First principles reaction discovery: from the Schrödinger equation to experimental prediction for methane pyrolysis
1. Introduction

Reaction discovery in complex chemical systems has recently developed into a thriving research topic. A complex reaction network typically contains reactants, products, and all relevant intermediate species, connected by elementary reactions. Fundamental understanding of these reactions involves quantitative analysis of the potential energy surface (PES), estimation of the rates of reaction, and modeling of the evolution of the chemical system over time. The rapid growth in computational capabilities in the past decades opens the possibility for systematic exploration of complex chemical mechanisms.1–4

Many of these approaches are knowledge-based, as they assume chemical rules that guide the reaction discovery. Examples include methods using tabulated reaction templates (Netgen,5–7 EXGAS,8 RMG,9,10 and RING11,12), connectivity graphs (Zstruct13–15 and others16,17), and heuristics-rule-based reaction discovery (HAQC18–20 and Chemoton21–23). Some studies also applied machine learning tools,24–29 taking advantage of modern computational capabilities to generate large training datasets from selected chemical reaction rules24–26 or first principles.27,28

Besides the knowledge-based reaction discovery approach, another systematic method for unraveling complex chemical mechanisms is reactive molecular dynamics (MD) simulations of freely reacting molecules.30–40 Perhaps the best way to allow for arbitrary bond rearrangements in the reactive MD is by using ab initio molecular dynamics.41–42 where the electronic Schrodinger equation is solved as needed to compute interatomic forces. The reactive MD approach bypasses the need for human heuristics by allowing chemical reactions to occur freely without predefined reaction rules. However, direct MD simulations under realistic chemical time scales and conditions are challenging, since the computationally accessible time scale is still too short compared to that of a realistic chemical reaction.43 A few studies have attempted to estimate rates of reaction events from the occurrence frequency over the simulation period.44–47,49 but it becomes very difficult to do this accurately because of the rarity of chemical events compared to molecular collisions.

A solution to the above issue is to decouple reaction rates from reaction discovery. In this strategy, it does not matter how reactions are discovered because the reaction rate is determined
independently (e.g., with transition state theory). One is therefore free to use any of the existing accelerated dynamics and enhanced sampling approaches, or even new ones which might not provide a direct connection to reaction rates. Our recently developed \textit{ab initio} nanoreactor\cite{30,31} introduces new acceleration techniques during \textit{ab initio} molecular dynamics (AIMD), including ones inspired by high-pressure\cite{39} and shockwave\cite{45,46} simulations. A virtual piston periodically pushes the molecules towards the reactor center, increasing the frequency of reactive events. Electronic structure calculations used to determine the interatomic forces in the nanoreactor are accelerated by graphical processing units (GPU),\cite{54,55} enabling on-the-fly AIMD simulations with density functional theory (DFT) in small to medium-sized basis sets. In a step that we call “refinement,” the reactions discovered from AIMD simulations are used to find minimum energy paths (MEPs), which include the minimum-energy geometries of both the reactant and product along with a transition state structure. From these refined paths, a kinetic model including reaction rate information can be constructed. Overall, the \textit{ab initio} nanoreactor combines reaction discovery, path refinement, and kinetic modeling in a computationally affordable manner. We have previously demonstrated the capability of the \textit{ab initio} nanoreactor for uncovering chemical pathways in the Urey–Miller experiment,\cite{90} acetylene polymerization,\cite{91} nitromethane decomposition,\cite{92} and multistep chemistry of a Donor–Acceptor Stenhouse Adduct (DASA) photoswitch.\cite{93} We have also developed a photochemical version which explores reaction dynamics on excited electronic states.\cite{94} Other workers have applied nanoreactor-like approaches to prebiotic formation of RNA precursors,\cite{95} terpene formation in nanocapsules,\cite{96} soot formation,\cite{97} and high-temperature graphene formation,\cite{98} highlighting its potential for reaction discovery across a broad range of chemistry.\cite{99,100} These previous studies demonstrate the nanoreactor’s advantage for systematic and efficient exploration of complex chemical systems. Most of the earlier studies focused on reaction discovery and path refinement, and only limited attempts were made to build kinetic models.\cite{101} In the present contribution, we build a reaction network entirely from nanoreactor discovery and refinement and then validate this against experimental data and other literature kinetic models.

In this work, we apply the \textit{ab initio} nanoreactor framework to methane combustion, specifically focusing on constructing and validating a chemical kinetic model for methane pyrolysis. Methane is the simplest hydrocarbon fuel for combustion. It is also an important practical fuel as it is the primary component (>85 molar%) in natural gas.\cite{79} Experimental and theoretical studies on methane combustion have been evolving for about half a century.\cite{71-78} One of the milestones is the development of GRI-Mech, a chemical kinetic model for methane\cite{79} (and later versions for natural gas\cite{80}) combustion. Recently, there has been a revival of interest in methane and liquified natural gas (LNG) as potential aerospace propellants due to their higher specific-impulse and less-concerning environmental effects compared to some of the conventional kerosene-based fuels.\cite{81,82} New research studies have emerged for experimenting with methane combustion at a higher-pressure, and/or under engine-relevant conditions.\cite{83-86} As a subset of methane combustion chemistry, the pyrolytic reaction process is critical to predicting methane’s combustion behavior and sooting propensity at fuel-rich conditions. Due to its importance, methane pyrolysis has also been extensively studied for many decades. Most of the studies are experimental efforts\cite{87-90} while theoretical studies in the literature are rare.\cite{96-97} Beyond the above reasons, we choose methane pyrolysis chemistry for its simplicity to test the validity of the nanoreactor methodology. This paper aims to achieve two major objectives. First, the accuracy of the nanoreactor-constructed kinetic model is validated against experimental data and modeling results from literature kinetic models. Second, we explore the suitability of sensitivity analysis and its potential for efficient identification of key reactions from a complex reaction network.

The paper is constructed as follows. Section 2 describes the nanoreactor framework and the methodology for reaction discovery, path refinement, rate coefficient calculation, kinetic modeling, and sensitivity analysis. Section 3 presents the results of the kinetic model validation and explores the applicability of the sensitivity analysis. Section 4 draws conclusions.

2. Methodology

The methane reaction network is automatically constructed by the \textit{ab initio} nanoreactor through a streamlined workflow, which is illustrated in Fig. 1. Briefly, the reaction events are discovered and extracted from accelerated \textit{ab initio} molecular dynamics simulations. The reaction paths detected during the discovery phase are further refined towards their corresponding minimum energy paths, including geometry optimization of reactants, products, and transition state configurations. Furthermore, the reaction rate parameters are calculated using transition state theory.\cite{102-105} The reaction kinetic model, consisting of all species and reactions discovered during the discovery/refinement phases (53 species and 134 reactions for the methane pyrolysis case studied here), is validated against experimental data and literature kinetic models. During the kinetic model validation, sensitivity analysis is performed to evaluate the relative importance of each reaction rate parameter. We describe the workflow in detail below.

2.1. Reaction discovery

Reaction discovery in the nanoreactor was explored in \textit{ab initio} molecular dynamics (AIMD) simulations using the unrestricted B3LYP density functional\cite{106,107} and a 3-21G basis set. All AIMD simulations are performed with GPU-accelerated energy and gradient calculations in the TeraChem electronic structure package.\cite{38,53,56,57,58,59} The simulations employed level shifting\cite{60} for open-shell states. Thermal smearing with fractional occupation number (FON) is enabled to accelerate self-consistent field (SCF) convergence.\cite{61} The electronic temperature for FON thermal smearing was set to 3160 K. The AIMD simulations started with 8 methane (CH\textsubscript{4}) molecules and 3 hydrogen (H) atoms as initial reactants with a spherical configuration (r = 4.2 Å) set by the Packmol program.\cite{108} Addition of the 3 H atoms can
speed up reaction discovery, since hydrogen atoms are highly reactive free radicals and also the product of the initiation step of CH₄ decomposition. The reason for the initial system size choice (i.e., 8 CH₄ molecules with 3 H atoms) is provided later. The equations of motion were integrated using Langevin dynamics with a friction coefficient of 41 ps⁻¹ and the velocity Verlet integrator at thermostat (equilibrium) temperatures (T₀) of 1500, 2000, 2500, 3000, and 3500 K, respectively (see Section S3 and Table S3† for details of all discovery runs). For each thermostat temperature condition, 10 discovery simulations covering 20 ps (with a 1 fs time step) were performed. This corresponds to a total of 1 ns discovery simulation time. In each simulation, a periodic piston compression potential was added to the ab initio potential energy surface in order to accelerate the reaction discovery:

\[ V_{\text{piston}}(\vec{r}, t) = f(t) U(\vec{r}, r_{\text{outer}}, k_{\text{outer}}) + (1 - f(t)) U(\vec{r}, r_{\text{inner}}, k_{\text{inner}}) \]

(1)

\[ U(\vec{r}, r_{\text{inner}}, k_{\text{inner}}) = \sum_{i} \frac{m_i k_i}{2} (r_i - r_0)^2 \theta(r_i - r_0) \]

(2)

\[ f(t) = \theta\left(\frac{t}{T} - \frac{1}{T} + \frac{1}{T}\right) \]

(3)

where \( \vec{r} \) collects the distance of each atom from the simulation origin (center of the spherical piston), and the parameters are \( k_{\text{outer}} = 4.5 \text{ kcal mol}^{-1} \text{ Å}^{-2} \text{ amu}^{-1}, r_{\text{outer}} = 5.0 \ Å, k_{\text{inner}} = 11.2 \text{ kcal mol}^{-1} \text{ Å}^{-2} \text{ amu}^{-1}, r_{\text{inner}} = 3.4 \ Å, \tau = 300 \text{ fs}, T = 450 \text{ fs}. \) The choice of these parameters and the initial system size (e.g., number of reactant molecules) are jointly determined by computational experiments in trial AIMD runs where two requirements should be fulfilled. First, the system should contain enough reactant molecules (yet not too many because of computational cost considerations) along with a sufficiently aggressive piston compression potential to ensure that chemical reactions occur during the lowest thermostat temperature simulations (i.e., 1500 K). Second, the piston potential must not be so aggressive that it hampers improbable processes such as spontaneous C–H bond rupture. As long as these parameters are reasonably determined from these two requirements, the discovery results should not be ultrasonic to the parameters chosen. Furthermore, the function \( f \) is the Heaviside step function, and \( | \) is the floor function. The function \( U(\vec{r}, r_{\text{inner}}, k_{\text{inner}}) \) describes a radial half-harmonic potential which vanishes for atoms inside the spherical boundary defined by \( r_{\text{inner}} \) and the force constant \( k \) is multiplied by the atomic mass to ensure equal acceleration on atoms with the same distance from the origin. In eqn (1), the harmonic potentials are switched between \( U(\vec{r}, r_{\text{outer}}, k_{\text{outer}}) \) with duration \( \tau \) and \( U(\vec{r}, r_{\text{inner}}, k_{\text{inner}}) \) with duration \( (T - \tau) \). The atoms within a radial position \( r_{\text{inner}} < r < r_{\text{outer}} \) experience a compression force towards the center of the reactor, leading to an increase in collision events. During each compression, the instantaneous temperature of the system can be much higher than the equilibrium temperature \( T_{\text{eq,}r} \). After the compression, the system is quenched to the target \( T_{\text{eq,}r} \) and the molecules quickly diffuse back to fill the larger reactor volume. An example temperature plot can be found in Section S3 (Fig. S1†) for further demonstration.

The simulated AIMD trajectories were analyzed to parse out reaction events. Detailed procedures can be found in our earlier work.46,51,62 Briefly, molecules in the AIMD trajectories were represented as a connectivity graph data structure where nodes correspond to atoms and edges correspond to covalent bonds. Reaction events were detected by changes in the connectivity graph as a function of time. Bond formation between atoms occurs when the bond order is greater than 0.7 and bond length is shorter than 1.2 covalent radii, while bond breaking happens when the bond order is below 0.1 and bond length exceeds 2.5 covalent radii. The raw signals of the time-dependent changes of the connectivity graph can be complicated by the large-

© 2023 The Author(s). Published by the Royal Society of Chemistry. Chem. Sci., 2023, 14, 7447–7464 | 7449
amplitude vibrations within a molecule that introduce significant noise. To address this issue, we apply a two-state hidden Markov model (HMM). The noise is removed by estimating the most likely sequence of the hidden states, which should only contain the chemical reaction. The resulting reaction event consists of 40 frames of coordinates of reactants, products, and other geometry frames across the reaction event time interval. The initial reaction path is extracted from the AIMD trajectory and saved in a database for subsequent path refinement purposes. Charge and spin multiplicity values for the reaction path and its reactant and product species are computed and saved to the database. Details of the charge and spin multiplicity assignment is included in Section S4 in the ESI†

2.2. Path refinement

The discovered reaction paths are refined using the same unrestricted B3LYP density functional and a larger 6-31G** basis set. Although using DFT may lead to inaccuracies in the optimized reaction path (e.g., reaction barrier height and reaction enthalpy) compared to a higher level of theory such as CCSD(T), we choose B3LYP/6-31G** based on two reasons. First, considering hundreds of reaction paths to be refined, DFT is computationally much cheaper than coupled-cluster theory. Second, we want to assess the accuracy of the later constructed kinetic model from a consistent method in both reaction discovery and path refinement. Nevertheless, we will see later that with kinetic modeling and appropriate sensitivity analysis, only a few reaction paths need to be refined with higher level of theory in the future work.

The path refinement follows a three-step procedure. First, the endpoints (i.e., reactants and products) are optimized using the L-BFGS algorithm with delocalized internal coordinates (DLC). Cartesian coordinates are used as a back-up option when the optimization does not converge within 50 iterations. Second, after the endpoints of each reaction path have been minimized, a geodesic smoothing algorithm is employed to produce an initial guess path for future minimum energy path (MEP) search. Specifically, for each discovered reaction path, geodesic smoothing finds the local minimum of the total path length by updating one geometry frame (i.e., its coordinates) at a time. The smoothing procedure is performed iteratively until it converges when the change of path length is smaller than the convergence threshold. Overall, geodesic smoothing is trying to preserve the information from a discovered path, and at the same time, the smoothed geodesic paths have been shown as good starting guesses for minimum energy path search. The smoothed path is saved in our database, labeled as a “relaxed path”.

In the third step, we take the relaxed path as an initial guess and carry out MEP optimizations using the growing string method. During the MEP search, each reaction path is optimized with the translation–rotation-internal coordinate (TRIC) system which explicitly includes the translations and rotations of molecules. Since the reactions detected from discovery are not necessarily all elementary reactions, a reaction splitting algorithm is employed during the path optimization. A converged path will be split if it is found to contain an intermediate minimum node that is at least 3 kcal mol⁻¹ lower in energy than the adjacent nodes. The newly split reactions are registered as relaxed paths and will be refined again. In each refined path, the highest energy point is used as the input for a transition state search, where the L-BFGSSDNA algorithm is used with DLC, and the norm of the energy gradient is minimized towards a saddle point. This approach may not guarantee a first-order saddle point on a reaction path, although this is most frequently the case for an elementary reaction that contains a transition state. In principle, one could carry out normal mode analysis at each putative transition state to ensure that all are first-order saddle points (and subject them to further optimization if they are not). Additional information about the transition state optimization is included in Section S5.3 of the ESI† (with distributions of imaginary frequency amplitudes in the transition state configurations in Fig. S2† and the refined path for 10 most important reactions for CH₄ destruction in Table S4†). The ESI† provides validation for the current transition state optimization workflow in the ab initio nanoreactor. As we will see later, the current transition state optimization framework may introduce some inaccuracy in rate constant estimations, but the overall impact on the kinetic modeling is not too significant. As an improvement, work is underway to incorporate some other transition state optimization algorithms, for example, the partitioned rational function optimization (PRFO) approach which keeps track of the Hessian information during optimization and follows the eigenvector pointing towards the desired transition state configuration.

2.3. Thermodynamic and reaction rate parameter calculation

After all paths are refined, the macroscopic thermochemical properties (i.e., heat capacities Cₚ, entropies S, enthalpies H, and free energies G) of molecular species and transition state geometries are further computed. The entire computation procedure is detailed in Section S5 of the ESI† We briefly describe the major points here. The thermochemical properties are derived from the expression for molecular partition function, which is assumed to be a product of translational, rotational, vibrational, and electronic partition functions.

\[ Q = Q_{\text{trans}}Q_{\text{rot}}Q_{\text{vib}}Q_{\text{elec}} \]  

(eqn 4)

In eqn (4), Qtrans is the translational partition function, whereas Qrot, Qvib, and Qelec are the rotational, vibrational, and electronic partition function, respectively. The contribution of Qtrans to the thermochemical properties is derived using the ideal gas law. The calculation of Qrot is based on the rigid rotor model. Qvib is computed as a classical vibrational partition function for polyatomic molecules, assuming each vibrational mode is a harmonic oscillator. For transition state geometries, we treat any small imaginary vibrational modes (which are lower than 300 cm⁻¹ and usually correspond to hindered rotations) as real modes with the same amplitude. For both the endpoints and transition state geometries, low frequencies are treated as rotational modes rather than vibrational modes.
through Grimme’s quasi-rigid rotor harmonic oscillator (RRHO) approach combined with a damping function to switch between the two alternate descriptions. The \( Q_{\text{vib}} \) contribution to enthalpy includes electronic energy and the Yamaguchi spin projection for correcting spin contamination in unrestricted DFT calculations (see Section S5.4 for details). The electronic contribution to entropy includes the electronic degeneracy which is based on the spin multiplicity of each species. The calculated thermochemical properties of molecular species are tabulated as a function of temperature in the form of NASA9 polynomial coefficients. The fitting is performed using least squares within the 250–3000 K temperature range and at 1 bar reference pressure.

The rate of each reaction in the kinetic model is calculated using transition state theory (TST), in which the forward rate of a reaction is given by

\[
k_f(T) = \frac{k_B T}{h} \left( e^0 \right)^{(1-n_c)} e^{\Delta G^f/RT}
\]

where \( k_B \) is the Boltzmann constant, \( h \) is the Planck constant, \( e^0 \) is the standard molar concentration referenced to 1 atm pressure, \( n_c \) is the number of reactant species, and \( \Delta G^f \) is the difference of standard Gibbs free energy of the transition state and the reactant. As usual, the contribution of the vibrational partition function \( Q_{\text{vib}} \) to the transition state Gibbs free energy neglects the largest magnitude negative force constant, which corresponds to the reaction coordinate. The reverse rate \( k_r(T) \) is computed using microscopic reversibility of the elementary reaction, taking the ratio of forward rate \( k_f(T) \) and the equilibrium constant \( k_e(T) \). In this work, we apply transition state theory to all bi-molecular reactions and reactions involving three or more reactants/products. For each unimolecular decomposition reaction (or association reaction in the reverse direction), the high-pressure limit rate coefficient \( k_{i,\infty}(T) \) is estimated using TST, while the low-pressure limit rate coefficient \( k_{i,0}(T) \) is estimated using Hinselwood theory, which takes into account the internal degrees of freedom of molecules. In the Cantera format of the kinetic model, equations of dissociation reactions in the form \( A \leftrightarrow B + C \) are modified to \( A (+\text{M}) \leftrightarrow B + C (+\text{M}) \) where \( \text{M} \) corresponds to the third body collision partner. With this modification, the pressure-dependent rate coefficient of a unimolecular reaction takes the form

\[
k_i(T, [\text{M}]) = \frac{k_{i,0}(T)[\text{M}]}{1 + k_{i,0}(T)[\text{M}]/k_{i,\infty}(T)}
\]

The final kinetic model is assembled with the reaction equations and their forward rate coefficients based on the modified Arrhenius equation (eqn (S40)). Details of the rate coefficient calculation procedure can be found in Section S6. The reaction kinetic model file in Cantera format (.cti) is also included in ESI.

Lastly, we note that all the rate parameters calculated and used in this work are referred to as phenomenological rate constants/coefficients, which are explicitly expressed to be dependent on macroscopic thermodynamic properties (e.g., temperature, and probably pressure), but independent of the time evolution of the reaction system. As Widom pointed out in 1965, a phenomenological rate constant should not be directly interpreted as the probability per unit time that reactants form products; rather, it encapsulates all the microscopic intermolecular and intramolecular transitions modes, regardless of whether the chemical system has reached a steady state energy distribution (SED). There has been some recent debate about whether it is appropriate to apply the phenomenological rate constants for modeling the gas-phase chemical reactions under high-temperature conditions. A practical solution to this problem is the master equation (ME) approach, which bypasses the use of phenomenological rate constants and directly simulates the species time-evolution in an energy-resolved chemical reaction system. A case study by Miller et al. suggests that the error induced by phenomenological rate constants seems to be small compared to the ME solutions, so long as they are estimated based on methods obeying detailed balance and taking into account both the internal energy relaxation eigenmodes (IERE) and chemically significant eigenmodes (CSE). Nevertheless, in this work, we simulate the kinetic evolution of the methane pyrolysis system using the phenomenological rate parameters.

### 2.4. Kinetic modeling and sensitivity analysis

Kinetic modeling is performed in Cantera, an open-source chemical kinetics numerical solver. The time evolution of species mole fractions (e.g., methane \( \text{CH}_4 \), hydrogen \( \text{H}_2 \), ethylene \( \text{C}_2\text{H}_4 \), and acetylene \( \text{C}_2\text{H}_2 \)) are simulated as initial value problems in a zero-dimensional (0D) homogeneous reactor assuming ideal gas mixtures. The estimated species mole fractions are validated against experimental data collected from two shock tube facilities and a flow reactor facility, respectively. For shock tube experiments, the 0D reactor was assumed to be adiabatic and constant volume; for the flow reactor experiment, constant-pressure and constant-temperature assumptions were used for the simulation. Additional details of the experimental conditions and simulation approach is documented in Section S7 in ESI. Because of the widespread interest in combustion reaction kinetics, numerous kinetic combustion models have been developed, based on a combination of experimental and theoretical information. We compare the simulation results from the \textit{ab initio} nanoreactor kinetic model with four of these semiempirical models: (1) GRI Mech 3.0, an optimized reaction model designed for natural gas combustion, (2) the Appel, Bockhorn, Frenklach (ABF) model, which predicts well the fuel rich combustion behavior of small hydrocarbons (e.g. ethane, ethylene, and acetylene), (3) USC Mech II, a kinetic model that captures the combustion behaviors of \textit{H}_2/\textit{CO}/\textit{C}_1–\textit{C}_4 hydrocarbons, and (4) AramcoMech 3.0, a recently developed kinetic model for \textit{H}_2/\textit{CO}/\textit{C}_1–\textit{C}_4 hydrocarbon combustion. Additional information regarding these four kinetic models is included in Section S2.

Sensitivity analyses presented in this work are performed with two approaches. The first approach is to explore the impact of each rate coefficient following a local sensitivity analysis
method. Briefly, it estimates the change of kinetic output with respect to the change of parameters in a kinetic model. For instance, the partial derivative \( \frac{\partial c_i}{\partial k_j} \) represents the first-order local sensitivity coefficient for the concentration of species \( i \) with respect to rate coefficient \( k_j \). In practice, the rate parameters and the output quantities of a kinetic model may carry different units and orders, and a more practical approach is to calculate the normalized sensitivity coefficient, which represents the fractional change in concentration \( c_i \) with respect to a fractional change of rate parameter \( k_j \):

\[
S_j = \frac{k_j \frac{\partial c_i}{\partial k_j}}{c_i \frac{\partial \ln c_i}{\partial \ln k_j}} = \frac{\Delta c_i / c_i}{\Delta k_j / k_j} = \frac{\Delta c_i / c_i}{(\frac{k_j}{f} - 1) / k_j} = \frac{\Delta c_i / c_i}{f - 1 / f}
\]  

Practically, a coarse-grained estimation of \( S_j \) can be accessed through a finite difference brute-force method, where \( S_j \) is expressed as

\[
S_j \approx \frac{\Delta c_i / c_i}{\Delta k_j / k_j} = \frac{\Delta c_i / c_i}{(\frac{k_j}{f} - 1) / k_j} = \frac{\Delta c_i / c_i}{f - 1 / f}
\]  

To estimate \( S_j \) using eqn (8), kinetic simulations need to be performed with each rate parameter \( k_j \) up-scaled by a given factor \( f \) (where \( f > 1.0 \)) and down-scaled by \( 1/f \). The change of concentration \( c_i \), \( \Delta c_i = c_{i,f} - c_{i,1/f} \), represents the change evaluated at perturbed rate constant \( k_j \) and \( 1/k_j \). Therefore, given a chemical kinetic model consisting of \( n \) reactions, \( 2n \) simulations of species concentration \( c_i \) are required to evaluate \( S_j \). In this work, we choose \( f = 2 \), which is reasonably effective yet small for most of the chemical reactions. Finally, the impact of each rate parameter \( k_j \) with respect to species concentration \( c_i \) are ranked according to the absolute value of \( S_j \). Details of this analysis will be discussed in the Results section.

The second sensitivity analysis approach is based on Monte Carlo (MC) sampling. While the brute-force sensitivity approach perturbs only one rate parameter at a time in a simulation, the MC-based sensitivity analysis allows all rate parameters to vary in one simulation. Specifically, for a particular set of MC sensitivity analysis, we perform 500 MC kinetic modeling calculations. In each MC kinetic modeling sample, the same target property is simulated (e.g., methane mole fraction at a given initial temperature, pressure, and simulation time), but all the rate parameters are randomly perturbed in their respective natural log space (i.e., \( \ln(k_j) \)) uniformly within a pre-defined upper-limit range (i.e., factor of 2, 10, and 100 in this work). In the MC-based sensitivity analysis, the modeling results are usually presented as an uncertainty band or a scatter of simulation points. The advantage of the MC-based sensitivity analysis will be shown and discussed later, as it takes into account the kinetic coupling among rate coefficients.

3. Results and discussions

3.1. Reaction network

After the reaction discovery and path refinement, the chemical kinetic model for methane pyrolysis is constructed. It consists of 53 species and 134 reactions. Fig. 2 shows a qualitative representation of the reaction network as a connectivity graph. In the graph, each node represents a unique species in the reaction network. Several common hydrocarbon species with less than four carbon atoms (C<4H) are highlighted with larger node sizes. The nodes are labeled with species chemical formula, and if necessary, some minimal information about molecular structure and spin multiplicity (e.g., in Fig. 2, c-C3H6 for cyclopropane, CH2(S) for singlet methylene, and C2H4(T) for triplet ethylene). Section S1 in ESI† lists all the 53 species with their labels in Fig. 2, modified SMILES string in the Cantera model, and their minimum-energy structures. In Fig. 2, each edge connecting two nodes indicates a reaction that involves these two species, one in reactant set and the other in product set. Reactions involving common (C<4H) species are also highlighted with heavier weights (displayed as thicker line widths). Note that one reaction may lead to multiple edges. For instance, dissociation of methane (CH4) into methyl radical (CH3) and hydrogen radical (H) results in an edge between nodes CH4 and CH3, and another edge between CH4 and H. Furthermore, there can be more than one reaction that involves two species connected by an edge. Taking the undirected edge between CH4 and CH3 as an example again, it represents any reaction that involves CH4 and CH3 as reactant and product or vice versa. Examples of these reactions are CH4(+M) ⇌ CH3 + H (+M), CH4 + H ⇌ CH3 + H2, C2H6 + CH3 ⇌ C2H5 + CH4, and C2H4 + CH3 ⇌ C2H3 + CH4.

The graph in Fig. 2 provides qualitative evidence of the ab initio nanoreactor’s ability to identify not only important species and reactions, but also secondary reactions in an automated, hypothesis-free manner. Some common hydrocarbons are highlighted with larger node sizes, and they are methane (CH4), methyl (CH3), ethane (C2H6), ethylene (C2H4), acetylene (C2H2), propane (C3H8), propene (C3H6), propyne (C3H4p) and allene (C3H4-a). Aromatic species, including single ring (e.g., benzene), and polycyclic aromatic hydrocarbons (PAHs) are not found in the present work because longer discovery trajectories are needed for these products to be found. The influence of the aromatic species on the kinetic evolution of the methane pyrolysis system is discussed later in the text. Furthermore, the nanoreactor not only captures almost all the important reactions impacting the global methane decomposition rates, as we will show later, but it also uncovers chemical reactions which are not considered by any literature kinetic models used in this work. We include two examples in ESI Section S8† Path refinement and transition state optimization suggest their transition state structures both exist (with one imaginary frequency each). Due to the high reaction barriers, neither reaction shows up in the sensitivity analysis as shown later, but this nevertheless highlights the nanoreactor’s potential for uncovering rare events.

3.2. Kinetic model validations for methane dissociation

The nanoreactor (NR) kinetic model is validated against experimental data collected from two recent methane pyrolysis experiments. The results are summarized in Fig. 3. Briefly, the top panels (Fig. 3a and b) show the comparison of simulated...
and experimental CH₄ mole fractions in two shock tube facilities. One series of simulations is performed at initial temperatures \( T₀ \) ranging from 1300 to 2400 K and initial pressure \( P₀ \) of 1.4 bar, and the other is carried out at \( T₀ \) ranging from 1100 to 2000 K and \( P₀ = 30 \) bar. The bottom panels (Fig. 3c and d) show the comparison of simulated and experimental CH₄ and H₂ mole fractions in a flow reactor experiment. Simulations are performed at atmospheric pressure and 1050–1500 K temperature range.

All kinetic modeling simulations are carried out as initial value problems in the 0D homogeneous reactor module in Cantera. Simulated species mole fractions are taken at the end of the simulation, which corresponds to the total reaction time \( t_{rec} \). Details of the simulation procedure and some discussions of initial conditions are included in Section S7.† Overall, the simulated CH₄ and H₂ mole fractions by the nanoreactor model are in reasonable agreement with the experimental data, though discrepancies are observed. For instance, the nanoreactor model suggests faster methane decay kinetics compared to experiment in the lower pressure 1.4 atm case (Fig. 3a), and slower dissociation kinetics in the higher pressure (30 bar) case (Fig. 3b). The nanoreactor model also underpredicts H₂ mole fractions as temperature rises. Here in this work, we do not wish to overemphasize the practice of kinetic model optimization towards experimental data as targets. Instead, we introduce (vide infra) two systematic sensitivity analysis approaches that can efficiently identify sets of rate parameters contributing to these discrepancies.

Simulations from the nanoreactor (NR) kinetic model are compared with results from four pre-existing literature models: GRI Mech 3.0, Appel, Bockhorn, Frenklach (ABF) model, USC Mech II, and AramcoMech 3.0. As seen in Fig. 3, the difference of predictions among four literature kinetic models is noticeable. This might be somewhat surprising, since some of these models share similar or even identical rate parameters (see Section S6† for detailed comparison of selected reaction rates across these models). The predictions from the NR kinetic model, obtained automatically and without empirical input, largely fall within the range of predictions from the four literature models. Fig. 3d suggests that none of the models tested herein (including the nanoreactor) can capture the H₂ production accurately.

Fig. 2  Reaction network displayed as a graph. Each node represents a unique species. Each edge connecting two nodes indicates a reaction that involves these two species, one in the reactant set and the other in the product set. Node labels are the species molecular formula and if necessary, some minimal information about molecular structures and spin multiplicity values (See Section S1 in ESI† for details). Common C₄ species (i.e., C₄H₄, C₃H₆, C₂H₆, C₂H₄, C₂H₂, C₃H₆, C₃H₄-p, C₃H₄-a) are highlighted with larger node sizes. Reactions connecting key species are highlighted with thicker edge widths. We note that H₂ molecule and H atom are important species as well, but they are not highlighted for better visualization since they are involved in most of the reactions in this network.
production at the higher end of the test temperature range (i.e., \( \sim 1480 \) K). Two possible reasons for this discrepancy were suggested in the original experimental work.\(^\text{148}\) First, the under-prediction of \( \text{H}_2 \) could be due to the inaccuracy of key \( \text{H} \)-abstraction reactions leading to \( \text{H}_2 \) formation (e.g., \( \text{CH}_4 + \text{H} \leftrightarrow \text{CH}_3 + \text{H}_2 \), \( \text{C}_2\text{H}_6 + \text{H} \leftrightarrow \text{C}_2\text{H}_5 + \text{H}_2 \), \( \text{C}_2\text{H}_4 + \text{H} \leftrightarrow \text{C}_2\text{H}_3 + \text{H}_2 \)).\(^\text{94}\) Secondly, there might be missing pathways related to formation and growth of large polycyclic aromatic hydrocarbons (PAHs).\(^\text{137,138}\) These PAH pathways are not discovered in the current nanoreactor reaction network because longer discovery trajectories are needed for such pathways to occur, and they are also missing in all the literature models except for the ABF model which includes only formation of PAHs up to four aromatic rings (i.e. pyrene). Monte Carlo sensitivity analyses (i.e., Sections 3.5 and S10\(^\text{†}\)) can shed light on this issue.

### 3.3. Local brute-force sensitivity analysis

In order to pinpoint the origin of the differences observed in Fig. 3, we need to determine (1) how different rate parameters contribute to a particular kinetic modeling outcome, and (2) the weights and rankings of rate parameter contributions in a kinetic model. One example is to understand the difference of rate parameter contributions to \( \text{CH}_4 \) predictions in Fig. 3a. A systematic approach to address these questions is the brute force local sensitivity analysis method.\(^\text{117,118}\) The calculation procedure for sensitivity coefficients was described above. All the brute-force local sensitivity analyses are performed with rate parameters up/down-scaled by a factor of 2 under the initial condition of experiments in Fig. 3a. We present results in Fig. 4 depicting the normalized \( \text{CH}_4 \) sensitivity coefficients (eqn (8)) for the top five most important rate coefficients at 1600 K initial temperature, where the nanoreactor (NR) model predicts lower methane concentration (i.e., faster methane decay) compared to the other four kinetic models. Results from a broader assessment will be shown later in Fig. 5. Among all the five kinetic models analyzed, Fig. 4 shows that only one or two distinct reactions dominate the ranked sensitivity charts. At the condition tested, the rate of \( \text{CH}_3 + \text{H} \leftrightarrow \text{CH}_2 \text{H}_3 \) and \( \text{H}_2 \) (d) mole fractions. The experimental/simulation conditions are (a) 10% \( \text{CH}_4/\text{Ar}, P_0 = 1.4 \) bar, and \( t_{\text{rec}} = 2.87 \) ms; (b) 10% \( \text{CH}_4/\text{Ar}, P_0 = 30 \) bar, and \( t_{\text{rec}} = 14.8 \) ms; (c) and (d) 10% \( \text{CH}_4/\text{N}_2, P_0 = 1.0 \) atm, and \( t_{\text{rec}} \) is reported in the experimental paper\(^\text{148}\) as a function of the initial temperature, as \( t_{\text{rec}} = 4.55 \times 10^3/T_0 \) (K) ms.

![Fig. 3](https://example.com/fig3.png)

**Fig. 3** Experimental and simulated \( \text{CH}_4 \) (a−c) and \( \text{H}_2 \) (d) mole fractions. The experimental/simulation conditions are (a) 10% \( \text{CH}_4/\text{Ar}, P_0 = 1.4 \) bar, and \( t_{\text{rec}} = 2.87 \) ms; (b) 10% \( \text{CH}_4/\text{Ar}, P_0 = 30 \) bar, and \( t_{\text{rec}} = 14.8 \) ms; (c) and (d) 10% \( \text{CH}_4/\text{N}_2, P_0 = 1.0 \) atm, and \( t_{\text{rec}} \) is reported in the experimental paper\(^\text{148}\) as a function of the initial temperature, as \( t_{\text{rec}} = 4.55 \times 10^3/T_0 \) (K) ms.
theory (e.g., CCSD(T)) could be necessary. Moreover, the use of Hinshelwood theory for the low-pressure limit rate and transition state theory for the high-pressure limit rate constant could result in inaccurate estimation of rate constants. Work is underway to exploit improved theories for pressure dependent reactions, such as RRKM. Lastly, note that the rate of \(2\text{CH}_3\leftrightarrow\text{C}_2\text{H}_5 + \text{H}\) does not show up in the NR sensitivity analysis, and we focus on this below.

The local sensitivity analysis is extended to more initial temperatures (1400, 1800, 2000, 2200, 2400 K) and the corresponding ranked bar charts are included in Section S10 in ESI. Here we summarize the rankings of reaction rate in Fig. 5, as a two-dimensional (2D) sensitivity map of \(\text{CH}_4\) mole fractions calculated by the nanoreactor, GRI Mech 3.0, ABF, USC Mech II, and AramcoMech 3.0 reaction models. Each column in the 2D map represents an initial simulation temperature ranging from

![Fig. 4](image-url) Ranked sensitivity bar chart for \(\text{CH}_4\) mole fraction with respect to reaction rates computed under condition 10\% \(\text{CH}_4/\text{Ar}, T_0 = 1600\) K, and \(t_{\text{rec}} = 2.87\) ms. Five reaction kinetic models are analyzed, including (a) the nanoreactor, (b) GRI Mech 3.0 & ABF, (c) USC Mech II & Aramco 3.0.

![Fig. 5](image-url) Two-dimensional (2D) sensitivity map for \(\text{CH}_4\) mole fraction with respect to reaction rates simulated by the nanoreactor (NR), GRI Mech 3.0, ABF, USC Mech II, and AramcoMech 3.0 reaction model, respectively. The sensitivity analyses are performed under condition 10\% \(\text{CH}_4/\text{Ar}, P_0 = 1.4\) bar, \(T_0 = 1400\) K to 2400 K. Each row shows the sensitivity coefficient ranking of the corresponding reaction listed on the left side (red: reactions involving only \(\text{CH}_4\) dissociation and H-abstraction, blue: reactions for \(\text{C}_2\) hydrocarbons, green: reactions for \(\text{C}_3\) hydrocarbons, purple: reactions for \(\text{C}_6\) hydrocarbons, mainly benzene \(\text{C}_6\text{H}_6\) and phenyl \(\text{C}_6\text{H}_5\)). The color map represents the rank of each reaction rate in the sensitivity map at a given \(T_0\) test. Note that the darkest color in the 2D map indicates the first rank for a reaction (i.e., the most sensitive one), while the lightest color means that the reaction either ranks below the tenth rank or is not included in the reaction model.
1400 K to 2400 K, increasing in 200 K increments. Each row shows the sensitivity coefficient ranking of the corresponding reaction listed on the left side. The color map represents the rank of each reaction in the sensitivity map at a given initial temperature. Note that the darkest color in the 2D map indicates the highest ranked reaction, while the lightest color means that the reaction either ranks below the tenth rank or does not exist in the reaction model. We present the rankings of the sensitivity coefficients instead of their values for two reasons. First, the values of sensitivity coefficient may vary among different kinetic models. Second, the ranking map provides qualitative yet straightforward evidence of the relative importance of rate parameters in a kinetic model.

The 2D sensitivity ranking map reveals a few interesting facts. The overall sensitivity landscapes vary among five kinetic models, and the variation itself increases with higher initial temperature. At initial temperatures below 1800 K, four literature models display a similar ranking map albeit with a few minor differences. The reaction (R0) CH₄ (+M) ⇌ CH₂ + H (+M) is the most impactful for CH₄ concentration prediction. The dominance of (R0) appears in all five kinetic models at 1400 and 1600 K, and starts to shift to the dominance of (R2) 2CH₃ ⇌ C₂H₂ + H at 1800 K in all four literature models. The NR kinetic model highlights the importance of (R1) CH₄ + H ⇌ CH₂ + H₂, as it ranks no. 2 in the NR sensitivity ranking while being a bit less critical in the other four models.

For initial temperatures exceeding 1800 K, the sensitivity ranking maps of CH₄ start to diversify within the five models (Fig. 5). The divergence of the ranking maps beyond T₀ = 1800 K also explains the enlarged variation of model predictions in Fig. 3a. Specifically, the top rank of reaction (R0) CH₄ (+M) ⇌ CH₂ + H (+M) is gradually taken over by (R2) 2CH₃ ⇌ C₂H₂ + H in GRI, ABF, and Aramco, while (R4) C₂H₆ + H ⇌ C₂H₅ + H₂ in the NR kinetic model and (R19) C₂H₅ + P ⇌ C₂H₅ + H in USC Mech II. We again notice that reaction (R2) 2CH₃ ⇌ C₂H₂ + H is not present in the NR sensitivity map over the entire range of temperatures tested. Discussion on this reaction is included in the ESI (Section S9). This reaction (R2) has historically been considered as a chemically activated reaction. Under high-temperature reaction conditions, the recombination of two methyl (CH₃) radicals forms a vibrationally activated adduct (C₂H₅*) which is so energetic that it can skip over the potential energy well of ethane (C₂H₆) and directly form the product of ethyl (C₂H₅) and hydrogen (H) radicals. This well-skipping phenomenon can also happen in the reverse reaction direction. Reaction (R2) is discovered during the AIMD simulations and is identified as a triplet-spin-multiplicity reaction. Therefore, the resulting path refinement and transition state optimizations found a structure with triplet spin multiplicity close to a first-order saddle point (Section S9†). The rate coefficient of (R2) in the NR model appears to be much lower than the rates in all four literature models (Fig. S8†), where the theoretical RRKM rate parameters from Stewart et al. are adopted. A small sensitivity test is also conducted by replacing the rate of (R2) in the nanoreactor by the rate from Stewart et al. and the result is illustrated in Section S9 and Fig. S9. The sensitivity of this rate coefficient replacement is small for CH₄, H₂, and C₂H₂ yet noticeable for C₂H₄. We infer this small sensitivity may not truly reflect the inaccuracy of the NR (R2) rate coefficient, because the rate of methane dissociation, (R0) CH₄ (+M) ⇌ CH₂ + H (+M) in the nanoreactor kinetic model is much higher than other literature models. The H radical generation rate, which partially reflects the global reactivity of the methane pyrolysis system, is only affected by the faster rate of CH₄ (+M) ⇌ CH₂ + H (+M). Therefore, the impact of 2CH₃ ⇌ C₂H₂ + H could be further underestimated. Nevertheless, future work should involve accurate calculations of well-skipping/chemically activated reaction rates.

Despite the observations above, the NR model also qualitatively captures the importance of some additional C₂ reactions, although their quantitative rankings and Sj values may vary. For instance, (R3) 2CH₃ (+M) ⇌ C₂H₆, (R4) C₂H₆ + H ⇌ C₂H₅ + H₂, (R5) C₂H₆ + CH₃ ⇌ C₂H₅ + CH₄, (R6) C₂H₆ (+M) ⇌ C₂H₅ + H (+M), (R8) C₂H₄ + H ⇌ C₂H₃ + H₂, and (R9) C₂H₄ + CH₃ ⇌ C₂H₃ + CH₄. Compared to the literature kinetic models, the NR also tends to noticeably overemphasize the rates of (R4) C₂H₆ + H ⇌ C₂H₅ + H₂ and (R9) C₂H₄ + CH₃ ⇌ C₂H₃ + CH₄. Such overemphasis is also observed in (R5), (R10), (R11), (R15), and (R17) but less pronounced. Another minor observation from USC Mech II and AramcoMech 3.0 is the significance of reaction rates (R19–26) leading to formation of aromatic species benzene (C₆H₆) and phenyl (C₆H₅), which also involves propyne (C₃H₄-p), allene (C₃H₄-a), propargyl (C₃H₃), this further explains their predictions of lower methane concentrations than GRI and ABF at higher temperatures in Fig. 3a–c, since a considerable amount of carbon is converted into benzene. Efforts to extend the reaction network to include benzene, propargyl, and propyne would be a good topic for future work.

Two major conclusions can be drawn from the results in Fig. 4 and 5. First, the rate of CH₄ (+M) ⇌ CH₂ + H (+M) is critical to accounting for CH₄ mole fraction prediction. The current approach in the nanoreactor tends to overestimate its rate. Second, the chemically activated reaction 2CH₃ ⇌ C₂H₂ + H has been discovered in the nanoreactor workflow, but the rate was extremely underestimated in the NR kinetic model. Future work should focus on improved theory (e.g., RRKM and/or master equations) for estimating the rates of pressure-dependent unimolecular reactions and chemically activated reactions.

3.4. Kinetic model validation for minor intermediates

We have presented the prediction accuracy of our nanoreactor methane pyrolysis model for methane and hydrogen. The model has been subsequently tested for secondary species as well. Fig. 6 shows the comparison of experimentally measured and simulated ethylene (C₂H₄, Fig. 6a) and acetylene (C₂H₂, Fig. 6b) mole fractions. The experimental data are collected from the same shock tube facility in Fig. 3a at initial temperatures between 1600 to 2400 K and pressures of 1.4 bar. Simulations are conducted again, using the nanoreactor model along with four literature kinetic models. The nanoreactor model overpredicts the ethylene concentration by a factor of 2.5 at its maximum experimental value, similar to USC Mech II and
AramcoMech 3.0. GRI and ABF models yield somewhat better agreement with the experimental C$_2$H$_4$ mole fractions. Fig. 6b suggests that the current nanoreactor model poorly predicts the C$_2$H$_2$ mole fractions over the temperature tested. The four literature models all agree reasonably well with the experimental C$_2$H$_2$ concentrations. Fig. 7 shows the ranked C$_2$H$_2$ sensitivity coefficient bar chart calculated at 1800 K initial temperature and 1.4 bar pressure. Clearly, the NR kinetic model shows a dominance of reaction C$_3$H$_5$:1-1 (+M) $\Rightarrow$ C$_3$H$_4$:p + H (+M) (see Section S1† for the structure of C$_3$H$_5$:1-1 and C$_3$H$_4$:p), as its S value (−0.54) is more than two times that of the no. 2 ranking reaction in the NR model (C$_2$H$_4$: + CH$_3$: $\Rightarrow$ C$_2$H$_3$: + CH$_4$:; S = 0.25). We attribute this dominance primarily to an insufficient discovery of C>3 reactions, particularly reactions related to propyne (C$_3$H$_4$:p) and probably its isomer allene (C$_3$H$_4$:a). As depicted in Fig. 2, propyne (C$_3$H$_4$:p) appears to be a terminal node in the graph, and it has only one parent node C$_3$H$_5$:1-1, one of whose parent nodes is C$_2$H$_2$. The kinetic outcome of such subtree structure would make C$_3$H$_4$:p a kinetic sink for C$_2$H$_2$ destruction. Without additional reaction pathways discovered, concentration of C$_3$H$_4$:p will accumulate during the kinetic modeling of the NR model due to its relative stability compared to radicals in the reaction system. The sinking flux from C$_3$H$_2$: to C$_3$H$_4$:p is realized by two sequential reactions, CH$_3$: + C$_2$H$_2$: (+M) $\Rightarrow$ C$_3$H$_5$:1-1 (+M) (written in the reverse direction opposed to Fig. 7a) and C$_3$H$_5$:1-1 (+M) $\Rightarrow$ C$_3$H$_4$:p + H (+M). Therefore, additional reaction discovery for C$_3$H$_4$:p (and its isomer C$_3$H$_4$:a) is necessary for more reasonable prediction of C$_2$H$_2$, and a more complete reaction network.

### 3.5. The Monte Carlo method

As discussed in the previous text, local brute-force sensitivity analysis is useful for quick sorting of rate parameters according to their influence on a particular kinetic output property. However, in a reaction kinetic model, rate coefficients can be strongly coupled within this high-dimensional parameter space. In this case, local brute-force sensitivity analysis may not be able to capture the kinetic couplings among reaction rates, as it only allows perturbation of one rate parameter at a time. Furthermore, it cannot map such degrees of coupling to the kinetic modeling outcomes (e.g., species concentration predictions). In this subsection, we further address the effect of reaction rate couplings on model predictions. Monte Carlo (MC) sensitivity analysis is used for this purpose and its description is included in the methods section. Fig. 8 shows scatter plots of CH$_4$: (Fig. 8a, d and g), C$_2$H$_4$: (Fig. 8b, e and h), and C$_2$H$_2$: (Fig. 8c, f and i) concentrations simulated by the nanoreactor model with MC sensitivity approach. Each filled circle represents a MC kinetic simulation in which all the rate parameters in NR model are randomly perturbed within their natural log space (i.e., ln(k$_j$:)) uniformly. All the rate parameters are perturbed under an
upper-limit factor of 2 (Fig. 8a–c), 10 (Fig. 8d–f), and 100 (Fig. 8g–i), respectively from the nominal values. The sensitivity analyses are performed under the condition of Fig. 3a (10% CH4/Ar, \( P_0 = 1.4 \) bar, and \( t_{\text{rec}} = 2.87 \) ms).

The choice of rate perturbation factors are roughly chosen according to the scale of inaccuracies in the reaction barrier and rate coefficient estimations. Taking the reaction barrier as an example, based on the Arrhenius law, a 1 kcal mol\(^{-1}\) inaccuracy in the barrier corresponds to a factor of \( \sim 1.5 \) perturbation of the reaction rate at the temperature range of 1000–1100 K, which is the initiation of the flame preheat zone during combustion. Moreover, a 5 kcal mol\(^{-1}\) inaccuracy in reaction barrier leads to about a factor of 10 rate uncertainty, and a 10 kcal mol\(^{-1}\) corresponds to a factor of 100 rate uncertainty, both at 1000–1100 K. In summary, the goal of this MC sensitivity test is to explore the parameter space in a kinetic model and assess the outcome of uncertainty and inaccuracy of rate coefficients.

As should be expected, the spread of predictions increases as the range of perturbations is increased in Fig. 8. For the CH4 and C2H4 mole fractions, one can see that small changes in key reaction rates could easily lead to better agreement with experiment. However, regardless of how much the rate parameters are perturbed, the simulated C2H2 mole fractions will not reproduce the experimental results. This suggests that the NR rate parameters are not the origin for the observed inaccurate prediction of C2H2. This bolsters our suggestion in Section 3.4, namely that the current reaction network requires additional expansion towards more reactions related to C2H2, and furthermore, propyne (C\(_3\)H\(_4\)-p), allen (C\(_3\)H\(_4\)-a), propargyl (C\(_3\)H\(_3\)), and probably higher carbon species. Additional MC results obtained at two other experimental conditions are included in Section S10 of the ESI.

Fig. 8 Monte Carlo (MC) sensitivity scatter plots of CH4 (a, d and g), C2H4 (b, e and h), and C2H2 (c, f and i) mole fractions simulated by the nanoreactor (NR) kinetic model along with experimental data and simulations obtained from other literature kinetic models. Each filled circle represents a simulation using one MC sample of NR model in which each rate coefficient is perturbed within an upper-limit factor of \( f = 2 \) (a–c), 10 (d–f), and 100 (g–i), respectively from the nominal values, assuming a uniform distribution in the log space of the rate coefficient. The sensitivity analyses are performed under condition 10% CH4/Ar, \( P_0 = 1.4 \) bar, and \( t_{\text{rec}} = 2.87 \) ms.
initio nanoreactor reaction discovery, we envision that both brute-force and MC sensitivity analyses can be essential components in the nanoreactor workflow, and instructive for the next round of reaction discovery and path refinement.

4. Conclusions

In the present work, a methane pyrolysis kinetic model is automatically constructed from first principles by the ab initio nanoreactor. The reactions during methane pyrolysis are discovered using GPU-accelerated ab initio molecular dynamics simulation, without human hypothesis or intervention. The reactions are subsequently refined towards their minimum energy paths using the growing string method. Reactants, products, and transition state geometries are optimized, and their thermochemical properties are computed. Reaction rate parameters are estimated using transition state theory. The kinetic model is validated against experimentally measured key species concentrations during methane pyrolysis over a wide range of thermodynamic conditions. The simulated results from the nanoreactor model are compared with the results from four literature combustion kinetic models. Overall, the nanoreactor model predicts major species (i.e., methane and hydrogen) with reasonable accuracy. We pointed out and discussed discrepancies between the predictions of the nanoreactor model and experiments, as well as among predictions of the nanoreactor model with four literature kinetic models.

We introduce the method of local sensitivity analysis which efficiently identifies the importance of rate coefficients in a kinetic model with respect to a prediction outcome. The local sensitivity analysis suggests that future work should focus on improving the theoretical approach for pressure-dependent reactions and chemically activated reaction rate constant calculations. Another sensitivity analysis, the Monte Carlo method is used to explore the rate parameter space of the kinetic models. In addition to the local sensitivity analysis, the Monte Carlo sensitivity analysis suggests the current reaction network may be incomplete for prediction of C₃ intermediate species such as acetylene. This leads to the future necessity of extending the subnetwork of acetylene into C₃ and C₄ species using ab initio molecular dynamics simulations. In the future, we envision the nanoreactor-based reaction discovery can be performed iteratively, leveraging kinetic modeling and sensitivity analysis as tools for efficient identification of critical reactions and rate parameters, and guidance for the next iteration of reaction discovery.

Data availability

The data supporting this article have been included as part of the ESI† including, in particular the kinetic model in Cantera format.

Author contributions

RX contributed to the conceptualization, data curation, formal analysis, investigation, methodology, software, visualization and writing (original draft) of the presented work. JM contributed to the data curation and formal analysis. AMC contributed to the formal analysis and software. KCT contributed to the methodology and project administration. TJM contributed to the conceptualization, funding acquisition, methodology, project administration, resources, and supervision. All the authors contributed to the review and editing phases of the draft.

Conflicts of interest

TJM is co-founder of PetaChem, LLC.

Acknowledgements

This work was supported by the Office of Naval Research (N00014-21-1-2151 and N00014-18-1-2659). RX and JM thank Dr. Xiaolei Zhu for helpful discussions. JM thanks the Dr Leni-Schöninger Foundation and the Deutsche Forschungsgemeinschaft (project number 419817859) for financial support. AMC acknowledges support from National Science Foundation Graduate Research Fellowship and the Stanford Graduate Fellowship.

References

9 C. W. Gao, J. W. Allen, W. H. Green and R. H. West, Reaction Mechanism Generator: Automatic construction of chemical


Edge Article


