334591
215. Rimoli JJ, Gürses E, Ortiz M (2010) Shock-induced subgrain microstructures as possible homogenous sources of hot spots and initiation sites in energetic polycrystals. Phys Rev B 81(1). https://doi.org/10.1103/physrevb.81.014112
216. An Q, Zybin SV, Goddard WA, Jaramillo-Botero A, Blanco M, Luo S-N (2011) Elucidation of the dynamics for hot-spot initiation at nonuniform interfaces of highly shocked materials. Phys Rev B 84(22). https://doi.org/10.1103/physrevb.84.220101
217. Todd SN, Caipen TL, Anderson MU, Lee BD (2011) Modeling damage induced initiation of explosives. Exp Mech 52(2):145–151. https://doi.org/10.1007/s11340-011-9533-9

218. Tsyshevsky RV, Sharia O, Kuklja MM (2016) Molecular theory of detonation initiation: insight from first principles modeling of the decomposition mechanisms of organic nitro energetic materials. Molecules 21(2). https://doi.org/10.3390/molecules21020236
219. Bdzil JB, Stewart DS (2007) The dynamics of detonation in explosive systems*. Annual Rev Fluid Mech 39(1):263–292. https://doi.org/10.1146/annurev.fluid.38.050304.092049
220. Baer MR, Gartling DK, DesJardin PE (2012) Probabilistic models for reactive behaviour in heterogeneous condensed phase media. Combust Theor Model 16(1):75–106. https://doi.org/10.1080/13647830.2011.606916
221. Ostoja-Starzewski M, Wang X (1999) Stochastic finite elements as a bridge between random material microstructure and global response. Comput Methods Appl Mech Eng 169:35–49
222. Ostoja-Starzewski M (2006) Material spatial randomness: from statistical to representative volume element. Probab Eng Mech 21(2):112–132. https://doi.org/10.1016/j.probengmech.2005.07.007
223. Yin X, Chen W, To A, McVeigh C, Liu WK (2008) Statistical volume element method for predicting microstructure–constitutive property relations. Comput Methods Appl Mech Engrg 197(43–44):3516–3529. https://doi.org/10.1016/j.cma.2008.01.008
224. Qidwai SM, Turner DM, Niezgoda SR, Lewis AC, Geltmacher AB, Rowenhorst DJ, Kalidindi SR (2012) Estimating the response of polycrystalline materials using sets of weighted statistical volume elements. Acta Mater 60(13–14):5284–5299. https://doi.org/10.1016/j.actamat.2012.06.026
225. Tripathy R, Bilionis I, Gonzalez M (2016) Gaussian processes with built-in dimensionality reduction: applications to high-dimensional uncertainty propagation. J Comput Phys 321:191–223. https://doi.org/10.1016/j.jcp.2016.05.039
226. Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16(3):199–231
227. Weirs VG, Kamm JR, Swiler LP, Tarantola S, Ratto M, Adams BM, Rider WJ, Eldred MS (2012) Sensitivity analysis techniques applied to a system of hyperbolic conservation laws. Reliab Eng Syst Safety 107:157–170. https://doi.org/10.1016/j.ress.2011.12.008
228. Sen O, Davis S, Jacobs G, Udaykumar HS (2015) Evaluation of convergence behavior of metamodeling techniques for bridging scales in multi-scale multimaterial simulation. J Comput Phys 294:585–604. https://doi.org/10.1016/j.jcp.2015.03.043
229. Bhattacharjee S, Matouš K (2016) A nonlinear manifold-based reduced order model for multiscale analysis of heterogeneous hyperelastic materials. J Comput Phys 313:635–653. https://doi.org/10.1016/j.jcp.2016.01.040

Index

N. Goldman (ed.), *Computational Approaches for Chemistry Under Extreme Conditions*, Challenges and Advances in Computational Chemistry and Physics 28, https://doi.org/10.1007/978-3-030-05600-1

Q

R

T

U