PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

PCCP

Reactive symbol sequences for a model of hydrogen combustion

Mohammad Alaghemandi[a] **and Jason R. Green**[*a,b]

Transient, macroscopic states of chemical disequilibrium are born out of the microscopic dynamics of molecules. As a reaction mixture evolves, the temporal patterns of chemical species encodes some of this dynamical information, while their statistics are a manifestation of the bulk kinetics. Here, we define a chemically-informed symbolic dynamics as a coarse-grained representation of classical molecular dynamics, and analyze the sequences of chemical species for a model of hydrogen combustion. We use reactive molecular dynamics simulations to generate the sequences and derive probability distributions for sequence observables: the reaction time scales and the chain length–the total number of reactions between initiation of a reactant and termination at products. The time scales and likelihood of the sequences depend strongly on the chain length, temperature, and density. Temperature suppresses the uncertainty in chain length for hydrogen sequences, but enhances the uncertainty in oxygen sequence chain lengths. This method of analyzing a surrogate chemical symbolic dynamics reduces the complexity of the chemistry from the atomistic to the molecular level and has the potential for extension to more complicated reaction systems.

1 Introduction

Away from equilibrium, simple theories for macroscopically observable phenomena may be hidden in the complicated dynamical motions of microscopic components of matter. Complex chemical reactions are a phenomenon where parameter-free theories remain elusive, in part, because the processes involved can be highly transient, dissipative, and irreversible.[1,2] Such behavior can obscure the essential microscopic dynamical details to include in a more macroscopic theory. Here we introduce a method for analyzing the information these dynamics encode on the statistical patterns of chemical species in reacting mixtures. In particular, we apply the method to combusting mixtures of hydrogen and oxygen. Through an analysis of the chemical symbol sequences produced during the reaction, we show this method is a route to extract information that is important to the search for concise macroscopic theories of complex chemistry away from equilibrium.

The nonequilibrium features of combustion chemistry have made it necessary to rely on chemical kinetic modeling for engineering applications.[3–9] Typical models of complex chemical kinetics can be massive, heavily-parameterized mathematical representations.[6,8] These models are extensively validated against experiments, and yet can still be inaccurate or incomplete, and subject to further reduction in some temperature or pressure regimes.[10] These kinetic models are the primary link between atomistic and bulk scale descriptions of chemistry, and so, are of significant importance in applications that include autoignition and the structure of detonations. From a fundamental perspective, these complex models use assumptions about partial equilibria or steady states to predict rate constants for the presumptive set of elementary reactions,[2] assumptions that intrinsically neglect molecular dynamics. The method we present here is a means of testing the validity of such assumptions and effectively coarse-graining the underlying microscopic dynamics.

Even within equilibrium assumptions, chemical kinetic models for combustion need the rate constants for all elementary reactions,[11,12] preferably derived from molecular properties. This fact presents a challenge because the complexity of the model explodes for fuels composed of molecules with increasing numbers of internal degrees of freedom. Hydrogen combustion, of interest here, involves eight chemical species and more than twenty elementary reactions,[11] but successful models for methane combustion involve around fifty chemical species and more than three-hundred elementary reactions.[13–15] An additional demand on these models is that the rate constants be predictable over a vast range of temperature and pressure, a direction in which there has been some recent success.[16] Both the high-dimensionality and nonequilibrium nature of complex chemistry are at the heart of current challenges to predictive theory with no free parameters to tune. In this paper, we use reactive molecular simulations to demonstrate an approach that is analogous to applied symbolic dynamics[17,18] in the field of dynamical systems theory. Specifically, we study the atomistic dynamics with a surrogate dynamics of the chemical species to quantify features that may be important in the search for parameter-free theory.

Molecular dynamics or classical trajectory simulations are an effective and efficient method to investigate the statistical behavior of molecular populations at and away from equilibrium.[19] Coupling classical molecular dynamics with the methods of dynamical systems theory has profited our understanding of small reactive molecular systems, in isolation,[21–25] and

[a] *Department of Chemistry, University of Massachusetts Boston, Boston, MA 02125.*
[b] *Department of Physics, University of Massachusetts Boston, Boston, MA 02125.*
[*] *Tel: 617-287-6136; E-mail: jason.green@umb.edu*

for high-dimensional, dissipative systems[26,27] in nonequilibrium environments.[28,29] Now it is possible to simulate mixtures undergoing complex chemical reactions with classical mechanics using reactive force fields.[20] It seems likely that new understanding of evolving reaction mixtures can also come from methods borrowed and adapted from dynamical systems theory. One possibility we explore here is symbolic dynamics, a method of coarse graining where one maps trajectories onto a surrogate symbolic dynamics through a partitioning of state space.[17] The dynamical symbol sequences that result encode, and can preserve, the key features of the nonlinear dynamical system of interest.[30]

Here we assume a force field captures the essential features of the chemistry that are necessary to formulate a statistical mechanical approach.[19,31] Rather than partitioning the high-dimensional phase space, we take a chemically-informed view, extracting chemical species from an ensemble of molecular dynamics trajectories. We apply the method to mixtures of hydrogen and oxygen that react and induce a transient dynamics of chemical symbol sequences composed of an alphabet of eight species. Our results show how the sequences of chemical symbols is a useful, but different, route to characterize the kinetics of complex chemical reactions. After presenting the details of our computational methods, we discuss the features of the overall reaction kinetics from simulation data. Then, we parse the chemical sequences extracted from the trajectories by their chain length and examine them over a wide range of temperature and density.

2 Methods and model

A combusting reaction mixture can generate a diverse set of chemical species. Representing a chemical species in the mixture at the time of a measurement as a symbol, x_i, the evolving mixture will produce a sequence of chemical symbols, $x_0 x_1, \ldots$. For a single mixture, many such sequences will result during the course of the overall reaction. This view motivates our use of the techniques of symbolic dynamics, some of which focus on quantifying statistical features of these symbol sequences (though the assignment of symbols to the underlying dynamics is done differently). To derive symbol sequences in computer simulations of detailed atomistic motions, one can track a particular atom and periodically ask to what chemical species it belongs. As the species undergoes transitions, a sequence of answers will result, for example $H_2 \rightarrow OH \rightarrow H_2O$. Here, we choose to only update symbol sequences when there is a reaction causing transitions between chemical species. This assumes purely irreversible transitions and a kinetic scheme in which there is no probability a chemical species will survive. Thus, a single classical trajectory generates a sample set of such sequences, $x_0 x_1, \ldots x_L$, one for each hydrogen (and oxygen), with variable length, L. We focus on the set of reactive sequences where $x_0 = H_2$ or O_2 and $x_L = H_2O$.

This symbolic dynamics is a step towards an alternative approach to the problems in combustion chemistry that stem from the high dimensionality of the classical phase space, rather than the quantum mechanical nature of the reaction mixture. In demonstrating the methodology here, we expand on simulation results for hydrogen combustion in the literature.[32,33] From each trajectory, we extract the sequences of chemical species by tracking each hydrogen and each oxygen. When identifying chemical species, we start from the last frame where the tracked atom was a part of H_2 (or O_2) and stop at the first frame where the tracked atom was a part of H_2O. Each atom in the reaction mixture generates its own symbol sequence. We group these sequences according to the element under observation. To accurately identify chemical species, and avoid including spurious atoms, we only accept species with a lifetime greater than 50 fs. Our primary aim here is to analyze the properties of these sequences. For example, the chain length–the total number of reactions between reactants and products.

Histories of chemical species are also the focus of other approaches in chemical kinetics. A recent example is the stochastic approach of Bai et al.[34] using a sum over histories representation for chemical kinetics, similar to Feynman's path integral formulation of quantum mechanics, to describe chemistry with reaction pathways. Chemical pathways are a time-ordered sequence of both chemical species and elementary reactions. Generating time-ordered sequences of the chemical species from an underlying dynamics, as we do here, is a complementary means of analyzing the overall reaction. For example, in the chlorination of methane, a two-step chemical pathway is $CH_4 \xrightarrow{R_1} CH_3 \xrightarrow{R_2} CH_3Cl$, and the corresponding chemical symbol sequence is $CH_4 \rightarrow CH_3 \rightarrow CH_3Cl$. Chemical symbol sequences are a further reduction of the description that does not distinguish between sequences brought about through different elementary reactions. This has the advantage that the analysis of the complex kinetics is independent of the notion of elementary reactions, which can breakdown in high-pressure environments and in the condensed phase.[35,36] Furthermore, in this work chemical symbols serve as a surrogate or reduced model for the deterministic classical dynamics.

Despite the accuracy of quantum mechanical methods to describe chemical reactions at the atomistic level, their computational expense is a limiting factor in their application to reaction mixtures. Instead, we simulate mixtures undergoing chemical reactions with a reactive force field, ReaxFF,[32,33,37]. Van Duin et al.[32] and Goddard III et al.[33] have used ReaxFF to simulate hydrogen combustion. Döntgen et al. used ReaxFF to determine both the reaction mechanism and rate constants of methane combustion,[38] using a method to identify and

quantify the rate coefficients of elementary reactions. We study the dynamics of the ReaxFF model of hydrogen oxidation with simulations of a stoichiometric mixture of H_2 and O_2. All molecular dynamics simulations were performed with the PuReMD-GPU simulation package[39] using the ReaxFF potential[32,33,37] and the parameters in Ref. 32.

Practical applications motivate an interest in combustion kinetics over a range of macroscopic parameters. In this study, we ran a set of 50 independent simulations, from unique initial conditions, at temperatures, T, from 2400 to 6800 K and densities, ρ, from 50 to 500 kg m^{-3} (the Supplementary Information, SI1, lists the exact temperatures and densities). To avoid simulating the time before an initiation reaction, we seeded the chemistry in each trajectory with a single OH radical. The cubic simulation box was also initially filled with 66 hydrogen and 33 oxygen molecules. The chosen densities at 3000 K correspond to initial pressures from 100 to 500 MPa; because we fixed the number of atoms, temperature, and volume, the pressure varied during the course of the reaction. For the Nosé-Hoover thermostat[40–42] we used a coupling time of 1 ps. Using this coupling time, the instantaneous simulation temperature was within 200 K of the target temperature at all times. In time steps of 0.1 fs, the simulation time ranged between 50 ps and 5 ns depending on the temperature and density. Simulations with density of 250 kg m^{-3} were run at temperatures below 4000 K for 2 ns, above 6000 K for 50 ps, and in between for 1 ns. Simulations with densities lower than 200 kg m^{-3} were run for 5 ns.

3 Results and discussion

3.1 Time evolution of chemical species

Before discussing the statistics of chemical symbol sequences, let us first discuss the overall hydrogen combustion reaction: 2 H_2 + O_2→ 2 H_2O. Fig. 1 shows simulation results for the number of reactant and product species versus time at 3000 K and 250 kg m^{-3}. Both reactant populations, H_2 and O_2, decay monotonically from their initial stoichiometric amounts as the reaction progresses. The reaction product, H_2O, monotonically increases in abundance before reaching a plateau that indicates the completion of the overall reaction. From these data, the average numbers from 50 parallel simulations are smooth functions of time. In what follows, we only show the average value of our simulation results.

The overall reaction hides the "complexity" of the underlying chemistry, which involves eight species: H_2, O_2, H, O, HO_2, OH, H_2O_2, and H_2O. As a reaction mixture initially composed of reactants evolves in time, it passes through compositions that can include five meta- or un-stable intermediates on the way to a product-dominated composition. We find in our results that the average number of OH, H, and O radicals are below one at all times during the reaction, while the mean number of HOOH and HO_2 increase slightly above one before tending towards zero; see Supplementary Information, SI2. Similar trends are seen with other simulations.[32,43]

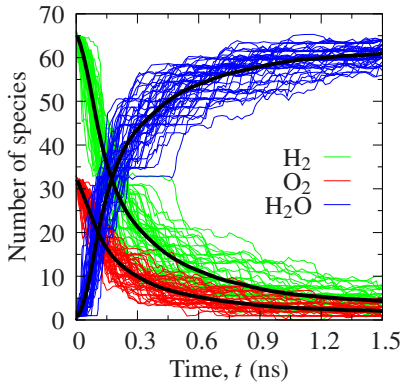

Fig. 1 The number of reactants and product versus simulation time, in nanoseconds, during the combustion of hydrogen. Data are from 50 independent simulations of a stoichiometric mixture (66 H_2, 33 O_2, 1 OH) at a temperature of $T = 3000$ K and a density of $\rho = 250$ kg m^{-3}. Simulation data at 21 temperatures and 11 densities show roughly the same trends, but on different time scales. Black solid lines are averages over trajectories.

Temperature and density have a strong influence on the reactivity of H_2/O_2 mixtures. Increasing either reduces the time that is necessary to reach the maximum number of water molecules. Fig. 2(a) shows the time to produce the maximum amount of H_2O over the range of temperatures above the flame ignition temperature, ≈ 800 K.[44,45] Note the maximum in the number of water molecules is temperature dependent, Fig. 2(b). Fig. 2(c) shows this reaction time over a range of densities. The reaction time decreases monotonically, and apparently exponentially, upon increasing both temperature and density. As we will show in the next section, shorter overall reaction time is a consequence of the "faster" completion of sequences producing water.

In chain reactions, the sum of the chain carriers cancel out in the overall reaction stoichiometry. However, this is not necessarily the case in these simulations because a pool of radicals can persist at sufficiently high temperatures. Fig. 2 shows the reduction in the time to reach 90% of the maximum water molecules produced at each temperature. There is more statistical uncertainty in our data at lower temperature and density, which nevertheless shows that at temperatures below 4000 K, there is more than 90% conversion to water. At temperatures above 4000 K, water dissociation lowers the water yield. The intermediate species are continually renewed by the dissociation of water; see Supplementary Information, SI2. Subtracting the intermediates from the maximum possible water molecules, 66, necessarily accounts for the lower water production at higher temperatures in Fig. 2(b).

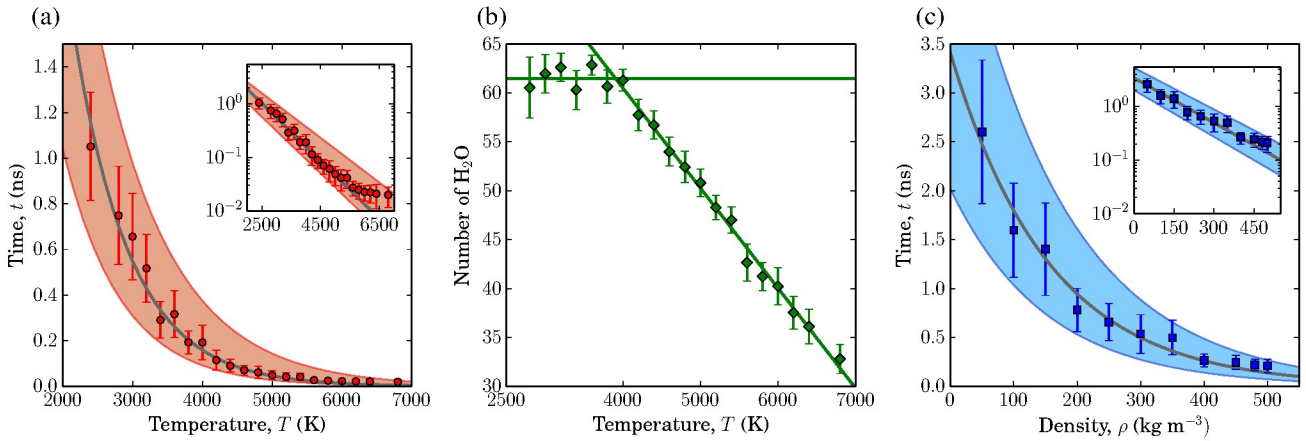

Fig. 2 Increasing temperature and density reduces the time to complete the overall reaction, but also the net amount of H_2O. (a) Mean time to reach 90% of the maximum possible number of water molecules as a function of temperature at a mass density of 250 kg m^{-3}. The inset is a log-linear plot of the same data demonstrating the roughly exponential dependence of the reaction time on temperature. Data is well fit to $e^{-\alpha T+\beta}$ with $\alpha = 0.001$ and $\beta = 3.120$, shown as a solid grey line. (b) The number of waters upon "completion" is independent of temperature below 4000 K. Above 4000 K, water dissociates readily, which maintains a persistent pool of radicals. Not shown is that the number of waters produced at 3000 K is independent of density within the uncertainty of our data. (c) Mean time to reach 90% the maximum number of waters versus density at a temperature of 3000 K. Best fit line is shown as a solid grey line via $e^{-\gamma\rho+\varepsilon}$ with $\gamma = 0.005$ and $\varepsilon = 1.029$.

3.2 Sequences connecting reactants and products

An important feature of combustion reactions is the chain length, L, the total number of reactions between initiation of a reactant and termination at products.[46] For all classical trajectories, we extract chemical symbol sequences with varying chain lengths, but only registering chemical species, or symbols, when a tracked atom transitions to another species; we also measure the simulation time at which the transition occurs and the waiting times between transitions. From each trajectory, there are roughly 169 hydrogen and 87 oxygen sequences (one per atom). Each sequence initiates at a reactant, but can pass through intermediates before terminating at the product. We call these sequences "reactive" to distinguish them from those that do not have these constraints on the end points. Even considering the constraints of the possible elementary reactions, there are many combinations of chemical species in this set of possible sequences. Our aim is to understand and simplify these complex combinations of reactive sequences as sampled directly from reactive molecular dynamics simulations. The chain length is a prominent feature.

From our sequence extraction procedure, based on tracking atoms, we identify separate sequences for hydrogen and oxygen. A state diagram is one representation of these sequences, taking chemical species as nodes and observed transitions as vertices (weighted according to transition probabilities or branching percentages). The state diagrams in Fig. 3 are for (a) hydrogen and (b) oxygen sequences at 3000 K and 250 kg m^{-3}. The vertex labels are the branching percentages, which are the averaged result of the 50 simulations at this temperature and density. The diagram in Fig. 3(a) reflects that for a tracked hydrogen atom, a sequence initiates when the last H_2 is observed in the classical trajectory. The sequence terminates when the first H_2O is observed, i.e., the reactive sequences neglect transitions back to H_2 (O_2) – they are starting states – or from H_2O – it is an absorbing state.

For each reactive chemical symbol sequence $x_0 x_1, \ldots x_L$ of length L, we estimate the probability $p(x_0, \ldots, x_L)$. The prob-

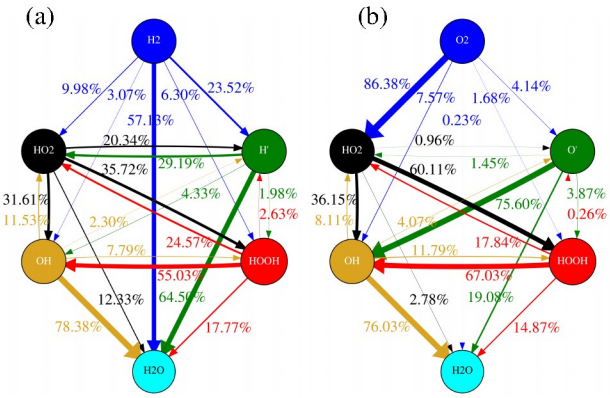

Fig. 3 State diagram for (a) hydrogen and (b) oxygen sequences at 3000 K and 250 kg m^{-3}. Arrows indicate the species-to-species transitions; labels and arrow widths indicate the observed branching percentages.

ability a sequence terminates after L reactions is

$$P(L) = \sum_{x_0} \dots \sum_{x_L} p(x_0, \dots, x_L). \quad (1)$$

The sum runs over all sequences with constraints on the end points: sequences must initiate at a reactant, $x_0 = H_2$ or O_2, and terminate at water, $x_L = H_2O$, after a variable number of L reactions. To simplify notation we do not specify that the probability estimates depend on temperature or density. As Fig. 3 shows, the most probable fate of H_2 is a direct transition to water ($> 57\%$), a sequence of chain length one. The next most probable sequence is $H_2 \rightarrow H \rightarrow H_2O$ with a chain length of two. Other sequences of chain length two and longer have lower probability. Oxygen though, has a longer sequence as the most probable, $O_2 \rightarrow HO_2 \rightarrow HOOH \rightarrow OH \rightarrow H_2O$; see Fig. 3(b). Direct conversion of oxygen to water is a rare event (0.23%) compared to hydrogen sequences, which is a constraint of the known elementary reactions.[47]

To simplify the reactive sequences between H_2 or O_2 and water, we group them according to their chain length. Only sequences with a length of 9 or shorter have a significant probability. Fig. 4 shows the probability, $P(L)$, of all sequences with a chain length less than four. The probabilities of longer sequences up to $L = 9$ are included in the Supplementary Information (SI3). Note that each $L > 1$ consists of different sequences with identical length. For example, both $H_2 \rightarrow H \rightarrow H_2O$ and $H_2 \rightarrow OH \rightarrow H_2O$ are possible two-step sequences. A particular feature of interest is the most probable chain length. For hydrogen sequences the most probable chain length is one because the most probable sequence is direct conversion of hydrogen to water, as also seen in the state diagram. A chain length of one is the most probable for all of the temperatures and densities studied. The probabilities of other chain lengths are also insensitive to temperature and density for hydrogen sequences. At higher temperatures $P(2)$ and $P(3)$ slightly increase at the cost of reducing $P(4)$ and longer sequences (not shown here); see the top panel of Fig. 4(a). By increasing the density at 3000 K the probabilities of $L = 1$, 3, and 4 slightly increase, while $P(2)$ decreases (Fig. 4(b), top panel).

Oxygen sequences differ significantly from those of hydrogen. Sequences with chain length one have the lowest probability. They are nearly zero at 2400 K and slightly increase to 0.1 at 6800 K. Sequences with a length of four are the most probable at temperatures up to 3400 K, where sequences with a chain length of three begin to dominate. Increasing the temperature above 3400 K continues to suppress the likelihood of four-step sequences, favoring three and then two-step sequences. We see two-step sequences are the most probable above 5800 K; see the bottom panel of Fig. 4(a). Overall, for both hydrogen and oxygen sequences, increasing temperatures in the range of 2400 K and 6800 K causes a decrease in the number of transitions that are necessary to reach water.

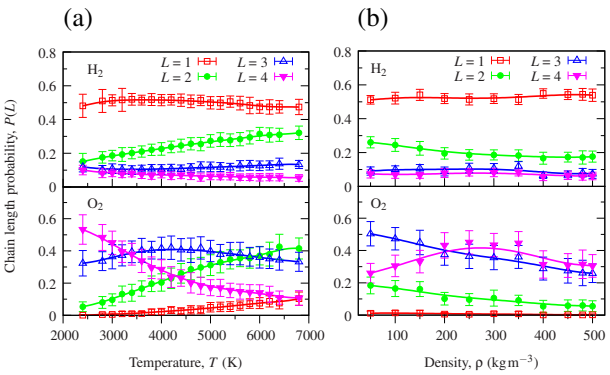

Fig. 4 Chain length probability, $P(L)$, of hydrogen (top) and oxygen (bottom) at (a) a range of temperature from 2400 to 6800 K and 250 kg m^{-3} and (b) 3000 K and densities from 50 to 500 kg m^{-3}.

The density also influences O_2 sequence lengths. The bottom panel of Fig. 4(b) shows our results for densities between 50 to 500 kg m^{-3} at 3000 K. From these data, we see $P(1)$ is close to zero for all densities, while $P(2)$ and $P(3)$ decrease with increasing density. The probability of four-step sequences is non-monotonic with density, passing through a maximum around 250 kg m^{-3}. This density marks the crossover of the most probable chain length from three to four. Of note in this figure is that the total probability of sequences with chain length less than four decreases. The probabilities of sequences with chain length greater than four increase by a corresponding amount. As shown in the Supplementary Information, SI3, this probability is mostly concentrated in five, six, and eight step sequences. A plausible interpretation of these results is that, at the highest densities, the diffusion rate of chemical species through the simulation volume is lower, and reactive collisions are more frequent because the reaction partners are more readily accessible in comparison to a more "dilute" environment, where species are more mobile. Consequently, it is reasonable to suspect longer sequences are more accessible in dense reaction mixtures on average.

The species-to-species transitions in a dynamically-generated reactive sequence can be caused by more than one elementary reaction. Furthermore, a single elementary reaction can simultaneously generate transitions for multiple tracked atoms. To illustrate this point, and given the common use of elementary reactions in combustion studies, we correlate the elementary reactions and chemical symbol sequences extracted from our trajectory calculations. While this correlation is possible for hydrogen oxidation, it will be a much more difficult task in the case of more complex fuels. The elementary reactions are identifiable at lower temperatures and densities by correlating the species undergoing a transition in successive time frames. All observed elementary reactions at

3000 and 6000 K and a density of 250 kg m^{-3} are tabulated in SI4 and SI5, respectively. From this complete set of observed elementary reactions, the reactions of interest are those associated with the most probable chain lengths for hydrogen and oxygen sequences, shown in Fig. 4. For hydrogen sequences, the most probable chain length is one, which corresponds to a single one-step sequence, $H_2 \rightarrow H_2O$, and two observed elementary reactions at 3000 K and four at 6000 K. These reactions are shown in Table 1 and 2 at 3000 K and 6000 K, respectively. The most probable elementary reaction observed overall studied temperature range is $H_2 + OH \rightarrow H + H_2O$. This finding is consistent with available kinetic data[47], where this elementary reaction has the highest rate coefficient in the temperature range.

Table 1 Elementary reactions leading to the sequence $H_2 \xrightarrow{T1} H_2O$ at 3000 K and 250 kg m^{-3}. Ti and Rj indicate the transition and elementary reaction, respectively. The complete set of elementary reactions is in SI4.

(Ti, Rj)	Reaction	Probability
(T1, R1)	$H_2 + OH \rightarrow H + H_2O$	0.264
(T1, R9)	$H_2 + HO_2 \rightarrow OH + H_2O$	0.026

Table 2 Elementary reactions leading to the sequence $H_2 \xrightarrow{T1} H_2O$ at 6000 K and 250 kg m^{-3}. The complete set of elementary reactions is in SI5.

(Ti, Rj)	Reaction	Probability
(T1, R1)	$H_2 + OH \rightarrow H + H_2O$	0.227
(T1, R7)	$H_2 + O \rightarrow H_2O$	0.038
(T1, R13)	$H_2 + HO_2 \rightarrow OH + H_2O$	0.023
(T1, R15)	$H_2 + O_2 \rightarrow O + H_2O$	0.015

Table 3 Elementary reactions leading to the sequence $O_2 \xrightarrow{T1} HO_2 \xrightarrow{T2} HOOH \xrightarrow{T3} OH \xrightarrow{T4} H_2O$ at 3000 K and 250 kg m^{-3}.

(Ti, Rj)	Reaction	Probability
(T1,R2)	$H + O_2 \rightarrow HO_2$	0.193
(T2,R4)	$H + HO_2 \rightarrow HOOH$	0.077
(T2,R5)	$H_2 + HO_2 \rightarrow H + HOOH$	0.052
(T3,R3)	$HOOH \rightarrow 2OH$	0.084
(T4,R1)	$H_2 + OH \rightarrow H + H_2O$	0.264
(T4,R6)	$H + OH \rightarrow H_2O$	0.045

The most probable sequence for oxygen has a chain length of four at 3000 K and two at 6000 K. The elementary reactions associated with each transition in these oxygen sequences are in Table 3 and 4. At 3000 K, for example, Table 3 indicates sequential reactions leading to the transitions in the most probable oxygen sequence, $O_2 \rightarrow HO_2 \rightarrow HOOH \rightarrow OH \rightarrow H_2O$. Elementary reactions are labeled in order of their probability of occurrence; see SI4 and SI5. Note that in both of the most probable oxygen sequences, at 3000 and 6000 K, the last transition is $OH \rightarrow H_2O$. Just as in the case of hydrogen, this transition is mostly due to the reaction $H_2 + OH \rightarrow H + H_2O$.

Table 4 Elementary reactions leading to the sequence $O_2 \xrightarrow{T1} OH \xrightarrow{T2} H_2O$ at 6000 K and 250 kg m^{-3}.

(Ti, Rj)	Reaction	Probability
(T1, R4)	$H + O_2 \rightarrow O + OH$	0.053
(T2, R1)	$H_2 + OH \rightarrow H + H_2O$	0.227
(T2, R3)	$H + OH \rightarrow H_2O$	0.144

Fig. 4 indicates the probability of each chain length, but does not contain information about the time to complete the sequences. Sequences can initiate at any time, but for the moment, we take the time zero to be when the simulation time at which the first transition from hydrogen (or oxygen) to another species is observed. The result is that the duration of sequences with a chain length of one is zero. The average time, $\bar{\tau}$, to complete longer sequences is shown in Fig. 5 as a function of chain length for both hydrogen and oxygen initiated sequences. The general trend is that more time is necessary on average to complete sequences with longer chain lengths, i.e., longer sequences need more time to convert from a reactant to the product. Increasing the temperature or density reduces the mean time for each L, as we saw for the overall reaction. Temperature also reduces the fluctuations around the mean time for each chain length; see SI6. The time to complete hydrogen and oxygen sequences, once begun, show a similar trend, except that the average time is higher for longer oxygen sequences.

Thus far, we have separately shown the most probable sequence length and the corresponding average duration at each temperature and density. Now, we consider both temporal and probabilistic information together to identify how quickly the most probable sequences terminate in simulation time. For each realization of a particular sequence there are distinct times, relative to the simulation time zero, at which each transition occurs. We use the average transition times, τ', over these specific sequences to represent the time evolution. Fig. 6 shows the three most probable hydrogen and oxygen sequences at two select temperatures. The times of each transition are using τ', with the clock starting at the beginning of simulation. Also note that the times in Fig. 5 are averages over all sequences with a given L, but in Fig. 6 the times are averages over the realizations of a specific sequence. Direct transformation from H_2 to H_2O is the most probable hydrogen sequence at both temperatures, as we saw in Fig. 4. Surprisingly, while it is the most probable sequence, it is not the first to complete on average. The first sequence to finish, among the most probable, is $H_2 \rightarrow H_2O_2 \rightarrow OH \rightarrow H_2O$ at 3000 K, and $H_2 \rightarrow OH \rightarrow H_2O$ at 6000 K. In addition, notice that the time for reactants to reach the product at 6000 K is more than one order of magnitude smaller than at 3000 K, which is consistent with the overall reaction time shown in Fig. 2. At both temperatures, oxygen sequences are most likely to be longer

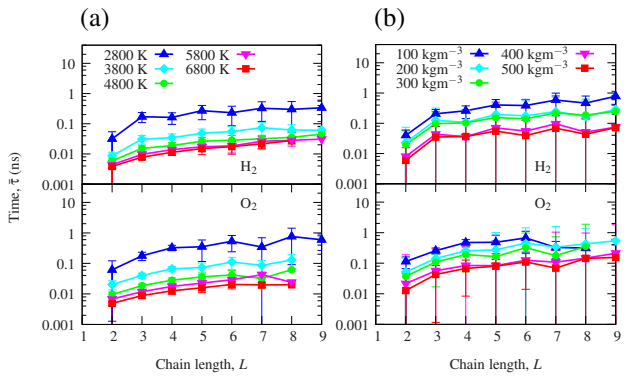

Fig. 5 Longer sequences take more time. Average time, $\overline{\tau}$, to complete a sequence for (top) hydrogen and (bottom) oxygen at (a) a range of temperature from 2800 to 6800 K and 250 kg m^{-3} and (b) 3000 K and densities from 50 to 500 kg m^{-3}.

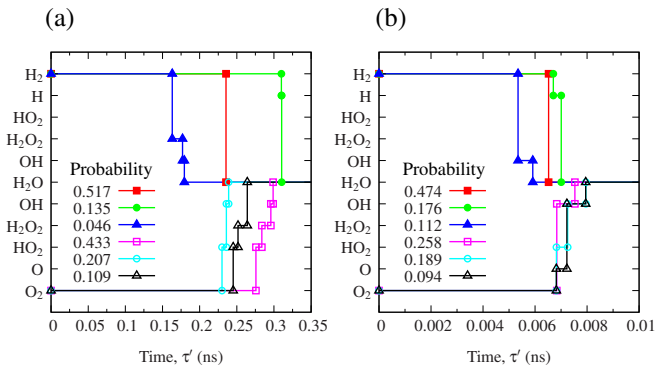

Fig. 6 The three most probable reactive hydrogen- and oxygen-initiated sequences averaged over time at (a) 3000 K and (b) 6000 K with a density of 250 kg m^{-3}. The most probable sequence is not necessarily the first to complete. The time τ' denotes the average time during the simulation at which the given transition was observed.

and more temporally compressed than hydrogen sequences.

The average uncertainty about the chain length is another feature that may be of interest for a macroscopic predictive theory that focuses on the set of reactive sequences. Generally, entropy quantifies the average degree of uncertainty associated with a probability distribution. Here we measure uncertainty in the number of reactions that will occur before chain termination, L, with the Shannon entropy

$$S_L(T) = -\sum_L P(L) \ln P(L). \qquad (2)$$

Fig. 7 shows the temperature dependence of the Shannon entropy at the density 250 kg m^{-3}. The chain length entropy $S_L(T)$ monotonically increases with temperature for O_2 sequences and monotonically decreases for H_2 sequences. The suppression of entropy for hydrogen-initiated sequences suggests the chain length becomes more certain with increasing temperature. In the context of Fig. 4 the decrease in uncertainty is the compression of the chain length distribution towards shorter sequences. The oxygen-initiated chain lengths, however, become less certain – the entropy increases. Even though the entropy increases, increasing temperature still has the net effect of decreasing the most probable chain length. Temperature consistently decreases the most probable sequence length. Of note in this data are the difference in statistics for oxygen sequences compared to those of hydrogen; this is a result of stoichiometry. For a stoichiometric mixture of H_2 and O_2, there are half as many oxygen atoms contributing to the estimate of the chain probabilities. Even within these statistical fluctuations, the trends are clear: raising the temperature decreases the uncertainty about the chain length for hydrogen sequences but increases our uncertainty about the chain length for oxygen sequences.

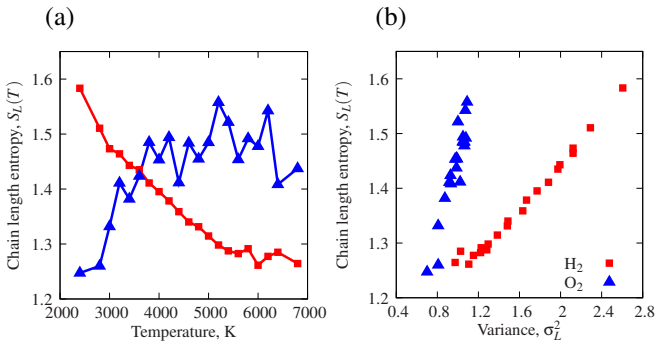

Fig. 7 Shannon entropy of chain length distribution, $S_L(T)$, as a function of (a) temperature for oxygen and hydrogen sequences and (b) the variance in the chain length probability distribution. Results are at a density of 250 kg m^{-3}. While temperature effectively shortens the most probable chain length, it can enhance or suppress the fluctuations in the chain length that determine the average uncertainty.

4 Summary

Dynamically generated chemical symbol sequences are a coarse-grained description of temporally evolving reaction systems. This approach, with connections to dynamical systems and information theory, is a means of analyzing diverse sets of reaction sequences. Each sequence represents the dynamical passage from reactants to products, potentially through many intermediates. Together, the symbol sequences are an alternative representation of chemical kinetics to traditional models for sets of elementary reactions. The species-to-species transitions in sequences relate to elementary reactions, but the mapping between these representations is not unique.

Our data for the reactive sequences from a model of hydrogen combustion show that oxygen-initiated sequences typically have a longer chain length and a longer time scale than hydrogen sequences. Increasing temperature effectively shortens the most probable chain length for all sequence types, but a shorter most-probable chain length, however, does not mean our uncertainty about the chain length necessarily decreases on average. Rather, it is the effect of temperature on chain length fluctuations that controls the average uncertainty, and the temperature dependence of these fluctuations is different for hydrogen and oxygen sequences. These findings, based upon thousands of sample trajectories and tens of thousands of sequences over temperatures between 2400 and 6800 K and densities between 50 and 500 kg m^{-3}, demonstrate the utility of the method. In sum, chemically-informed symbolic dynamics reduces the description of complex chemistry and encodes details of molecular-scale events.

Acknowledgments

This material is based upon work supported by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant number W911NF-14-1-0359. We acknowledge the use of the supercomputing facilities managed by the Research Computing Group at the University of Massachusetts Boston as well as the University of Massachusetts Green High Performance Computing Cluster. We gratefully acknowledge Ananth Grama and Sudhir Kylasa for assistance with PuReMD-GPU. We also thank Richard West for comments on the manuscript.

References

1 V. V. Lissianski, V. M. Zamansky and W. C. Gardiner Jr., *Gas-Phase Combustion Chemistry*; Springer-Verlag, W. C. Gardiner Jr. Ed., 2nd ed., 2004.
2 I. Glassman and R. A. Yetter, *Combustion*; Fourth Edition, Elsevier Inc. U.S.A. 2008.
3 S. M. Sarathy, S. Vranckx, K. Yasunaga, M. Mehl, P. Osswald, W. K. Metcalfe, C. K. Westbrook, W. J. Pitz, K. Kohse-Hoinghaus, R. X. Fernandes and H. J. Curran, *Combust. Flame*, 2012, **159**, 2028.
4 V. R Katta, S. K. Aggarwal and W. M. Roquemore, *Fuel*, 2012, **93**, 339.
5 R. Zhao, H. Liu, X. J. Zhong, Z. J. Wang, H. Hu and J. R. Qiu, *Fuel Process. Technol.*, 2011, **92**, 939.
6 I. Iliuta, R. Tahoces, G. S. Patience, S. Rifflart and F. Luck, *AIChE J.*, 2010, **56**, 1063.
7 C. Liu and G. A. Karim, *Int. J. Hydrogen Energ.*, 2008, **33**, 3863.
8 Q. Tang, W. Zhao, M. Bockelie and R. O. Fox, *Combust. Theor. Model.*, 2007, **11**, 889.
9 J. A. Miller, R. J. Kee and C. K. Westbrook, *Annu. Rev. Phys. Chem.*, 1990, **41**, 345.
10 R. van de Vijver, N. M. Vandewiele, P. L. Bhoorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, H.-H. Carstensen, M. F. Reyniers, G. B. Marin, R. H. West and K. M. van Geem, *Int. J. Chem. Kin.*, 2015, **47**, 199.
11 K. Shimizu, A. Hibi, M. Koshi, Y. Morii and N. Tsuboi, *J. Propul. Power*, 2011, **27**, 383.
12 D, Lee and S. Hochgreb, *Int. J. Chem. Kinet.*, 1998, **30**, 385.
13 DA. D. e Bortoli, G. S. L Andreis and F. N. Pereira, *Modeling and Simulation of Reactive Flows*; Elsevier Inc. U.S.A., 2015.
14 S. van Loo and J. Koppejan, *The Handbook of Biomass Combustion and Co-firing*; Earthscan U.K. and U.S.A., 2008.
15 G. Stiesch, *Modeling Engine Spray and Combustion Processes*; Springer-Verlag Berlin Heidelberg, 2003.
16 Jasper, A. W.; Pelzer, K. M.; Miller, J. A.; Kamarchik, E.; Harding, l. B.; Klippenstein, S. J. *Science* **2014**, *346*, 1212.
17 H. Bailin, *Elementary Symbolic Dynamics And Chaos In Dissipative Systems*, World Scientific, 1989.
18 C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems: An Introduction* Cambridge University Press, 1985.
19 R. Hernandez and A. V. Popov, *WIREs Comput. Mol. Sci.*, 2014, **4**, 541.
20 K. Farah, F. Müller-Plathe and M. C. Böhm, *ChemPhysChem*, 2012, **13** 1127.
21 C. Cerjan and W. P. Reinhardt, *J. Chem. Phys.*, 1979, **71**, 1819.
22 R. Kosloff and S. A. Rice, *J. Chem. Phys.*, 1981, **74**, 1947.
23 D. J. Wales and R. S. Berry, *J. Phys. B: At. Mol. Opt. Phys.*, 1991, **24**, 351.
24 C. Amitrano and R. S. Berry, *Phys. Rev. Lett.*, 1992, **68**, 729.
25 J. R. Green, T. S. Hofer, R. S. Berry and D. J. Wales, *J. Chem. Phys.*, 2011, **135**, 184307.
26 H. A. Posch and W. G. Hoover, *Phys. Rev. A*, 1989, **39**, 2175.
27 J. R. Green, A. B. Costa, B. A. Grzybowski and I. Szleifer, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **110**, 16339.
28 G. T. Craven, T. Bartsch and R. Hernandez,*J. Chem. Phys.*, 2015, **142**, 074108.
29 S. Kawai and T. Komatsuzaki, *Phys. Chem. Chem. Phys.*, 2011, **13**, 21217.
30 J. P. Crutchfield and N. H. Packard, *Int. J. Theor. Phys.*, 1982, **21**, 433.

31 G. E. Uhlenbeck and G. W. Ford, *Lectures on Statistical Mechanics*, American Mathematical Society, 1963.

32 S. Agrawalla and A. C. T. van Duin, *J. Phys. Chem. A.*, 2011, **115**, 960.

33 T. Cheng, A. Jaramillo-Botero, W. A. Goddard III and H. Sun, *J. Am. Chem. Soc.*, 2014, **136**, 9434.

34 S. Bai, D. Zhou, M. J. Davis and R. T. Skodje, *J. Phys. Chem. Lett.*, 2015, **6**, 183.

35 S. W. Flynn, H. C. Zhao and J. R. Green, *J. Chem. Phys.*, 2014, **141**, 104107.

36 J. W. Nichols, S. W. Flynn and J. R. Green, *J. Chem. Phys.*, 2015, **142**, 064113.

37 K. Chenoweth, A. C. T. van Duin and W. A. Goddard III, *J. Phys. Chem. A.*, 2008, **112**, 1040.

38 M. Döntgen, M. D. Przybylski-Freund, L. C. Kröger, W. A. Kopp, A. E. Ismail and K. Leonhard, *JCTC*, 2015, **11**, 2517.

39 S. B. Kylasa, H. M. Aktulga and A. Y. Grama *J. Comput. Phys.*, 2014, **272**, 343.

40 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511.

41 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

42 M. P. Allen and T. J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press: Oxford. U.K., 1987.

43 D. G. Vlachos, *Chem. Eng. Sci.*, 1998, **53**, 157.

44 S. J. Klippenstein, V. S. Pande and D. J. Truhlar, *J. Am. Chem. Soc.*, 2014, **136**, 528.

45 R. W. Schefer, W. D. Kulatilaka, B. D. Patterson and T. B. Settersten, *Combust. Flame*, 2009, **156**, 1234.

46 S. W. Benson, *Ind. Eng. Chem.*, 1964, **56**, 18.

47 M. P. Burke, M. Chaos, Y. Ju, F. L. Dryer and S. J. Klippenstein, *Int. J. Chem. Kin.*, 2012, **44**, 444.