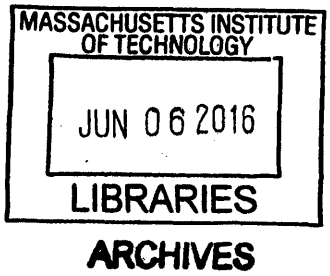
Automatic Reaction Mechanism Generation: High Fidelity Predictive Modeling of Combustion Processes

by

Connie W. Gao

M. S. Chemical Engineering Practice
Massachusetts Institute of Technology, 2013

B. S. Chemical Engineering
California Institute of Technology, 2010



Submitted to the Department of Chemical Engineering
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN CHEMICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Signature redacted

Author .....
Department of Chemical Engineering
May 9, 2016

Signature redacted

Certified by .....
William H. Green
Hoyt C. Hottel Professor of Chemical Engineering
Thesis Supervisor

Signature redacted

Accepted by .....
Richard D. Braatz
Edward R. Gilliland Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Automatic Reaction Mechanism Generation: High Fidelity Predictive Modeling of Combustion Processes

by

Connie W. Gao

Submitted to the Department of Chemical Engineering
on May 9, 2016 in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Chemical Engineering

Abstract

Growing awareness of climate change and the risks associated with our society's dependence on fossil fuels has motivated global initiatives to develop economically viable, renewable energy sources. However, the transportation sector remains a major hurdle. Although electric vehicles are becoming more mainstream, the transportation sector is expected to continue relying heavily on combustion engines, particularly in the freight and airline industries. Therefore, research efforts to develop cleaner combustion must continue. This includes the development of more efficient combustion engines, identification of compatible alternative fuels, and the streamlining of existing petroleum resources. These dynamic systems have complex chemistry and are often difficult and expensive to probe experimentally, making detailed chemical kinetic modeling an attractive option for simulating and predicting macroscopic observables such as ignition delay or CO_2 concentrations.

This thesis presents several methods and applications towards high fidelity predictive modeling using Reaction Mechanism Generator (RMG), an open source software package which automatically constructs kinetic mechanisms. Several sources contribute to model error during automatic mechanism generation, including incomplete or incorrect handling of chemistry, poor estimation of thermodynamic and kinetics parameters, and uncertainty propagation. First, an overview of RMG is presented along with algorithmic changes for handling incomplete or incorrect chemistry. Completeness of chemistry is often limited by CPU speed and memory in the combinational problem of generating reactions for large molecules. A method for filtering reactions is presented for efficiently and accurately building models for larger systems. An extensible species representation was also implemented based on chemical graph theory, allowing chemistry to be extended to lone pairs, charges, and variable valencies.

Several chemistries are explored in this thesis through modeling three combustion related processes. Ketone and cyclic ether chemistry are explored in the study of diisoproyl ketone and cineole, biofuel candidates produced by fungi in the decomposition of cellulosic biomass. Detailed kinetic modeling in conjunction with engine experiments and metabolic engineering form a collaborative feedback loop that efficiently screens biofuel candidates for use in novel engine technologies. Next, the challenge of modeling constrained cyclic geometries is tackled in generating a combustion model of JP-10, a synthetic jet fuel used in propulsion technologies. The model is validated against experimental and literature data and succeeds in capturing key product distributions, including aromatic compounds, which are precursors to polyaromatic hydrocarbons (PAHs) and soot. Finally, oil-to-gas cracking processes under geological conditions are studied through modeling the low temperature pyrolysis of the heavy oil analog phenyldodecane in the presence of diethyldisulfide. This system is used to gather mechanistic insight on the observation that sulfur-rich kerogens have accelerated oil-to-gas decomposition, a topic relevant to petroleum reservoir modeling. The model shows that free radical timescales matter in low temperature systems where alkylaromatics are relatively stable.

Local and global uncertainty propagation methods are used to analyze error in automatically generated kinetic models. A framework for local uncertainty analysis was implemented using Cantera as a backend. Global uncertainty analysis was implemented using adaptive Smolyak pseudospectral approximations to efficiently compute and construct polynomial chaos expansions (PCE) to approximate the dependence of outputs on a subset of uncertain inputs. Both local and global methods provide similar qualitative insights towards identifying the most influential input parameters in a model. The analysis shows that correlated uncertainties based on kinetics rate rules and group additivity estimates of thermochemistry drastically reduce a model's degrees of freedom and can have a large impact on model outputs. These results highlight the necessity of uncertainty analysis in the mechanism generation workflow.

This thesis demonstrates that predictive chemical kinetics can aid in the mechanistic understanding of complex chemical processes and contributes new methods for refining and building high fidelity models in the automatic mechanism generation workflow. These contributions are available to the kinetics community through the RMG software package.

Thesis Supervisor: William H. Green
Title: Hoyt C. Hottel Professor of Chemical Engineering

Dedication

I dedicate this thesis to my parents, Xiaoming Gao and Ruilian Wu,
who dreamed of me completing a PhD at MIT.

Acknowledgments

My PhD has been a long journey filled with many challenges and successes. It would not have been possible without the support and mentorship from many people in my academic and personal life.

First, I'd like to express my gratitude towards my advisor, Professor William Green. Throughout my PhD, he has provided the feedback and encouragement I needed to grow professionally. His positive outlook and ability to bring people together also taught me the importance of interpersonal skills in academics and the work place. I thank Bill for his email six years ago recruiting me into the Green Group. I am now positive that I made the best choice.

I am also extremely grateful to my collaborator and unofficial second mentor, Professor Richard West. I first met Richard when he was a postdoc in the Green Group. Since then, he has been an invaluable source of ideas for RMG and a provider of hands on advice and code review. In particular, I thank Richard for the idea to use Anaconda to package RMG, which allowed us to release the first binary for the Python version of RMG.

I also thank my thesis committee, Professors Paul Barton, George Stephanopoulos, and Youssef Marzouk. Their knowledge of numerical analysis and uncertainty propagation was extremely helpful in guiding my research.

Now, I'd like to thank several Green Group members who I have collaborated with during my PhD. First, I'd like to thank my predecessor and the previous lead developer of RMG, Josh Allen. Josh is a brilliant chemical engineer and computer scientist who ported over much of the RMG-Java code to Python. He also worked on the fungi biofuels project with me.

I also thank Shamel Merchant and Aäron Vandeputte for their mentorship and friendship both in and out of lab. Shamel helped guide and support me through the toughest moments of my PhD. He was always able to make me see the bigger picture and helped me prioritize my research. Aäron is a quantum chemist and dedicated researcher who was invaluable on the JP-10 and cineole projects. I will remember our group's times together very fondly– running along the river, watching videos together, and playing Magic in the lab.

I thank Nathan Yee, who has been a fellow RMG developer, deskmate, and friend. I also thank Nick Vandewiele and Kehang Han for their commitment as RMG developers and peers. Nathan and Nick both helped work on the JP-10 project, and Kehang has been a collaborator on the phenyldodecane project. I will miss our coffee breaks and RMG chats. I also thank the older members of the group who I have had the pleasure of working with: Michael Harper, Greg Magoon, Amrit Jalan, Caleb Class, Beat Buesser, Jorge Aguilera-Iparraguirre and Enoch Dames.

I would also like to thank my scientific collaborators outside of the Green Group. I thank Prof. Subith Vasu, Dr. Adam Scheer, and Dr. Craig Taatjes from Sandia National Laboratory, who I worked with on the fungi biofuels project. I had the pleasure of living in Livermore, CA and working at the Combustion Research Facility for 3 months in 2013 under their mentorship. I also thank Robin Bonomi, Greg Magoon, and Professor Kevin Van Geem for collaborating with me on the JP-10 project. I gratefully acknowledge the work of Eoghan Reeves, who has done the experimental work on the phenyldodecane project. I also appreciate the guidance offered by Professor Shuhei Ono on this project. I thank Professor Youssef Marzouk for helping provide advice for my uncertainty work, as well as his students Andrew Davis, Dr. Matthew Parno, and Dr. Daniele Bigoni, who helped introduce me to the MIT Uncertainty Quantification (MUQ) library. I also thank Dr. Ray Speth for his help working with Cantera.

Our daily tasks and meetings would not have been possible without the logistical support provided by the administrative staff in the Green Group. I'd like to give a special thank you to Gwen Wilcox and Barbara Balkwill for their tireless efforts.

Finally, I'd like to thank all the people in my personal life who have supported me during my PhD. I am thankful for all the friendships I made during Practice School and will cherish the fond memories I've had exploring Europe and working on food science with Kristen Whaley, Zachary Waxman, Sayalee Mahajan, Vishnu Sresht, Qing Xu, Li Tan, and Yuran Wang. I'm especially grateful to Yuran for putting up with me as a roommate in Minnesota and for all of our homecooked Chinese dinners. I am thankful for my other roommates while at MIT: Mariah Mandt, who is now a rockstar at Shell working in the Iraqi oil fields, and Ted Steiner, who is a close friend, rocket scientist, and hiking buddy. I also thank Nathan Yee, Winston Chern, Sakul Ratanalert, Alex Bourque, and Dariusz Murakowski for all the hot pot and board game nights we've had together. I also had the pleasure of meeting Shireen Goh, who was truly a big sister to me before she returned to Singapore. I am grateful for my friendship with Sivaraman Ramaswamy and our times philosophizing at Starbucks,

and Poyen Chen for our times working together in the first year. I'm grateful for the McKinsey Insight Program for bringing Jouha Min, Lulu Li, and Alex Guo into my life. I thank Yuqing Cui for our book club discussions and outdoor escapes. I am also grateful for the 6 years of friendship with Noemie Dorval and our longstanding traditions of hosting parties and celebrating Halloween together. I thank my college friends- Heejae, Jon, Lil Jon, Jennifer, and Michelle for supporting me from both near and far. I also want to thank my boyfriend Ben Renner for his companionship over the last three years. I will always treasure our memories of biking riding around Cambridge and exploring the MIT libraries and tunnels together. Finally, I thank my family: my parents who pushed me to complete my PhD, Baodi, DD, and MM. I have made it this far because of their love.

Contents

List of Figures

List of Tables

1

INTRODUCTION

Growing awareness of climate change and the risks associated with our society's dependence expansion on a resource-limited planet has motivated global initiatives to develop technologies that decrease our dependency on fossil fuels. Particularly imminent is the need to curb greenhouse gas emissions. A limit of 2 °C warming as measured from preindustrial times has been set in the 2009 Copenhagen Accord; any further warming would lead to "dangerous" implications on the global climate system. Rising global temperatures could lead to the break up of the West Antarctic ice sheet, which may lead to devastating ripple effects such as the rise of sea levels by as much as 12 feet, flooding most coastal cities. A 2013 report by a United Nations panel on climate change [1] claimed a rise of up to a meter in global mean sea level by 2100 if current greenhouse gas emissions rates continue. More ominous is the recent 2016 climate change model [2] which doubles that projection. Although the 2015 Paris Climate Deal signed by both the United States and China has resolved to curb emissions, a full throttle global effort is necessary to develop economically viable renewable energy for replacing carbon-intensive energy sources.

In 2015, renewable energy represented the majority of new electricity generation capacity added around the world, due to strong investments in developing countries such as China, India, and Brazil. [3] However, a major hurdle remains in transportation fuels. Although fuel cell technology and electric vehicles are becoming more prominent, the transportation sector is expected to rely heavily on combustion engines in the near future, particularly in the freight and airline industries. Shown in Figure 1.1, the transportation sector contributes to more than 28% of total energy consumption in the United States. Despite the introduction of ethanol and biodiesel, renewables contributed to only 5% of the overall US transportation sector energy usage in 2013. [4] The steep rise of oil prices in 2008 spawned not only renewable energy research but also the increased refining of heavier crudes, extraction of oil from

tar sands, and fracking of shale oil and gas. The refining of these unconventional oil sources have actually led to an increase in emissions associated with extraction and refining of oil by nearly 33% over the last decade. [5] Consequently, the crude oil price drop of 2014 has done little to curb fossil fuel demand. Therefore, it is up to policy-makers to implement new measures to combat climate change and for researchers to develop affordable, cleaner combustion.

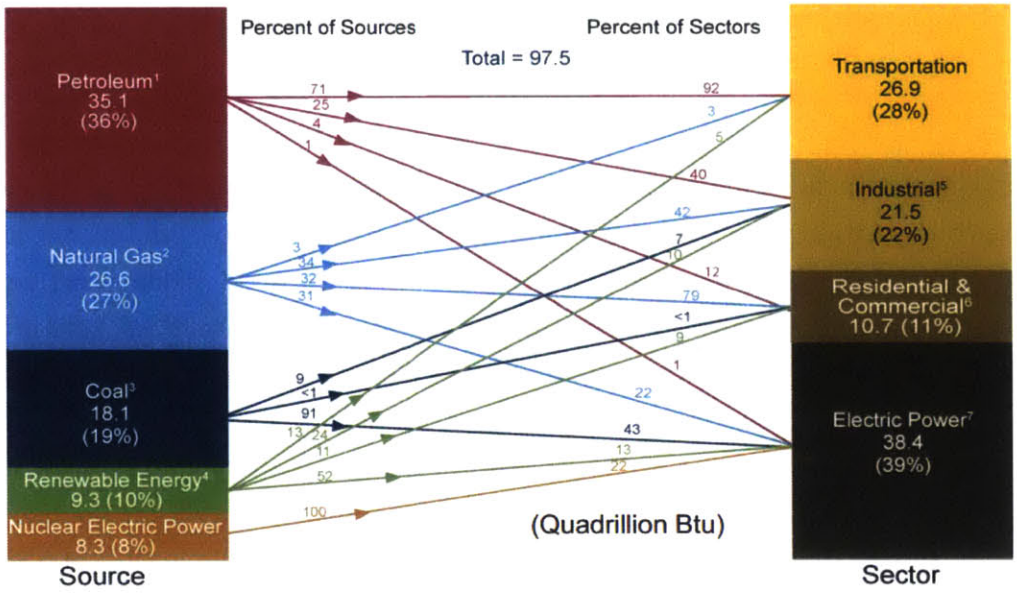


Figure 1.1: United States primary energy consumption by source and sector in 2013. Total usage was 97.5 quadrillion Btu. Figure taken from the Energy Information Administration.

A multi-pronged approach is required to tackle the clean energy problem: (1) management of existing petroleum resources, (2) development of alternative green fuels, (3) improvement of emissions controls in existing engine designs, and (4) development of more efficient and cleaner novel engine technologies. New technologies such as the homogeneous charge compression ignition (HCCI) engine can operate at up to 15% improved fuel efficiency over the spark ignition engine while maintaining cleaner emissions. [6] In addition, the HCCI engine is compatible with a wide variety of alternative fuels unlike conventional diesel and spark ignition engines, providing new avenues in the search for renewable fuels. Ignition in HCCI engines occurs via auto-ignition after the mechanical compression of a homogeneous gas mixture. The challenge associated with auto-ignition in HCCI engines and similar novel engine technologies is that their ignition timing is fuel dependent. Extensive knowledge of fuel chemistry and combustion kinetics is necessary for such engines to operate suc-

cessfully, and engine characteristics must be adjusted to the fuel used. Although the combustion kinetics of conventional fuels such as gasoline is well known, detailed investigation of new fuel candidates is required.

One approach that ameliorates the expense of expensive engine tests and experimental data collection is predictive *detailed kinetic modeling.* Detailed kinetic models consist of reaction networks with rate coefficients and equilibrium constants that describe the system at a set of reaction conditions. These models can be used to simulate macroscopic behaviors such as ignition timing or the formation of soot. Being able to predict the behavior of fuels within engines provides a means to quickly screen fuel candidates.

1.1 Automatic mechanism generation

Detailed combustion models become very tedious to construct manually as the number of reaction species and intermediates increase. These processes involve complex free-radical reactions between hundreds of reaction intermediates, coupled with thermodynamic and heat and mass transfer effects. In the past, models were often greatly simplified due to lack of computing power as well as lack of understanding of underlying chemistry. Today, numerical solvers and computational chemistry has advanced to the point where detailed kinetic models can now be constructed and applied to complex systems. Some detailed kinetic models are constructed by hand through carefully keeping track of all species and reactions and incorporating relevant chemistry. This process is often tedious and error-prone and additionally requires expert and up-to-date understanding of chemistry. These challenges can be avoided by using automatic reaction mechanism generation, which comprehensively constructs kinetic models using a set of reaction templates and kinetics rules, relying on a database that is continually updated. Automatic reaction mechanism generation can be used as a high throughout method for screening potential fuel targets for advanced engines, as well as for developing kinetic models for regimes where it is expensive or unwieldy to perform experiments.

1.2 High fidelity predictive kinetic modeling

Several sources can contribute to model error during automatic mechanism generation, including (1) incomplete or incorrect handling of chemistry, (2) poor thermodynamic and kinetics parameter estimation, and (3) uncertainty propagation (Figure 1.2).

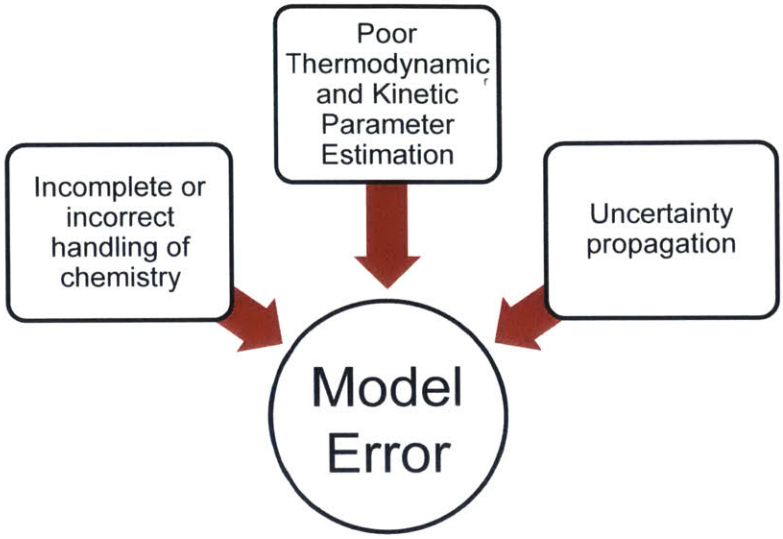


Figure 1.2: The three major sources of model error in automatic reaction mechanism generation.

Uncertainty analysis can help identify the largest contributing sources of output error from the model inputs. Thermodynamic and kinetics parameter estimation can be improved through algorithmic updates to chemistry treatment and continued additions to the chemistry database. Identifying incorrect or incomplete chemistry is the most difficult, as there are no systematic methods for identifying unknown chemistry without extensive model development and experimental validation. However, through modeling of new molecules and physical systems, computer understanding of chemistry will gradually improve. Addressing all three sources of error is necessary for the generation of high fidelity predictive kinetic models.

1.3 Thesis overview

My thesis has made several contributions toward developing methodologies for high fidelity predictive modeling within RMG. To put these methodologies into practice, my thesis also highlights three combustion related process modeled by RMG, including understanding ketone and cyclic ether chemistry through the study of biofuel candidates diisopropyl ketone and cineole, combustion modeling of the fused tricyclic jet fuel JP-10, and modeling the geological oil to gas cracking process via the heavy oil analog phenyldodecane.

Chapter 2 discusses the framework for automatic mechanism generation, focusing on the software package **R**eaction **M**echanism **G**enerator (RMG) developed at MIT

and used throughout this thesis. The features of the Python version of RMG are summarized in this chapter, including several new features that differentiate it from the former Java version: a new species representation that is capable of describing lone pairs and formal charges on molecules, a modification to the rate-based model enlarging algorithm which filters reaction to speed up model generation, and a web interface that allows users to draw molecules and generate adjacency lists, search RMG's databases, and visualize RMG-generated models online. RMG is capable of generating reaction mechanisms for C, H, O, N, S, and Si chemistry and modeling both gas and liquid-phase kinetics, thermochemistry, and transport.

Chapter 3 presents the role of automatic mechanism generation in a biofuel collaboration that uses mutual feedback between biologists, combustion experimentalists and theorists to screen fungi-derived biofuel candidates for use in HCCI engines. RMG is used to generated detailed low temperature oxidation models for diisopropyl ketone and cineole. Quantum chemical calculations are used to derive the initial hydrogen abstraction reaction rate coefficients and used as a base reaction network for which the final reaction mechanism is generated. The model is evaluated against chlorine-initiated oxidation experiments and is qualitatively able to identify a variety of species intermediates, demonstrating the capability of automatic mechanism generation to model new chemistries in practice.

Chapter 4 presents a comprehensive combustion model of JP-10, a tricyclic synthetic aviation fuel prized for its high volumetric energy density and low freezing point. Significant ongoing research is focused on the development of ramjet, scramjet and pulse detonation engines that burn JP-10 fuel. However, the testing of these engines is very expensive and time-consuming and have been prone to ignition problems, prompting the development of kinetic models to better understand the fuel's chemical behavior. A comprehensive detailed model was generated using RMG and extensively validated against a new set of shock tube experimental data and flow tube pyrolysis and ignition delay data from literature. Since JP-10 is a fused tricyclic compound, this presented additional challenges for estimating thermochemistry and kinetics accurately during automatic mechanism generation.The final model succeeds in capturing key trends in product distributions. We are also able to predict speciation of aromatic compounds, which are the precursors to polyaromatic hydrocarbons (PAH's) and soot, both potent atmospheric pollutants. Finally, we use our model to predict ignition delays over multiple experimental datasets, with general agreement within a factor of 4.

Chapter 5 discusses ihe influence of non-hydrocarbon compounds on the thermal

decomposition of kerogen and oil at geological conditions, a topic of great interest for petroleum reservoir modeling and prospecting. Organic sulfur is thought to be an accelerant in kerogen decomposition due to the facile nature of mercapto radical generation; however, the mechanism by which this occurs is relatively unknown. In this study, we investigate the influence of organic sulfur (diethyl disulfide, DEDS) on the decomposition of a heavy oil analog (phenyldodecane, PDD) through a combination of confined anhydrous pyrolysis experiments in gold capsules and automatic mechanism generation. RMG was used to automatically construct a full decomposition mechanism of PDD neat pyrolysis and decomposition in the presence of DEDS. We show that DEDS does not substantially accelerate PDD decomposition at 350°C (after 72h), in contrast to previous investigations in fixed volume reactors, but does exhibit statistically significant acceleration effects on decomposition at 250°C (after 1001h), hinting that free radical timescales may affect oil decomposition.

Chapter 6 presents methods for uncertainty analysis within RMG. Uncertainties within a kinetic model can become exaggerated for some input parameters while becoming irrelevant for others, due to the nonlinearity of chemical kinetic systems. Uncertainty analysis is the most efficient way to improve a model because it seeks to identify the greatest sources of model output error. This is especially significant in automatically generated mechanisms, where uncertainties are correlated because kinetics rate rules and thermodynamic group values can be used to estimate values in multiple input parameters. Local first order uncertainty propagation was implemented in RMG with respect to kinetic parameters and species thermochemistry. A cythonized wrapper PyDAS [7] was written for the DASPK3.1 [8] fortran code, a solver with native sensitivity analysis, to compute sensitivities in RMG. Global uncertainty propagation was also implemented using adaptive Smolyak pseudospectral approximations [9] to generate global sensitivity indices for the input parameters. A toy model was used to evaluate results from the two uncertainty propagation methods. Local uncertainty analysis was also used to investigate the effect of correlated uncertainties.

Finally, Chapter 7 discusses several recommendations for future work in predictive chemical kinetics.

1.4 References

[1] J. Church, P. Clark, A. Cazenave, J. Gregory, S. Jevrejeva, A. Levermann, M. Merrifield, G. Milne, R. Nerem, P. Nunn, A. Payne, W. Pfeffer, D. Stammer, and A. Unnikrishnan. "Sea Level Change." *Climate Change 2013: The Physical Science Basis.*

Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 2013, pp. 1137–1216.

[2] R. M. Deconto and D. Pollard. "Contribution of Antarctica to past and future sea-level rise." *Nature* 531, 2016, pp. 591–597. DOI: 10.1038/nature17145.

[3] F. S. F.-U. C. C. for Climate & Sustainable Energy Finance. *Global Trends in Renewable Energy Investment.* 2016, pp. 1–84.

[4] U.S. Energy Information Administration. *Annual Energy Outlook*, 2013.

[5] D. Gordon, A. Brandt, J. Bergerson, and J. Koomey. "Know Your Oil: Creating a Global Oil-Climate Index", 2015.

[6] D. Flowers, N. Killingsworth, and R. Dibble. "In Pursuit of New Engine Dynamics." *Mech. Eng.* 129, 2007.

[7] J. W. Allen and C. W. Gao. *PyDAS: A Python wrapper for the DASSL, DASPK, and DASKR differential algebraic system solvers.* 2016. URL: https://github.com/ReactionMechanismGenerator/PyDAS.

[8] L. Petzold and S. Li. *DASPK3.1.* URL: http://www.cs.ucsb.edu/~cse/software.html.

[9] P. R. Conrad and Y. M. Marzouk. "Adaptive Smolyak Pseudospectral Approximations." *SIAM J. Sci. Comp.* 35 (6), 2013, A2643–A2670. DOI: 10.1137/120890715.

2

Reaction Mechanism Generator: an open source software package for automatic mechanism generation

This chapter introduces the primary software used in this thesis for modeling the chemistry of combustion processes: **R**eaction **M**echanism **G**enerator. RMG is an automatic reaction mechanism generator which uses known chemistry knowledge stored in a database along with parameter estimation methods to generate detailed chemical kinetic mechanisms which can then be used as input to third party reactor software (*e.g.* CHEMKIN, Cantera, ANSYS Fluent) to simulate predictions for macrovariables of interest such as product composition, ignition behavior, or flame speed.

The four principal capabilities required for any automatic reaction mechanism generation code are [1]: a way to uniquely and unambiguously represent chemical species, a method to determine what reactions can occur between species, a means to estimate the kinetic and thermodynamic parameters, and a metric by which to include or exclude species and reactions in the model.

RMG uses a functional group based methodology to work with species and reactions. In this approach, reaction families are defined by templates that manipulate matching functional groups to convert molecules from reactants to products. Chemical graph theory is used to represent molecules and functional group substructures, with vertices representing atoms and edges representing bonds. This allows for graph isomorphism comparisons: for identifying functional groups when estimating parameters, and for comparing species against one another. Thermodynamic parameters are estimated for chemical structures using the Benson group contribution method [2, 3] or on-the-fly quantum chemistry calculations [4]. The species and reactions included in the final model are chosen by expanding the model using the rate-based algorithm of Susnow *et al.* [5]

2.1 Background

Several automatic reaction mechanism generation codes have been developed in the past, some proprietary and some open-source, including MAMOX, NetGen, REACTION, and EXGAS. Broadbelt and Pfaendtner [6] provide an introduction to the general concepts and terminology of kinetic model generation, and several recent reviews describe the commonalities, differences, and histories of these software projects [7–10].

The open-source software package RMG was developed in the Green Group at MIT to to help researchers model physical processes through automatic mechanism generation. All 60,000 lines of Python code are open-source and hosted on Github (`https://github.com/ReactionMechanismGenerator/RMG-Py`). RMG was originally developed in Java by Jing Song [11] in 2004, following approaches pioneered by NetGen and the ExxonMobil Mechanism Generator (XMG) in the 1990s [12, 13]. An object-oriented programming style was used prioritizing flexibility and extensibility of chemical rules and code re-usability. Over the years, several detailed kinetic models generated by RMG have been published in literature, including models for butanol [14], ketone biofuels [15], JP-10 jet fuel [16], and neopentane [17]. The source code for the Java version of RMG can be found both on Github (`https://github.com/ReactionMechanismGenerator/RMG-Java`) and Sourceforge (`http://rmg.sourceforge.net/`), with over 7000 downloads from Sourceforge alone over the last two years. In 2008, Joshua Allen and Richard West began writing a Python version of RMG, known as RMG-Py [18]. This was motivated by improved code readability, better error handling, and broader access to a variety of existing chem-informatics libraries. This chapter presents the features and usage of the new Python version of RMG.

2.2 Overview

An overview of the RMG workflow is presented in Figure 2.1. First, *ab initio* calculations or experimental data is inputted into the RMG databases. RMG uses a combination of chemistry knowledge, functional group-based assumptions, and the flux-based algorithm to build detailed kinetic models for use in simulating macrobehaviors of interest.

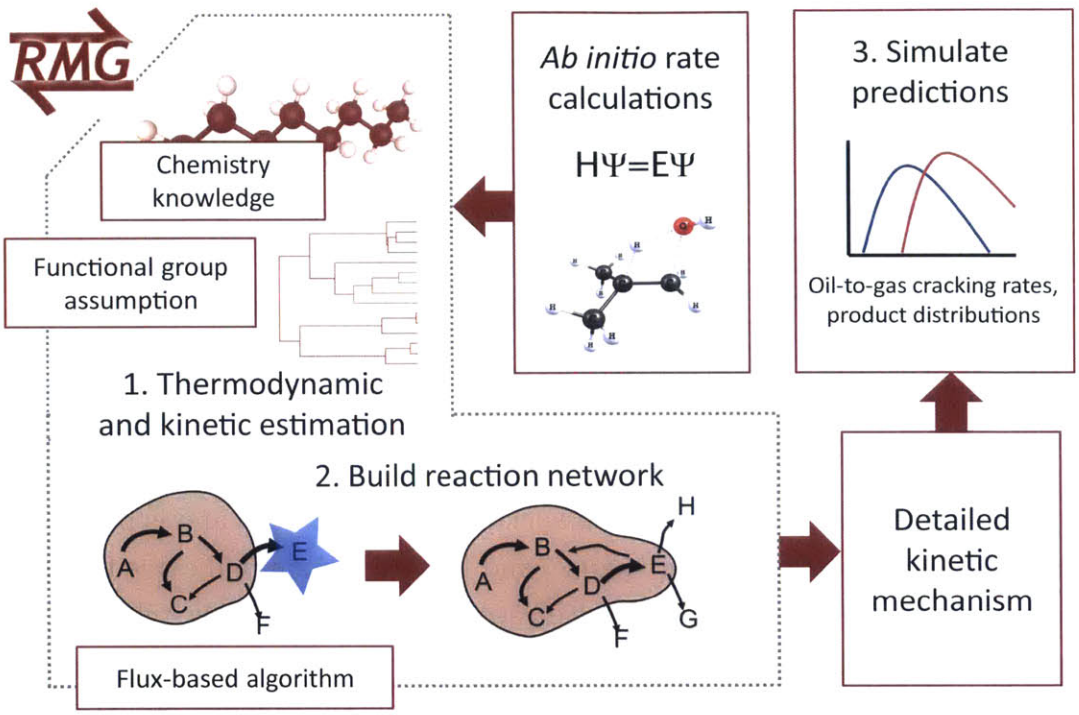


Figure 2.1: Overview of the RMG workflow.

2.2.1 Species and functional group representation

Chemical graph theory is used in for species and functional group representation, with graph vertices representing atoms and edges representing bonds. Graph isomorphism checks between reacting molecules and group definitions can quickly identify functional group-specific reaction sites, which allows for functional-group based thermo estimation and kinetics. Currently, the VF2 algorithm [19] for graph and subgraph isomorphisms is implemented.

2.2.1.1 Molecule adjacency lists

Molecules are described using "adjacency lists," a graph representation of the atoms and bonds that connect them. A set of molecule objects which are resonance isomers form a single species. This species contains its own thermochemical (i.e. enthalpy, entropy, and heat capacities) and statistical mechanical (i.e. frequencies and energies) information. The adjacency list for a methyl radical CH_3 is depicted in Figure 2.2.

The first column indicates the atom index, the second column indicates the atomic element, the 3rd column indicates the number of unpaired electrons associated each atom and is preceded by the lowercase letter u representing "unpaired", the 4th col-

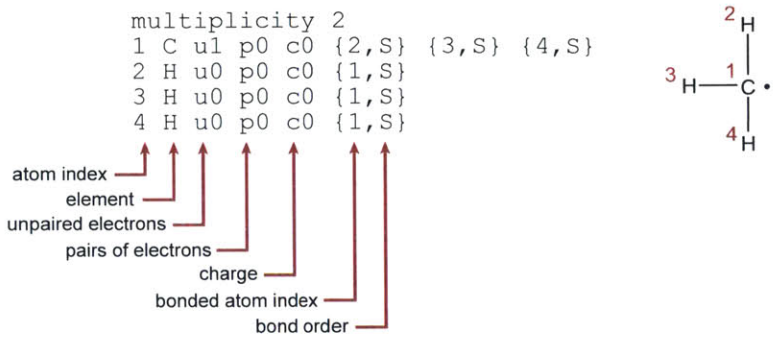


Figure 2.2: Adjacency list (left) and graph (right) of a methyl radical.

umn indicates the number of lone pairs associated with each atom and is preceded by the lowercase letter `p` representing "pairs", and the 5th column indicates the formal charge on the atom preceded by the lowercase letter `c` representing "charge". The values in brackets indicate the presence of a bond, with the first value in brackets indicating the atom index of the atom to which the current atom is bonded, and the second value indicating whether the bond is single `S`, double `D`, triple `T`, or benzene `B`. Finally, the molecule has an overall spin multiplicity defined above the adjacency list. In the above adjacency list, the carbon atom has a single unpaired electron and 3 single bonds to hydrogen atoms, forming a methyl radical.

2.2.1.2 Group adjacency lists

Similar to molecule adjacency lists, functional groups can also be described by adjacency lists, but group atom types are used in the adjacency list instead of atomic elements. These atom types can describe more a more general set of elements and can sometimes provide additional local bond structure requirements. The use of atom types accelerates graph isomorphism, or equivalency, checks and helps define both broad and general functional groups. The list of atom types used in RMG are defined in Table 2.1. Shown in Figure 2.3 is an example of a group adjacency list, utilizing the `R!H`, `Cd`, and `Od` atomtypes described in Table 2.1.

Note that in a group adjacency list, multiplicity, bonds, atom types, and even unpaired electrons can be a set of values. In order to distinguish from the notation used for bonds, square brackets are used for a grouping of values. In a group, only the number of unpaired electrons and bond information are required. Values that are unspecified, such as the number of lone pairs and charges on each atom in the adjacency list above, are assumed to be wildcards. The notation `x` can also be used in place of a wildcard.

Table 2.1: Atom types used in RMG group definitions

Atom type	Description
R	any atom with any local bond structure
R!H	any non-hydrogen atom with any local bond structure
H	hydrogen atom with any local bond structure
Cl	chlorine atom with any local bond structure
He	helium atom with any local bond structure
Ne	neon atom with any local bond structure
Ar	argon atom with any local bond structure
	Carbon atom types
C	carbon atom with any local bond structure
Cs	carbon atom with only single bonds
Cd	carbon atom with one double bond
CO	carbon atom with one double bond (to oxygen)
CS	carbon atom with one double bond (to sulfur)
Cdd	carbon atom with two double bonds
Ct	carbon atom with one triple bond
Cb	carbon atom with two benzene bonds
Cbf	carbon atom with three benzene bonds
	Silicon atom types
Si	silicon atom with any local bond structure
Sis	silicon atom with only single bonds
Sid	silicon atom with one double bond
SiO	silicon atom with one double bond (to oxygen)
Sidd	silicon atom with two double bonds
Sit	silicon atom with one triple bond
Sib	silicon atom with two benzene bonds
Sif	silicon atom with three benzene bonds
	Nitrogen atom types
N	nitrogen atom with any local bond structure
N1d	nitrogen atom with one double bond and two lone pairs
N3s	nitrogen atom with up to three single bonds
N3d	nitrogen atom with one double bond and up to one single bond
N3t	nitrogen atom with one triple bond
N3b	nitrogen atom with two benzene bonds
N5s	nitrogen atom with four single bonds
N5d	nitrogen atom with one double bond two single bonds
N5dd	nitrogen atom with two double bonds
N5t	nitrogen atom with one triple bond and one single bond
N5b	nitrogen atom with two benzene bonds and one single bond
	Oxygen atom types
O	oxygen atom with any local bond structure
Os	oxygen atom with only single bonds
Od	oxygen atom with one double bond
Ot	oxygen atom with one triple bond
Oa	oxygen atom with no bonds
	Sulfur atom types
S	sulfur atom with any local bond structure
Ss	sulfur atom with only single bonds
Sd	sulfur atom with one double bond
Sa	sulfur atom with no bonds

Group

```
multiplicity [1,2,3]
1 R!H      ux      {2,S} {3,D}
2 H        u0      {1,S}
3 [Cd,Od]  u[0,1]  {1,D}
```

Possible Molecular Structures

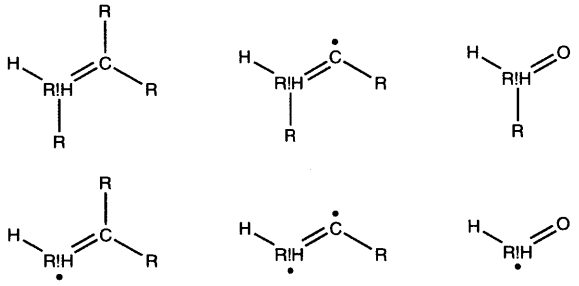


Figure 2.3: Adjacency list for a functional group (top) and its possible molecular structures (bottom).

2.2.1.3 Shift to new-style adjacency list: inclusion of lone pairs, charge, and multiplicity

RMG-Java was developed for C, H, O elements only, which generally required only sp^3 orbital chemistry and fixed valencies. These old style adjacency lists contained only information about atom to atom connectivity and radical electrons per atom. An old style adjacency list is shown in Figure 2.4 depicting a singlet carbene species. Without spin multiplicity descriptors, there is no way to distinguish a singlet carbene from a triplet form carbene species. Similarly, variable valencies cannot be implemented because bonds are calculated implicitly based on an 8 electron configuration for all atoms, whereas the new style adjacency list independently assigns unpaired electrons, lone pairs, and partial charges on each atom, allow far more flexibility for accurate species representations (Figure 2.5).

The implementation of new lone pair attributes for group adjacency lists also distinguishes the different reactivities of singlets vs triplets. Previously, any type of radical electron could participate in a reaction family involving radicals. With the new style group adjacency lists, triplets instead participate in radical step reactions (i.e. hydrogen abstraction), while singlets participate as electrophiles and nucleophiles that react with p orbitals. Three new lone pair reaction families were

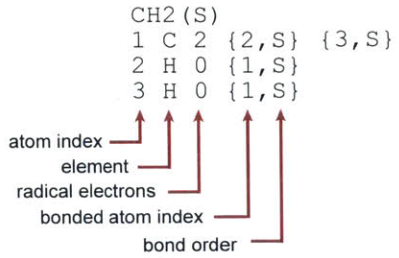


Figure 2.4: Old style molecular adjacency list for singlet carbene.

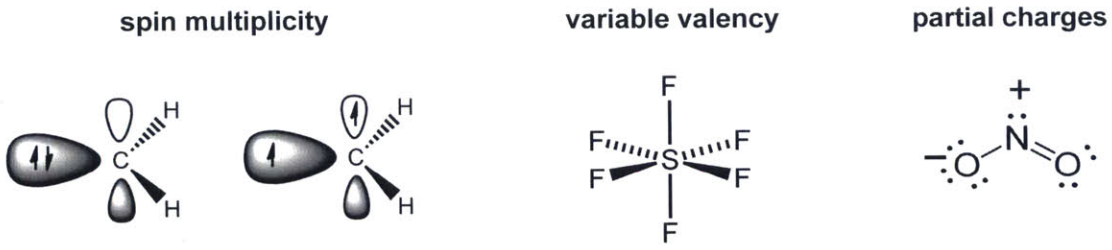


Figure 2.5: Three new species representation capabilities of the new style adjacency list: spin multiplicity as a species attribute (left, i.e. singlet vs. triplet carbene), variable valency (middle, i.e. sulfur hexafluoride), and partial charges (right, i.e. nitrogen dioxide).

added to RMG as a result: `1+2_Cycloaddition`, `1,2_Insertion_carbene`, and `lone_electron_pair_bond`.

2.2.2 Thermodynamic parameter estimation

Benson-style group additivity [2, 3] is used to estimate thermochemical parameters, including enthalpy ΔH_f°, entropy S°, and heat capacities C_p. For free radicals, we use the hydrogen bond increment (HBI) method of Lay *et al.* [20]. RMG uses hierarchical trees in its database for organizing functional group data in order to improve the speed of identifying group contributions. Trees are organized by placing general functional groups as top nodes, then creating more specific functional groups as children. Identifying the group contribution requires traversing down the tree from general to specific functional group.

The algorithm for estimating the thermodynamic parameters for a species is shown in Figure 2.6. First, resonance isomers of the species are generated, including aromatic forms of the species. Then, the thermodynamic parameters for each individual isomer

is calculated, first by checking whether the isomer is a free radical species in which HBI corrections are needed. After this step, group contributions to the enthalpy, entropy, and heat capacities are applied to the the saturated compound. Then, symmetry algorithms are used to apply a total symmetry number σ correction to the entropy of formation:

$$S^\circ = S^\circ_{GA} - R \ln \sigma \tag{2.1}$$

where S° is the standard corrected entropy of formation at 298K, S°_{GA} is the standard entropy at 298K calculated by the group additivity method, and R is the gas constant. Finally, cyclic and polycyclic ring corrections and gauche corrections are made to the thermodynamic parameters. Once the algorithm finishes iterating through all the isomers, RMG chooses the thermochemistry of the isomer with the most stable enthalpy to represent the thermochemistry for the overall species.

2.2.2.1 On-the-fly quantum mechanics for cyclic species

Benson group additivity is known to poorly estimate the enthalpy and entropy of cyclic and fused cyclic compounds due to the lack of appropriate ring strain corrections. RMG includes a **Q**uantum **M**echanics **T**hermodynamic **P**roperty (QMTP) interface [4] which allows it to perform on-the-fly quantum calculations to determine thermodynamic parameters for cyclic and polycylic species. This interface utilizes three-dimensional molecular structures in force field or quantum mechanical calculations to obtain thermodynamic parameter estimates. First, RMG sends molecular connectivities derived form its internal molecular graph representations to RDKit [21], which converts them to 3D coordinates using a distance geometry algorithm. Then, an input file containing the 3D molecular structure is sent to an external quantum mechanics program such as MOPAC or GAUSSIAN. RMG derives the thermodynamic properties from parsing the relevant frequencies and energies from the output files. Currently, RMG supports several semi-empirical methods such as PM3, PM6, and PM7.

2.2.3 Kinetic parameter estimation

RMG generates elementary reactions from chemical species using an extensible set of 45 reaction families. A reaction family consists of a template that describes the reactive sites, as well as a reaction recipe which dictates how the bond connectivity changes when the reaction proceeds to products. Associated with each reaction family is a hierarchical tree of rate estimation rules, classifying reactions according to neighboring functional groups. The rate estimation trees can be modified and

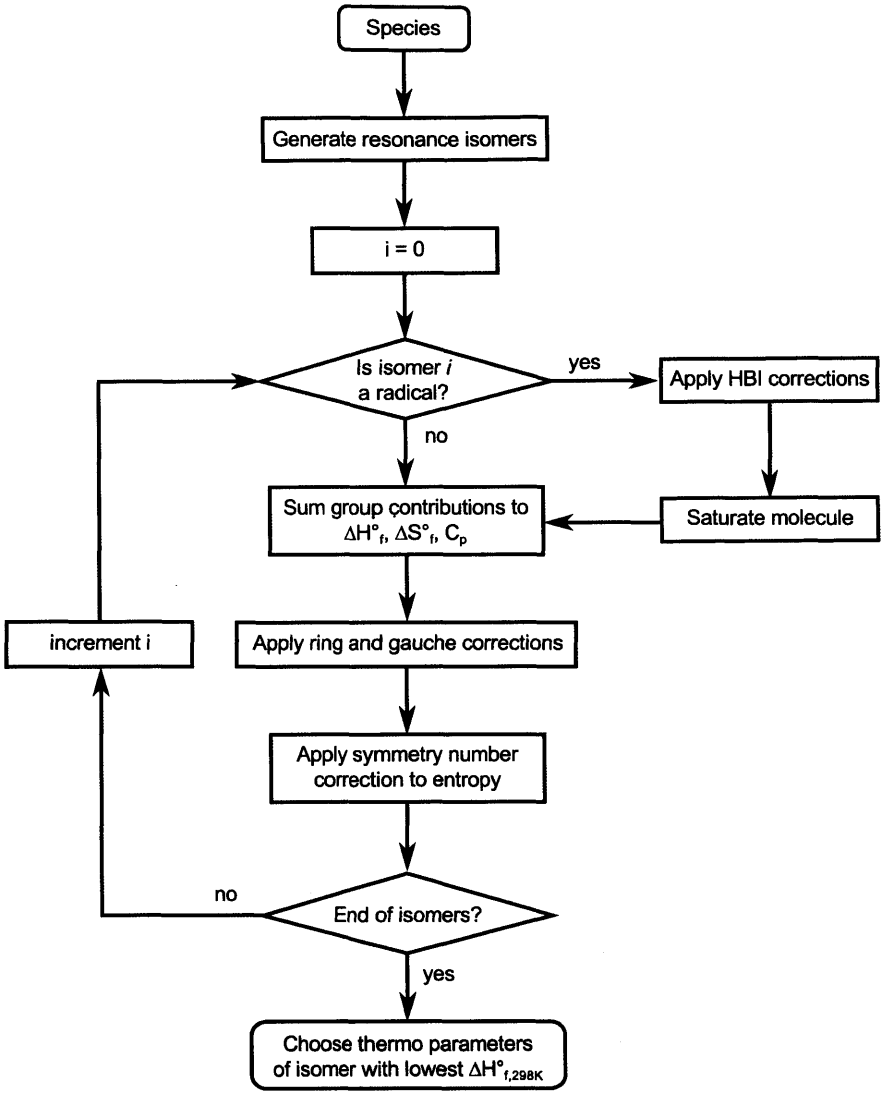


Figure 2.6: Flowchart of the group additivity-based thermodynamic parameter estimation algorithm as implemented in RMG.

extended without editing or recompiling the software, making it much more practical for chemists to add new information.

For example, the reaction family `H_abstraction` dictates the hydrogen abstraction from species XH by a radical species Y, as shown in Figure 2.8 along with its reaction recipe.

For any reversible reaction $Reactant(s) \underset{k_r}{\overset{k_f}{\rightleftharpoons}} Product(s)$, thermodynamic consistency is maintained through the following relation between the forward and back-

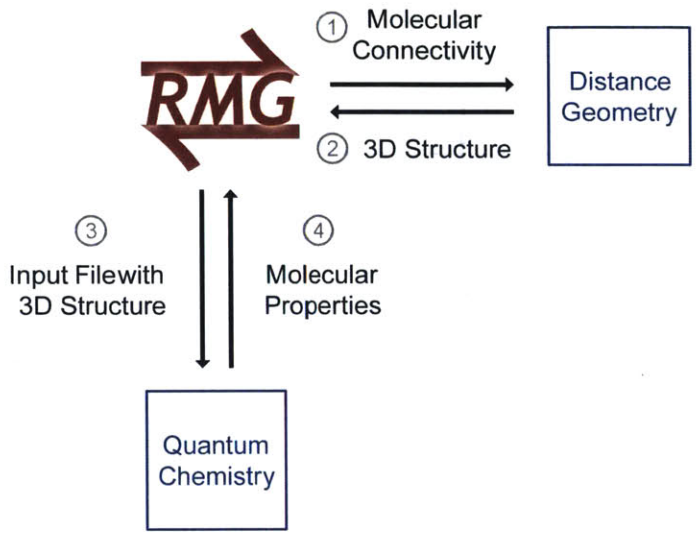


Figure 2.7: Schematic depicting the QMTP interface for calculating thermodynamic parameters on-the-fly using quantum mechanics in RMG. Adapted from Magoon *et al.* [4]

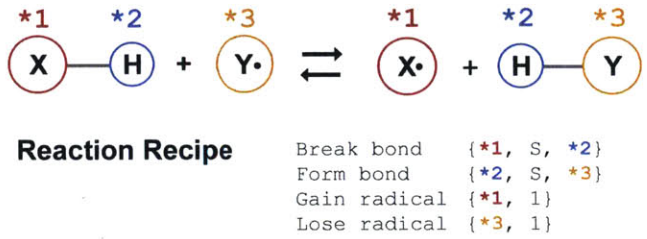


Figure 2.8: Reaction template and recipe for the `H_Abstraction` family.

wards reaction rate:

$$\frac{k_f}{k_r} = K_{eq} = \left(\frac{RT}{P^\circ}\right)^{-\Delta n} \exp\left(\frac{-\Delta G^\circ_{rxn}(T)}{RT}\right) \tag{2.2}$$

where K_{eq} is the equilibrium constant of the reaction, T is the reaction temperature, R is the gas constant, P° is the standard pressure (1 bar), $\Delta G^\circ_{rxn}(T)$ is the standard reaction free energy, and Δn is the change in moles in the reaction.

For most reaction families in RMG, the rates are defined in the forward direction. The reverse kinetics are calculated through the relation $k_r = k_f/K_{eq}$ using the thermodynamic parameters estimated for the reaction species. The organization of kinetic rate parameters is done, like for thermodynamic parameters, using hierarchical trees based on the principle that reactions between similar reacting sites in a

family will have similar kinetic rates. For each reactant site, a hierarchical tree is constructed with a general functional group that subdivides into children that are mutually exclusive and more specific than the parent. Partial representative trees of the two reactants from the `H_abstraction` family, XH and Y, are shown in Figure 2.9. Individual "rate rules" are defined for a set of functional groups for each reaction site through a temperature dependent kinetic paramter $k(T)$ described by the modified Arrhenius expression:

$$k(T) = A\left(\frac{T}{1\text{ K}}\right)^n \exp\left(-\frac{E_a}{RT}\right) \tag{2.3}$$

where A is the pre-exponential factor, n is the temperature exponential factor, E_a is the activation energy, R is the gas constant, and T is the temperature. Alternatively, the activation energy E_a can be related to the enthalpy of reaction ΔH_{rxn} through the constrained Evans-Polanyi relationship:

$$E_a = max(0,\ \alpha\Delta H_{rxn} + E_0) \tag{2.4}$$

where α, and E_0 are constants. Finally, for endothermic reactions in RMG, the activation energy E_a is raised to at least the endothermicity of the reaction using the following relationship:

$$E_a = max(E_a,\ \Delta H^\circ_{rxn}) \tag{2.5}$$

Upon loading the database, RMG fills in data within the hierarchical tree through an averaging algorithm. It locates sets of parent functional groups which have children containing data, and generates an geometrically averaged rate rule using the formula:

$$\log k(T) = \sum_{i=1}^{n} \frac{\log k_i(T)}{n} \tag{2.6}$$

where $k_i(T)$ is the ith child rate rule and n is the total number of children. Parents which are higher up in the tree may include children that are averaged rate rules themselves. To estimate the kinetic rate parameter for an individual reaction, the most specific functional groups that describe the reaction are determined by descending the reactant trees as far as possible. This set of functional groups is then used to match rate rules in the database. Consider the simple hierarchical trees for functional groups in a bimolecular reaction shown in Figure 2.10. If a reaction matches the functional group pair (A_3,B_3), but the database is missing this particular rate rule, then RMG will first search the nearest distance parent pairs (A_1,B_3) and (B_1,A_3) to see if these pairs contain data. If one or more of these pairs contains data, the geometric average of those rates will be used as the kinetics estimate. If neither contain data,

RMG will move on to the distance 2 pairs: (A_1,B_1), (A,B_3), (B,A_3), and once again check for data and use the geometric average of these data if it exists, continuing until it reaches the topmost general set of groups. Given the use of the averaging approximation for kinetic parameter estimation, RMG's accuracy is highly dependent on the amount of data present in its hierarchical trees.

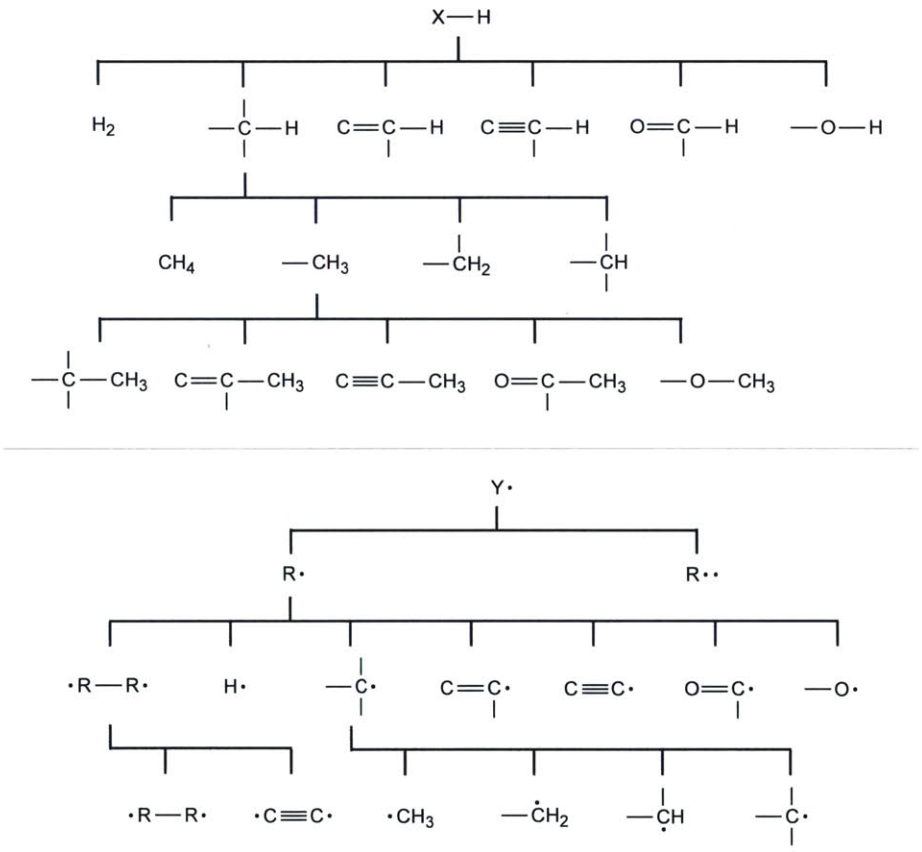
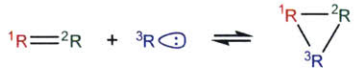
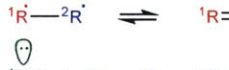
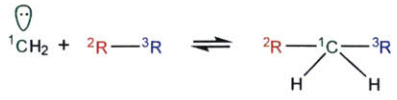


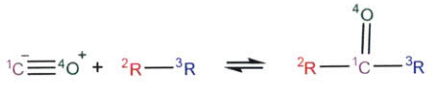
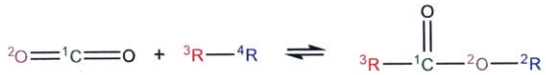
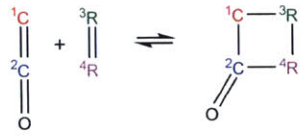
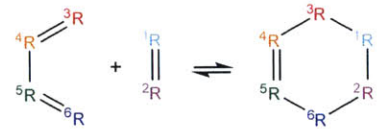
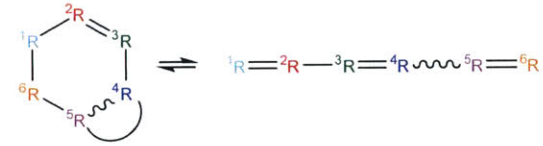
Figure 2.9: Hierarchical trees for the reactants in the `H_Abstraction` family. *Top*: Partial tree of the X-H reactant. *Bottom*: Partial tree of the Y reactant.

Table 2.2: RMG Reaction Families

`1+2_Cycloaddition`	
`1,2-Birad_to_alkene`	
`1,2_Insertion_carbene`	

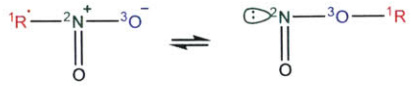
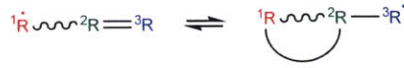
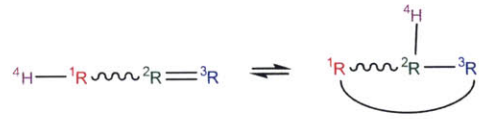
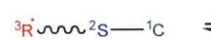
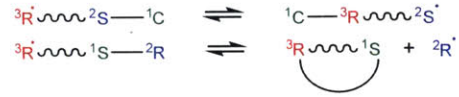
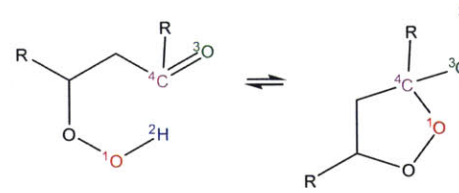
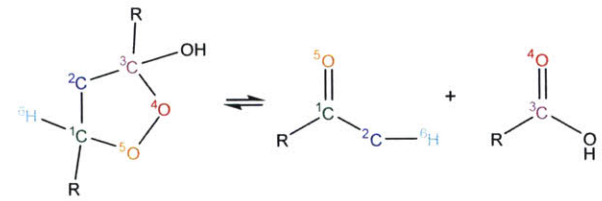
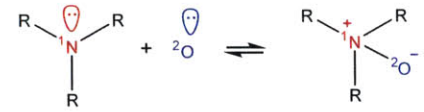
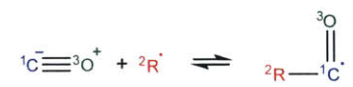
– Continued on next page –

– Table 2.2: continued from previous page –

1,2_Insertion_CO	
1,2_shiftS	
1,3_Insertion_CO2	
1,3_Insertion_ROR	
1,3_Insertion_RSR	
1,4_Cyclic_birad_scission	
1,4_Linear_birad_scission	
2+2_cycloaddition_CCO	
2+2_cycloaddition_Cd	
2+2_cycloaddition_CO	
Birad_recombination	
Cyclic_Ether_Formation	
Diels_alder_addition	
Disproportionation	
H_Abstraction	
H_shift_cyclopentadiene	
HO2_Elimination_from_PeroxyRadical	
Intra_Diels_alder	
Intra_Disproportionation	
intra_H_migration	

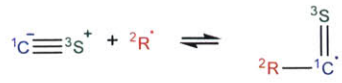
– Continued on next page –

– Table 2.2: continued from previous page –

intra_NO2_ONO_conversion	
intra_OH_migration	
Intra_R_Add_Endocyclic	
Intra_R_Add_Exocyclic	
Intra_R_Add_ExoTetCyclic	
Intra_RH_Add_Endocyclic	
Intra_RH_Add_Exocyclic	
intra_substitutionCS_cyclization	
intra_substitutionCS_isomerization	
intra_substitutionS_cyclization	
intra_substitutionS_isomerization	
ketoenol	
Korcek_step1	
Korcek_step2	
lone_electron_pair_bond	
Oa_R_Recombination	
R_Addition_COm	

– Continued on next page –

– Table 2.2: continued from previous page –

R_Addition_CSm	
R_Addition_MultipleBond	
R_Recombination	
Substitution_O	
SubstitutionS	

2.2.3.1 Training reactions

One drawback of a rate rule based database is the loss of molecular information. The user inputting the rate rule into the database selects the most fitting groups to represent the reaction; however, this provides only functional group information rather than complete information regarding the specific reactants and products. If the functional group hierarchical trees were to be altered, it becomes very difficult to reassign the rate rules without knowing the original reaction. Therefore, in RMG-Py, new kinetic parameter data is encouraged to be added in the database through "training" reactions, which retains information about the real molecules. When RMG finds the training reaction during model generation, it uses the exact kinetics from the training reaction. In addition, RMG uses the training reaction to generate rate rule data for the set of functional groups that match that reaction in the current hierarchical tree so that it can improve the kinetics of similar reactions.

2.2.3.2 Reaction libraries

Sometimes users will wish to use reactions from literature or calculations that they do not wish to influence other kinetics. In this case, they can create a "reaction library" which contains individual reactions and kinetic data that overrides RMG's native kinetic parameter estimation scheme. RMG will use the reaction library's kinetic data to determine how to proceed in generating the model. In RMG, multiple reaction libraries can be used with user-assigned priority.

2.2.3.3 Seed mechanisms

If an entire submechanism is desired to be included in the model, the user can use a "seed mechanism" in the model generation process. By doing so, the seed mechanism's kinetic parameters will both override RMG's native parameter estimation as well as

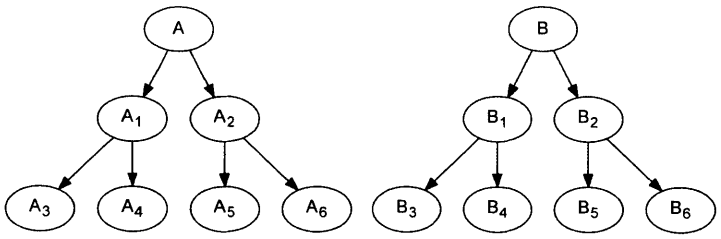


Figure 2.10: Simple hierarchical trees for a bimolecular reaction family containing reactants A and B.

be forced to be included in the model, whereas reactions from a reaction library will only enter the model if the reaction is deemed to be important through the flux-based model expansion algorithm. Multiple seed mechanisms can be specified in RMG with user-assigned priority.

2.2.4 Rate-based algorithm

RMG uses the rate-based algorithm of Susnow *et al.* [5] to determine which species and reactions to include in the model. The flow chart shown in Figure 2.11 demonstrates the model generation process, which begins with a user-specified set of initial species and conditions (i.e. temperature, pressure) and some termination criteria for which to end the simulation (i.e. a specified end time or goal conversion for some initial species). The initial species in the reaction system are placed into the "core" of the model and RMG determines all the possible reactions that can result from the core species, generating a list of possible product species on the "edge." The reactor is initialized at $t = 0$ and integrated in time until the flux $R_i = \frac{dC_i}{dt}$ to an edge species i exceeds ϵR_{char}, where ϵ is the user-specified error tolerance, and R_{char} is the characteristic flux of the system, defined by:

$$R_{char} = \sqrt{\sum_j R_j^2} \qquad \text{species } j \in \text{core} \tag{2.7}$$

The edge species with the largest flux is brought into the core, and the reaction generation and integration steps are repeated until the termination criteria is satisfied, generating the final kinetic model, which now contains all the species and reactions that have significant fluxes at the reaction conditions. The expansion of the model core is depicted in Figure 2.12.

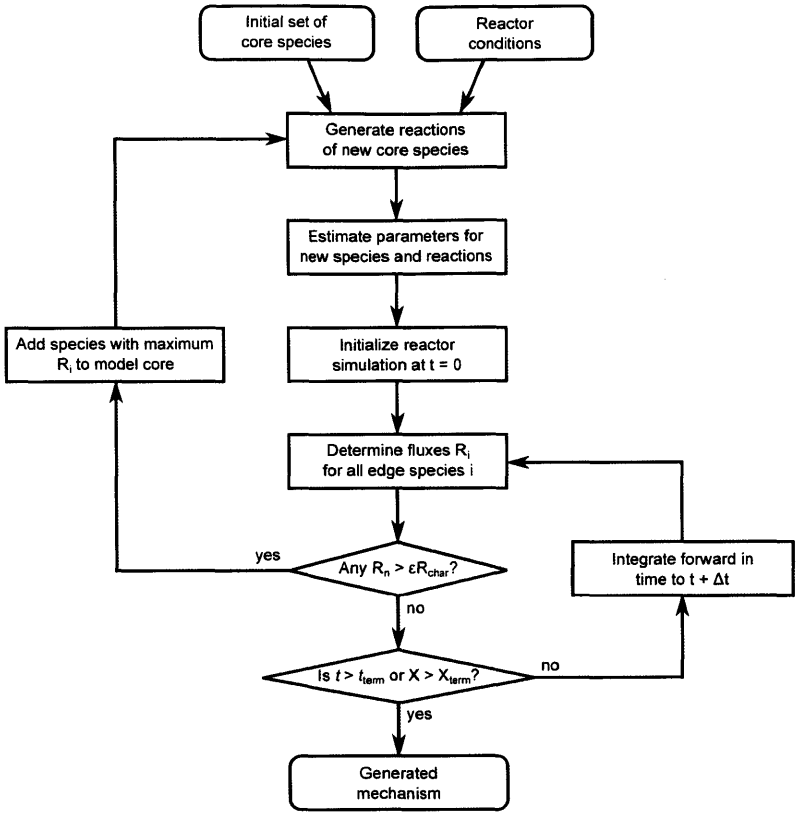


Figure 2.11: Flowchart of the rate-based algorithm as implemented in RMG. The generated mechanism contains the final set of species and reactions in the core.

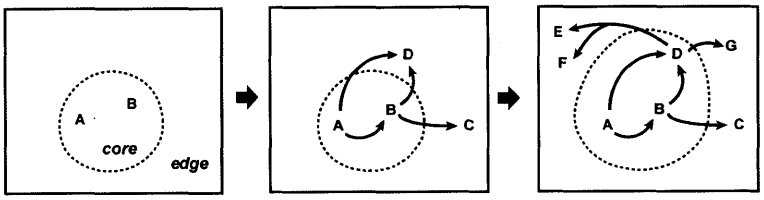


Figure 2.12: Schematic depicting expansion of the model core as RMG proceeds.

Currently, there are two reactor types that can be simulated within RMG. The first is the `SimpleReactor`, which is an isothermal, isobaric reactor in the gas phase. The second is the `LiquidReactor`, which is isothermal and isochoric reactor in the liquid phase. More information regarding liquid phase solvation and diffusion estimation is detailed in the Additional Features section.

2.2.4.1 *Filtering reactions within the rate-based algorithm*

Model convergence through the rate-based algorithm is often difficult when generating models for large molecules. For instance, model generation can take several days to converge for a C_{18} molecule such as phenyldodecane with very loose tolerances. The pain point in the rate-based algorithm is the generation of reactions, particularly within the bimolecular reaction families, where new edge reactions are created on the order of $m_{\text{reaction sites}} \cdot n_{\text{core species}} \cdot$ number of bimolecular reaction families per iteration, where m is the number of reaction sites on the newly added core species, and n is the number of total core species.

Reaction generation can be thought of as a two step process. For each new edge reaction, RMG must create the appropriate product species objects in memory based on the reaction family template as well as retrieve the kinetics from the rate rule estimation scheme (step 1). Then, the fluxes of all new and old edge reactions are evaluated to determine the next species to include in the model core (step 2). In terms of CPU speed, step 1 is extremely slow in comparison to the simple numerical evaluation performed in step 2. Therefore, a method was devised to filter the reactions generated in step 1, based on the expectation of their fluxes in step 2. This can be done by applying a conservative bound on the reaction fluxes based on the maximum kinetic rate possible $k_{threshold}$ and turning off reaction generation when fluxes are expected to be negligible.

For unimolecular reactions, the bound is applied to the reaction flux $R_{unimolecular}$

$$\begin{aligned} R_{unimolecular} &= k_{unimolecular,\,threshold}\, C_A > \epsilon R_{char} \quad \text{for all} \quad t \\ k_{unimolecular,\,threshold} &= \frac{k_B T}{h} \end{aligned} \tag{2.8}$$

where C_A is the concentration of the unimolecular reactant species A, ϵ is the model error tolerance, R_{char} is the characteristic flux defined in Equation 2.7, k_B is the Boltzmann constant, T is temperature, h is Planck's constant, and t is time.

For bimolecular reactions, the bound is applied to the reaction flux $R_{bimolecular}$

$$\begin{aligned} R_{bimolecular} &= k_{bimolecular,\,threshold}\, C_A\, C_B > \epsilon R_{char} \quad \text{for all} \quad t \\ k_{bimolecular,\,threshold} &= 1 \times 10^{13} \frac{cm^3}{mol \cdot s} \end{aligned} \tag{2.9}$$

where C_A and C_B are the concentrations of the bimolecular reactant species A and B, ϵ is the model error tolerance, R_{char} is the characteristic flux defined in Equation 2.7, and $k_{bimolecular,\,threshold}$ is set to the diffusion-limit rate coefficient.

Table 2.3: Comparison of model generation times for n-heptane pyrolysis with reaction filtering turned off and on in the rate-based enlarging algorithm.

	Reaction filtering off	Reaction filtering on
Model generation time (hr)	24	2.5
Memory usage (MB)	2500	2000
Core species	128	248
Core reactions	1410	2420

Within the code, binary arrays `unimolecularThreshold` and `bimolecularThreshold` are used to store flags for whether a species or a pair of species have concentrations that satisfy the bounds applied in Equation 2.8 and 2.9 evaluated at each time step in the simulation.

Two additional binary arrays `unimolecularReact` and bimolecularReact store flags detecting when the `unimolecularThreshold` or `bimolecularThreshold` flags shift from `False` to `True`. These flags signal for RMG to react species together in the reaction generation step, filtering reactions that have expected fluxes that are negligible.

Reaction filtering results in more than ten times CPU speed up as well as a decrease in RAM usage when applied to a n-heptane pyrolysis model, as shown in the comparison in Table 2.3. The model generated when reaction filtering is turned on produces contains nearly twice as many species as the original model with no loss of accuracy in output species profiles.

2.3 Features

A number of additional features in RMG include the ability to automatically generate pressure-dependent rate coefficients, reactions in the liquid phase, estimation of transport properties, and sensitivity analysis. In addition, the latest version of CanTherm is bundled within RMG and can be used to calculate thermochemical and kinetic quantities through transition state theory when used in conjunction with quantum chemistry software. In order to assist browsing the database and working with the species representation within RMG, a web front end has been developed and hosted on `http://rmg.mit.edu`, where many tools and graphical interfaces for RMG are available.

2.3.1 Estimation of pressure-dependent rate coefficients

Thermal unimolecular reactions proceed via nonreactive collisions with an inert third body to provide or remove the energy necessary for reaction. The reaction rate for these unimolecular reactions depend on the number of nonreactive collisions, which in turn is dependent on the pressure of the system. Under conditions where such collisions are rate-limiting, the observed phenomenological rate coefficient $k(T, P)$ is a function of both temperature T and pressure P. A unimolecular system is shown in Figure 2.13.

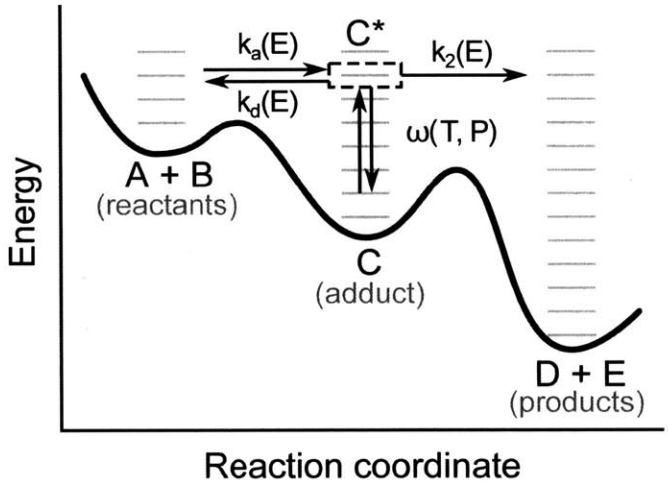


Figure 2.13: A typical unimolecular system. An activated species $C*$ can be formed either from chemical activation (as the product of an association reaction) or thermal activation (via collisional excitation). Once activated, multiple isomerization and dissociation reactions may become competitive with one another and with collisional stabilization; these combine to form a network of unimolecular reactions described by a set of phenomenological rate coefficients $k(T, P)$ that connect each pair of configurations, not just those directly adjacent.

A framework for estimating these pressure-dependent rate coefficients using high-pressure-limit kinetic data has been implemented in RMG and is described thoroughly in a separate paper [22]. The master equation model describes the unimolecular reaction network mathematically but is very computationally intensive. Thus, three methods for reducing the master equation and estimating the phenomenological rate coefficients have been implemented within RMG in addition to the master equation model: the modified strong collision method [23], the reservoir state method [24], and the chemically-significant eigenvalues method [25]. In the case of automatic generation of pressure-dependent rate coefficients, the modified strong collision

method is recommended for its speed and robustness. However, detailed investigation of individual reaction networks should be refined using either the reservoir state or chemically-significant eigenvalues method as they are more accurate.

2.3.2 Liquid phase solvation and diffusion

A framework for modeling solvent effects in RMG has been implemented [26] and is described briefly here. To model solution phase chemistry, we must estimate the changes in the thermochemical properties of a species going from the gas phase to the solvent phase. The thermodynamics of solvation for a species can be modeled through the partition coefficient K, which is defined as the ratio of the concentration of the species in the solvent phase to that in the gas phase at equilibrium:

$$K = \left(\frac{C_{solv}}{C_{gas}}\right)_{eq} \tag{2.10}$$

Assuming the chemical potential of a species i in each phase may be modeled using the equation:

$$\mu_i = \mu_i^\circ + RT \ln C_i \tag{2.11}$$

the change in the standard Gibbs free energy of a species in going from the gas phase to the solvent phase may be written as

$$\Delta G^\circ_{i,solv} = -RT \ln \left(\frac{C_{i,solv}}{C_{i,gas}}\right)_{eq} = -RT \ln K_i \tag{2.12}$$

In other words, the free energy of solvation can be directly calculated from the partition coefficient. Thus, RMG must be able to estimate the free energy change using estimated partition coefficient data as well as assign the free energy change to enthalpic and entropic contributions in a suitable theoretical scheme.

Linear Solvation Energy Relationships (LSERs) have been developed in order to understand the fundamental nature of solute-solvent interactions. In particular, the Abraham model [27, 28] uses molecular descriptors to predict the partition coefficient of a species in a large number of solvents:

$$\log_{10} K = c + aA + bB + sS + eE + lL \tag{2.13}$$

where K is the partition coefficient, the upper case parameters A, B, E, S and L are properties of the solute, and the lower case letters c, s, a, b, e and l are properties of the solvent.

The Abraham model is an empirical model that relies on experimental partition coefficient data to fit the model parameters. The aA and bB account for the free

energy change associated with the formation of hydrogen bonds between the solute and the solvent, the sS and eE terms account for intermolecular interactions such as dipole moments, the lL term accounts for the free energy change associated with the cavity formation process, and c is a correction factor.

The solvent parameters are obtained through multiple linear regression techniques on partition coefficient data of several solutes in the solvent of interest. These parameters are known and found in RMG's database as a library. The technique used to obtain solute parameters A, B, E, S, L for a compound for which experimental data are available is similar to the method used for the solvent parameters; however, in order to use the model for a large variety of solutes where experimental data are unavailable, a predictive method is necessary. RMG uses the Group Additivity based scheme for the estimation of Abraham solute parameters published by Platts *et al.* [29].

In the absence of a quantitative understanding of the temperature dependence of solvation thermodynamics [30], we use the simple approximation to model first order temperature dependence of ΔG_{solv}:

$$\Delta G_{solv}(T) = \Delta H^{\circ}_{solv} - T\Delta S^{\circ}_{solv} \tag{2.14}$$

In RMG we use the Mintz [31–33] correlations to estimate ΔH°_{solv} empirically:

$$\Delta H^{\circ}_{solv} = c' + a'A + b'B + e'E + s'S + l'L \tag{2.15}$$

where A, B, E, S and L are the same solute descriptors used in the Abraham model for the estimation of ΔG_{solv}. The lowercase coefficients c', a', b', e', s' and l' characterize the solvent and were obtained by regression to experimental data in a manner similar to that employed for the Abraham correlations.

Generation of kinetic models requires an understanding of solvation effects on elementary reaction rates. Solution phase reactions can be limited by transport of the reacting species towards each other (known as diffusive limits) and the cage-effect, which describes the increased probability of reaction between species trapped in a solvent cage. The theory behind diffusive limits in solution phase reactions is well established [34] and gives the the expression of the effective rate constant k_{eff} for a diffusion limited reaction:

$$k_{eff} = \frac{4\pi RDk_r}{4\pi RD + k_r} \tag{2.16}$$

where k_r is the intrinsic reaction rate, R is the sum of radii of the reactants and D is the sum of the diffusivities of the reacting species. This expression represents the

simplest equilibrium treatment of diffusive limits in the solution phase and is based on the Smoluchowski theory with corrections made by Collins and Kimball [35]. The effect of diffusive limits on reaction rates depends on the relative magnitudes of the intrinsic reaction rate k_r and the diffusive limit $4\pi RD$. Estimation of diffusive limits for a given reaction requires estimates of the species radii and diffusivities in different solvents. In RMG, we use the McGowan scheme [36] to estimates the species volume and its effective radius and the Stokes-Einstein correlation to estimation species diffusivities, which requires the solvent viscosity as input. Temperature-dependent viscosity correlations are included for a variety of solvents.

In order to maintain thermodynamic consistency, the forward rate constant, k_f in the reaction scheme $A \underset{k_{r,\text{eff}}}{\overset{k_f}{\rightleftharpoons}} B + C$ shall be affected if the reverse process $k_{r,\text{eff}}$ is slowed down by diffusion. In cases where both the forward and the reverse reaction rates are bimolecular, both diffusive limits are evaluated and the direction with the larger effect is used.

2.3.3 Transport property estimation

RMG is capable of estimating the transport properties of chemical species in a reaction mechanism automatically [37]. The transport data are saved in a CHEMKIN compatible format and can be used to run transport-dependent simulations such as laminar flames. RMG includes the GRI-Mech3.0 [38] transport library and estimates the transport properties for other molecules. The transport properties outputted are the parameters for the Lennard-Jones potential, which describes the intermolecular potential between two molecules or atoms:

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.17}$$

where V is the intermolecular potential, ϵ is the well depth and measures the strength of attraction between the two particles, σ is the internuclear distance at which the intermolecular potential is zero, and r is the internuclear distance between the two particles. RMG estimates σ (in Ångstroms) and ϵ (in Joules) using the properties of the fluid at the critical point (c) through empirical correlations taken from Tee *et al.* [39]:

$$\sigma = 2.44 \left(\frac{T_c}{P_c}\right)^{\frac{1}{3}} \tag{2.18}$$

$$\frac{\epsilon}{k_B} = 0.77 T_c \tag{2.19}$$

where k_B is the Boltzmann constant, T_c is the critical temperature in Kelvins, and P_c is the critical pressure in bar. The critical temperature and pressure, as well as boiling point T_b, for each molecule are estimated using the Joback group additivity method [40, 41]:

$$T_c = \frac{T_b}{0.584 + 0.965 \sum_i T_{c,i} - \left(\sum_i T_{c,i}\right)^2} \tag{2.20}$$

$$P_c = \frac{1}{\left(0.113 + 0.0032 n_a - \sum_i P_{c,i}\right)^2} \tag{2.21}$$

$$T_b = 198 + \sum_i T_{b,i} \tag{2.22}$$

where n_a is the total number of atoms in the molecule, and $T_{c,i}$, $P_{c,i}$, and $T_{b,i}$ are the group contributions to critical temperature, critical pressure, and boiling point, respectively.

RMG also provides the shape index, which indicates whether the molecule is monatomic (shape index = 0), linear (shape index = 1) or nonlinear (shape index = 2) in geometry. Currently, RMG sets the dipole moment, polarizability, and rotational relaxation collision number to zero.

2.3.4 Sensitivity analysis

Sensitivity analysis can be performed within RMG with respect to either the forward kinetic rate parameters or thermochemistry $\Delta G_f^\circ(T)$ of an individual species. The kinetic model can be described as a set of ordinary differential equations of the form:

$$\begin{aligned} \frac{dy}{dt} &= f(y, t; \lambda) \\ y(t_0) &= y_0 \end{aligned} \tag{2.23}$$

where y is the solution vector, t is time, λ is the time-independent vector of input parameters, and y_0 is the initial value of y. In this definition the first order sensitivity coefficient of output y_i with respect to parameter λ_j is given as:

$$S_{i,j} = \frac{\partial y_i}{\partial \lambda_j} \tag{2.24}$$

We are particularly interested in the normalized sensitivity of species i with respect to the rate coefficient k_j of reaction j:

$$\frac{\partial \ln c_i}{\partial \ln k_j} = \frac{k_j}{c_i}\left(\frac{\partial c_i}{\partial k_j}\right) \tag{2.25}$$

and the semi-normalized sensitivity with respect to the the ΔG_j° of species j:

$$\frac{\partial \ln c_i}{\partial \Delta G_j^\circ} = \frac{1}{c_i}\left(\frac{\partial c_i}{\partial \Delta G_j^\circ}\right) \tag{2.26}$$

Both types of sensitivities can be automatically simulated within RMG, either in a stand-alone analysis or at the end of an RMG job. RMG relies on the differential equation solver DASPK3.1 [42] to compute the sensitivities automatically.

2.3.5 CanTherm

The most up to date version of CanTherm is bundled as a subprogram of RMG and contains additional features and improvements on the original program [43]. It is a tool used for computing the thermodynamic properties of chemical species and the high-pressure limit rate coefficients for chemical reactions using the results of quantum chemistry calculations. Thermodynamic properties are computed using the Rigid Rotor Harmonic Oscillator approximation with optional corrections for hindered internal rotors. Kinetic parameters are computed using canonical transition state theory with optional tunneling corrections. CanTherm is compatible with output files from several well known quantum chemistry software programs, including Gaussian, MOPAC, QChem, and MOLPRO. For different methods, CanTherm applies additional atom, bond, and spin-orbit coupling energy corrections to adjust the computed energies to the usual gas-phase reference states.

2.3.6 Cantera support

Cantera [44] support allows RMG to access a variety of reactor system solvers beyond RMG's native isothermal, isobaric batch reactor. Several new classes and functions to help the user work with Cantera models and simulations were implemented in `rmgpy.tools.canteraModel`. Cantera is written in C++ but contains Python wrappings, making it highly compatible with RMG. Support includes both the capability to convert a RMG-generated CHEMKIN model to a Cantera format `.cti` file, as well as direct conversion of RMG `Species` and `Reaction` objects to Cantera python objects.

2.3.7 Web front-end

Since RMG is a software program with a command-line interface, there is a learning curve for many experimental chemists and kineticists. To familiarize users with RMG, we have created a web interface for working with the software, located at `http://rmg.mit.edu`.

A visual display and molecular search tool available for the all the databases (kinetics, thermodynamics, solvation, statistical mechanics, and transport), providing live queries for thermodynamic and kinetic parameter estimates valuable for working with the software as well as for comparing with literature. The demand for such comprehensive databases is high in the chemistry community. For example, the NIST Chemical Webbook has more than 10,000 unique visitors per month. With an user-inputted molecule or reaction, the website can return RMG's estimate for kinetics or other properties, along with the original sources for those estimates. This helps provide transparency for RMG's databases and methodology.

Since the RMG adjacency list format for molecules can be difficult to construct by hand, the website also provides tools to convert SMILES, InChI, CAS number, and common species names into the adjacency list format.

Visualization of the molecular structures and reactions within an RMG-generated model is also possible by uploading the outputted CHEMKIN file and associated species dictionary file. Visualization through the interactive website also enables filtering of the reactions by family and species, and automatic display of evaluated kinetics and heats of reaction. Additional web tools include visual model comparison, model merging, input file construction, and kinetics plotting.

The RMG interactive website is written on a Django python-based framework and is also open source. The most up to date version of the source code can be found on the Github repository `https://github.com/ReactionMechanismGenerator/RMG-website/`.

2.4 Example: n-heptane pyrolysis

The following example uses an n-heptane model generated by RMG to simulate recent pyrolysis experimental data gathered by Yuan *et al.* [45] at Hefei. The experimental study was performed at low pressure (400 Pa) with temperatures ranging from 780-1780 K.

The `input.py` file for the n-heptane pyrolysis model is shown below:

```
database(
    thermoLibraries = ['primaryThermoLibrary'],
    reactionLibraries = [],
    seedMechanisms = [],
    kineticsDepositories = ['training'],
    kineticsFamilies = 'default',
```

```
    kineticsEstimator = 'rate rules',
)
generatedSpeciesConstraints(
    maximumRadicalElectrons = 1,
)
species(
    label='n-heptane',
    reactive=True,
    structure=SMILES('CCCCCCC'),
)
species(
    label='Ar',
    reactive=False,
    structure=SMILES('[Ar]'),
)
simpleReactor(
    temperature=(1000,'K'),
    pressure=(400,'Pa'),
    initialMoleFractions={
        'n-heptane': 0.02,
        'Ar': 0.98,
    },
    terminationConversion={
        'n-heptane': 0.99,
    },
    terminationTime=(1e6,'s'),
)
simpleReactor(
    temperature=(1500,'K'),
    pressure=(400,'Pa'),
    initialMoleFractions={
        'n-heptane': 0.02,
        'Ar': 0.98,
    },
    terminationConversion={
        'n-heptane': 0.99,
```

```
    },
    terminationTime=(1e6,'s'),
)
simpleReactor(
    temperature=(2000,'K'),
    pressure=(400,'Pa'),
    initialMoleFractions={
        'n-heptane': 0.02,
        'Ar': 0.98,
    },
    terminationConversion={
        'n-heptane': 0.99,
    },
    terminationTime=(1e6,'s'),
)
simulator(
    atol=1e-16,
    rtol=1e-8,
)
model(
    toleranceMoveToCore=0.02,
    toleranceInterruptSimulation=0.02,
)
pressureDependence(
    method='modified strong collision',
    maximumGrainSize=(0.5,'kcal/mol'),
    minimumNumberOfGrains=250,
    temperatures=(300,3000,'K',8),
    pressures=(0.001,100,'bar',5),
    interpolation=('Chebyshev', 6, 4),
)
```

This input file first describes databases to be used: the specific libraries and estimation method. A `generatedSpeciesConstraints` option is used to constrain the maximum number of electrons in the model to 1 or fewer, excluding biradical species from appearing in the model. This parameter is used to speed up model

convergence by restricting the types of species that RMG considers within the model.

Then, the reactive and nonreactive species are declared: n-heptane and argon, followed by a series of one or more reaction systems that describe the initial quantities of the reactants and the temperature and pressure conditions. Since the experimental conditions spanned a wide range of temperatures, 3 reactor systems were used with temperatures of 1000, 1500, and 2000 K. Each reactor has a conversion termination criteria of 99% conversion of n-heptane with a fallback time termination criteria.

Then the numerical simulation tolerances are given, followed by RMG's desired model generation error tolerance ϵ described in Section 2.4. This tolerance $\epsilon = 0.02$ can be further tightened to a smaller value if the user wishes to obtain a larger and more comprehensive model.

Pressure dependence in this example is turned on because the reaction conditions are at low pressure and high temperatures, making pressure dependence highly relevant. Additional options for the job for drawing molecules, saving simulation data, etc. follow. Note that these input options are incomplete; the complete set of input file options can be found in the documentation. Several example input files are included within the `RMG-Py/examples/rmg/` folder.

The completed n-heptane pyrolysis RMG model contains 49 species and 638 reactions. Simulations of the flow tube pyrolysis experiments were carried out in CHEMKIN-PRO [46] for the RMG-generated model and the LLNL n-heptane detailed mechanism version 3.1 [47]. The major species found in experiment along with the simulations are shown in Figure 2.14. The RMG-generated model is a first-pass model constructed automatically solely using RMG's databases. It is able to match the LLNL model for predicting the species conversion of n-heptane. Sensitivity analysis and refinement of thermodynamic and kinetic parameters within the database typically follow in the model development cycle.

2.5 Design principles

RMG-Py is implemented using a new modularization design that improves upon its predecessor, RMG-Java, using smaller modules and packages grouped by more specific functionality. The database is distinctly separate from the code and stored in a separate RMG-database folder. RMG-Py has also been developed using the software principles of unit testing, strong error handling, integrated documentation, and distributed version control through GitHub. The shift to Python allows us to work with a number of existing cheminformatics libraries which provide a number

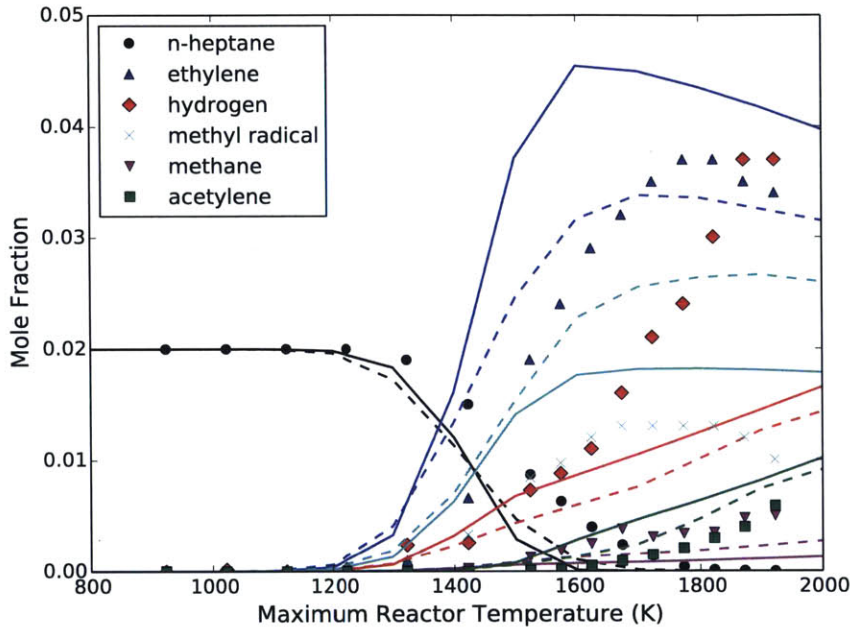


Figure 2.14: N-heptane flow tube experiments at P = 400 Pa for an initial mixture of 2% n-heptane and 98% argon. Simulated mole fraction profiles by the RMG-generated model (lines) and LLNL v3.1 model (dotted lines.)

of advanced features to RMG. Selective optimization of performance-critical areas of the code has been done through Cython [48, 49]. Cython compiles Python-style code to C through static typing, leading to over an order of magnitude or higher speed up for numerically intensive code. We are currently developing parallelization methods to further increase the speed and reduce memory requirements. The new version of RMG is designed with the user in mind, with a web front end to improve ease of use and transparency within the kinetics community.

2.6 Binary packaging using the Anaconda Python distribution

The advantage of migrating to the Python language is wide availability of Python-based cheminformatics libraries and dependencie. A major drawback of relying on a variety of libraries is the increased difficulty of installing a working version of the software given dependency compatibility issues across different operating systems and software versions. This is especially difficult when a user develops or compiles other

software using Python, which may lead to conflicting dependencies with RMG.

The Anaconda Python distribution [50], developed in 2013, was selected to address the problem of managing the Python environment for RMG. Anaconda has local desktop environment management and segregation, allowing software to be compiled with an isolated set of dependencies. It also has the capability to compile Anaconda-compatible software packages using simple recipes. Package management is freely accessible to individuals and groups, allowing us to upload RMG-specific dependency packages onto the Anaconda RMG channel found at `http://anaconda.org/rmg/`. These packages are then accessible to anyone by simply adding an extra flag to the installation commands referring to the rmg channel. Common dependencies such as numpy, scipy, and matplotlib are managed by Anaconda, and a variety of other libraries can be found precompiled, available to copy and distribute onto any number of channels. Using the Anaconda Python distribution allows us to maintain and manage RMG's dependencies on a single channel. Anaconda not only gives us the capability to create RMG binaries, it also enables developers to compile RMG by source without having to reinstall dependencies and easily switch between isolated Python environments when working on different projects. From installations that often took several hours or fail, RMG installation now takes less than 20 minutes across all major operating systems: Linux (32 and 64-bit), Mac OSX (64-bit), and Windows (32 and 64-bit).

More information on installing RMG through the Anaconda platform can be found in the RMG documentation online: `http://reactionmechanismgenerator.github.io/RMG-Py/`. Additional guides for compiling Anaconda binaries can be found on the RMG wiki: `https://github.com/ReactionMechanismGenerator/RMG-Py/wiki`.

2.7 Conclusion

RMG is one of the most widely used automatic reaction mechanism generation codes currently available. With the shift to Python, it now takes advantage of several existing chemistry libraries and is capable of constructing mechanisms for species involving carbon, hydrogen, oxygen, sulfur, and nitrogen. RMG's features include thermodynamic and kinetic parameter estimation, automatic generation of pressure-dependent rate coefficients, liquid phase solvation networks, transport property estimation, and sensitivity analysis. RMG has been developed for over a decade and has generated numerous validated reaction networks. This new version of RMG provides the most

advanced features for reaction mechanism generation in a single open-souce and easily extensible package, with a convenient web-based graphical user interface.

2.8 References

[1] A. S. Tomlin, T. Turányi, and M. J. Pilling. "Mathematical Tools for the Construction, Investigation and Reduction of Combustion Mechanisms." *Low-Temperature Combustion and Autoignition.* Vol. 35. Comprehensive Chemical Kinetics. Elsevier, 1997, pp. 293–437. DOI: 10.1016/S0069-8040(97)80019-2.

[2] S. W. Benson and J. H. Buss. "Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties." *J. Chem. Phys.* 29 (3), 1958, pp. 546–572. DOI: 10.1063/1.1744539.

[3] S. W. Benson. *Thermochemical Kinetics.* New York: John Wiley and Sons, 1968.

[4] G. R. Magoon and W. H. Green. "Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation." *Comp. & Chem. Eng.* 52, 2013, pp. 35–45. DOI: 10.1016/j.compchemeng.2012.11.009.

[5] R. G. Susnow, A. M. Dean, W. H. Green, P. Peczak, and L. J. Broadbelt. "Rate-based construction of kinetic models for complex systems." *J. Phys. Chem. A* 101 (20), 1997, pp. 3731–3740. DOI: 10.1021/jp9637690.

[6] L. J. Broadbelt and J. Pfaendtner. "Lexicography of Kinetic Modeling of Complex Reaction Networks." *AIChE J.* 51 (8), 2005, pp. 2112–2121. DOI: 10.1002/aic.10599.

[7] F. Battin-Leclerc, E. Blurock, R. Bounaceur, R. Fournet, P.-A. Glaude, O. Herbinet, B. Sirjean, and V. Warth. "Towards cleaner combustion engines through groundbreaking detailed chemical kinetic models." *Chem. Soc. Rev.* 40, 2011, pp. 4762–4782. DOI: 10.1039/C0CS00207K.

[8] R. Vinu and L. J. Broadbelt. "Unraveling Reaction Pathways and Specifying Reaction Kinetics for Complex Systems." English. *Annu. Rev. Chem. Biomol. Eng.* 3 (1), 2012, pp. 29–54. DOI: 10.1146/annurev-chembioeng-062011-081108.

[9] E. Blurock, F. Battin-Leclerc, T. Faravelli, and W. H. Green. "Automatic Generation of Detailed Mechanisms." *Cleaner Combustion.* Green Energy and Technology. Springer-Verlag London, 2013, pp. 59–92. DOI: 10.1007/978-1-4471-5307-8_3.

[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. "Automatic Mechanism and Kinetic Model Generation for Gas- and Solution-Phase Processes: A Perspective on Best Practices, Recent Advances, and Future Challenges." *Int. J. Chem. Kin.* 47 (4), 2015, pp. 199–231. DOI: 10.1002/kin.20902.

[11] J. Song. "Building Robust Chemical Reaction Mechanisms: Next Generation of Automatic Model Construction misctware." PhD thesis. Massachusetts Institute of Technology, 2004.

[12] W. H. Green, P. I. Barton, B. Bhattacharjee, D. M. Matheu, D. A. Schwer, J. Song, R. Sumathi, H. H. Carstensen, A. M. Dean, and J. M. Grenda. "Computer Construction of Detailed Chemical Kinetic Models for Gas-Phase Reactors." *Ind. Eng. Chem. Res.* 40 (23), 2001, pp. 5362–5370. DOI: 10.1021/ie001088s.

[13] J. M. Grenda, I. P. Androulakis, A. M. Dean, and W. H. Green. "Application of computational kinetic mechanism generation to model the autocatalytic pyrolysis of methane." *Ind. Eng. Chem. Res.* 42 (5), 2003, pp. 1000–1010. DOI: 10.1021/ie020581w.

[14] M. Harper, K. M. Van Geem, S. P. Pyl, G. B. Marin, and W. H. Green. "Comprehensive reaction mechanism for n-butanol pyrolysis and combustion." *Combust. Flame* 158 (1), 2011, pp. 16–41. DOI: 10.1016/j.combustflame.2010.06.002.

[15] J. W. Allen, A. M. Scheer, C. W. Gao, S. S. Merchant, S. S. Vasu, O. Welz, J. D. Savee, D. L. Osborn, C. Lee, S. Vranckx, Z. Wang, F. Qi, R. X. Fernandes, W. H. Green, M. Z. Hadi, and C. A. Taatjes. "A coordinated investigation of the combustion chemistry of diisopropyl ketone, a prototype for biofuels produced by endophytic fungi." *Combust. Flame* 161 (3), 2014, pp. 711–724. DOI: 10.1016/j.combustflame.2013.10.019.

[16] G. R. Magoon, J. Aguilera-Iparraguirre, W. H. Green, J. J. Lutz, P. Piecuch, H. Wong, and O. O. Oluwole. "Detailed Chemical Kinetic Modeling of JP-10 (exo-Tetrahydrodicyclopentadiene) High-Temperature Oxidation: Exploring the Role of Biradical Species in Initial Decomposition steps." *Int. J. Chem. Kin.* 44 (3), 2012, pp. 179–193. DOI: 10.1002/kin.20702.

[17] S. V. Petway, H. Ismail, W. H. Green, E. G. Estupinan, L. E. Jusinski, and C. A. Taatjes. "Measurements and automated mechanism generation modeling of OH production in photolytically initiated oxidation of the neopentyl radical." *J. Phys. Chem. A* 111 (19), 2007, pp. 3891–3900. DOI: 10.1021/jp0668549.

[18] J. W. Allen. "Predictive Chemical Kinetics: Enabling Automatic Mechanism Generation and Evaluation." PhD thesis. Massachusetts Institute of Technology, 2013.

[19] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. "A (sub)graph isomorphism algorithm for matching large graphs." *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (10), 2004, pp. 1367–1372. DOI: 10.1109/TPAMI.2004.75.

[20] T. H. Lay, J. W. Bozzelli, A. M. Dean, and E. R. Ritter. "Hydrogen Atom Bond Increments for Calculation of Thermodynamic Properties of Hydrocarbon Radical Species." *J. Phys. Chem.* 99 (39), 1995, pp. 14514–14527. DOI: 10.1021/j100039a045.

[21] G. Landrum. *RDKit.* URL: http://rdkit.org.

[22] J. W. Allen, C. F. Goldsmith, and W. H. Green. "Automatic estimation of pressure-dependent rate coefficients." *Phys. Chem. Chem. Phys.* 14 (3), 2012, pp. 1131–1155. DOI: 10.1039/C1cp22765c.

[23] A. Y. Chang, J. W. Bozzelli, and A. M. Dean. "Kinetic Analysis of Complex Chemical Activation and Unimolecular Dissociation Reactions using QRRK Theory and the Modified Strong Collision Approximation." *Z. Phys. Chem.* 214, 2000, pp. 1533–1568. DOI: 10.1524/zpch.2000.214.11.1533.

[24] N. J. B. Green and Z. A. Bhatti. "Steady-State Master Equation Methods." *Phys. Chem. Chem. Phys.* 9, 2007, pp. 4275–4290. DOI: 10.1039/b704519k.

[25] J. A. Miller and S. J. Klippenstein. "Master Equation Methods in Gas Phase Chemical Kinetics." *J. Phys. Chem. A* 110, 2006, pp. 10528–10544. DOI: 10.1021/jp062693x.

[26] A. Jalan, R. H. West, and W. H. Green. "An Extensible Framework for Capturing Solvent Effects in Computer Generated Kinetic Models." *J. Phys. Chem. B* 117 (10), 2013, pp. 2955–2970. DOI: 10.1021/jp310824h.

[27] M. J. Kamlet, J. L. M. Abboud, M. H. Abraham, and R. W. Taft. "Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π, α, and β, and some methods for simplifying the generalized solvatochromic equation." *J. Org. Chem.* 48 (17), 1983, pp. 2877–2887. DOI: 10.1021/jo00165a018.

[28] R. W. Taft, J.-L. M. Abboud, M. J. Kamlet, and M. H. Abraham. "Linear solvation energy relations." *J. Solution Chem.* 14 (3), 1985, pp. 153–186. ISSN: 0095-9782. DOI: 10.1007/BF00647061.

[29] J. A. Platts, D. Butina, M. H. Abraham, and A. Hersey. "Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach." *J. Chem. Inf. Comput. Sci.* 39 (5), 1999, pp. 835–845. DOI: 10.1021/ci980339t.

[30] C. J. Cramer and D. G. Truhlar. "Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics." *Chem. Rev.* 99 (8), 1999, pp. 2161–2200. DOI: 10.1021/cr960149m.

[31] C. Mintz, M. Clark, K. Burton, W. E. Acree Jr., and M. H. Abraham. "Enthalpy of solvation correlations for gaseous solutes dissolved in toluene and carbon tetrachloride based on the Abraham model." *J. Solution Chem.* 36 (8), 2007, pp. 947–966. DOI: 10.1007/s10953-007-9163-0.

[32] C. Mintz, M. Clark, K. Burton, W. E. Acree Jr., and M. H. Abraham. "Enthalpy of solvation correlations for gaseous solutes dissolved in benzene and in alkane solvents based on the Abraham model." *QSAR Comb. Sci.* 26 (8), 2007, pp. 881–888. DOI: 10.1002/qsar.200630152.

[33] C. Mintz, T. Ladlie, K. Burton, M. Clark, W. E. Acree Jr., and M. H. Abraham. "Enthalpy of solvation correlations for gaseous solutes dissolved in alcohol solvents based on the Abraham model." *QSAR Comb. Sci.* 27 (5), 2008, pp. 627–635. DOI: 10.1002/qsar.200730128.

[34] S. A. Rice. *Comprehensive Chemical Kinetics: Diffusion-limited reactions.* Elsevier, 1985.

[35] F. C. Collins and G. E. Kimball. "Diffusion-controlled reaction rates." *J. Colloid Sci.* 4 (4), 1949, pp. 425–437. ISSN: 0095-8522. DOI: 10.1016/0095-8522(49)90023-9.

[36] M. H. Abraham and J. C. McGowan. "The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography." English. *Chromatographia* 23 (4), 1987, pp. 243–246. ISSN: 0009-5893. DOI: 10.1007/BF02311772.

[37] M. R. Harper Jr. "Automated reaction mechanism generation: data collaboration, heteroatom implementation, and model validation." PhD thesis. Massachusetts Institute of Technology, 2011.

[38] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, R. K. H. C. T. Bowman, S. Song, J. W. C. Gardiner, V. V. Lissianski, and Z. Qin. *GRI-Mech 3.0.* 2011. URL: http://www.me.berkeley.edu/gri_mech/.

[39] L. S. Tee, S. Gotoh, and W. E. Stewart. "Molecular Parameters for Normal Fluids. Lennard-Jones 12-6 Potential." *Ind. Eng. Chem. Res.* 5 (3), 1966, pp. 356–363. DOI: 10.1021/i160019a011.

[40] K. G. Joback. "A unified approach to physical property estimation using multivariate statistical techniques." MA thesis. Massachusetts Institute of Technology, 1984.

[41] K. G. Joback and R. C. Reid. "Estimation of pure-component properties from group-contributions." *Chem. Eng. Comm.* 57 (1-6), 1987, pp. 233–243. DOI: 10.1080/00986448708960487.

[42] L. Petzold and S. Li. *DASPK3.1.* URL: http://www.cs.ucsb.edu/~cse/software.html.

[43] S. Sharma, M. R. Harper, and W. H. Green. *CanTherm v1.0.* URL: http://cantherm.sourceforge.net/.

[44] D. G. Goodwin, H. K. Moffat, and R. L. Speth. *Cantera: An Object-oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes.* Version 2.2.1. 2016. URL: http://www.cantera.org.

[45] T. Yuan, L. Zhang, Z. Zhou, M. Xie, L. Ye, and F. Qi. "Pyrolysis of n-Heptane: Experimental and Theoretical Study." *J. Phys. Chem. A* 115 (9), 2011, pp. 1593–1601. DOI: 10.1021/jp109640z.

[46] *CHEMKIN-PRO 15131.* Reaction Design: San Diego. 2013.

[47] M. Mehl, W. J. Pitz, C. K. Westbrook, and H. J. Curran. "Kinetic modeling of gasoline surrogate components and mixtures under engine conditions." *Proc. Combust. Inst* 33 (1), 2011, pp. 193–200. DOI: 10.1016/j.proci.2010.05.027.

[48] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. S. Seljebotn, and K. Smith. "Cython: The Best of Both Worlds." *Comp. Sci. Eng.* 13 (2), 2011, pp. 31–39. DOI: 10.1109/MCSE.2010.118.

[49] R. Bradshaw, S. Behnel, D. S. Seljebotn, G. Ewing, et al. *The Cython compiler.* URL: http://cython.org/.

[50] C. Analytics. *Anaconda Software Distribution.* Vers. 2-2.4.0. 2015. URL: http://continuum.io.

3

Screening fungi-derived biofuel candidates: a low temperature oxidation model for cineole

Conversion of lignocellulosic biomass to fuels has been a subject of interest driven by an era of high oil prices and a desire for a sustainable, carbon-neutral fuel economy. The first generation of biofuels have been derived primarily from sugar, starch and oil crops, producing biodiesel, ethanol, and syngas. Due to the nature of these feedstocks being tied to global food production, research has been directed toward second-generation biofuels derived from lignocellulosic biomass. [1] Although biofuels are derived from renewable sources and thus have lower net greenhouse gas emissions than fossil fuels, their economic viability for large-scale commercialization remains an open question. Furthermore, biomass-derived biofuel candidates are not typically compounds found in conventional fuels, leading to the question of their compatibility with current and future engine technologies. In particular, advanced clean, efficient combustion engines which rely on compression ignition are very sensitive to fuel oxidation chemistry. [2, 3]

Recently, several strains of endophytic fungi which can directly break down cellulose into potential fuel compounds have been discovered by researchers at Montana State University. [4] This discovery could potentially eliminate the expensive decomposition step required in most conventional processes for lignocellulosic biofuel production. Isolating and scaling up the fungi metabolic processes through biological engineering may be a promising route toward cellulosic biofuel generation, especially given their relatively minimal fungi genomes. [5] These isolated metabolic pathways can then be expressed in more tractable organisms such as *E. coli* to maximize fuel production. [6]

The fungi produce a large selection of novel hydrocarbon compounds with widely unknown fuel chemistry, including a variety of ketones, ethers, and terpenoids. [7, 8] Conventional diesel and gasoline cannot be obtained in pure form. It is believed that

bioengineering can improve the selectivity of certain products, but product optimization cannot occur without knowledge of which fuel targets are viable. Fundamental combustion experiments on the representative compounds and detailed theoretical kinetics are needed to provide feedback on the desirability of specific molecules as fuels.

To address these issues comprehensively, coordinated efforts towards metabolic engineering and biofuel-engine co-development are required. Figure 3.1 depicts the collaborative framework. Combustion researchers develop fundamental mechanisms for the combustion of potential biofuels that have been identified by synthetic biologists. Ignition and engine trials then provide feasibility tests for the fuels and yield recommendations for the bioengineering scale-up of specific metabolic pathways. This coupling of fundamental and applied combustion chemistry with synthetic biology is a strategy to identify and investigate the most promising fuel compounds through mutual feedback. [6, 9]

First, the spectrum of volatile organic compounds are profiled from the natural fungal metabolization of biomass. From these compounds, candidates for further investigation are chosen based on both (1) selectivity and prominence in the product yield and (2) lack of existing combustion chemistry knowledge. In this chapter, two model compounds that fit this criteria were studied: DIPK (diisopropyl ketone, Figure 3.2) and cineole (1,3,3-trimethyl-2-oxabicyclo[2,2,2]octane, Figure 3.3). Fundamental experimental measurements of these two compounds are combined with detailed theoretical kinetics efforts using RMG to develop predictive combustion models that can simulate these fuels in engines. These models are intended to validate against HCCI engine performance measurements and provide feedback in the screening and recommendation of viable biofuel candidates.

The detailed kinetic modeling of DIPK and experimental validation has been published in *Combustion and Flame* [9] and also described in Josh Allen's thesis [10]. Strong NTC (negative temperature coefficient) behavior was observed in the ignition behavior owing to the stable energy of the fuel's resonance-stabilized tertiary alkyl radical. HCCI engine experiments were conducted for DIPK and compared against conventional gasoline and neat ethanol performance in an *SAE* paper. [11] DIPK demonstrated high temperature and pressure sensitivity compared to gasoline and ethanol, while demonstrating lower autoignition reactivity. In summary, DIPK operates at a higher thermal efficiency than gasoline at comparable loads and does not require combustion timing retardation, making it suitable for use in HCCI engines.

The experimental and kinetic modeling study of cineole is the primary focus of

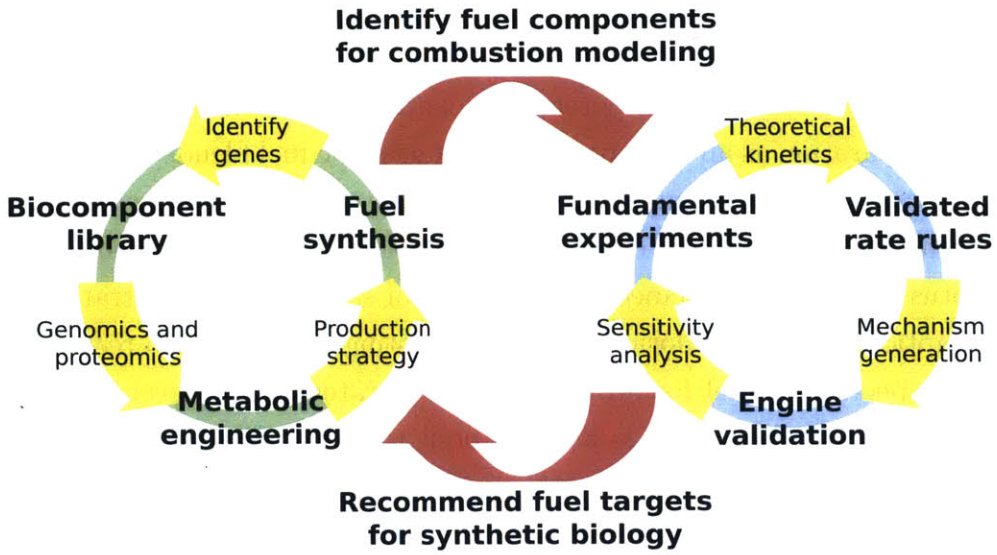


Figure 3.1: The collaborative lignocellulosic biofuel research effort framework for integrating combustion chemistry and engine performance feedback into the metabolic engineering of fungi-generated biofuels. From Allen *et al.* [9]

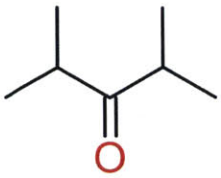


Figure 3.2: DIPK ($C_7H_{14}O$); diisopropyl ketone; 114 amu

Figure 3.3: Cineole ($C_{10}H_{10}O$); 1,3,3-trimethyl-2-oxabicyclo[2,2,2]octane; 154 amu

this chapter. Experiments were performed by collaborators Adam M. Scheer and Craig A. Taatjes at the Combustion Research Facility at Sandia National Laboratories. Cineole, sometimes referred to as eucalyptol, is a saturated cyclic ether and monoterpenoid found naturally in the essential oil of the Eucalyptus tree. Due to its pleasant aroma, cineole is traditionally used in fragrances and flavorings at low

concentrations [12, 13] and is also used in non-prescription pharmaceuticals [14, 15] and solvent degreasers. [16] Cineole is relatively non-toxic and fully biodegradable, making it a favorable candidate in potential fuel applications. It has been shown to prevent phase separation as an additive in ethanol-gasoline fuel blends in the presence of water, [17] as well as boosting octane number due to its limited reactivity. [18, 19] Widespread industrial use of cineole is limited due to its primary source being the Eucalyptus tree. However, cineole is produced in significant concentrations from fungi consumption of lignocellulosic biomass, prompting researchers to reconsider cineole as a prospective biofuel for advanced engine strategies. However, despite the interest in cineole as a fuel, little is known about the oxidation chemistry of fused cyclic ethers.

3.1 Methodology

The low temperature oxidation pathways of cineole were studied through a combination of experiment and theory. Chlorine-initiated oxidation using multiplexed photoionization time-of-flight mass spectrometry (MPIMS) [20] was used to identify products as a function of mass, time, and photoionization energy, allowing for the isolation of species intermediates. Qualitative mechanistic information is available from the experimental observations, and quantitative determinations can be made for some reaction products. Nevertheless, the complexity of the cineole oxidation reaction pathways makes complete detailed analysis of the intermediates and products infeasible and must therefore include input from theory. Quantum chemical calculations were performed to determine thermochemical parameters of key species and kinetic rates for initial abstraction reactions and major pathways. These theoretical values were used in conjunction with RMG to generate a detailed kinetic model for the low temperature oxidation of cineole.

The detailed kinetic model will assist the identification of peaks generated by the MPIMS experiments. The RMG-generated model will provide guesses for isomers found in experiment. Conflicts between the mechanism and data will be investigated by performing quantum calculations for specific isomers to deduce photoionization spectra or by making recommendations for further experimentation.

3.1.1 Chlorine-initiated oxidation

Low temperature oxidation of fuels can be probed using Cl-initiated oxidation. First, a flow tube reactor is filled with a homogeneous mixture containing a small concentration of target fuel, excess oxygen, trace Cl_2, and inert bath gas at low pressure. A

pulsed photolysis laser beam photolyzes a small amount of Cl_2 into Cl atoms. The Cl atoms quickly react with the target fuel, generating an R radical and propagating a series of reactions between the R radical and excess oxygen molecules. The resulting products can be analyzed through a multiplexed photoionization time-of-flight mass spectrometry (MPIMS) as a function of mass, time and photoionization energy. Previous publications describe the MPIMS methodology and 3-dimensional data analysis in detail [20, 21] and is briefly described here.

The flow tube reactor is side-sampled during the reaction and irradiated with tunable photoionizing radiation to form ions that are detected using a time-of-flight mass spectrometer. This allows isomers to be differentiated not only by molecular weight but also their photoionization spectra. However, the 3D data must first be decoupled to give quantitative concentrations of individual isomers. The signal intensity for a given ion mass peak z^+ at any given time t and photoionization energy eV can be represented as:

$$I_{z^+}(t, eV) = \sum_i c_i(t) g_{i \to i^+}(eV) f_{i^+ \to z^+}(eV) R(z^+) \tag{3.1}$$

where c_i is the concentration of isomer i, g is the function describing the fraction ionized, and f is the function describing the fraction of ions that convert to ions of mass peak z^+, and R is a mass-dependent response factor that accounts for the different sampling and detection efficiencies for ions with different masses. Signals are difficult to transform back to concentrations due to the unknown functions g and f. Lack of known photoionization spectra for most isomers (particularly radical intermediates) make it difficult to construct g without isolated experiments using known standards. Due to the nature of the photoionizing radiation, isomers also fragment through dissociative ionization during sampling, leading to a non-unity function f. The combined g and f terms can be collectively understood to be the absolute photoionization cross section σ for a single isomer:

$$\sigma_{i,z^+}(eV) = g_{i \to i^+}(eV) f_{i^+ \to z^+}(eV) \tag{3.2}$$

which is known for only a small number of compounds and difficult and time-consuming to construct for others.

The cineole experiments were carried out by Dr. Craig Taatjes and coworkers at the Combustion Research Facility in Sandia National Laboratories. The tunable photoionization radiation was provided by the Synchrotron at the Advanced Light Source (ALS) in Lawrence Berkeley National Laboratory (LBNL).

Table 3.1: Initial concentrations (molecules/cm^3) in the cineole Cl-initiated oxidation experiments. Helium was added to reach a total pressure of 4 Torr in all experiments.

Temperature (K)	[Cineole]	[O_2]	[Cl_2]	[Cl]
550	2.6×10^{13}	1.8×10^{16}	5.9×10^{13}	1.8×10^{12}
650	2.2×10^{13}	1.5×10^{16}	5.0×10^{13}	1.5×10^{12}

The vapor pressure of cineole is low (1.34 Torr at 20 °C) [22] and a bubbler held at 20 °C backed by He was used to introduce cineole into the flow tube. Data was recorded with both single photoionization energies and energy scans at 550 and 650 K. The initial concentrations used for the 550 and 650 K oxidation experiments are given in Table 3.1. In all experiments helium was added to a total pressure of 4 Torr.

Cineole is a large molecule prone to dissociative photoionization. Therefore, it is important to identify all fragment ions of the parent cineole before an analysis of oxidation data. Figure 3.4 gives the partial photoionization cross-section of cineole as a function of photoionization energy. This spectrum was obtained by flowing a known concentration of cineole along with calibration gas standards of known photoionization cross-sections. The cineole signal at 10.4 eV was pinned to that of a propene standard at the same energy and the rest of the spectrum is normalized accordingly. The largest source of error anticipated in such a measurement involves uncertainty in the flow rates maintained by mass flow controllers used to introduce gases into the reactor tube. This uncertainty is estimated at ±20%. It is observed that the ionization onset for cineole begins at 8.5 eV followed by a steady increase until approximately 10.4 eV before leveling off.

Figure 3.5 shows a calibration mass spectrum of cineole and its fragment ions observed by integrating the ion signal from 8.5 – 11.0 eV. Prominent fragment ions appear at m/z = 139, 136, 125, 96 and 84. No fragment ions below m/z = 59 are observed for photon energies up to 11.0 eV.

Mass spectra shown in the remainder of this chapter for the photolytically initiated reactions are difference mass spectra, subtracting the pre-photolysis signal of cineole and its fragment ions from the post-photolysis ion-signal. This analysis isolates the changes in concentration because of photolysis and chemical reaction; removal of precursor molecules appears as negative-signal peaks (e.g., for cineole), and products appear as positive peaks.

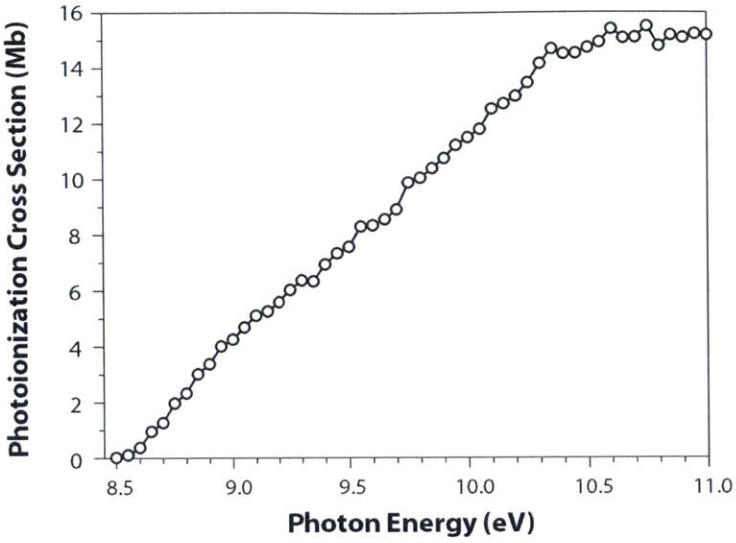


Figure 3.4: Absolute partial photoinization spectrum of cineole (m/z =154). The ionization onset is observed near 8.5 eV followed by a steady rise before leveling off at approximately 10.4 eV.

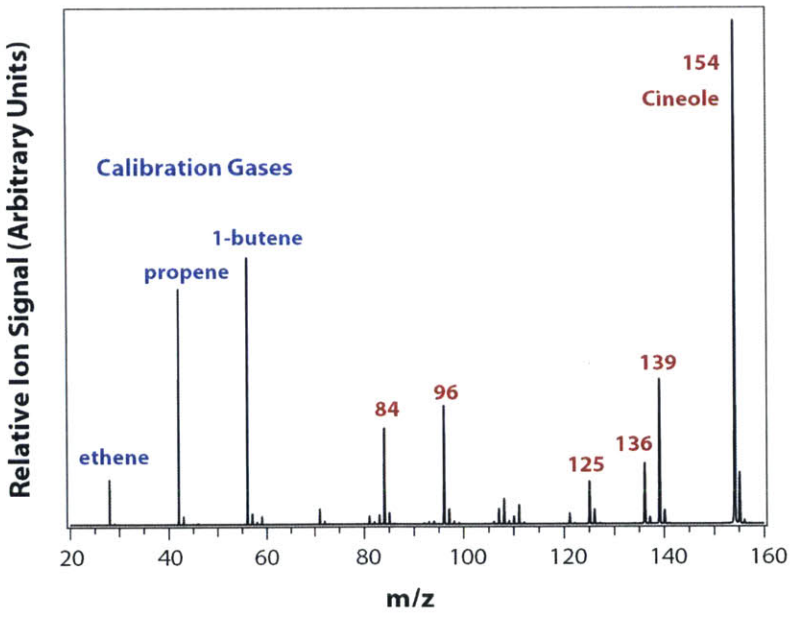


Figure 3.5: Integrated ion signal (8.5 – 11.0 eV) for cineole and its fragment ions. Calibration standards run simultaneously are also shown.

3.1.2 Quantum chemical calculations

Prior to developing a kinetic mechanism for the reactions between cineole and Cl, the thermodynamics of key species and kinetics of initial abstraction reactions were obtained from quantum calculations at the CBS-QB3 level of theory. [23] CBS-QB3 is a reliable and relatively low-cost composite method that has proven throughout literature to yield accurate thermochemical and kinetic data. The CSB-QB3 composite method consist of a geometry optimization and frequency analysis at the B3LYP/6-311G(2d,d,p) level, followed by single-point calculations on the CCSD(T)/6-31+G(d') and MP4SDQ/CBSB4 levels. The total energy is obtained from a basis-set extrapolation using MP2/CBSB3 pair natural orbital energies. For reactions having more than 11 heavy atoms (mainly oxidation reactions of cineole), the BMK/6-311G(2d,d,p) density functional method was used instead of CBS-QB3. [24] This hybrid meta GGA-functional has been recommended for main-group thermochemistry and kinetics. [25]

For reactions having a well-defined barrier, rate coefficients were calculated with classical transition state theory using the harmonic oscillator formalism for rovibrational motions. Tunneling corrections were obtained from the Eckart formula. [26] For some reactions, mainly hydrogen abstraction reactions by Cl from cineole, submerged barriers were found. In some cases saddle points were found that were close to 10 kcal/mol below the energy of the reactants. For these specific cases, a canonical two transition state theory model was used:

$$\frac{1}{k_{H_{abstraction}}} = \frac{1}{k_{inner}(T)} + \frac{1}{k_{outer}(T)} \tag{3.3}$$

The two transition state model is used because for the reaction to occur, the reactants must pass two bottlenecks. A first one, the outer transition state, is characterized with the formation of a van der Waals minimum while the second bottleneck, or inner transition state, corresponds to the actual breaking and forming of new bonds. The inner transition state can easily be located by a saddle point search on the potential energy surface. The outer transition state can only be obtained using a variational approach, in which care needs to be taken to accurately treat the modes that describe the relative position of the two interacting fragments. At larger separations these modes can no longer be approximated as harmonic oscillators. To circumvent this problem k_{outer} was estimated to be temperature independent and equal to $8 \times 10^{13}\,\mathrm{cm}^3/(\mathrm{mol}\cdot\mathrm{s})$ for each equivalent C site that can participate in the abstraction reaction. This value agrees well with the measured value of $8.97 \times 10^{13}\,\mathrm{cm}^3/(\mathrm{mol}\cdot\mathrm{s})$

for hydrogen abstraction of a tertiary hydrogen atom from *iso*-butane [27] by Cl and leads to rate coefficients for hydrogen abstraction from a primary C atom that agree within a factor of 2 with those obtained for hydrogen abstraction from ethane. [28] At the temperatures of interest for this paper (600 K), the rate coefficients for hydrogen abstraction reactions by Cl are mainly limited by k_{outer}.

The initial hydrogen abstraction of cineole can occur at multiple sites, as shown in Figure 3.6, leading to alkyl radicals R1-R5. All possible hydrogen abstraction kinetics of cineole with chlorine radical and with methyl radical were computed using the CBS-QB3 method. The resulting rate coefficients can be found in Table 3.2. These rate coefficients are important as they will determine the initial alkyl radical pool in the system. It is found that at the temperature of interest (600 K) the highest rate coefficients are obtained for hydrogen abstraction by Cl of one of the secondary H atoms, i.e. leading to R2 and R4. These hydrogen abstraction reactions are approximately a factor 2 faster than for abstraction from the tertiary site (R5). As all these abstraction reactions are barrierless, this difference is mainly caused by the reaction path degeneracy. For abstraction by CH_3 significant barriers are obtained for the reactions, and in this case the tertiary site is preferred.

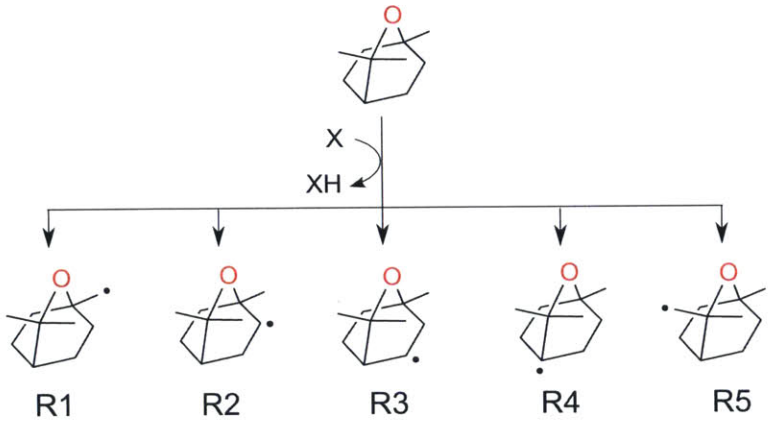


Figure 3.6: Hydrogen abstractions from cineole lead to five possible alkyl radicals.

Decomposition pathways for the dominant R2, R4 and R5 alkyl radicals have also been theoretically studied. These β-scissions and isomerization reactions are shown in Figure 3.7 to 3.11 while the calculated CBS-QB3 Arrhenius parameters can be found in Table 3.3. The calculated rates are compared with the RMG estimated rate for the addition of O_2 to the radical site. A direct comparison is possible when this rate coefficient is multiplied with the O_2 concentration. The rate coefficients at 600 K presented in Table 3.3 show that the various β-scission reactions in R2 are

Table 3.2: CBS-QB3 Arrhenius parameters and rate coefficients at 600K for hydrogen abstraction reactions by chlorine and methyl radical from cineole. The resulting radicals R1 to R5 are shown in Figure 3.6. A two-transition state approach was used to obtain data for reactions involving Cl radicals. For hydrogen abstraction from cineole leading to R5 no inner transition state could be found.

Reaction	A ($\frac{cm^3}{mol \cdot s}$)	n	E_a (kcal/mol)	k(600 K) ($\frac{cm^3}{mol \cdot s}$)
Cineole + Cl $\rightleftarrows$ R_1 + HCl	2.9×10^{11}	0.65	-1.1	4.7×10^{13}
Cineole + Cl $\rightleftarrows$ R_2 + HCl	1.7×10^{15}	-0.33	0.30	1.6×10^{14}
Cineole + Cl $\rightleftarrows$ R_3 + HCl	6.0×10^{9}	1.19	-1.9	6.0×10^{13}
Cineole + Cl $\rightleftarrows$ R_4 + HCl	1.7×10^{16}	-0.67	0.5	1.5×10^{14}
Cineole + Cl $\rightleftarrows$ R_5 + HCl	8.0×10^{13}	0.00	0.0	8.0×10^{13}
Cineole + CH_3 $\rightleftarrows$ R_1 + CH_4	1.7×10^{-4}	4.90	7.6	1.2×10^{7}
Cineole + CH_3 $\rightleftarrows$ R_2 + CH_4	1.5×10^{-3}	4.62	5.3	1.2×10^{8}
Cineole + CH_3 $\rightleftarrows$ R_3 + CH_4	1.6×10^{-4}	4.91	7.4	1.4×10^{7}
Cineole + CH_3 $\rightleftarrows$ R_4 + CH_4	5.3×10^{-4}	4.66	5.2	6.0×10^{7}
Cineole + CH_3 $\rightleftarrows$ R_5 + CH_4	3.9×10^{-3}	4.58	5.4	2.2×10^{8}

slow. The fastest reaction is the β-scission leading to R2c for which the CBS-QB3 rate coefficient at 600 K is $3.5 \times 10^3\ s^{-1}$. R2c will in a subsequent step lead to the formation of acetone and a 1-methyl-cyclohex3-en-1-yl radical. Due to the relatively slow β-scission reactions in R2, addition reactions of O_2 to this radical competes with the non-oxidative decomposition path. Hydrogen abstraction reactions in R2OO followed by a substitution reaction lead to cyclic ethers with a mass of 168 amu. A slower rate coefficient for $R_2 \rightleftarrows R_2c$ will hence lead to more cyclic ethers. Because the network is very sensitive to this reaction, and CBS-QB3 is known to overestimate rate coefficients for β-scissions of this type, [29] the BMK/6-311G(2d,d,p) rate coefficient was used instead for network generation. This value is shown in Table 3.3 and is approximately one order of magnitude smaller than the CBS-QB3 value. R4 will mainly decompose by β-scission to R4b. The data in Table 3.3 show that at 600 K β-scissions of a C−O bond ($1.6 \times 10^6\ s^{-1}$) is approximately 3 orders of magnitude faster than scission of a C−C bond ($6.0 \times 10^3\ s^{-1}$). The R4b radical swiftly decomposes to acetone and cycC7H11-2 ($1.5 \times 10^{11}\ s^{-1}$). Also R5 will predominantly decompose by β-scission reactions, as a C−O bond in a β position is available. The resulting radical R5a is found to further decompose to R5a_bscis, which will in subsequent steps react with O_2.

Rate coefficients for the hydrogen abstractions and decomposition reactions shown

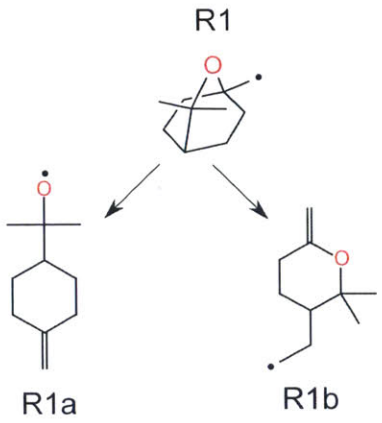


Figure 3.7: Major Reaction Pathways for R1 Alkyl Radical.

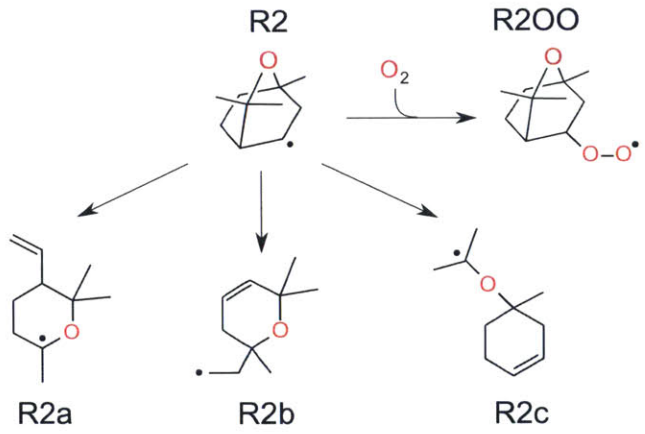


Figure 3.8: Major Reaction Pathways for R2 Alkyl Radical.

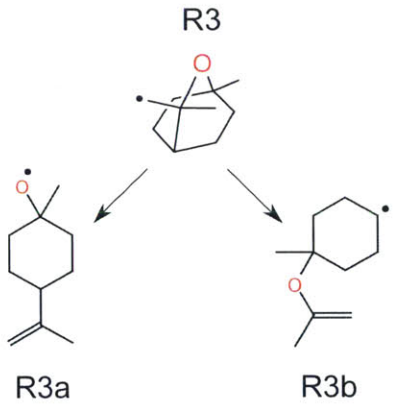


Figure 3.9: Major Reaction Pathways for R3 Alkyl Radical.

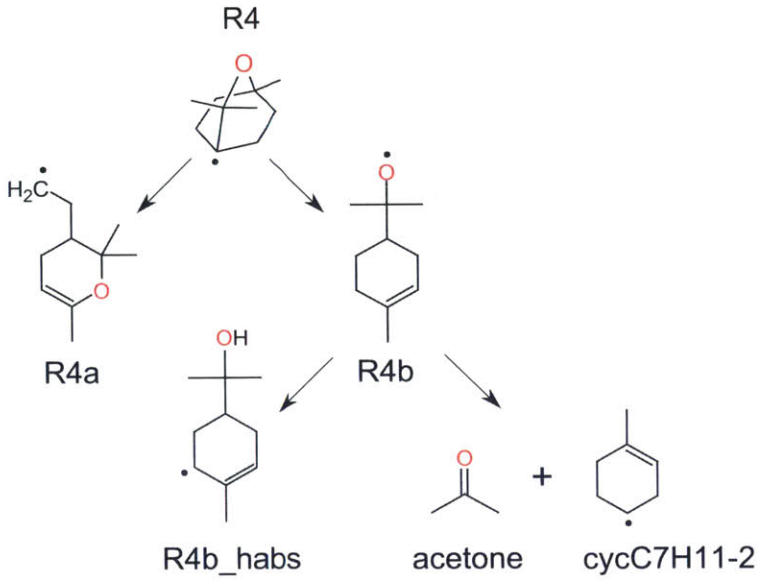


Figure 3.10: Major Reaction Pathways for R4 Alkyl Radical.

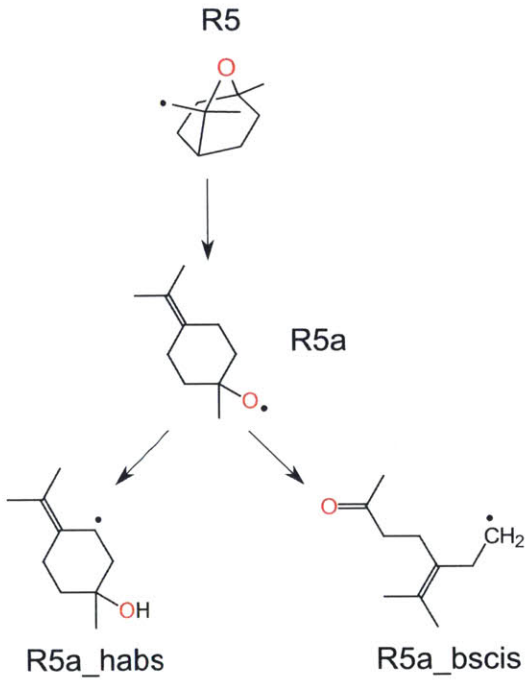


Figure 3.11: Major Reaction Pathways for R5 Alkyl Radical.

in Table 3.2 and Table 3.3 were gathered in a seed mechanism for RMG mechanism generation.

The thermochemistry of the major radicals and decomposition products were com-

Table 3.3: Important reactions for major alkyl radicals of cineole. High-pressure limit Arrhenius parameters and rate coefficients at 600 K for the β-scission reactions were calculated using the CBS-QB3 method. Values in parentheses were calculated using BMK/6-311G(2d,d,p).

Reaction	A (s^{-1})	n	E_a (kcal/mol)	k(600 K) (s^{-1})
$R_1 \rightleftarrows R_1a$	2.5×10^{12}	0.080	24.3	5.9×10^3
$R_1 \rightleftarrows R_1b$	3.6×10^{11}	0.446	26.0	2.1×10^3
$R_2(^+O_2) \rightleftarrows R_2OO$	1.9×10^5 *	0.0	0.0	1.9×10^5
$R_2 \rightleftarrows R_2a$	1.4×10^{12}	0.323	32.2	2.1×10^1
$R_2 \rightleftarrows R_2b$	4.3×10^{11}	0.596	28.9	5.8×10^2
$R_2 \rightleftarrows R_2c$	3.5×10^{12}	0.261	26.7	3.4×10^3
	(4.32×10^{12})	(0.18)	(29.8)	(2.0×10^2)
$R_3 \rightleftarrows R_3a$	3.8×10^{12}	-0.058	21.8	3.2×10^4
$R_3 \rightleftarrows R_3b$	1.1×10^{12}	.310	27.8	6.0×10^2
$R_4 \rightleftarrows R_4a$	6.1×10^{11}	0.500	25.8	6.0×10^3
$R_4 \rightleftarrows R_4b$	1.2×10^{13}	0.071	19.4	1.6×10^6
$R_4b \rightleftarrows R_4b\_habs$	5.6×10^7	1.37	3.5	1.9×10^{10}
$R_4b \rightleftarrows$ acetone + $cycC_7H_{11}-2$	8.2×10^{11}	0.63	6.8	1.5×10^{11}
$R_5 \rightleftarrows R_5a$	2.0×10^{12}	0.46	18.4	7.8×10^5
$R_5a \rightleftarrows R_5a\_habs$	2.5	3.81	7.5	9.8×10^7
$R_5a \rightleftarrows R_5a\_bscis$	4.4×10^{11}	0.75	9.5	9.7×10^9

* Assuming an oxygen concentration of 1.5×10^{16} molecules/cm^3 (see Table 3.2). From Curran *et al.* [30]

puted at the CBS-QB3 level and listed in Table 3.4.

3.1.3 Mechanism generation

The model was generated in RMG-Java at experimental conditions at temperatures of 550 and 650 K and 4 Torr pressure. The reacting species were subject to a constraint of a maximum of 10 carbon atoms, 5 oxygen atoms, and two radical electrons per species. Additionally, a thermodynamic library for key species in cineole decomposition was used to override RMG's estimates. Two seed mechanisms were used to provide a reaction framework from which RMG would proceed: the first seed mechanism contains small molecule reactions taken from the EFRCv0.3 mechanism [31] which are highly influential and strongly pressure-dependent; the second seed mechanism contains initial cineole decomposition and oxidation rate parameters computed via quantum chemistry. The final model contains 211 species reacting in 537 reactions.

Table 3.4: Thermochemistry of important radicals in cineole oxidation chemistry. ΔH in units of kcal/mol, and ΔS, C_p in units of cal/(mol · K)

Species	$\Delta H^{\circ}_{f,298K}$	$\Delta S^{\circ}_{f,298K}$	$C_p(300K)$	$C_p(400K)$	$C_p(500K)$	$C_p(600K)$	$C_p(800K)$	$C_p(1000K)$	$C_p(1500K)$
Cineole	-82.51	99.96	48.22	64.02	78.05	89.87	108.38	122.01	141.90
R1	-30.76	103.90	49.22	64.43	77.67	88.74	106.24	119.05	137.58
R2	-36.84	102.49	49.42	64.34	77.30	88.48	106.27	119.14	137.52
R3	-31.61	104.56	49.35	64.56	77.73	88.76	106.29	119.10	137.61
R4	-35.70	102.56	49.21	64.39	77.38	88.47	106.18	119.05	137.53
R5	-30.86	100.49	48.19	63.54	76.79	87.93	105.70	118.67	137.37
R2a	-15.44	109.39	50.96	65.01	77.76	88.94	106.38	118.74	137.49
R2b	-15.19	105.79	51.07	65.24	77.56	88.22	105.26	117.69	135.69
R2c	-17.01	112.26	50.86	64.75	77.47	88.74	106.60	118.97	137.58
R2d	-52.20	97.67	47.17	62.062	74.60	85.36	102.48	114.86	132.55
R2e	-5.10	97.74	47.62	62.55	75.054	85.76	102.79	115.11	132.70
R4a	-20.46	111.13	52.07	66.18	78.83	89.81	106.83	119.10	137.62
R4b	-17.46	108.70	50.07	64.57	77.63	88.88	106.16	118.76	137.55
R4c	3.46	91.50	41.25	54.78	66.28	76.00	91.33	102.30	117.69

RMG estimates thermodynamic quantities using the Benson group additivity scheme; however, group additivity is known to poorly estimate the thermochemistry of cyclic and fused cyclic compounds, which are highly important in the cineole mechanism. Therefore, on-the-fly semi-empirical PM7 calculations were performed for cyclic and polycyclic species for which an exact ring correction cannot be found. These calculations are performed by outsourcing molecular geometries to MOPAC2012 and returning the relevant thermodynamic parameters to RMG. For radical species, HBI (hydrogen bond increment) corrections are applied to PM7 calculations for the saturated parent compounds. For more details, refer to Chapter 2.

3.2 Results and discussion

The general oxidation pathways of fuel radicals (R) have been extensively investigated in the literature. [2, 32] A partial schematic of the low temperature oxidation pathways for cineole is shown in Figure 3.12. Coproducts of the chain-propagating OH-elimination channels from QOOH are expected to appear at m/z = 168. Similarly, the unsaturated coproducts of the chain-terminating HO2-elimination channels are expected at m/z = 152. The initial alkyl radical R can also decompose through β-scission reactions, forming additional radicals that can react with O_2 and further decompose. The ROO can abstract a hydrogen atom. for example from HO_2, and form a hydroperoxide ROOH. Dissociation of this hydroperoxide could form OH and an oxy radical RO (m/z = 169) that would likely rapidly dissociate or react with O_2.

Results from MPIMS verify that the expected products at m/z = 152, 153, and 168 form during low temperature chlorine-initiated oxidation of cineole. Figure 3.13 demonstrates the 3-dimensional data collection scheme where photoionization spectra, product time profiles, and mass spectra are acquired simultaneously. The mass peak at m/z = 168 is likely one or more of the cyclic di-ether ring-closure coproducts resulting from OH-elimination. The time profile of the m/z = 153 signal displays a sharp rise followed by a gradual increase reflective of a stable product. Both R and ROO radicals form and decompose rapidly after a few ms. Therefore, m/z = 153 is assigned to a combination of the initial fuel radical as well as possible fragment ions from ROO or fragment ions resulting from loss of a neutral methyl radical from the m/z = 168 product(s). The peak at m/z = 152 has an onset at 8.4 eV, and in accordance with previous discussion, is assigned to an unsaturated coproduct of HO_2-elimination.

The full experimental and simulated product mass spectra for Cl-initiated oxidation of cineole at 550 K and 650 K are shown in Figure 3.14. The experimental

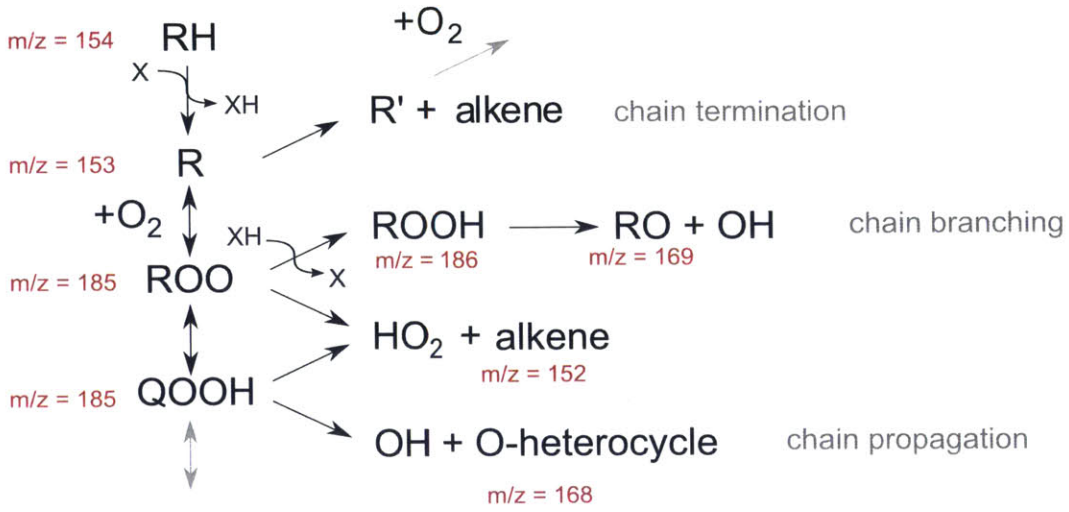


Figure 3.12: Schematic of low temperature oxidation pathways for cineole. Expected masses of species involved in cineole (m/z=154) oxidation are indicated in red.

spectrum is obtained by integrating the post-photolysis ion signal over the first 40 ms of reaction, then subtracting the averaged pre-photolysis signal from the post-photolysis signal to exclude cineole and its fragment ions. Major peaks are observed at m/z = 43, 58, 70, 82, 94, 110, 124, 152, 153, and 168 along with a number of additional minor features. The RMG-generated model was simulated at experimental conditions in CHEMKIN-PRO. [33] In order to visually compare the results against the mass spectra given by experiment, the product mole fractions were integrated over the first 40 ms of reaction, and the parent fuel cineole and chlorine-containing compounds were omitted for clarity. Like found in experiment, the simulated spectra are very similar at 550 K and 650 K, with minor temperature dependence in the appearance of certain mass peaks. However, quantitative differences exist between the experiment and the simulations in the relative yields of the various products.

Product time profiles alongside photoionization spectra and deduced species at 550 K are given in Figure 3.16 and 3.17. Typically, fragment ions exhibit gradual onsets and gradual increases in ion signal since they are formed through dissociation of true products. From this observation, m/z = 82, 94, 110, and 124 are possible fragment ions of other products or mixtures of stable products and fragment ions. The m/z =70 product is matches the molecular formula C_4H_6O. With an ionization onset of 9.55 eV, this excludes dimethyl ketene and ethyl ketene, which have much lower onsets. The model assigns the m/z = 70 product to a mixture of methyl vinyl ketone and methacrolein (see Figure 3.16). Methacrolein has a measured AIE of 9.91

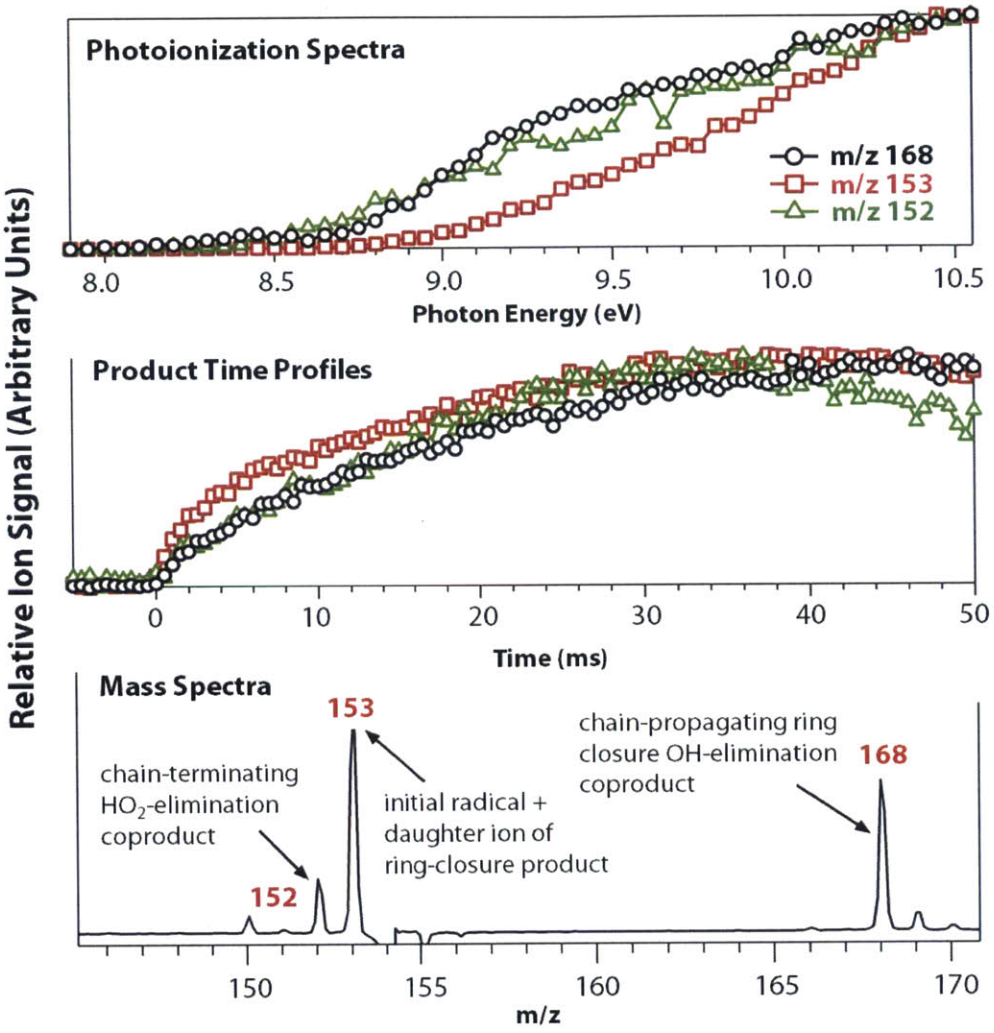


Figure 3.13: An example of the 3-dimensional data collection scheme employed in this study with cineole, O_2, chlorine, and chlorine radical reacting at 550 K. Top: Photoionization spectra for m/z = 152, 153, and 168 signals. Middle: Product time profiles for m/z = 152, 153, and 168 signals. Bottom: Mass spectrum showing the coproducts of the chain-propagating and chain-terminating channels associated with the oxidation of cineole radicals. The three sets of data are acquired simultaneously.

eV, [34] while methyl vinyl ketone has a measured AIE between 9.61-9.66 eV, [34–36] which is very close to the ionization onset found in our experiment. Shown in Figure 3.15, fitting the two PIE spectrums for methacrolein and methyl vinyl ketone against the m/z = 70 product PIE spectrum shows confirms that these two are likely products.

The m/z = 58 photoionization spectrum matches that of acetone and is confirmed

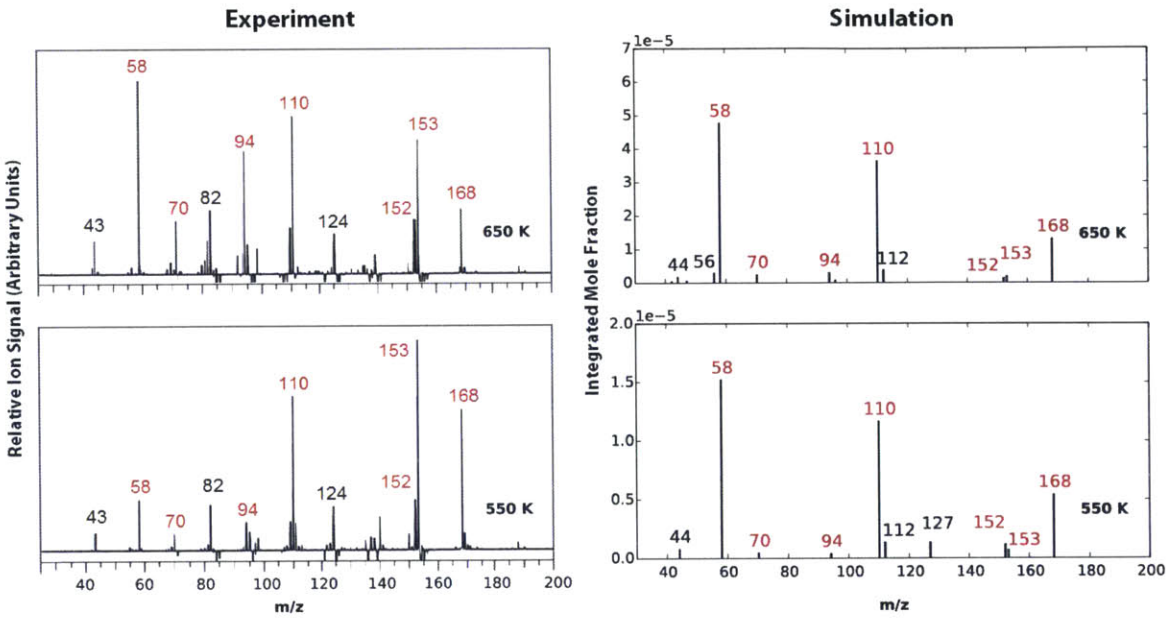


Figure 3.14: A comparison of experimental (left) vs. simulated (right) product mass spectra for the Cl-initiated oxidation of cineole at 550 K (bottom) and 650 K (top). The experimental difference mass spectra is normalized to photocurrent resulting from integrating the ion signal for 40 ms immediately following photolysis over ionizing photon energies 7.9 - 10.6 eV, then subtracting the averaged pre-photolysis signal. The simulated mass spectra gives the product mole fractions integrated over the first 40 ms of reaction, omitting cineole and chlorine-containing compounds for clarity. Common mass peaks found in both experiment and simulation are labeled in red.

to be acetone in the model as well. The m/z = 70 and 58 ion signals grow as a function of temperature, consistent with the assignment of β-scission products of the initial radicals that undergo thermal decomposition. However, in comparing the signals relative to m/z = 110, the experimentally observed increase of the m/z = 58 and m/z = 70 channels with temperature appears slightly larger than that predicted by the RMG model. Signals for m/z = 152, 153, and 168 are all found in the model, confirming some of the previously discussed expected pathways. Finally, m/z = 43, 95, and 135 display sharp temporal onsets followed by varying degrees of decay in their time profiles. These signals are likely radical intermediates which produced early in reaction post-photolysis and then decay quickly. The model has assigned species to most of these product peaks; these structures are shown in Figures 3.16 and 3.17. m/z = 43 and 95 are found within the model at low concentration, but m/z = 135 remains

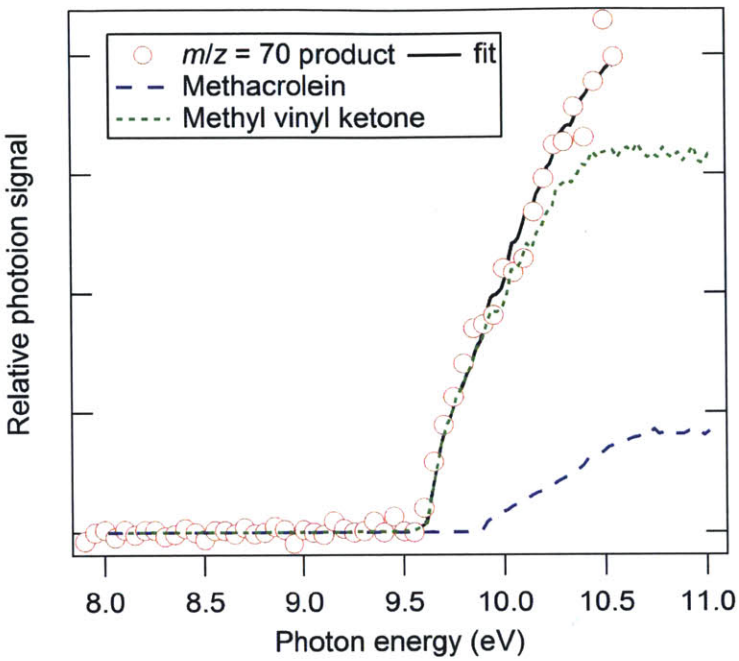


Figure 3.15: Comparison of the relative photoionization signal for the product peak found at m/z = 70 against the individual photoionization spectra of methacrolein and methyl vinyl ketone. The black solid line shows the fitted weighted sum of the two species' photionization spectra.

an unknown species. Several of the species found in simulation but not in experiment have photoionization onsets above the ionizing photon energies used, which explains why they do not appear in experiment. However, it may be worthwhile to investigate experimentally using higher photoionization energies or other techniques to validate the existence of these intermediates in the model.

Rate of production analysis was performed on the simulated model at 650 K and 40 ms of reaction and shown in Figure 3.18, identifying the major product pathways leading to mass peaks found in experiment. Chlorine radical abstraction of hydrogens from the parent fuel molecule occurs rapidly and exhausts the chlorine radical pool within the first 1 ms of reaction. The alkyl radicals R1-R5 then proceed to react and decompose, producing OH radicals that lead to additional parent fuel consumption. The figure shows OH as the major radical attacking cineole because instantaneous fluxes are shown at 40 ms towards the end of reaction time, when Cl radicals have already been consumed.

The branching ratio of hydrogen abstraction from cineole by OH radical leans heavily towards the secondary R2 and R4 alkyl radicals. R2 and R4 react predomi-

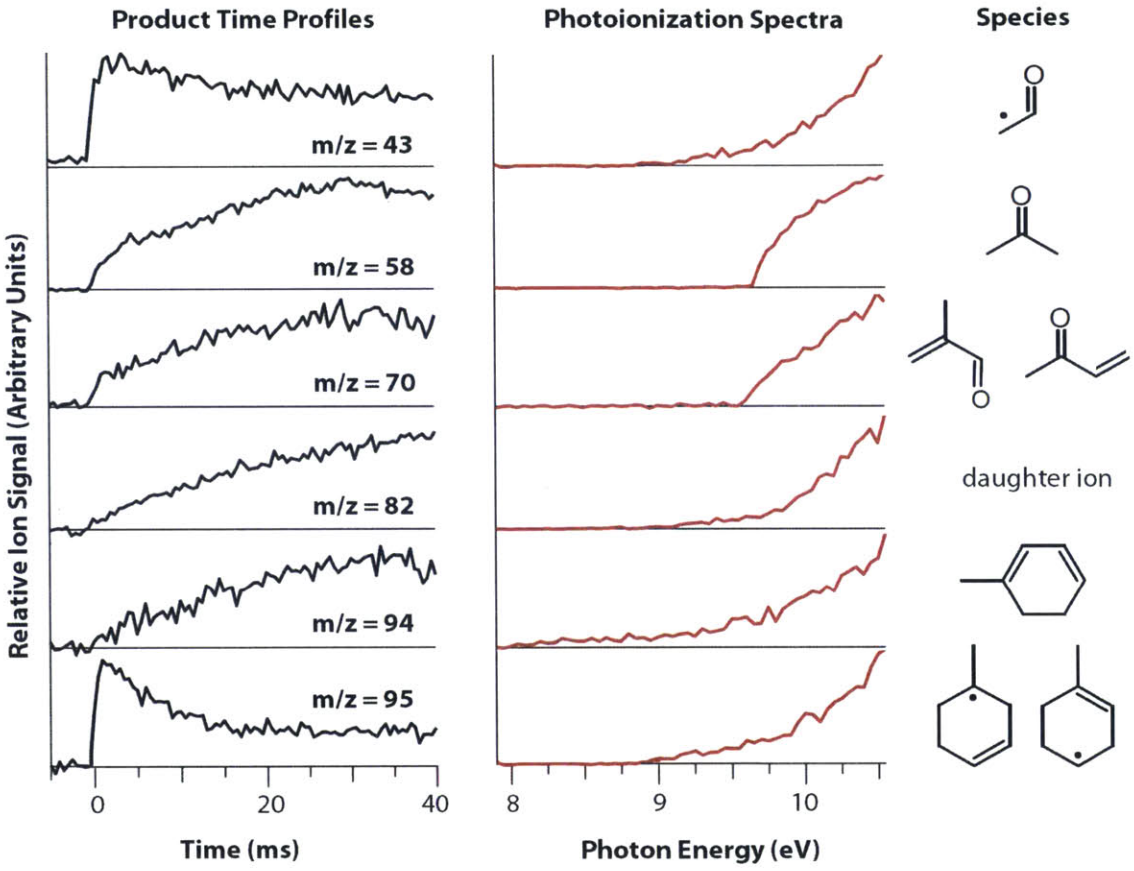


Figure 3.16: Product time profiles (left) and corresponding photoionization spectra (middle) and deduced species (right) for mass peaks m/z = 43, 58, 70, 82, 94, and 95 at 550 K.

nantly via a ring-opening β-scission and subsequent β-scission decomposition leading to large quantities of acetone (m/z = 58). The radical coproduct of acetone generates the rapidly decaying m/z = 95 peak found in experiment, which proceeds to react via O_2 addition, QOOH isomerization, and OH elimination to produce large quantities of cyclic ether products, observed at m/z = 110. The m/z =168 cyclic ether product found in experiment originates from cyclic ether formation via O_2 addition with the R3 alkyl radical followed by QOOH isomerization and OH elimination, with a minor channel originating from the R2 alkyl radical (not pictured in Figure 3.18). The model shows that the major m/z = 152 product is not formed via HO_2 elimination from an ROO radical resulting from O_2 addition on a parent alkyl radical, but rather elimination from an ROO radical formed by O_2 addition to a β-scission

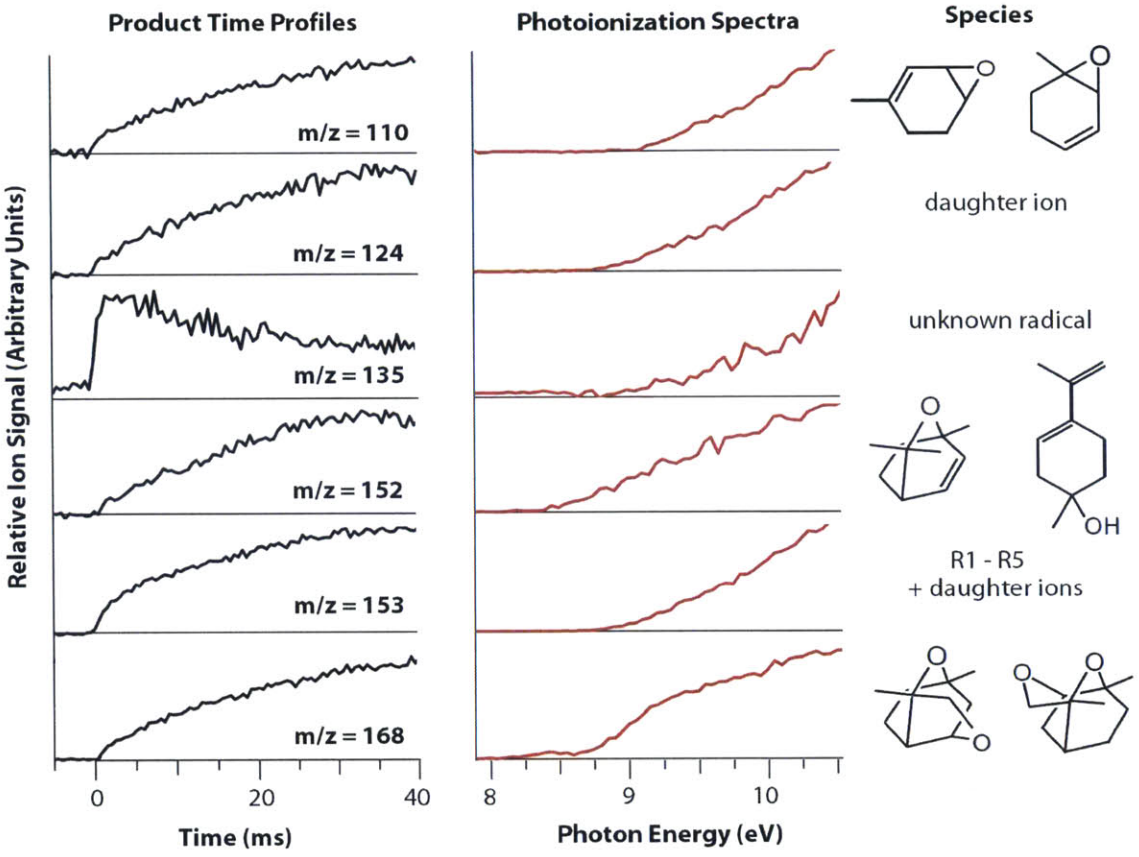


Figure 3.17: Product time profiles (left) and corresponding photoionization spectra (middle) and deduced species (right) for mass peaks m/z = 110, 124, 135, 152, 153, and 168 at 550 K.

product of R5. Ring-opening scissions in the parent alkyl radicals tend to be favored over O_2 addition, making the prima facie alkene formation channels unfavorable. The relative experimental yields of the m/z = 152, m/z = 168, and m/z = 110 products are not directly determined because the photoionizaton spectra and cross sections of the products are not known. However, the relative signal for the m/z = 152 product appears substantially more prominent in the experiment than in the RMG prediction, so that if the photoionization cross sections for the various products are similar, the experiment suggests more facile formation of HO_2 in reactions of the initial fuel radical (or its ring-opening products) with O_2.

Minor experimental products m/z = 70 and 94 are predicted to originate from side channels in R4 decomposition. In fact, all experimentally observed peaks were

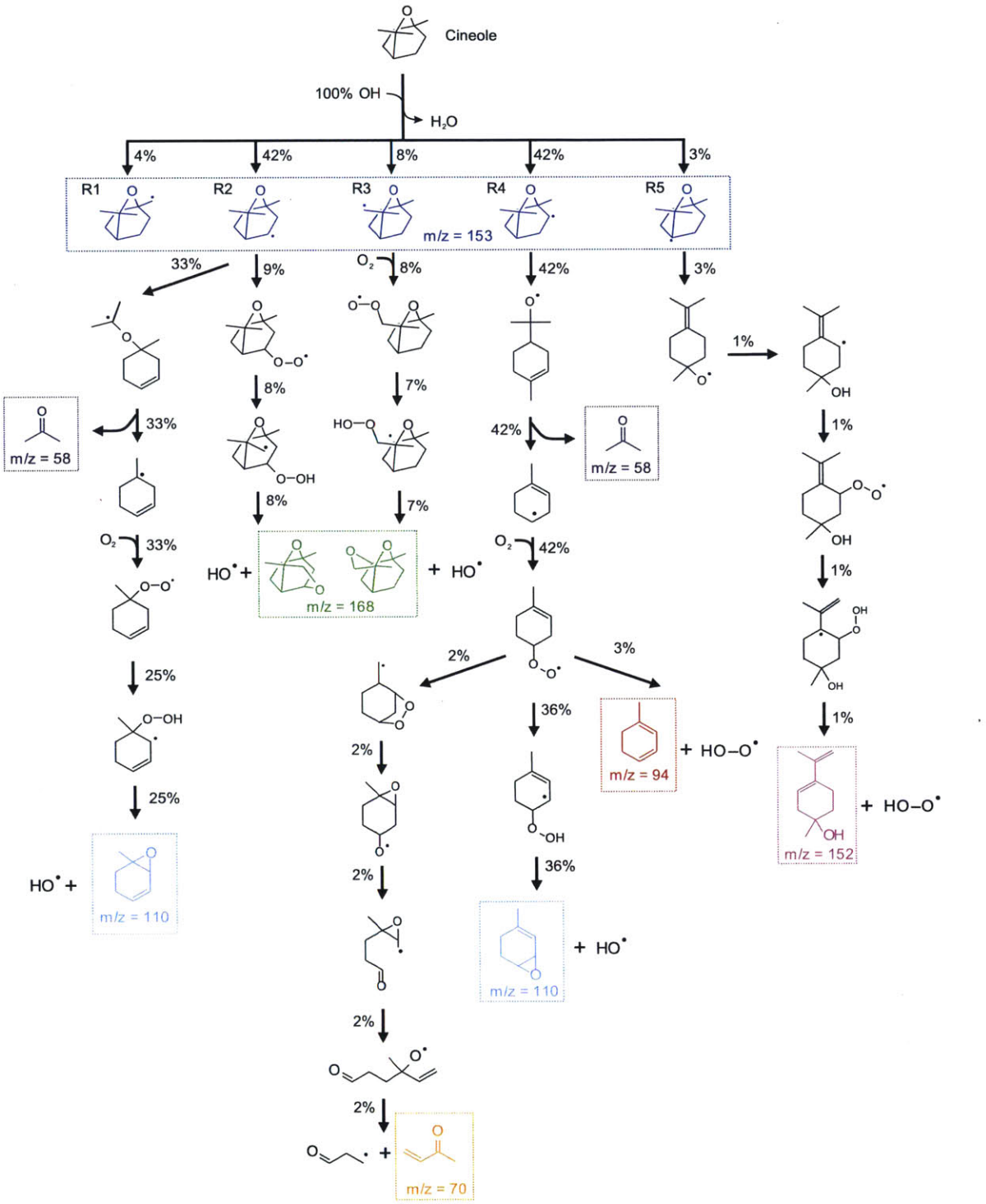


Figure 3.18: Reaction fluxes relative to the total consumption of cineole at 40 ms, T = 650 K, P = 4 Torr. Key reaction species found in experiment are boxed and highlighted.

predicted in the model except for an unknown radical species at m/z = 135.

For species intermediates, fused ring configurations lead to unstable thermochemistry. Thus, ring-opening reactions are favored over oxygen additions of the parent alkyl radicals, in contrast with typical oxidation pathways exhibited in alkanes. In addition, many intramolecular hydrogen abstractions are slowed by the constrained geometry of cineole's fused rings. This leads to many interesting and unusual reaction channels in the low temperature oxidation of cineole.

3.3 Conclusion

We have generated a detailed kinetic model for the low temperature oxidation of cineole that is qualitatively consistent with experiment. The model corroborates major product peaks found in experiment and helps identify unknown species intermediates and product formation pathways. This preliminary study on a fused bicyclic ether shows that cyclic ethers can exhibit significantly different oxidation chemistry than alkanes due to the limitations of their geometric configurations and ring-strain dependent thermochemistry and transition states. Quantum chemical theory can help identify the primary oxidation and decomposition pathways for cyclic compounds, which often diverge from their alkane counterparts. This coordinated study of cineole, a promising prospective biofuel, using a combination of experiment and theory serves as a a basis for future investigations of exotic fuel compounds.

3.4 References

[1] H. An, W. E. Wilhelm, and S. W. Searcy. "Biofuel and petroleum-based fuel supply chain research: A literature review." *Biomass and Bioenergy* 35 (9), 2011, pp. 3763–3774. DOI: 10.1016/j.biombioe.2011.06.021.

[2] D. K. Manley, A. McIlroy, and C. A. Taatjes. "Research Needs for Future Internal Combustion Engines." *Phys. Today* 61 (11), 2008, pp. 47–51. DOI: 10.1063/1.3027991.

[3] K. Kohse-Höinghaus, P. Oßwald, T. A. Cool, T. Kasper, N. Hansen, F. Qi, C. K. Westbrook, and P. R. Westmoreland. "Biofuel Combustion Chemistry: From Ethanol to Biodiesel." *Angewandte Chemie Intl. Ed.* 49 (21), 2010, pp. 3572–3597. DOI: 10.1002/anie.200905335.

[4] G. A. Strobel, B. Knighton, K. Kluck, Y. Ren, T. Livinghouse, M. Griffin, D. Spakowicz, and J. Sears. "The production of myco-diesel hydrocarbons and their derivatives by the endophytic fungus Gliocladium roseum." *Microbiology* 154, 2008, pp. 3319–3328. DOI: 10.1099/mic.0.2008/022186-0.

[5] G. A. Strobel, S. Spang, K. Kluck, W. M. Hess, J. Sears, and T. Livinghouse. "Synergism Among Volatile Organic Compounds Resulting in Increased Antibiosis in *Oidium* sp." *FEMS Microbiol. Lett.* 283 (2), 2008, pp. 140–145. DOI: 10.1111/j.1574-6968.2008.01137.x.

[6] J. M. Gladden, C. A. Taatjes, C. W. Gao, G. O 'Bryan, A. J. Powell, A. M. Scheer, and W. Wu. *Tailoring Next-Generation Biofuels and their Combustion in Next-Generation Engines.* Tech. rep. SAND2013-10094. Sandia National Laboratories, 2013.

[7] M. T. Mends, E. Yu, G. A. Strobel, S. Riyaz-Ul-Hassan, E. Booth, B. Geary, J. Sears, C. A. Taatjes, and M. Z. Hadi. "An Endophytic Nodulisporium sp. Producing Volatile Organic Compounds Having Bioactivity and Fuel Potential." *J. Pet. Environ. Biotechnol.* 3 (3), 2012, pp. 1–8. DOI: 10.4172/2157-7463.1000117.

[8] S. K. Singh, G. A. Strobel, B. Knighton, B. Geary, J. Sears, and D. Ezra. "An Endophytic Phomopsis sp Possessing Bioactivity and Fuel Potential with its Volatile Organic Compounds." *Microbial Ecology* 61 (4), 2011, pp. 729–739. DOI: 10.1007/s00248-011-9818-7.

[9] J. W. Allen, A. M. Scheer, C. W. Gao, S. S. Merchant, S. S. Vasu, O. Welz, J. D. Savee, D. L. Osborn, C. Lee, S. Vranckx, Z. Wang, F. Qi, R. X. Fernandes, W. H. Green, M. Z. Hadi, and C. A. Taatjes. "A coordinated investigation of the combustion chemistry of diisopropyl ketone, a prototype for biofuels produced by endophytic fungi." *Combust. Flame* 161 (3), 2014, pp. 711–724. DOI: 10.1016/j.combustflame.2013.10.019.

[10] J. W. Allen. "Predictive Chemical Kinetics: Enabling Automatic Mechanism Generation and Evaluation." PhD thesis. Massachusetts Institute of Technology, 2013.

[11] "Bio-Ketones: Autoignition Characteristics and Their Potential as Fuels for HCCI Engines." *SAE Int. J. Fuels Lubr.* 6 (3), 2013, pp. 713–728. DOI: 10.4271/2013-01-2627.

[12] M. De Vincenzi, M. Silano, A. De Vincenzi, F. Maialetti, and B. Scazzocchio. "Constituents of aromatic plants: eucalyptol." *Fitoterapia* 73 (3), 2002, pp. 269–275. DOI: 10.1016/S0367-326x(02)00062-X.

[13] D. L. J. Opdyke. "Monographs on Fragrance Raw-Materials." *Food and Cosmetics Toxicology* 13 (4), 1975, pp. 449–457. DOI: 10.1016/S0015-6264(75)80165-9.

[14] U. R. Juergens, U. Dethlefsen, G. Steinkamp, A. Gillissen, R. Repges, and H. Vetter. "Anti-inflammatory activity of 1.8-cineol (eucalyptol) in bronchial asthma: a double-blind placebo-controlled trial." *Respiratory Medicine* 97 (3), 2003, pp. 250–256. DOI: 10.1053/rmed.2003.1432.

[15] F. A. Santos and V. S. N. Rao. "Antiinflammatory and antinociceptive effects of 1,8-cineole a terpenoid oxide present in many plant essential oils." *Phytotherapy Research* 14 (4), 2000, pp. 240–244. DOI: 10.1002/1099-1573(200006)14:4<240::Aid-Ptr573>3.0.Co;2-X.

[16] M. Soh and G. W. Stachowiak. "The application of cineole as a grease solvent." *Flavour and Fragrance Journal* 17 (4), 2002, pp. 278–286. DOI: 10.1002/ffj.1103.

[17] A. F. M. Barton and J. Tjandra. "Eucalyptus Oil as a Cosolvent in Water - Ethanol Gasoline Mixtures." *Fuel* 68 (1), 1989, pp. 11–17. DOI: 10.1016/0016-2361(89)90004-5.

[18] A. Tomsheck, G. Strobel, E. Booth, B. Geary, D. Spakowicz, B. Knighton, C. Floerchinger, J. Sears, O. Liarzi, and D. Ezra. "Hypoxylon sp., an Endophyte of Persea indica, Producing 1,8-Cineole and Other Bioactive Volatiles with Fuel Potential." *Microbial Ecology* 60 (4), 2010, pp. 903–914. DOI: 10.1007/s00248-010-9759-6.

[19] P. K. Devan and N. V. Mahalakshmi. "A study of the performance, emission and combustion characteristics of a compression ignition engine using methyl ester of paradise oil-eucalyptus oil blends." *Appl. Energy* 86 (5), 2009, pp. 675–680. DOI: 10.1016/j.apenergy.2008.07.008.

[20] C. A. Taatjes, N. Hansen, D. L. Osborn, K. Kohse-Hoinghaus, T. A. Cool, and P. R. Westmoreland. "Imaging Combustion Chemistry via Multiplexed Synchrotron-Photoionization Mass Spectrometry." *Phys. Chem. Chem. Phys.* 10, 2008, pp. 20–34. DOI: 10.1039/b713460f.

[21] D. L. Osborn, P. Zou, H. Johnsen, C. C. Hayden, C. A. Taatjes, V. D. Knyazev, S. W. North, D. S. Peterka, M. Ahmed, and S. R. Leone. "The Multiplexed Chemical Kinetic Photoionization Mass Spectrometer: A New Approach to Isomer-Resolved Chemical Kinetics." *Rev. Sci. Instrum.* 79 (10), 2008, p. 104103. DOI: 10.1063/1.3000004.

[22] B. Gimeno, S. Martinez, J. S. Urieta, and P. Perez. "Vapor Pressures and Activity Coefficients of (1-Propanol plus 1,8-Cineole) at 10 Temperatures between 278.15 K and 323.15 K." *J. Chem. Eng. Data* 57 (11), 2012, pp. 3026–3031. DOI: 10.1021/je300736k.

[23] J. Montgomery John A., M. J. Frisch, J. W. Ochterski, and G. A. Petersson. "A complete basis set model chemistry. VI. Use of density functional geometries and frequencies." *J. Chem. Phys.* 110 (6), 1999, pp. 2822–2827. DOI: 10.1063/1.477924.

[24] A. D. Boese and J. M. L. Martin. "Development of density functionals for thermochemical kinetics." *J. Chem. Phys.* 121 (8), 2004, pp. 3405–3416. DOI: 10.1063/1.1774975.

[25] Y. Zhao and D. G. Truhlar. "Density functionals with broad applicability in chemistry." *Acc. Chem. Res.* 41 (2), 2008, pp. 157–167. DOI: 10.1021/Ar700111a.

[26] C. Eckart. "The penetration of a potential barrier by electrons." *Phys. Rev.* 35 (11), 1930, pp. 1303–1309. DOI: 10.1103/PhysRev.35.1303.

[27] J. D. Smith, J. D. DeSain, and C. A. Taatjes. "Infrared laser absorption measurements of HCl(upsilon=1) production in reactions of Cl atoms with isobutane, methanol, acetaldehyde, and toluene at 295 K." *Chem. Phys. Lett.* 366 (3-4), 2002, pp. 417–425. DOI: 10.1016/S0009-2614(02)01621-4.

[28] M. G. Bryukov, I. R. Slagle, and V. A. Knyazev. "Kinetics of reactions of Cl atoms with ethane, chloroethane, and 1,1-dichloroethane." *J. Phys. Chem. A* 107 (34), 2003, pp. 6565–6573. DOI: 10.1021/Jp0275138.

[29] M. K. Sabbe, A. G. Vandeputte, M. F. O. Reyniers, V. Van Speybroeck, M. Waroquier, and G. B. Marin. "Ab initio thermochemistry and kinetics for carbon-centered radical addition and beta-scission reactions." *J. Phys. Chem. A* 111 (34), 2007, pp. 8416–8428. DOI: 10.1021/Jp072897t.

[30] H. J. Curran, P. Gaffuri, W. J. Pitz, and C. K. Westbrook. "A comprehensive modeling study of iso-octane oxidation." *Combust. Flame* 129 (3), 2002, pp. 253–280. DOI: 10.1016/S0010-2180(01)00373-X.

[31] E. E. Dames. *Energy Frontier Research Center.* Personal Communication. 2014.

[32] J. Zador, C. A. Taatjes, and R. X. Fernandes. "Kinetics of Elementary Reactions in Low-Temperature Autoignition Chemistry." *Prog. Energy Combust. Sci.* 37 (4), 2011, pp. 371–421. DOI: 10.1016/j.pecs.2010.06.006.

[33] *CHEMKIN-PRO 15131.* Reaction Design: San Diego. 2013.

[34] G. Pan, C. J. Hu, M. Q. Huang, Z. Y. Wang, Y. Cheng, Z. Liu, X. J. Gu, W. X. Zhao, W. J. Zhang, J. Chen, F. Y. Liu, X. B. Shan, and L. S. Sheng. "A VUV photoionization mass spectrometric study on the OH-initiated photooxidation of isoprene with synchrotron radiation." *Journal of Environmental Sciences-China* 24 (12), 2012, pp. 2075–2082. DOI: 10.1016/S1001-0742(11)61051-6.

[35] W. C. Tam, D. Yee, and C. E. Brion. "Photoelectron-Spectra of Some Aldehydes and Ketones." *J. Electron. Spectrosc. Relat. Phenom.* 4 (1), 1974, pp. 77–80. DOI: 10.1016/0368-2048(74)80045-9.

[36] J. K. Terlouw, W. Heerma, J. L. Holmes, and P. C. Burgers. "Structure and Formation of Gaseous [C4h6o]+ Ions .1. The Enolic Ions [Ch2=C(Oh)-Ch=Ch2]+. And [Ch2=Ch-Ch=Ch(Oh)]+. And Their Relationship with Their Keto Counterparts." *Org. Mass Spectrom.* 15 (11), 1980, pp. 582–586. DOI: 10.1002/oms.1210151108.

4

An experimental and theoretical study of the pyrolysis and oxidation of JP-10

This chapter presents the combined experimental and theoretical modeling study on the pyrolysis and combustion of JP-10. A more detailed publication containing the CHEMKIN format mechanism and supporting information can be found elsewhere [1]. New shock tube studies were conducted at Aerodyne, Inc. at pyrolysis conditions and oxidation equivalence ratios between $\Phi = 0.14$ to 1.0 and temperatures between 1000-1600 K. This new dataset enabled the tracking of product yield dependence on equivalence ratio as well as the identification of several new decomposition intermediates. A detailed, comprehensive model of JP-10 combustion and pyrolysis kinetics was constructed using RMG and extensively validated against the new shock tube data as well as flow tube pyrolysis data from Ghent. The model yields new insights on JP-10's complex decomposition and oxidation chemistry and identifies key pathways towards aromatics formation.

4.1 Background

JP-10 is a synthetic aviation turbine fuel that is almost entirely composed of one molecule, *exo*-tetrahydrodicyclopentadiene, $C_{10}H_{16}$ (Figure 4.1). Its low freezing point, high volumetric energy density, and high specific impulse make it one of the most promising air-breathing missile fuels [2–4]. Due to its success, JP-10 prices skyrocketed and almost doubled from 2006 to 2010, driving further research into more efficient use of this expensive fuel, such as blending JP-10 with cheaper alternatives. Major research efforts have also been focused on the development of propulsion technologies, including ramjet, scramjet and pulse detonation engines, using JP-10 as the predominant fuel. JP-10 is often regarded as being easier to study than other jet propulsion fuels in that it is essentially a single-component jet fuel, making it a favor-

able target for kinetic modeling. Component analyses indicate that JP-10 consists of approximately 96.5 wt% *exo*-tetrahydrodicyclopentadiene, 1.0 wt% adamantane and 2.5 wt% *endo*-tetrahydrodicyclopentadiene (Figure 4.1). Due to the complexity of the molecules involved, we assume in the remainder of this work that JP-10 can be regarded as its primary component.

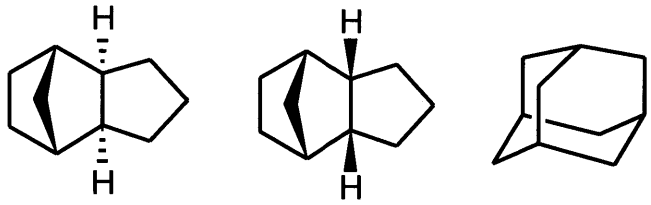


Figure 4.1: Main constituents of JP-10: (from left to right) *exo*-tetrahydrodicyclopentadiene, *endo*-tetrahydrodicyclopentadiene and adamantane.

The current literature data provide an incomplete picture of JP-10's complex pyrolysis and combustion behavior. Moreover, available datasets often show significant discrepancies amongst each other. One of the first extensive studies on the pyrolysis of JP10 was conducted by Davidson *et al.* [5] using high speed UV adsorption behind reflected shock waves at 1400 K and 1 atm. The authors identified cyclopentene as one of the major initial decomposition products amongst other major decomposition products: ethene, propylene, benzene, butadiene and cyclopentadiene. Nakra *et al.* [6] studied the pyrolysis of JP-10 in a micro flow tube reactor at roughly 2×10^{-3} atm with electron impact and ion impact mass spectra analysis. Their studies covered high temperature pyrolysis and millisecond residence times but did not include the high-pressure regime. Benzene and cyclopentadiene were identified as main decomposition products but the authors failed to identify large concentrations of cyclopentene, which would be expected from the Davidson *et al.* experiments. Striebich and Lawrence [7] focused on higher pressure experiments, working at supercritical conditions of 34 atm and 373 - 873 K to illustrate the thermal stability of JP-10, which was found to be more stable than other jet fuels tested. They found that JP-10's decomposition products were predominantly alkanes and alkenes, but they also detected smaller concentrations of cyclohexene and dimethylcyclopentadiene. Xing *et al.* [8] studied the decomposition of JP-10 at 900 K and pressures ranging from 1 to 40 atm, contributing to an enhanced understanding of reaction pressure on product selectivities. That work points to secondary reactions of alkenes, formed during the initial decomposition phase, as the major contributing pathways to aromatic species formation. However,

this conclusion contradicts Herbinet et al.'s observation that some aromatics such as benzene and toluene must also have direct reaction channels, based on experiments conducted in a jet stirred reactor [9].

The combustion of JP-10 has primarily been studied under conditions that mimic those applicable in engines: residence times are typically very short (i.e. ms range) with temperatures exceeding 1000 K. These studies generally focused on measuring ignition delays and do not report detailed product distributions, which are important for understanding engine performance and predicting the fuel's sooting behavior. Davidson *et al.* [10] reported ignition delay times and OH concentration time histories behind reflected shock waves in the temperature range 1300 - 1700 K with pressures ranging from 1 - 6 atm. Colket and Spadacinni [11] reported ignition delays in the temperature range 1250 - 1500 K. Mikolaitis *et al.* [12] conducted ignition delay measurements at higher temperatures up to 2000K and higher pressures of 10 - 25 atm. They also found that JP-10's ignition delays were generally insensitive to fuel additives. More recently, Wang *et al.* [13] reassessed ignition delays for JP-10 at pressures of 1.5 - 5.5 atm and temperatures ranging from 1000 - 2100 K to fill gaps between the other datasets.

Many computational studies have focused on solving specific problems related to understanding JP-10's decomposition chemistry. Zehe and Jaffe [14] studied the thermochemistry of *exo*-tetrahydrodicyclopentadiene using composite methods such as CBS-QB3 and the Gx family of methods. Bozzelli *et al.* [15] extended this to all possible radicals formed after abstracting a hydrogen from JP-10 and used a reaction scheme approach to obtain enthalpies of formation for various radicals with approximately 1 kcal/mol accuracy. Chenoweth *et al.* [16] used a reactive force field (ReaxFF) to identify major initial decomposition pathways including (a) a path leading to the formation of ethene and a C_8 hydrocarbon and (b) a path producing two C_5 hydrocarbons. Various reactions paths have also been studied by Yue *et al.* [17] using the M06-2X functional. Recent work by Magoon *et al.* [18] focused on calculating energy barriers for intramolecular disproportionation reactions that are important during the pyrolysis of JP-10.

Despite the broad range of experimental and computational studies that have been conducted in an effort to increase our understanding of JP-10's decomposition chemistry, the development of an accurate, detailed combustion mechanism has remained elusive. The pyrolysis mechanism developed by Herbinet *et al.* [9] lacks oxidation chemistry and is not publicly available. The first JP-10 combustion mechanism was developed by Li and coworkers [19] a decade ago; however, the initial decomposition

steps were modeled using lumped reactions to capture a simple model which largely avoided the kinetics of any species larger than C_3. The resulting lumped model contained 36 species and 174 reactions, with no decomposition products larger than C_5. This preliminary model was admittedly highly speculative for kinetics involving species larger than C_3 but yielded satisfactory ignition delay predictions relative to available shock tube data. For many years, the Li *et al.* mechanism was, to our knowledge, the only published mechanism available for modeling JP-10 combustion. Previous Green group members developed and refined the first detailed JP-10 combustion mechanism [18] using RMG-Java. This first-generation mechanism characterized the initial decomposition kinetics of JP-10 combustion but included very rough estimates for many important reactions and had little experimental validation due to lack of speciation data. More recently, Vandewiele *et al.* [20] generated an improved pyrolysis model validated by new experimental data obtained at Ghent University.

In this chapter, RMG was used to generate a new comprehensive combustion model for JP-10, extending on the Vandewiele pyrolysis model with updates to rate coefficients and thermochemistry based on quantum chemistry calculations and validated through new shock tube experiments. These experiments were performed by collaborators at Aerodyne Research, Inc. The aim of this new set of experiments is to fill in the gap existing in the current literature and better understand JP-10's decomposition and oxidation chemistry. Products of decomposition were identified and measured behind reflected shock waves at pyrolysis and pre-ignition oxidation conditions ranging from very lean ($\Phi = 0.14$), lean ($\Phi = 0.24$), to stoichiometric ($\Phi = 1.0$). The experiments were conducted at temperatures ranging between 1000 - 1600 K, pressures of 6 - 8 atm, and a residence time of about 500 μs. These data were further used to validate a new detailed kinetic model generated by RMG that captures both the pyrolysis and oxidation chemistry of JP-10. Modeling advances were accomplished by supplementing the RMG database with higher fidelity computational estimates and reported literature values for critical kinetics and thermodynamic parameters, in addition to algorithmic improvements to the RMG software. This work highlights the current capabilities of computer-based kinetic model generation with respect to cyclic chemistries such as JP-10 as well as opportunities for further research.

4.2 Methodology

The shock tube experiments described here were performed by the following collaborators at Aerodyne, Inc.: Robin E. Bonomi, Hsi-Wu Wong, Oluwayemisi O. Oluwole,

and David K. Lewis. The modeling work was assisted by Aäron G. Vandeputte, Gregory R. Magoon, Nick M. Vandewiele, and Nathan Yee.

4.2.1 Shock tube experiments

The JP-10 pyrolysis and oxidation experiments were carried out in a 2.54 cm internal diameter pressure-driven single-pulse shock tube of the design described by Lifshitz *et al.* [21] Shown in Figure 4.2, the driven (sample) section is 1.75 m long, and the driver section (adjustable for tuning) is 0.91 m long. A Mylar diaphragm, 2.0 μm thick, separates the two sections. A $3.2 \times 10^{-2}\,\mathrm{m}^3$ dump tank is positioned just downstream from the diaphragm for the purpose of preventing multiple reflections of the shock. Two rapid-response pressure sensors are mounted in the wall of the driven section to detect the arrival of the incident and reflected shocks; the output is amplified and recorded on digitizing storage oscilloscopes, and reaction temperatures are determined from the recorded shock velocities.

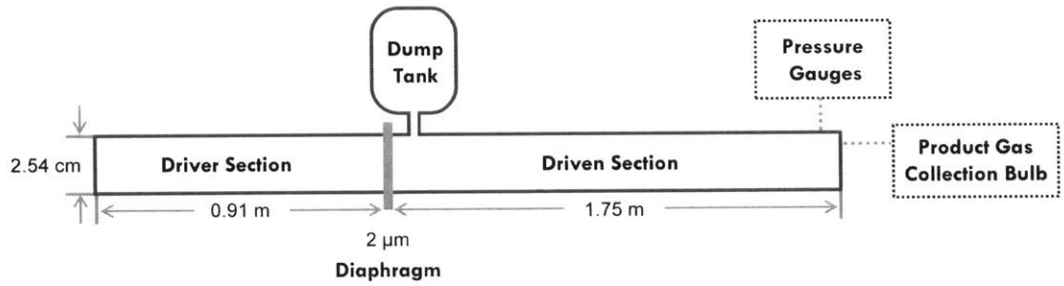


Figure 4.2: Schematic of the shock tube used for the experimental study.

Before each experiment, the entire system is evacuated by a molecular drag pump that is backed by a diaphragm pump, to a pressure < 0.1 Pa. Helium is used as the driver gas, and research-grade argon is used as the diluent for the reactant samples. Reactant samples were prepared in 1.0 L glass bulbs (sealed with Teflon stopcocks) containing glass beads to enhance mixing. A few drops of liquid JP-10 were added, and the bulb was evacuated briefly (some liquid JP-10 remained). After this, the bulb was filled to 850 torr with either argon (for pyrolysis) or a pre-mixed O_2/Ar mixture (for oxidation) and swirled for at least 5 minutes to establish a saturated JP-10/Ar or JP-10/O_2/Ar mixture. Reactant samples were stored for at least an hour at 20°C before use. Heated tubes ($\sim$ 50°C) connecting the sample bulb and shock-tube prevent JP-10 condensation, and 2.5×10^4 to 5.0×10^4 Pa of the reactant sample was eventually transferred to the shock-tube. As soon as the desired sample pressure had been measured in the shock tube the shock was initiated by exceeding

the diaphragm's bursting pressure with the He driver gas. Immediately after each shock, the product sample was withdrawn through a small outlet in the side-wall 7.5 mm upstream of the end wall of the driven section into a 100 cm^3 glass bulb that had been previously evacuated and then filled with a small amount (1.3×10^3 Pa) of a calibration mixture containing argon and measured amounts of two standards, isobutene and cis-cyclooctene. These two compounds were chosen as calibration standards since neither had been detected among the JP-10 pyrolysis or oxidation products in earlier qualitative test runs, and because both were well-resolved from the pyrolysis and oxidation products.

The gas collection bulbs containing product gas samples and the calibration standards were analyzed on two GC-MS systems: an Agilent 6890 GC with an Agilent 7850 MS and a Hewlett-Packard 5890 GC with a Hewlett-Packard 7820 MS. The Agilent GC-MS was used to analyze compounds greater or equal to C_5 with an Agilent HP5-MS (5% Phenyl Methyl Siloxane) column of dimensions 30 m $\times$ 0.25 mm $\times$ 0.25 μm. The HP GC analyzed compounds smaller than C_5 using a Restek PLOT S-Bond column of dimensions 30 m $\times$ 0.25 mm $\times$ 0.53 μm. Previously, a set of 10 compounds of known concentration, ranging from C_3 to C_{10} in size, had been analyzed and quantified on each GC-MS. Using calibration analyses of cis-cyclooctene, isobutene and the ten other compounds, standard curves were made to relate the MS response to the molar concentration of each species. This gave a set of MS response calibration curves that were offset according to molecular weight and molecular structure. For species found in the JP-10 reaction products that were specifically calibrated for with standards, the product quantification was done based on direct response factors for those specific compounds. For species found in the reaction products that were not in the calibration standards, the MS response factors were interpolated from the standard curves for species of similar molecular weight and structure. In all cases the peak areas were first normalized by the MS response to cis-cyclooctene or isobutene in order to correct for day-to-day variations in the responses of the two GC-MS instruments. Every sample was analyzed three times before taking an average mole fraction of each species.

Under pyrolysis and stoichiometric oxidation conditions, carbon balances of 80 to 110% were obtained. This is within the accepted error in GC-MS quantitative analysis, i.e. $\pm$ 20%, and averages on the low side of 100% mainly due to our inability to detect CO, CO_2, CH_4 and HCHO because of the peaks' overlapping retention times with the diluent gas Ar. Due to the excess argon in the samples, it was very difficult to detect and quantify most of the lighter species, including all C_1 com-

pounds. Under leaner conditions, carbon balances based on the species we were able to quantify dropped to approximately 50%, due to larger relative amounts of CO and CO_2 formation. The pressure, temperature and density of the reacting shock-heated gas immediately after passage of the reflected shock was calculated from measured incident shock speeds, the initial (pre-shock) sample temperature and pressure, conservation equations (mass, momentum, energy), and heat capacity polynomials for JP-10, O_2 and Ar. The reaction time for the sample was determined from the pressure versus time traces at the downstream end of the tube, as the time interval between initiation of the reflected shock and arrival of the cooling expansion wave. This time is longer for gas at the end of the tube, but is less for gas that was further upstream. Assuming that the composition of each 100 cm^3 product sample collected is representative for the product gas at a certain position in the shock tube, we can derive the average reaction time from its average location under the compressed reaction conditions. Samples had an effective average reaction time of 500 ($\pm$ 10%) μs.

4.2.1.1 Experimental error analysis

A single-pulse shock tube can be, in many respects, an ideal choice for the study of rapid thermally-induced reactions under homogeneous conditions. Gas samples can be brought rapidly ($t \sim 10^{-11}$ s) and homogeneously to temperatures ranging from hundreds to thousands of degrees above ambient temperature, at pressures ranging from $< 10^3$ to $> 10^7$ Pa, then quenched back to room temperature and recovered for analysis. Since the heating is brought about by the passage of a compression shock wave through the gas sample, the walls of the apparatus remain essentially at room temperature throughout the sub-millisecond reaction time; a thin boundary layer of cold gas forms along the walls insulating them from the hot reacting gas. This effectively eliminates the potential complications of catalytic surfaces but also means that some of the recovered gas for analysis was not heated to the temperature of the gas in the center of the tube. In the present study, that percentage was determined by running JP-10/Ar samples to temperatures above 1600 K, under which JP-10 consumption should have been close to 100%. These experiments indicated that the actual residual reactant concentration amounted to approximately 1/3 of the original JP-10 feed (i.e. 67% $\pm$ 5% of the feed reacted).

Another limitation arises because of the inherent differences between compression (heating) and expansion (cooling) waves. In an ideal temperature jump experiment, both the heating and the cooling should be instantaneous. Shock waves are by nature convergent processes; as the shock front is formed, successive stages of compression

travel through increasingly hot gas, at progressively faster velocities, until they coalesce into a single, abrupt pressure discontinuity. Expansion waves travel through progressively cooler gas and therefore diverge. Thus the desired instantaneous heating of the reactant sample is achieved, but the cooling of the reaction products at the end of the reaction time is not instantaneous. In the shock tube used in this study, the typical initial cooling rate was determined, from the recorded pressure trace, to be $> 5 \times 10^5$ °C/s; so 100 μs into the cooling process (one fifth of the reaction time) the temperature drops at least 50 degrees. The rate at which each reaction is quenched depends on the activation energy and the reaction order (since the gas density also drops, lowering collision rates). For the initial pyrolysis reactions of JP-10, with estimated activation energies > 80 kcal/mol, a temperature decrease from 1200 to 1150 K should lower the rate of JP-10 consumption by more than a factor of four. However, abstraction or addition reactions of radicals, such as would form alkenes, alkynes and aromatic species, will be slowed down less rapidly. By the time the product sample is collected for analysis, it will contain only neutral, non-radical species. Thus, the concentrations of benzene, toluene, cyclopentene, cyclopentadiene and similar product species are likely to be greater than they were in the shock tube at the end of the reaction time but before the cooling process occurred. Considering all these uncertainties, the total estimated uncertainty in species concentrations is approximately 30%. For this reason, we believe that the pyrolysis and oxidation experiments performed in this study are most useful for qualitative and semi-quantitative comparisons with the model.

The final source of uncertainty in the experimental data is the error associated with the reaction temperature, which is back-calculated from the recorded shock wave velocities. These times are determined from the pressure-jump found in the oscilloscope readings, which have an uncertainty of approximately 0.4 μs (the oscilloscopes have a resolution of 0.2 μs). The temperature error arises from the propagation of the oscilloscope error and from the assumption of complete translational, rotational, and vibrational equilibration of kinetic energy when converting shock speed to temperature. We estimate the total temperature uncertainty to be about 3 - 4%, or 50 K. This is a significant uncertainty, since the chemistry can vary greatly with a 50 K change in temperature.

4.2.2 Detailed chemical kinetic modeling

RMG-Java was used to generate the kinetic models in JP-10 combustion. The modeling of JP-10 represents the first time RMG has been used to investigate the com-

Table 4.1: RMG model generation conditions for the four subnetworks in the JP-10 mechanism.

Network	Equivalence Ratio (Φ)	Temperature (K)	Pressure (atm)
Low temperature pyrolysis	n/a	1000	11.7
Medium temperature pyrolysis	n/a	1400	7
Medium temperature oxidation	0.14	1400	7
High temperature oxidation	0.14	2000	1 and 20

bustion kinetics of a fused cyclic molecule. The model development process led to multiple improvements to RMG's databases.

The final JP-10 combustion model presented in this work was obtained by merging four smaller networks generated at the reaction conditions of interest. The final model contains 691 species and 15518 reactions. A low temperature pyrolysis model was generated to cover the reaction conditions associated with the Ghent pyrolysis data. Medium temperature pyrolysis and oxidation models were generated to cover the reaction conditions of the shock tube data presented in this study. Finally, a high temperature oxidation model was generated to span the range of experimental conditions associated with the ignition delays. The conditions associated with these four models span a broad range of temperatures, pressures and equivalence ratios as detailed in Table 4.1.

For all models, the JP-10 goal conversion was set to 99% with an error tolerance of $\epsilon = 0.05$. RMG reached full convergence under all simulated reaction conditions. In order to control the radical addition reactions proceeding to much larger hydrocarbons, all jobs restricted the maximum number of carbon atoms per species to 10, the number of heavy atoms per species to 12, and the number of unpaired electrons to 2 or less. All networks were generated within the high pressure limit; however, rates for small molecule reactions involving OH, H, H_2O, CO, etc. which are highly influential and strongly pressure dependent were taken from the EFRC v0.3 mechanism [22] and seeded into the RMG-generated models. RMG also had access to a set of calculated rate coefficients (mainly CBS-QB3 [23] and BMK/6-311G(2d,d,p) [24]. The calculated reactions are mainly hydrogen abstraction reactions from JP-10 and β-scissions in the most important JP-10 radicals formed after hydrogen abstraction from JP-10. For the hydrogen abstraction reactions, the 1-D hindered rotor (1D-HR) approach was used to obtain a more accurate treatment of low vibrational internal rotations. The methodology for this approach has been described in the work of Sharma *et*

al. [25] Due to their importance in the initial decomposition steps, rate coefficients for intramolecular disproportionation reactions presented in the work of Magoon *et al.* [18] calculated using various multireference methods were also included in this database.

Compared to previous models this current model benefits from major upgrades to the RMG thermochemical database which are (a) revision of the ring strain corrections for polycyclic compounds, (b) improved databases for hydrogen abstraction reactions and β-scission reactions that use rate rule estimates based on the group additivity concept for Arrhenius parameters [26, 27] and (c) updated estimates for important intramolecular reactions. RMG traditionally estimates thermodynamic quantities using a Benson group additivity scheme, which is unable to deal with many cyclic and fused cyclic compounds accurately without requiring a large number of *ad hoc* ring corrections. To avoid inaccuracies, when an exact polycyclic ring correction cannot be found, we used the on-the-fly quantum mechanics (QM) feature [28] within RMG to perform quantum calculations for cyclic species based on estimating explicit 3D geometries. In particular, on-the-fly semi-empirical PM7 calculations were performed for cyclic species using MOPAC2012 [29].

After model generation, additional third body reactions involving Ar and Ne were added manually as RMG currently cannot deal with them. These reactions involve recombination reactions of atoms such as H and O. The initial model however tended to over-predict the JP-10's decomposition rate and yielded too much propene. A sensitivity analysis showed that this was likely caused by an over-prediction of the rate coefficient for hydrogen abstraction reactions from JP-10 by allyl estimated by group additivity [27]. To correct for this we lowered the n factor in the three parameter Arrhenius equation for hydrogen abstraction reactions from JP-10 by allyl radical from 4.24 to 4.00.

4.3 Results and discussion

4.3.1 The reaction network

The pyrolysis of JP-10 can be understood as a three step process. In the initial step, JP-10 isomerizes to other components, typically by a ring opening followed by an intramolecular disproportionation. The model predicts in this initial stage the isomerization of JP-10 to 2-cyclopentylcyclopentene (2CC) and 4-ethylbicyclo[3.3.0]oct-2-ene (4EBO), as shown in Figure 4.3. The shock tube experiments confirm the presence of 2CC during pyrolysis and oxidation. This is also supported by experi-

mental data presented by Vandewiele *et al.* [30], where many unidentified C_9 and C_{10} species with similar volatility to JP-10 were found at low conversion. The species undergo bond scissions which lead to the formation of two radicals that initiate the radical mechanism. Those radicals undergo β-scissions until a radical is obtained that has a sufficiently long lifetime to participate in hydrogen abstraction reactions. In agreement with Herbinet *et al.* [9], we find that the main radical initiation reaction is bond scission in 2-cyclopentylcyclopentene (2CC in Figure 4.3) with the formation of cyclopentenyl and cyclopentyl radicals. Cyclopentenyl radicals subsequently decompose by β-scission to cyclopentadiene and hydrogen radicals. Cyclopentyl radicals decompose mainly by ring opening and β-scission to ethene and allyl.

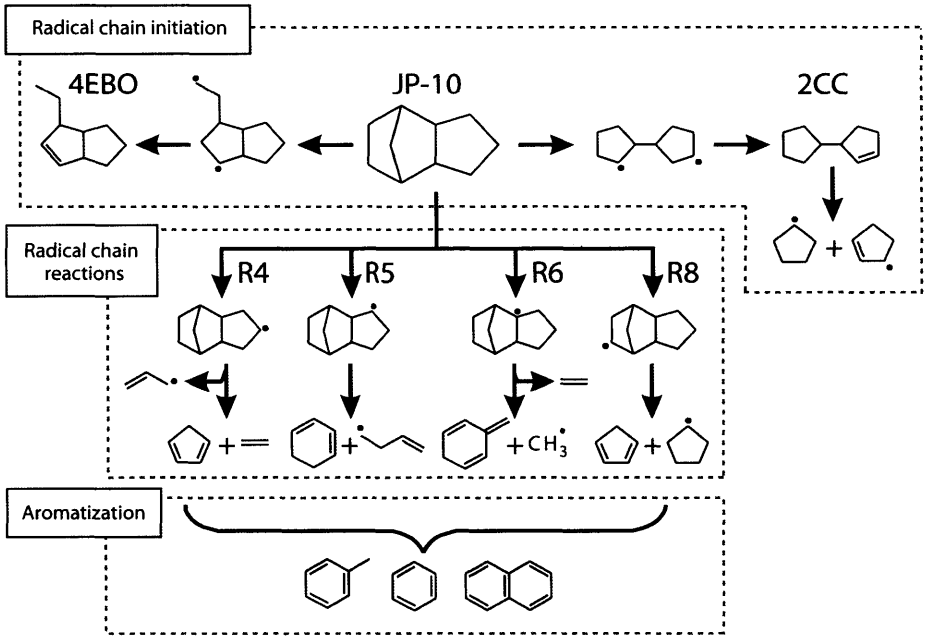


Figure 4.3: The three step mechanism of JP-10 thermal decomposition: radical chain initiation reactions (top), radical chain reactions (middle), and finally aromatization (bottom).

Radical chain reactions constitute the second step of the decomposition process, which begins with hydrogen abstractions from JP-10. Allyl and H radicals formed in step 1 are responsible for most of the hydrogen abstraction reactions, which can occur at 6 different sites on JP-10. As intramolecular hydrogen abstraction reactions in JP-10 radicals are strongly hampered due to its fused ring structure, each primary hydrogen abstraction reaction initiates a distinct decomposition channel with its own set of products. The main decomposition products for the four dominant channels

are presented in Figure 4.3. It can be seen that the main products are ethene, 1,3-cyclohexadiene, cyclopentadiene and propene. Once a sufficient amount of products are formed, secondary reactions will lead to the formation of aromatic components from the main decomposition products. This can be seen as an important third step in the pyrolysis mechanism and is often regarded one of the most difficult ones to model. Under oxidation conditions, the decomposition of JP-10 will be very similar to the one observed during the pyrolysis, except that oxygen will participate in chain branching reactions, speeding up the radical reactions and reacting with the pyrolysis products to form CO, H_2O and CO_2.

In Figure 4.4 the predicted decomposition pathways for the pyrolysis and stoichiometric oxidation of JP-10 is shown. Both conditions for the model were simulated in a plug flow reactor operated at 1000 K and a pressure of 1.7 atm. At 10% conversion, almost 8% of the JP-10 decomposes by ring opening reactions leading to three different biradicals (BR1, BR2, and BR3 in Figure 4.4). These biradicals undergo intramolecular disproportionation reactions leading to saturated products that further decompose to two radicals each. Shown in Figure 5, BR1 leads to the formation of two cyclic C_5 radicals that further react to form cyclopentadiene and hydrogen radicals or allyl and ethene. BR3 has a decomposition path leading to the formation of ethyl and a bicyclo[3.3.0]oct-2-enyl radical. Approximately 92% of the decomposition of JP-10 is governed by hydrogen abstraction reactions at the pyrolysis conditions. The main abstractors are the hydrogen radical and allyl, each accounting for approximately one fourth of the total decomposition rate of JP-10. The other main abstractors from JP-10 are the methyl radical, phenyl and cyclopent-2-enyl. The favored abstraction site is the secondary carbon atom of the norbornane-type ring structure forming R8 in Figure 5. R8 decomposes mainly to cyclopentadiene and cyclopentyl. Hydrogen abstractions leading to R4, R5 and R6 are nearly equally important. R4 decomposes to an allyl radical and norbornene, which then decomposes by a reverse Diels-Alder reaction to cyclopentadiene plus ethene. This reaction path has also been proposed by Herbinet *et al.* [9] However, the reaction network developed by RMG predicts alternative decomposition channels for radicals R5 and R6. Almost 50% of the R5 radicals form benzene, by (a) a path that leads to the formation of butadiene and cyclohexenyl through a bicyclo(3.1.0)hexane ring intermediate or (b) a path that leads to the formation of two ethene molecules and cyclohexadienyl that involves the formation of intermediate tricyclo[4.2.1.1(2,5)]decyl radicals. Channel (a) is favored at higher temperatures while at lower temperatures (< 1000 K) channel (b) is dominant. The rate of production analysis further shows that R6 primarily decomposes to

toluene or important toluene precursors such as 3-methylidenecyclohexa-1,4-diene.

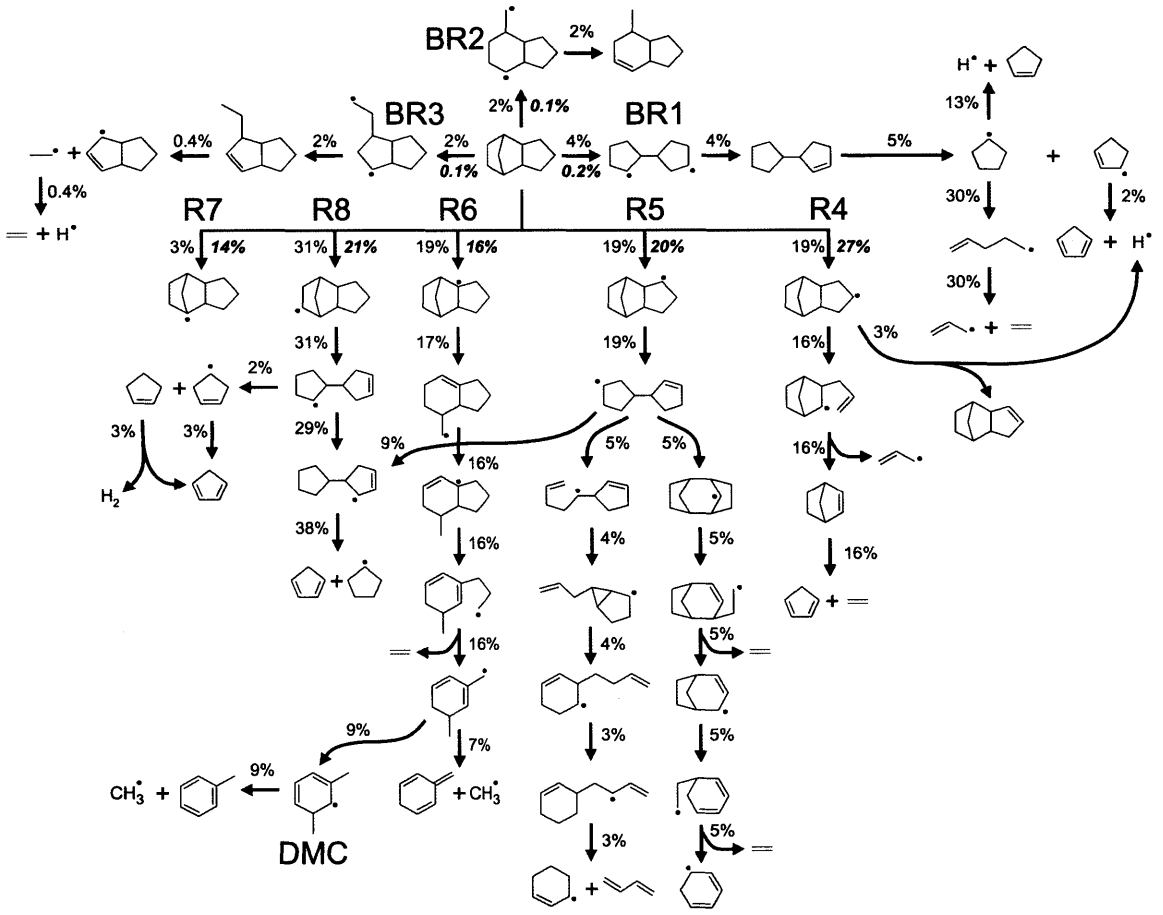
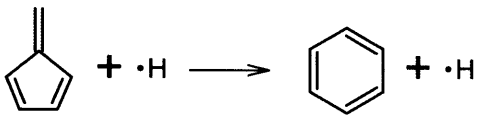


Figure 4.4: Predicted major decomposition pathways of JP-10 pyrolysis (9.34 mol% JP-10 in N_2) and stoichiometric combustion (0.62 mol% JP-10 and 8.72 mol% O_2 in N_2) in an isothermal batch reactor operated at 1000 K and 1.7 atm. Rates are expressed relative to the total JP-10 consumption rate and are reported at a conversion of 10%. The values for the pyrolysis are written in normal font while those for the stoichiometric combustion are indicated in italic and bold.

Under stoichiometric oxidation conditions at 1.7 atm and 1000 K, it was found that at 10% conversion, ring opening reactions contribute to a mere 0.4% of the total consumption rate of JP-10. This is almost a factor of 20 lower than observed for pyrolysis. Chain branching reactions, particularly $H + O_2 \rightleftarrows OH + O$ and $H_2O_2 \rightleftarrows OH + OH$, lead to an increase in the radical count and thus a faster radical mechanism. Abstractions by OH and OOH account for almost 80% of the

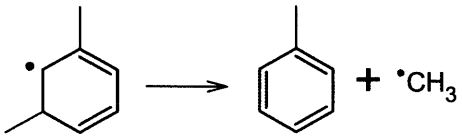
total consumption rate of JP-10. Hydrogen abstractions predominantly occur from secondary carbon sites in JP-10, leading to alkyl radicals R4, R5, and R8. In contrast to pyrolysis, hydrogen abstraction reactions from the tertiary carbon atom of the norbornane structure become important and lead to a significant reaction channel for R7 formation. This channel accounts for almost 14% of JP-10's decomposition under stoichiometric combustion whereas it contributes to only 3% of the consumption in pyrolysis. R7 primarily decomposes through β-scissions to butadiene and C_6 products passing through an intermediate 9-membered ring structure.

RMG identified reaction paths to the following aromatic products: benzene, toluene, ethylbenzene, styrene and indene. As shown in Figure 4.4, a large part of R5 decomposes to cyclohexenyl and cyclohexadienyl radicals, which can easily dehydrogenate to form benzene. Most of the benzene (approximately 70% at 1000 K, 1.7 atm and a conversion of 10%) is hence formed by the β-scission reaction in cyclohexadienyl leading to benzene and a hydrogen atom. Other reactions leading to the formation of benzene are well-skipping reactions as studied by Sharma *et al.* [25] Approximately 25% of the benzene is computed to be formed by RXN 4.1, the well-skipping reaction involving fulvene:



(RXN 4.1)

The formation of fulvene is initiated by the recombination reaction of a methyl radical with cyclopentadienyl, which forms methylcyclopentadiene. Hydrogen abstraction reactions from this component followed by β-scission/disproportionation reactions yield fulvene. Toluene is predominantly formed by decomposition of R6. The model predicts that approximately 50% of the toluene is formed by RXN 4.2, β-scission reactions of dimethylcyclohexadienyl (DMC in Figure 4.4):

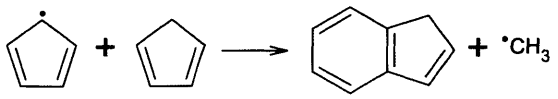


(RXN 4.2)

The other 50% is formed by hydrogen abstraction reactions involving benzyl radicals, which are formed by hydrogen abstraction from 3-methylidenecyclohexa-1,4-diene, another decomposition product of R6. As benzyl is fairly abundant in the system, the main reaction path leading to ethylbenzene is the recombination of benzyl with methyl radical. Ethylbenzene easily dehydrogenates and forms styrene in

successive reactions. Indene is predominantly formed from a path initiated by the addition of cyclopentadienyl to cyclopentadiene (RXN 4.3). This model uses the rate coefficient for this well-skipping reaction computed by Cavallotti *et al.* [31]

The reaction network presented here can be regarded as an extension of the Vandewiele *et al.* [20] JP-10 pyrolysis model to include combustion. Due to the scarcity of rate rules in certain reaction families in the older version of RMG, Vandewiele *et al.* opted not to include intramolecular hydrogen abstraction reactions and *exo-* and *endo*-cycloadditions. The new network reported here includes these reaction families with the addition of new rate rules and additional updates to the thermo and kinetics databases. As shown in Figure 4.4, these reaction families are important in the paths towards toluene and benzene formation. The inclusion of these pathways corrects for the underestimation of ethene, coproduct in the formation of benzene, and overestimation of cyclopentadiene yields in the Vandewiele *et al.* mechanism.



(RXN 4.3)

4.3.2 Pyrolysis and oxidation shock tube speciation

CHEMKIN-PRO [32]was used to simulate the experimental shock tube conditions in an isothermal isobaric batch reactor with an initial loading of 0.2% mol JP-10 with oxygen at fuel-to-oxidation equivalence ratios of 1.0, 0.24, and 0.14 diluted in Ar. The main decomposition products were found to be ethene, cyclopentadiene, cyclopentene, propene, allene and aromatics such as benzene and toluene. Figure 4.5 shows that as expected, JP-10 decomposes faster under leaner conditions. The main reason for this is that oxygen speeds up initiation of the radical mechanism by abstracting hydrogen atoms from JP-10 and also participates in chain branching ratios which increase the total number of radicals in the system. An important chain branching reaction in this regard is shown in RXN 4.4, in which one H radical reacts with O_2 to form two reactive radicals:

$$H + O_2 \rightleftarrows OH + O \qquad \text{(RXN 4.4)}$$

The model is able to predict the decomposition rate of JP-10 as a function of the temperature but tends to underpredict the yields of the major decomposition products (ethene and cyclopentadiene) by a factor of three. Propene and allene, on the other

hand, are predicted fairly well by the model. In particular, the dependence of the propene yield on equivalence ratio is captured accurately for all studied equivalence ratios. Since most propene is formed through hydrogen abstractions by allyl from the parent molecule, the model proves to accurately capture the hydrogen abstractions from the parent fuel and manages to accurately describe the influence of oxygen on the radical pool, responsible for those hydrogen abstraction reactions. This is also supported by the good agreement for the JP-10 decomposition profiles, which are dominated by hydrogen abstraction reactions.

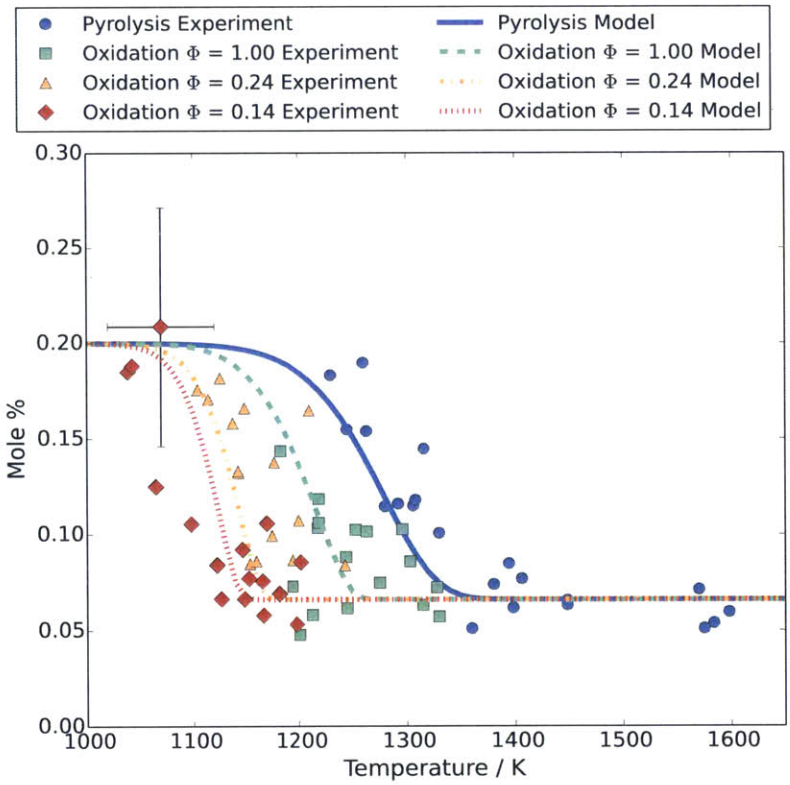


Figure 4.5: Experimentally measured (data points) and simulated concentration profiles (lines) of JP-10 as a function of temperature. Pyrolysis and oxidation experiments were conducted in a shock tube with an initial JP-10 loading of 0.2 mol% diluted in Ar, P = 7 atm, and τ = 500 μs. Experimental error is shown on a single data point to illustrate uncertainties in concentration ($\pm$ 30%) and in temperature ($\pm$ 50 K).

Figure 4.7 illustrates that the model accurately captures the formation of most aromatic compounds and the important intermediates leading to aromatic products, such as 1,3-cyclohexadiene. Under pyrolysis conditions and a residence time of 500

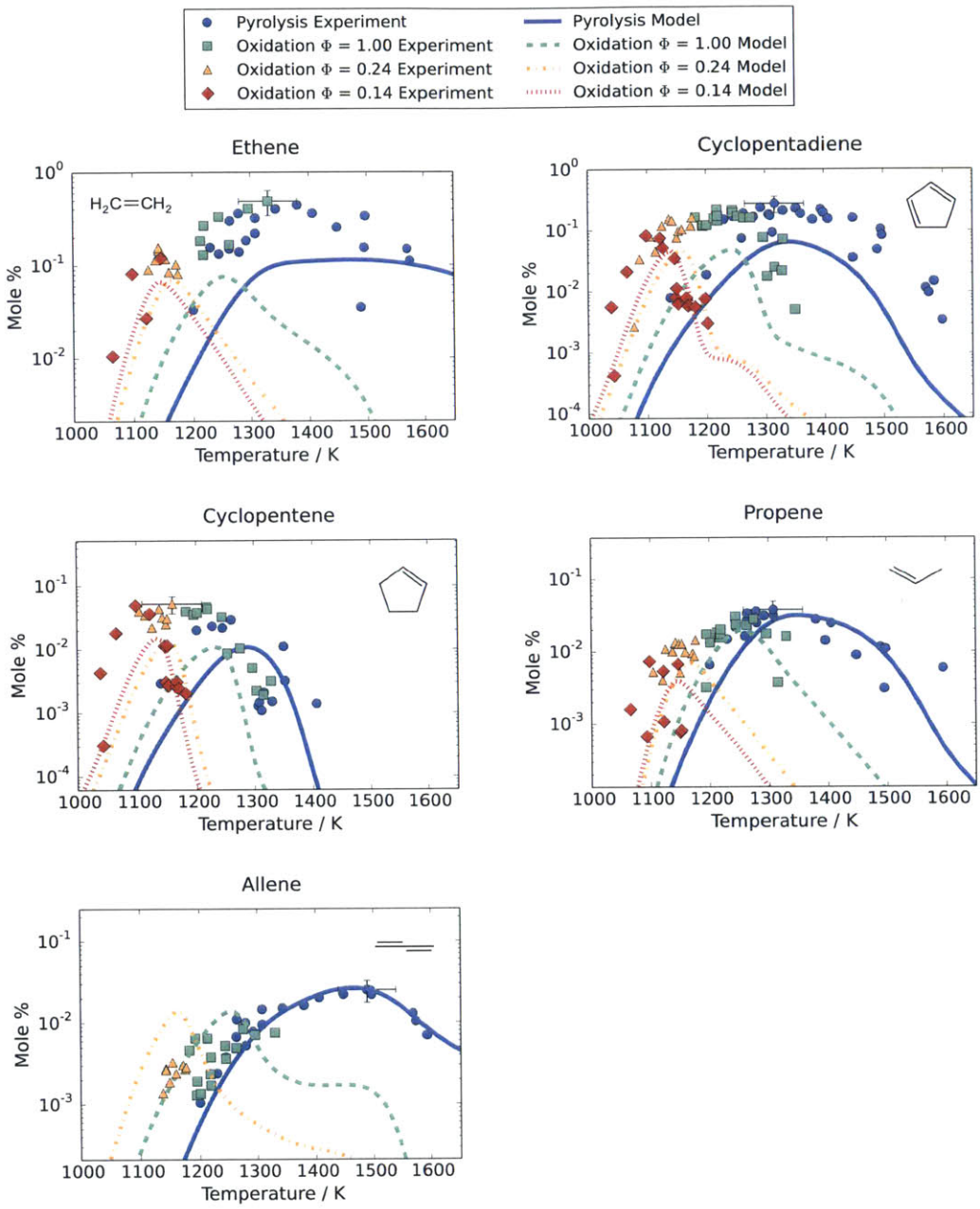


Figure 4.6: Experimentally measured (data points) and simulated concentration profiles (lines) of JP-10's main decomposition products as a function of temperature. Pyrolysis and oxidation experiments were conducted in a shock tube with an initial JP-10 loading of 0.2 mol% diluted in Ar, P = 7 atm, and τ = 500 μs. Experimental error is shown on a single data point to illustrate uncertainties in concentration ($\pm$ 30%) and in temperature ($\pm$ 50 K).

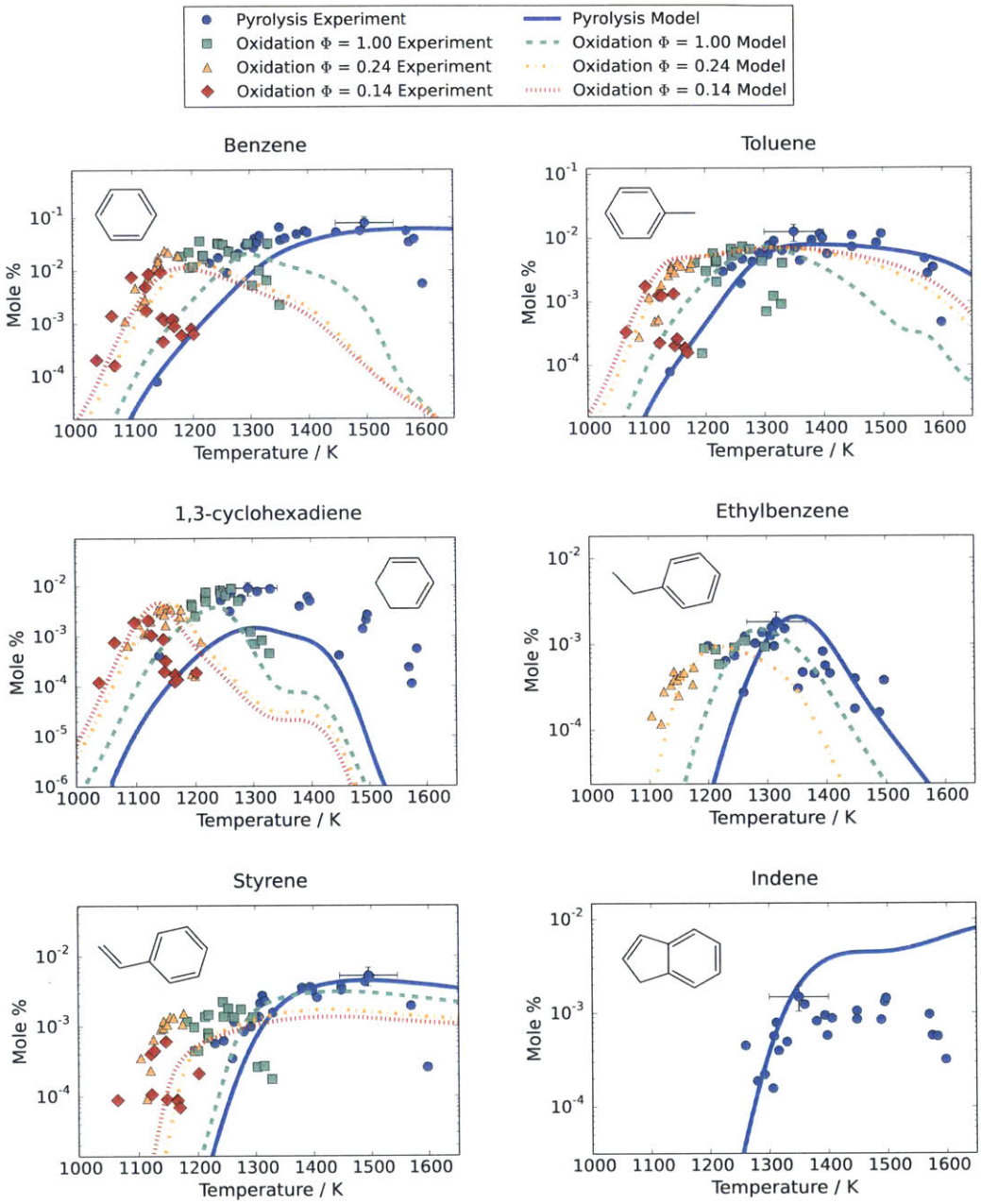


Figure 4.7: Experimentally measured (data points) and simulated concentration profiles (lines) of aromatics and related species from JP-10's decomposition as a function of temperature. Pyrolysis and oxidation experiments were conducted in a shock tube with an initial JP-10 loading of 0.2 mol% diluted in Ar, P = 7 atm, and τ = 500 μs. Experimental error is shown on a single data point to illustrate uncertainties in concentration ($\pm$ 30%) and in temperature ($\pm$ 50 K).

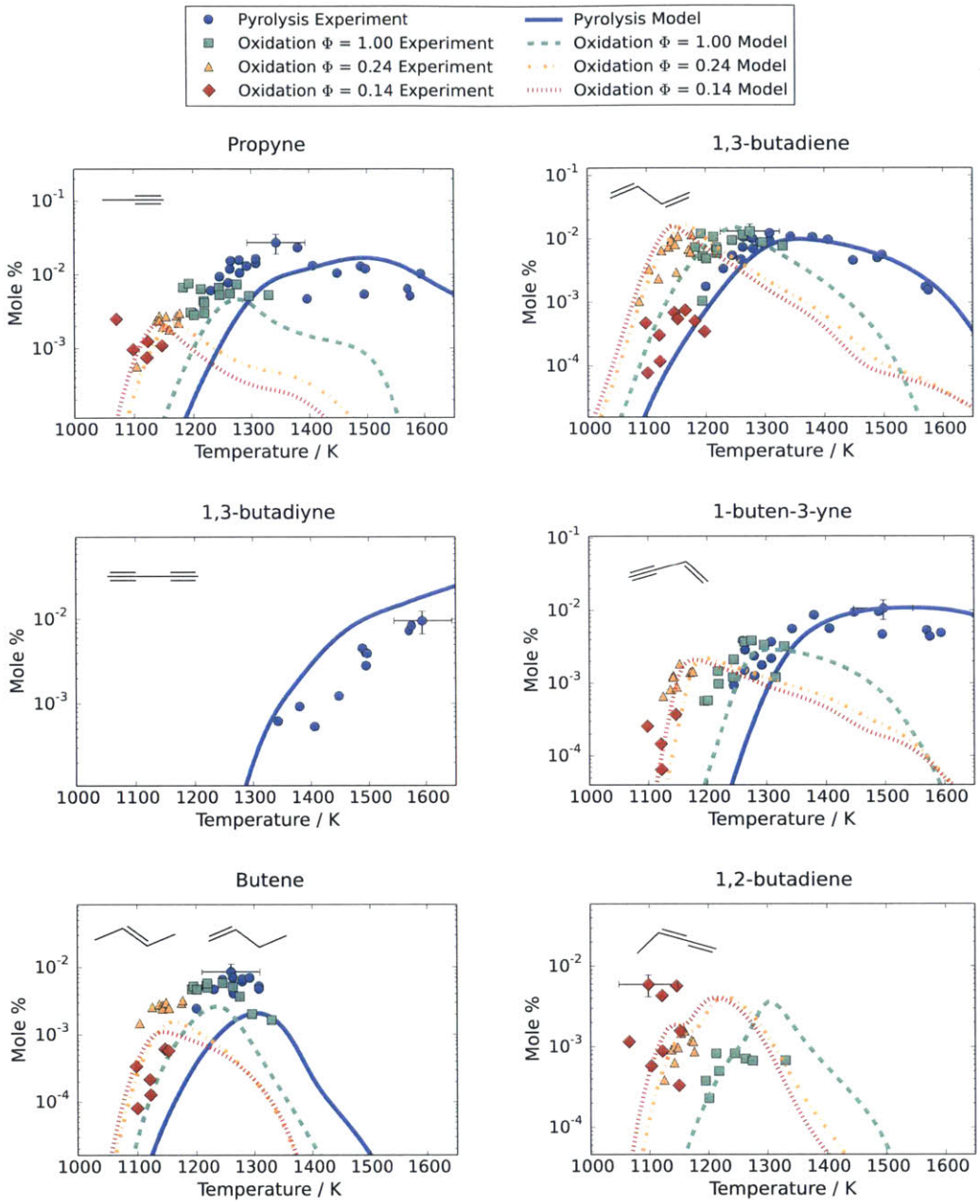


Figure 4.8: Experimentally measured (data points) and simulated concentration profiles (lines) of C_3 and C_4 species from JP-10's decomposition as a function of temperature. Pyrolysis and oxidation experiments were conducted in a shock tube with an initial JP-10 loading of 0.2 mol% diluted in Ar, P = 7 atm, and τ = 500 μs. Experimental error is shown on a single data point to illustrate uncertainties in concentration ($\pm$ 30%) and in temperature ($\pm$ 50 K).

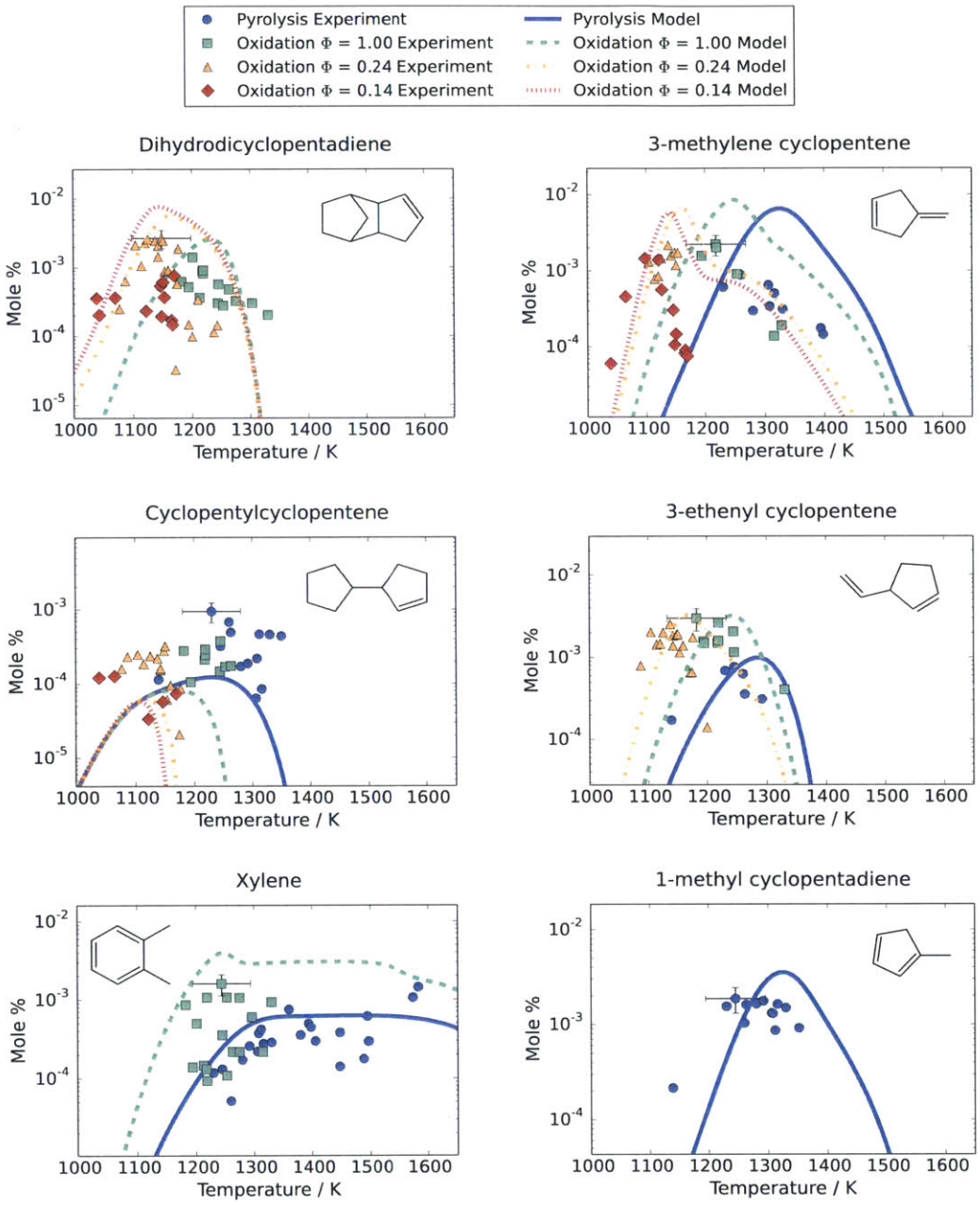


Figure 4.9: Experimentally measured (data points) and simulated concentration profiles (lines) of other cyclic species from JP-10's decomposition as a function of temperature. Pyrolysis and oxidation experiments were conducted in a shock tube with an initial JP-10 loading of 0.2 mol% diluted in Ar, P = 7 atm, and τ = 500 μs. Experimental error is shown on a single data point to illustrate uncertainties in concentration ($\pm$ 30%) and in temperature ($\pm$ 50 K).

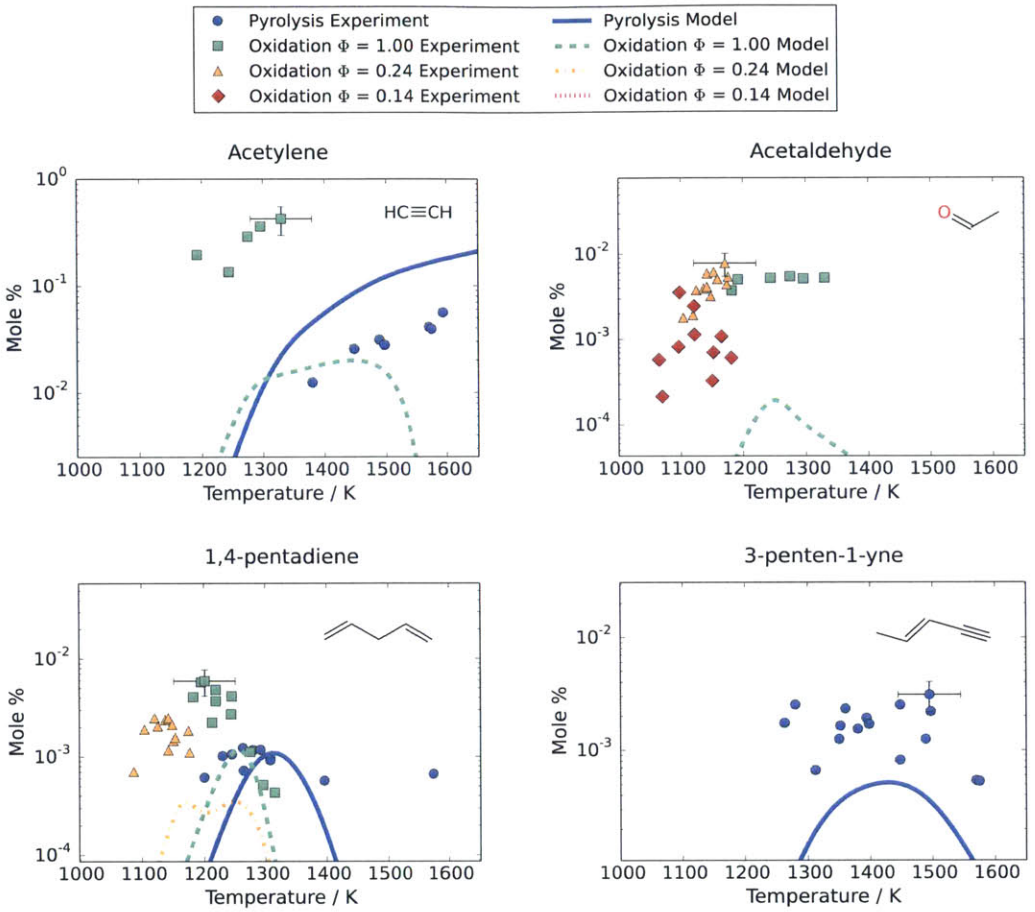
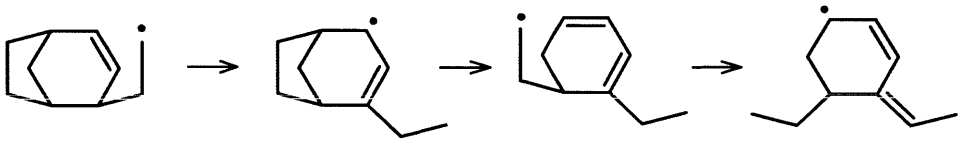


Figure 4.10: Experimentally measured (data points) and simulated concentration profiles (lines) of C_2 and C_5 species from JP-10's decomposition as a function of temperature. Pyrolysis and oxidation experiments were conducted in a shock tube with an initial JP-10 loading of 0.2 mol% diluted in Ar, P = 7 atm, and τ = 500 μs. Experimental error is shown on a single data point to illustrate uncertainties in concentration ($\pm$ 30%) and in temperature ($\pm$ 50 K).

μs, most aromatic components (indene, toluene, styrene ...) have a maximum yield around 1400 to 1500 K. The 1,3-cyclohexadiene and ethylbenzene concentrations peak at slightly lower temperatures as they are the main intermediates leading to benzene and styrene, respectively. Although most oxidation experiments were performed at lower temperatures (1000 - 1200 K) where pre-ignition products can be recovered, the experimental data suggest that the observed maximum in the yield curves for aromatics shifts to lower temperatures. This is also supported by the model. In Figure 4.8, the profiles of the main C_3 and C_4 species are shown. Once again, the model succeeds to reproduce most components within a factor of two to three. The

largest discrepancies between the model and experiments are found for 1,3-butadiene, in particular for the leanest oxidation case ($\Phi = 0.14$), and for butene. The model systemically underestimates the butene yield by a factor of 2 to 3 but seems to be able to accurately reproduce the temperature at which the butene yield will be maximal.

Species that have been systematically observed during the pyrolysis experiments but are missing in the model are summarized in Table 4.2. The main missing components are 1,3,5-cycloheptatriene and 2-propenylidene cyclobutene. Both products are structural isomers of toluene. We expect that 1,3,5-cycloheptatriene and 2-propenylidene cyclobutene are predominantly formed by isomerization reactions on the C_7H_7 surface followed by hydrogen abstraction reactions. Other missing components are mostly (de)hydrogenated forms of some of the major reactions products. Phenylethyne is obtained by dehydrogenation of styrene while indane and 1,3-cycloheptadiene are the hydrogenated products of indene and 1,3,5-cycloheptatriene. The formation of 4-ethyl 3-ethylidene cyclohexene and 1,2-divinyl cyclohexane further support the newly discovered reaction paths towards benzene. Intramolecular hydrogen shifts in the major intermediates, such as for 4-ethylbicyclo[3.2.1]oct-2-ene shown in RXN 4.5, can easily lead to the observed species.



(RXN 4.5)

4.3.3 Pyrolysis flow tube speciation

A second validation for the model compares our predictions with the JP-10 pyrolysis flow tube experimental data reported by Vandewiele *et al.* [30] These experiments were obtained in a bench scale setup recently used for studies on butanol and aromatic hydrocarbons. The experimental setup is extensively described in the previous papers [33–35] and described briefly here. The reactor is a 1.475-m long, 6-mm internal diameter tube made of Incoloy 800HT (Ni, 30-35; Cr, 19-23; and Fe > 39.5 wt%). The analysis section consists of a GC×GC equipped with a mass spectrometer to analyze the C_5^+ fraction, while the C_4^- fraction is analyzed by a Refinery Gas Analyzer (RGA) equipped with two TCD's and a FID. Residence times varied around 0.4 s while conversions between 4 and 94% were obtained.

Two dilutions (7 and 10 mol %) of JP-10 in nitrogen were studied and the results are presented in comparison with model predictions in Figure 4.11 and Figure 4.12,

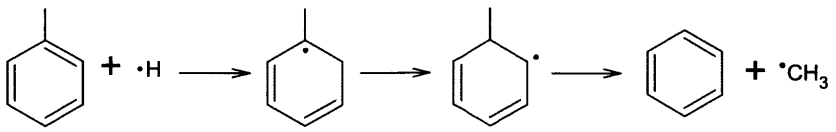
Table 4.2: Species that have been experimentally observed but are missing in the model. The peak concentrations listed are from the shock tube pyrolysis experiments. The initial JP-10 feed was 0.2 mol%. Structures are deduced from experiment based on mass spectrum and retention time in GC/MS.

Name	Chemical Formula	Peak Concentration (mol %)
2-propenylidene cyclobutene	C_7H_8	1.2×10^{-2}
1,3,5-cycloheptatriene	C_7H_8	2.4×10^{-2}
1,3-cycloheptadiene	C_7H_{10}	4.4×10^{-3}
bicyclo(4.1.0)hept-2-ene	C_7H_{10}	4.0×10^{-3}
phenylethyne	C_8H_6	5.3×10^{-3}
naphthalene	C_8H_8	4.7×10^{-4}
indane	C_9H_{10}	5.6×10^{-4}
4-ethyl 3-ethylidene cyclohexene	$C_{10}H_{16}$	6.5×10^{-3}
1,2-divinyl cyclohexane	$C_{10}H_{16}$	2.1×10^{-3}

with time traces computed as normalized mole percentages excluding nitrogen. It can be seen that the model is able to accurately reproduce the decomposition rate of JP-10 for both dilution conditions. At higher temperatures 1020 to 1080 K, the model tends to overestimate the conversion of JP-10 by 20% and consequently slightly over predicts the yields of the main decomposition products.

The largest discrepancy between the model and simulated data is the methane yield. Experimentally it is observed that up to 15% of the products consist of methane (mole based), while the model only predicts values up to 10%. This is directly related to a too low methyl concentration in the simulated radical pool, as methane is predominantly formed by hydrogen abstraction reactions involving the methyl radical. This discrepancy can have various origins, i.e. missing pathways that produce methyl or recombination/addition reactions involving the methyl radical that are too fast. Due to the overestimation of the toluene concentration, we suspect that part of the underestimation of the methyl concentration is due to missing either some of the addition reactions of hydrogen radicals to methylated aromatics (toluene, xylene ...) or missing some of the elimination reactions leading to CH_3 radicals, *e.g.* (RXN 4.6). Due to breaking aromaticity in the ring, some of the independent addition reactions are deemed too slow to be included by RMG, however, the sum of their contributions

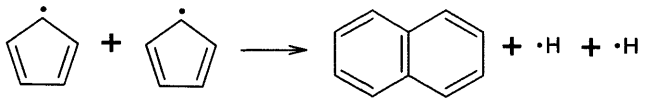
can be significant.



(RXN 4.6)

Figure 4.11 and Figure 4.12 illustrate that temperature dependencies and relative yields for the major components are reproduced well by the model. Both the experimental data and model predict a maximum for propene and cyclopentadiene at temperatures around 1050 K. Most aromatic and minor components are reproduced within a factor of two, again capturing most temperature dependencies. Trace components such as ethyne, propyne and allene are also predicted by the model and generally agree within a factor of 4 with the experimental data. Contradicting the model is the higher concentration 10 mol% JP-10 pyrolysis experiment showing a decreasing propyne yield with increasing temperature. However, this trend is reversed in the 7 mol% experiment, where the propyne yield increases with temperature as predicted by the model.

An important missing species in the model is naphthalene, which accounts for approximately 1 mol% of the measured product distribution excluding N_2 at a temperature of 1080 K in the experiments. We found that in the lower temperature region (1000 to 1100 K), the rate coefficient reported by Murakami *et al.* [36] for RXN 4.7:



(RXN 4.7)

is able to predict the observed naphthalene quite well. However, we opted not to include this reaction in the network since at higher temperatures (1300 to 1400K), this rate leads to an over prediction of the naphthalene yield by a factor of 5. This systematic over prediction of the naphthalene yield by inclusion of RXN 4.7 using the Murakami *et al.* rate coefficient has also been observed by other authors. [37]

4.3.4 Ignition delay

Four experimental datasets for JP-10 ignition delay can be found in literature. [10–13] The summary of the experimental conditions and ignition delay measurement metrics can be found in Table 4.3. The expression $\tau = AP^n \exp(B/T)$ can be used

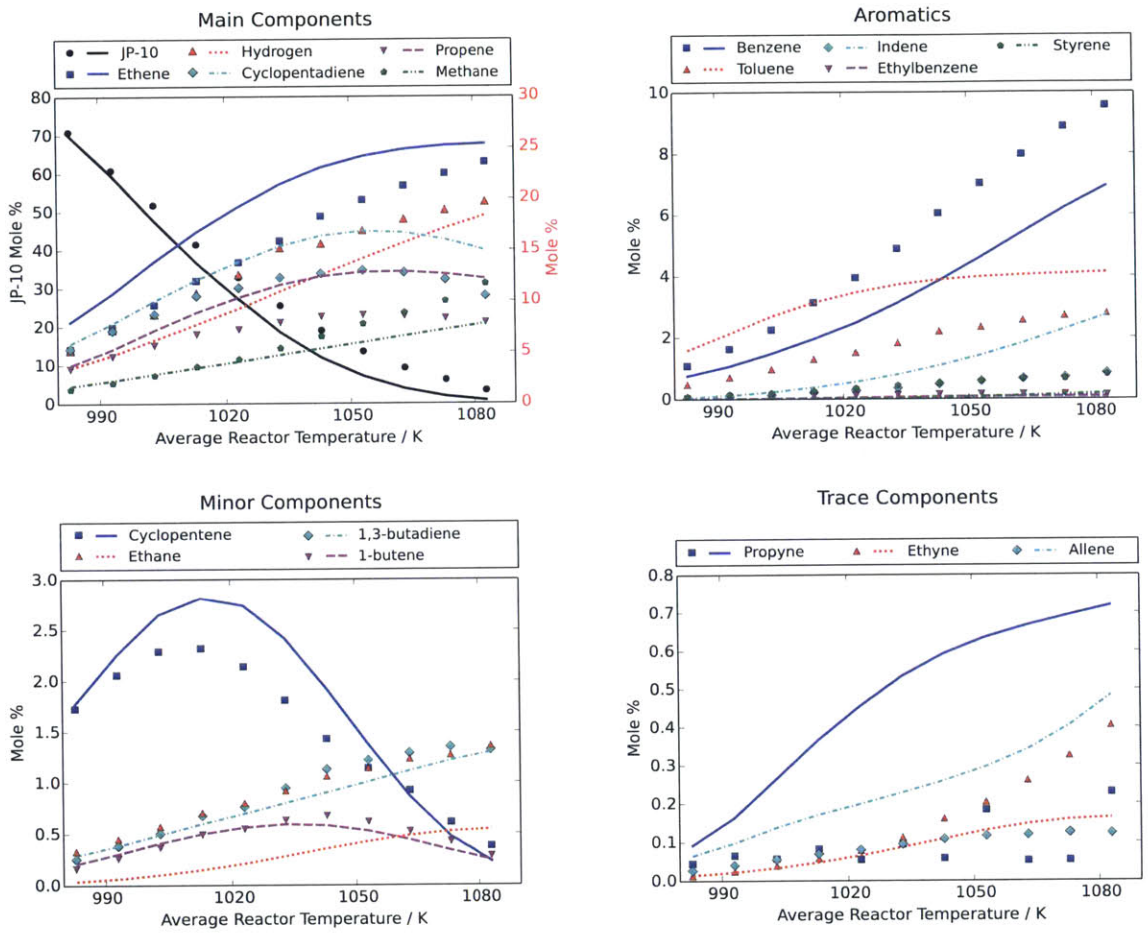


Figure 4.11: Experimentally measured (data points) and simulated concentration profiles (lines) as a function of temperature for the pyrolysis of JP-10 as reported by Vandewiele *et al.* Experimental conditions: 7 mol% JP-10 diluted in N_2, P = 0.14 MPa, mass flow rate of JP-10 = 2.33×10^{-2} g/s. Mole percentages are normalized excluding N_2.

to correlate the experimental data (where P is in atmospheres and T in Kelvin). By fitting a unique correlation for each dataset at the given mole % JP-10 and stoichiometric ratio, pressure corrected ignition delays $\tau_P = e.g.(P/P_0)^n \tau_0$ can be derived, allowing datasets amongst multiple pressures to be compared. The fitted parameters for these correlations can be found in the Supporting Information of reference [1]. In Figure 4.3, the stoichiometric ignition delays from all four datasets are pressure-corrected to $P_0 = 7$ atm and compared with simulated ignition delays. It can be seen from this figure that there are some discrepancies amongst the existing

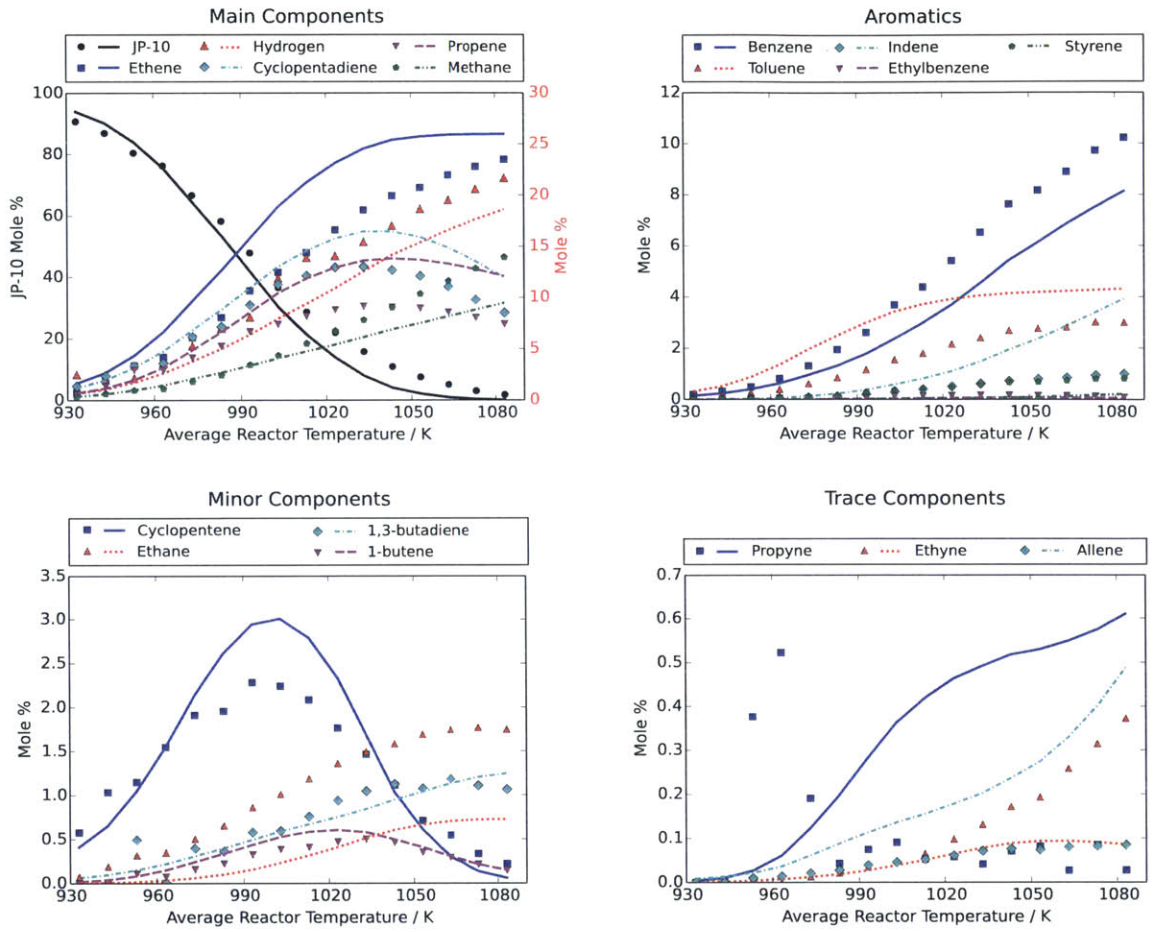


Figure 4.12: Experimentally measured (data points) and simulated concentration profiles (lines) as a function of temperature for the pyrolysis of JP-10 as reported by Vandewiele *et al.* Experimental conditions: 10 mol% JP-10 diluted in N_2, P = 0.14 MPa, mass flow rate of JP-10 = 2.33×10^{-2} g/s. Mole percentages are normalized excluding N_2.

datasets. Though the Colket and Spadaccini [11] experiments were conducted at lower feed concentration (0.15 mol% JP-10), the ignition delays were comparatively fast compared to the Davidson *et al.* [10] experiments conducted at slightly higher feed concentrations (0.20 mol% JP-10). The Mikolaitis *et al.* [12] dataset has the fastest ignition delays, as expected from a higher feed concentration (1.48 mol% JP-10). The apparent activation energy of ignition (the slope of the correlation) for the Wang *et al.* [13] ignition delay data appears to be smaller than the for the other datasets. The discrepancies between these datasets may be due to slight differences

Table 4.3: Summary of experimental conditions for JP-10 ignition delay datasets available in literature.

Dataset	Temperature range	Pressure range	Equivalence ratios (Φ)	Bath gas	Ignition delay metric
Colket and Spadaccini [11]	1250 - 1500 K	3 - 7 atm	0.5 - 1.5	Ar	Initial rise onset in OH* emission
Davidson *et al.* [10]	1300 - 1700 K	1 - 6 atm	0.5 - 2.0	Ar	Peak CH* emission
Mikolaitis *et al.* [12]	1200 - 2000 K	10 - 25 atm	1.0	N_2	Peak CH* emission
Wang *et al.* [13]	1000 - 2100 K	1.5 - 5.5 atm	0.25 - 2.0	Ar	Initial rise onset in CH* emission for $\Phi = 1.0$ experiments, initial rise onset in OH* emission for all other experiments

in the ignition delay metrics employed, as well as accuracy of the JP-10 concentration measurements. The Davidson and Mikolaitis datasets used the peak CH* emission as the indicator of ignition, while both Colket and Wang experiments used initial rise onsets of the CH* emission, leading to faster measured ignition times. Mikolaitis and coworkers [12] found that at intermediate temperatures, two emission peaks for ignition delay were present. The first peak ignition delays from the CH emission were recommended by the authors the true indicator of ignition and are compared with the model predictions here.

Ignition delays at stoichiometric conditions were simulated using the model at 7 atm and experimental feed conditions and compared with experiment in Figure 4.13. The maximum in the $[C_2H][O]$ product was used as the metric for ignition delay in the simulations. Since CH* emissions occur primarily via the reaction $2H + O \rightleftharpoons CH^* + CO$, the $[C_2H][O]$ product most effectively simulates experimental CH emissions used to measure ignition delay. [10] In Figure 4.13, we see that the simulated ignition delays are slower than experiment for the Davidson and Colket datasets, but systematically faster than the Mikolaitis ignition delays. However, the predicted apparent activation energy for ignition agrees well with all datasets except

the Wang experiments.

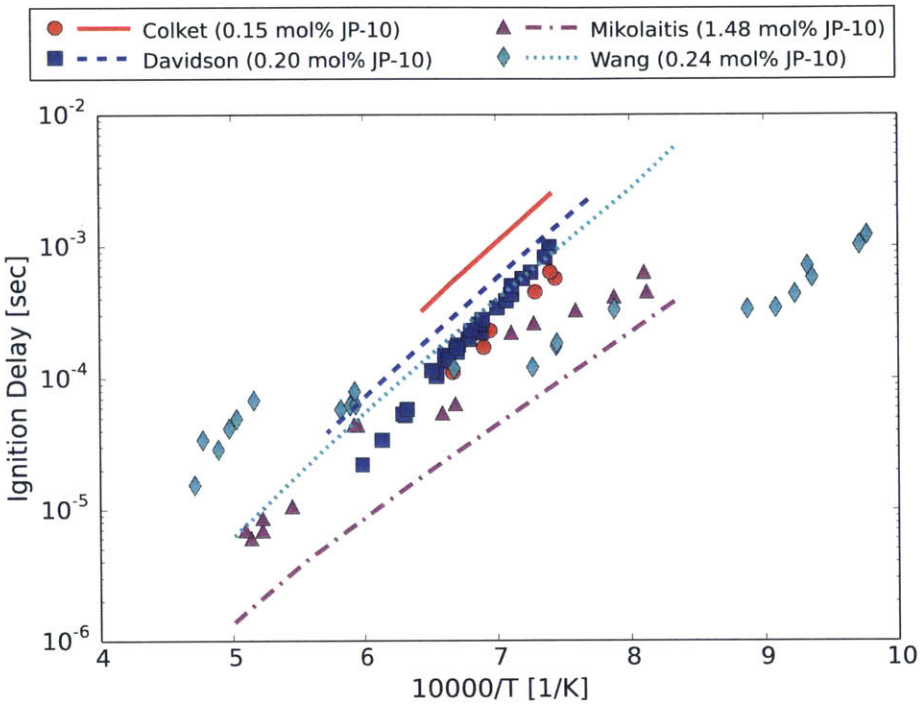


Figure 4.13: Pressure corrected (7 atm) experimental ignition delays for Colket (0.15 mol% JP-10 feed, circles), Davidson (0.20 mol% JP-10 feed, squares), Wang (0.24 mol% JP-10 feed, diamonds), and Mikolaitis (1.48 mol% JP-10 feed, triangles) at stoichiometric conditions, along with simulated ignition delays (lines) at $\Phi = 1.0$ and P = 7 atm.

In Figure 4.14, a parity plot is shown between experimentally observed and modeled ignition delay times for all four experimental datasets over all experimental conditions. It can be seen from Figure 4.14 that the model succeeds to reproduce all reported ignition delays covering a broad range of temperatures to mostly within a factor of 4.

In particular, the experimental ignition delays measured by Davidson and by Colket are mostly reproduced by the model (within a factor of 2 to 3).

OH sensitivity analysis was performed for the model at stoichiometric conditions (1500 K and 7 atm for a 0.20 mol% feed JP-10 in argon, 50% JP-10 conversion), with the most important reactions presented in Figure 4.15. In agreement with Davidson *et al.* [5], we find that the OH concentration is most sensitive to the branching reaction $H + O_2 \rightleftarrows O + OH$. Davidson's model used a modified version of the Lindstedt and Maurice mechanism [38] and found that the OH concentration is pri-

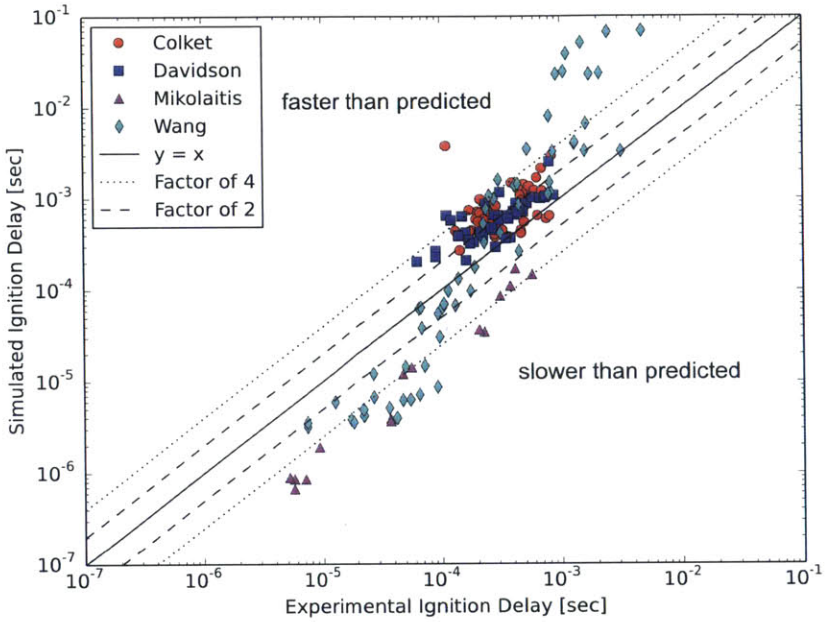


Figure 4.14: Parity plot comparing simulated and measured ignition time delays for JP-10 for existing experimental datasets: Colket (circles); Davidson (squares); Mikolaitis (triangles); and Wang (diamonds).

marily sensitive to C_2 oxidation reactions. Figure 4.15 illustrates that besides the C_2 oxidation chemistry, ignition delays are also sensitive to reactions with some of the decomposition products. These reactions are mainly recombination reactions or hydrogen abstraction reactions involving long-lived, resonantly stabilized radicals such as cyclopentadienyl and allyl. The largest negative sensitivity coefficient is obtained for the recombination of H with allyl forming propene, followed by the recombination of H with cyclopentadienyl forming cyclopentadiene and the recombination of H with propargyl radical to produce allene. Particularly in oxidation conditions, resonantly stabilized radicals can accumulate in larger concentrations and combine with the relatively unreactive HO_2 radical, contributing to chain branching reactions that accelerate ignition. [39] For instance, the sensitive reaction of cyclopentadienyl radical with HO_2 can lead to active OH radical generation through the dissociation of the RO−OH bond in the hydroperoxide product. This shows that a predictive model for ignition delays requires (i) correct small molecule chemistry and (ii) a correct treatment of the main stable radicals and their interaction with the radical pool.

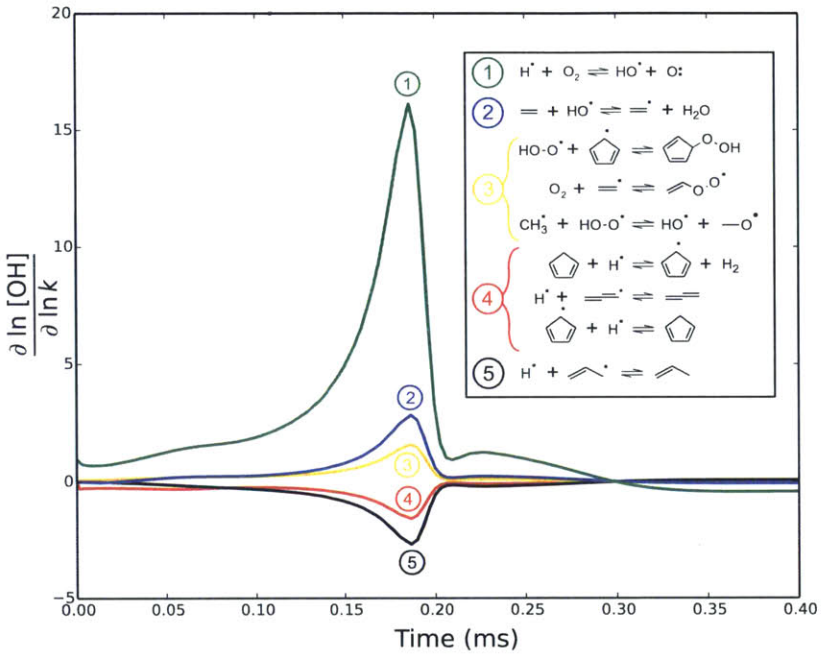


Figure 4.15: Most sensitive reaction kinetics found in normalized OH sensitivity analysis (dln[OH]/dln k) conducted at 1500 K, 7 atm, Φ = 1.0, 0.20 mol% JP-10 diluted in Ar in an adiabatic batch reactor. Reactions grouped together have nearly equal importance.

4.4 Conclusion

Despite JP-10's popularity as a jet propellant, few detailed kinetics models for JP-10 combustion exist, and most are unvalidated against pre-ignition product distributions, mainly due to lack of experimental data. This work aims to gain new insights on JP-10's complex decomposition chemistry through a combination of experimental and modeling efforts. Experiments were carried out in single-pulse shock tubes at pyrolysis and oxidation conditions, ranging from very lean (Φ = 0.14) to stoichiometric (Φ = 1.0), with products analyzed using GC/MS. The experiments show that JP-10 decomposes primarily to ethene, propene, cyclopentadiene, and aromatics consisting mostly of benzene and toluene. GC/MS allowed for the identification of a variety of trace components such as 1,2-divinylcyclohexane, butadiene and 1,3-cyclohexadiene. Though minor, these components are relatively important from a modeling viewpoint as they help to elucidate intermediate pathways in JP-10's decomposition mechanism.

The kinetic model was developed using RMG, an automatic reaction mechanism

generator which uses a rate based termination criteria to include species with the most significant fluxes to a reaction network. Four models were generated, covering temperatures from 1000 K up to 2000K with equivalence ratios ranging from very lean ($\Phi = 0.14$) to pyrolysis conditions. The final merged network contains almost 700 species reacting through more than 15000 reactions. Despite many similarities with existing mechanisms, this is the first model that contains multiple reaction paths from the parent fuel towards benzene. At lower temperatures (< 1000K), benzene formation involves the formation of a tricyclo[4.2.1.1(2,5)]decyl radical that undergo four successive β-scissions to form cyclohexadienyl and two ethene molecules. Higher temperatures favor an alternate reaction path that involves the formation of a bicyclo(3.1.0)hexane ring structure leading to butadiene and cyclohexenyl. These reaction paths are supported by the identification of intermediate species by experiment and also backed up by theory.

The generated model was thoroughly validated against the new experimental data and literature data and succeeds in accurately simulating JP-10 conversion rates under all equivalence ratios. The model captures most temperature dependencies in the shock tube product distributions, with agreement between the modeled and experimentally observed product concentrations generally within a factor of 3. The model was also validated using the speciation data from recent pyrolysis experiments conducted by the Ghent group, succeeding in predicting the major products, ethene and cyclopentadiene, within 30% and reproducing most other product concentrations within a factor of 2. Experimentally observed ignition delays over four existing literature datasets are reproduced by the model mostly within a factor of 4. Sensitivity analyses show that small molecule chemistry is controlling for the ignition delays at high temperature. We conclude that the combined efforts of experimentalists, theoreticians and modelers has led to a new JP-10 model that is able to reliably reproduce speciation data and ignition delays over a wide range of temperatures and equivalence ratios, as well as elucidate new pathways towards aromatics formation.

4.5 References

[1] C. W. Gao, A. G. Vandeputte, N. W. Yee, W. H. Green, R. E. Bonomi, G. R. Magoon, H.-W. Wong, O. O. Oluwole, D. K. Lewis, N. M. Vandewiele, and K. M. Van Geem. "JP-10 combustion studied with shock tube experiments and modeled with automatic reaction mechanism generation." *Combust. Flame* 162 (8), 2015, pp. 3115–3129. DOI: 10.1016/j.combustflame.2015.02.010.

[2] T. Bruno, M. Huber, A. Laesecke, E. Lemmon, and R. Perkins. "Thermochemical and thermophysical properties of JP-10." NISTIR 6640, 2006.

[3] H. S. Chung, C. S. H. Chen, R. A. Kremer, J. R. Boulton, and G. W. Burdette. "Recent developments in high-energy density liquid hydrocarbon fuels." *Energy & Fuels* 13 (3), 1999, pp. 641–649. DOI: 10.1021/Ef980195k.

[4] C. Joppin. "Cooling Performance of Storable Propellants for a micro rocket engine." PhD thesis. Massachusetts Institute of Technology, 2002.

[5] D. Davidson, D. Horning, M. Oehlschlaeger, and R. Hanson. "The Decomposition Products of JP-10." *AIAA 01-3707 37th Joint Propulsion Conference*, 2001.

[6] S. Nakra, R. J. Green, and S. L. Anderson. "Thermal decomposition of JP-10 studied by micro-flowtube pyrolysis-mass spectrometry." *Combust. Flame* 144 (4), 2006, pp. 662–674. DOI: 10.1016/j.combustflame.2005.08.035.

[7] R. C. Striebich and J. Lawrence. "Thermal decomposition of high-energy density materials at high pressure and temperature." *J. Anal. Appl. Pyrol.* 70 (2), 2003, pp. 339–352. DOI: 10.1016/S0165-2370(02)00181-X.

[8] Y. Xing, X. Yang, W. J. Fang, Y. S. Guo, and R. S. Lin. "Measurement on vapor pressure, density and viscosity for binary mixtures of JP-10 and methylcyclohexane." *Fluid Phase Equilib.* 305 (2), 2011, pp. 192–196. DOI: 10.1016/j.fluid.2011.03.030.

[9] O. Herbinet, B. Sirjean, R. Bounaceur, R. Fournet, F. Battin-Leclerc, G. Scacchi, and P. M. Marquaire. "Primary mechanism of the thermal decomposition of tricyclodecane." *J. Phys. Chem. A* 110 (39), 2006, pp. 11298–11314. DOI: 10.1021/Jp0623802.

[10] D. F. Davidson, D. C. Horning, J. T. Herbon, and R. K. Hanson. "Shock tube measurements of JP-10 ignition." *Proc. Combust. Inst.* 28, 2000, pp. 1687–1692. DOI: 10.1016/S0082-0784(00)80568-8.

[11] M. B. Colket and L. J. Spadaccini. "Scramjet fuels autoignition study." *J. Propul. Power* 17 (2), 2001, pp. 315–323. DOI: 10.2514/2.5744.

[12] D. W. Mikolaitis, C. Segal, and A. Chandy. "Ignition delay for jet propellant 10/air and jet propellant 10/high-energy density fuel/air mixtures." *J. Propul. Power* 19 (4), 2003, pp. 601–606. DOI: 10.2514/2.6147.

[13] S. Wang, H. J. Gou, B. C. Fan, Y. Z. He, S. T. Zhang, and J. P. Cui. "Shock Tube Study of JP-10 Ignition Delay Time." *Chin. J. Chem. Phys.* 2007. DOI: 10.1360/cjcp2007.20(1).48.5.

[14] M. J. Zehe and R. L. Jaffe. "Theoretical Calculation of Jet Fuel Thermochemistry. 1. Tetrahydrodicylopentadiene (JP10) Thermochemistry Using the CBS-QB3 and G3(MP2)//B3LYP Methods." *J. Org. Chem.* 75 (13), 2010, pp. 4387–4391. DOI: 10.1021/jo100050w.

[15] J. W. Bozzelli, A. Castillo, and J. M. Hudzik. "Exo-tricyclo[5.2.1.0(2,6)]decane (TCD or JP-10) decomposition: Thermochemostry and kinetics of ring opening and formed intermediates." *Abstr. Pap. Am. Chem. S.* 244, 2012.

[16] K. Chenoweth, A. C. T. van Duin, S. Dasgupta, and W. A. Goddard. "Initiation Mechanisms and Kinetics of Pyrolysis and Combustion of JP-10 Hydrocarbon Jet Fuel." *J. Phys. Chem. A* 113 (9), 2009, pp. 1740–1746. DOI: 10.1021/Jp8081479.

[17] L. Yue, H. J. Xie, X. M. Qin, X. X. Lu, and W. J. Fang. "A DFT study on the thermal cracking of JP-10." *J. Mol. Model.* 19 (12), 2013, pp. 5355–5365. DOI: 10.1007/s00894-013-2029-z.

[18] G. R. Magoon, J. Aguilera-Iparraguirre, W. H. Green, J. J. Lutz, P. Piecuch, H. Wong, and O. O. Oluwole. "Detailed Chemical Kinetic Modeling of JP-10 (exo-Tetrahydrodicyclopentadiene) High-Temperature Oxidation: Exploring the Role of Biradical Species in Initial Decomposition steps." *Int. J. Chem. Kin.* 44 (3), 2012, pp. 179–193. DOI: 10.1002/kin.20702.

[19] S. C. Li, B. Varatharajan, and F. A. Williams. "Chemistry of JP-10 ignition." *AIAA Journal* 39 (12), 2001, pp. 2351–2356. DOI: 10.2514/2.1241.

[20] N. M. Vandewiele, G. R. Magoon, K. M. Van Geem, M.-F. Reyniers, W. H. Green, and G. B. Marin. "Kinetic Modeling of Jet Propellant-10 Pyrolysis." *Energy & Fuels* 29 (1), 2015, pp. 413–427. DOI: 10.1021/ef502274r.

[21] A. Lifshitz, S. H. Bauer, and E. L. Resler. "Studies with a Single-Pulse Shock Tube .1. Cis-Trans Isomerization of Butene-2." *J. Chem. Phys.* 38 (9), 1963, pp. 2056–2063. DOI: 10.1063/1.1733933.

[22] E. E. Dames. *Energy Frontier Research Center.* Personal Communication. 2014.

[23] J. W. Ochterski, G. A. Petersson, and J. A. Montgomery. "A complete basis set model chemistry. V. Extensions to six or more heavy atoms." *J. Chem. Phys.* 104, 1996, pp. 2598–2619. DOI: 10.1063/1.470985.

[24] A. D. Boese and J. M. L. Martin. "Development of density functionals for thermochemical kinetics." *J. Chem. Phys.* 121 (8), 2004, pp. 3405–3416. DOI: 10.1063/1.1774975.

[25] S. Sharma, M. R. Harper, and W. H. Green. "Modeling of 1,3-hexadiene, 2,4-hexadiene and 1,4-hexadiene-doped methane flames: Flame modeling, benzene and styrene formation." *Combust. Flame* 157 (7), 2010, pp. 1331–1345. DOI: 10.1016/j.combustflame.2010.02.012.

[26] M. K. Sabbe, M.-F. Reyniers, V. Van Speybroeck, M. Waroquier, and G. B. Marin. "Carbon-Centered Radical Addition and β-Scission Reactions: Modeling of Activation Energies and Pre-exponential Factors." *ChemPhysChem* 9 (1), 2008, pp. 124–140. DOI: 10.1002/cphc.200700469.

[27] M. K. Sabbe, A. G. Vandeputte, M.-F. Reyniers, M. Waroquier, and G. B. Marin. "Modeling the influence of resonance stabilization on the kinetics of hydrogen abstractions." *Phys. Chem. Chem. Phys.* 12 (6), 2010, pp. 1278–1298. DOI: 10.1039/b919479g.

[28] G. R. Magoon and W. H. Green. "Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation." *Comp. & Chem. Eng.* 52, 2013, pp. 35–45. DOI: 10.1016/j.compchemeng.2012.11.009.

[29] Computer Program. 2012. URL: http://openmopac.net.

[30] N. M. Vandewiele, G. R. Magoon, K. M. Van Geem, M. F. Reyniers, W. H. Green, and G. B. Marin. "Experimental and Modeling Study on the Thermal Decomposition of Jet Propellant-10." *Energy & Fuels* 28 (8), 2014, pp. 4976–4985. DOI: 10.1021/ef500936m.

[31] C. Cavallotti, D. Polino, A. Frassoldati, and E. Ranzi. "Analysis of Some Reaction Pathways Active during Cyclopentadiene Pyrolysis." *J. Phys. Chem. A* 116 (13), 2012, pp. 3313–3324. DOI: 10.1021/jp212151p.

[32] *CHEMKIN-PRO 15131.* Reaction Design: San Diego. 2013.

[33] M. Harper, K. M. Van Geem, S. P. Pyl, G. B. Marin, and W. H. Green. "Comprehensive reaction mechanism for n-butanol pyrolysis and combustion." *Combust. Flame* 158 (1), 2011, pp. 16–41. DOI: 10.1016/j.combustflame.2010.06.002.

[34] Q. Chen and G. F. Froment. "Thermal cracking of substituted aromatic hydrocarbons. I. Kinetic study of the thermal cracking of i-propylbenzene." *J. Anal. Appl. Pyro.* 21, 1991, pp. 27–50. DOI: 10.1016/0165-2370(91)80014-Y.

[35] Q. Chen and G. F. Froment. "Thermal cracking of substituted aromatic hydrocarbons. II. Kinetic study of the thermal cracking of n-propylbenzene and ethylbenzene." *J. Anal. Appl. Pyro.* 21, 1991, pp. 51–77. DOI: 10.1016/0165-2370(91)80015-Z.

[36] Y. Murakami, T. Saejung, C. Ohashi, and N. Fujii. "Investigation of a new pathway forming naphthalene by the recombination reaction of cyclopentadienyl radicals." *Chem. Lett.* 32 (12), 2003, pp. 1112–1113. DOI: 10.1246/Cl.2003.1112.

[37] R. Robinson and R. Lindstedt. "On the chemical kinetics of cyclopentadiene oxidation." *Combust. Flame* 158 (4), 2011, pp. 666–686. DOI: 10.1016/j.combustflame.2010.12.001.

[38] R. P. Lindstedt and L. Q. Maurice. "Detailed kinetic modelling of n-heptane combustion." *Combust. Sci. and Technol.* 107 (4-6), 1995, pp. 317–353. DOI: 10.1080/00102209508907810.

[39] X. Zhong and J. W. Bozzelli. "Thermochemical and Kinetic Analysis of the H, OH, HO2, O, and O2 Association Reactions with Cyclopentadienyl Radical." *J. Phys. Chem. A* 102 (20), 1998, pp. 3537–3555. DOI: 10.1021/jp9804446.

5

UNDERSTANDING GEOLOGICAL OIL TO GAS CRACKING PROCESSES USING PHENYLDODECANE AS A HEAVY OIL ANALOG

The influence of non-hydrocarbon compounds on the thermal decomposition of kerogens and oils at geologic temperature conditions is a subject of great interest for petroleum reservoir modeling and prospecting.

In particular, organic sulfur (OS) is implicited in accelerating the onset of oil generation in Type II-S kerogen, which contain between 8-14% OS. Type II-S kerogen appears to generate oil at much lower temperatures than Type II kerogens which contain less than 6% sulfur. There is speculation that weak sulfur bonds lead to cleavage at labile sulfur functional groups, unlike in low sulfur kerogen, where oil is generated by the cleavage of oxygen containing functional groups and C−C bonds. [1] However, the effects of OS at low temperatures representative of natural systems have rarely been studied in experiment, and the mechanism by which OS accelerates thermal decomposition is poorly understood. Predictive capabilities at geological conditions are invaluable for understanding petroleum systems and developing more accurate reservoir models.

We examine the influence of organosulfur in the petroleum decomposition by using a model OS compound DEDS (diethyldisulfide, Figure 5.1) on the anhydrous decomposition of a heavy oil analog PDD (phenyldodecane, Figure 5.2).

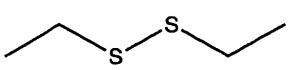


Figure 5.1: DEDS ($C_4H_{10}S_2$); diethyldisulfide; 122 amu

We aim to elucidate the mechanistic role of DEDS on the thermal decomposition of PDD through a combination of confined pyrolysis experiments and chemical kinetic

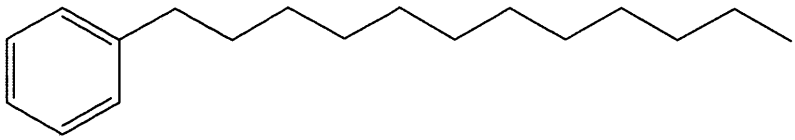


Figure 5.2: PDD ($C_{18}H_{30}$); phenyldodecane; 246 amu

modeling. Experiments were performed by Dr. Eoghan P. Reeves from the Stable Isotope Geobiology Laboratory in the Department of Earth, Atmospheric and Planetary Sciences at MIT. Our goal is to develop an initial framework for understanding the influence of C−S and S−S bonds on alkylaromatic thermal stability, leading to broader implications on the generation and fate of petroleum.

A preliminary DEDS model [2] has been generated in RMG-Java by Dr. Caleb A. Class, a former graduate student in the Green Group who worked extensively on sulfur chemistry. RMG-Java was also used to generate an initial PDD pyrolysis model; however, it lacked critical species found in experiment and never ran to completion. In light of the new low temperature Reeves experiments, we tackle the challenges of modeling the PDD pyrolysis using RMG-Py in this chapter. Due to the sheer number of reaction sites on PDD, memory and CPU speed become limiting during model generation. These issues were ameliorated through reaction filtering in the model enlarging algorithm. The new PDD model was combined with the DEDS chemistry previously generated to understand the implications of sulfur on alkylaromatic decomposition.

5.1 Background

Predictive capability in modeling reaction networks at geological conditions is sought to understand the long term fate of sedimentary organic matter and aid the management of existing hydrocarbon resources. Differences in the initial composition of the kerogen, a complex matrix of organic compounds in sedimentary rocks, and the temperature profiles (typically 100 - 200 °C) under the basin affect the survival and composition of the final hydrocarbon mixture in the reservoir. These hydrocarbons crack and pyrolyze at low temperature for millions of years. Mimicking these high pressure, low temperature geological conditions are impractical for experiment, necessitating models of hydrocarbons representative of heavy oils such as PDD. A kinetic mechanism developed using first principles that corroborates near-geologic temper-

ature experiments can be extrapolated to predict the long term stability of oils in petroleum systems.

Numerous observations from both pyrolysis experiments and natural systems strongly support the argument that the chemical mechanisms of petroleum formation and decomposition are dominated by free radical reactions. [3–12] These chain reactions typically occur as a sequence: initiation reactions from C-C bond scissions form initial radicals, followed by propagation reactions that further decompose the parent compound, and finally termination reactions where radicals quench and recombine to form more stable products. Understanding and predicting factors that can accelerate or hinder these networks in natural systems has been difficult due to their inherent complexity and the immense number of coupled radical reactions involved. [11] Some molecular classes (e.g. aromatics, cyclic structures), for example, can inhibit these networks, thereby reducing overall decomposition rates. [9, 13–17]

Alkylaromatics are a part of the family of natural oils present in petroleum systems. We chose to work with PDD due to the extensive experimental and mechanistic investigations found in literature. [18–22] Aliphatic (often long) alkyl chains bonded to aromatics are common structural moieties naturally occurring in heavy oils, asphaltenes and coals. [23] In addition, a dominant product of long chain alkyl aromatic decomposition is toluene, [18–20, 23, 24] a known inhibitor of free radical decomposition of aliphatic hydrocarbons. Toluene exerts a scavenging effect on alkyl radicals that increases with decreasing temperature, by forming and subsequently terminating resonance-stabilized benzylic radicals. [15–17] Investigating mechanistic effects of OS on alkyl aromatic decomposition is therefore important for understanding the stability of the petroleum hydrocarbon mixtures affected by similar constituents at geologic conditions.

5.1.1 Effects of sulfur on the thermal stability of petroleum systems

There is some general knowledge of sulfur reactivity in crude oils. Gas phase pyrolysis experiments suggest that certain OS functional groups may accelerate free radical reactions. [13] Aromatic sulfur, or thiophenic ring systems are very stable and typically do not affect the thermal stability of kerogen, whereas aliphatic sulfide groups R−S−R′ and disulfide groups R−S−S−R′, in particular, are fairly reactive and can readily decompose to form thiols, hydrogen sulfide, or other aliphatic sulfides. OS contents of mature oils tend to be dominated by less reactive thiophene and benzothiophene (aromatic S) moieties, but alkyl sulfides R−S−R′ and disulfides (including DEDS) are much more common in immature petroleum. [25, 26] These

disulfide and sulfide moieties have been implicated in enhancing the rate of petroleum formation from OS-rich kerogen under geologic conditions, [1, 21, 27] and are considered highly reactive constituents of immature OS-rich oils at reservoir conditions [25], which potentially reduces petroleum stability. [28]

Using probe molecules to illustrate this phenomenon, Lewan [21] reacted phenyldodecane (PDD) and diethyldisulfide (DEDS) liquid mixtures in stainless steel autoclaves and observed increased PDD decomposition with increasing DEDS concentrations at 350°C. DEDS contains both C−S and S−S moieties, both considered to be relatively reactive in petroleum and sedimentary organic matter in general at elevated temperatures. [21, 25] Lewan thus hypothesized that similar sulfide moieties in OS-rich kerogens (Types I-S and II-S) could preferentially generate higher initial concentrations of propagating radicals relative to OS-poor kerogen at a given thermal stress. This would provide a mechanistic explanation for observations of petroleum formation at lower than expected thermal stresses, rather than the relative weakness of C−S versus C−C bond strengths, in OS-rich kerogens. [1, 27, 29, 30]

Despite the importance of OS-rich kerogen in generating a large proportion of the world's petroleum resources, much of it OS-rich crude oil, [1] few experimental studies have reexamined Lewan's working hypothesis in further detail or tested the efficacy of OS-acceleration mechanisms under conditions approaching geologic petroleum formation and decomposition. Despite some pioneering works on the application of automated reaction networks to petroleum systems, [11, 28, 31, 32] there have also been no major efforts to investigate such accelerant effects from a theoretical perspective. We argue that the hypothesis warrants reexamination, given the concerns with the original experimental design. The stainless steel of the pressure vessel used, for example, is a known catalyst for decomposition reactions. [33, 34] Iron sulfide (FeS) evidently also formed on the reactor walls due to reaction of DEDS-derived H_2S with the iron, [21] which also has some potential to catalyze organic transformations. [35, 36]

New experiments conducted by Dr. Eoghan P. Reeves revisit Lewan's hypothesis under confined pressure (single phase) conditions more representative of geologic conditions, while minimizing the possible catalytic effects of steel reactors and unrealistic low pressures of Lewan's studies. [21]

5.1.2 Effects of water on the thermal stability of petroleum systems

While extensive evidence exists for the role of water, minerals (e.g. iron oxides and sulfides) and fluid redox in enhancing the transformation of petroleum to natural

gas, [12, 37–40] the experiments and model presented in this chapter do not include any water or mineral phases for simplicity. We wish to fully isolate the mechanistic effects of DEDS alone on the decomposition of PDD at moderate conversion levels as a function of temperature. We therefore do not attempt to simulate the complex natural inorganic-organic environment, where water and minerals are invariably present. Since RMG is not yet capable of integrating mineral phase reactivity, associated potential mineral catalysis would also complicate comparisons of model outputs with experimental results.

The effects of subcritical water on the PDD-DEDS system will be investigated in a subsequent study to address the implications of natural systems where water and organic matter coexist. We consider our results to be applicable, however, to environments where the activity of water is extremely low, such as within kerogen organic matrices and discrete oil phases.

5.2 Experimental and analytical methods

5.2.1 Model compounds

Several experimental factors make PDD a suitable model compounds for this study. The larger products of PDD cracking are more easily recoverable than that from shorter chain alkyl substituted aromatics. From a practical standpoint, PDD and the majority of its liquid and gaseous products (with the exception of semi-volatile benzene and C_5 to C_8 aliphatic hydrocarbons) are also readily recoverable and quantifiable by gas chromatography. PDD was reacted both alone and in the presence of DEDS in confined pyrolysis experiments.

The mole fractions of DEDS ($\chi_{DEDS} \approx .087$, in PDD) used this study closely match the most S-rich compositions of Lewan ($\chi_{DEDS} \approx .084$), allowing for direct comparison with Lewan's results. Experimentally, DEDS is easy to work with because its decomposition products (e.g. ethane, H_2S) are also distinguishable from the major products of PDD decomposition. [21]

5.2.2 Confined pyrolysis

A series of confined pyrolysis experiments were conducted at 252 - 353°C in sealed gold tubes with PDD both in the presence and absence of DEDS. The moderate organic sulfur contents chosen ($\sim$ 2.5 wt.% S of total hydrocarbon weight) mimic both natural S abundances in moderate to OS-rich oils, [26] and the PDD-DEDS mixture compositions reacted by Lewan. Although there is debate over the role of

pressure in organic matter maturation, petroleum generation and oil cracking, [41–44] a constant pressure (35 MPa) reasonably within the range for petroleum formation and destruction [45] was employed in order to focus on compositional effects alone.

Confined pyrolysis in gold capsules has been widely utilized in studies of organic matter maturation and probe compound reactivity in recent decades. [22, 46–49] The closed system technique allows for external control of pressure and therefore reactant-product phase relations, as well as preventing products (which may be reactive to starting materials) from exiting the reaction network. Gold is considered relatively impermeable to gases such as H_2, [50] and is the most inert material that can fulfill the requirements of the experimental technique. There is some evidence that suggests catalysis by gold capsule walls is possible, but this is suspected to be catalysis by residual iron contamination from tube fabrication and inadequate/inconsistent cleaning procedures prior to use, rather than the gold itself. [51]

All gold capsules in this study (99.99% purity, 5.0-9.0mm OD, 0.2mm wall thickness, $\sim$ 35mm length) were acid-cleaned in concentrated HCl (10 M, trace metal analytical grade) for exactly 6 h after arc welding one end closed, thoroughly rinsed in Milli-Q then oven-dried. Prior to loading reactants, capsules were heated in a propane flame both to anneal the gold and combust any residual organic material. Known masses of either PDD (97% purity, Sigma Aldrich) or a gravimetrically prepared mixture of the same PDD with added DEDS (99% purity, Sigma Aldrich) were then loaded into the capsules using glass-stainless steel syringes (previously rinsed in dichloromethane (DCM) and oven-dried). The headspace of all capsules was then flushed with Ar gas to remove air, crimped and cut to form an initial seal before final arc weld closure. Capsules loaded with PDD-DEDS were chilled in dry ice powder during Ar flushing to prevent evaporative losses of the more volatile DEDS. No mass losses were discernable as a result.

All capsules were placed into stainless steel cold seal autoclaves using H_2O as a pressurizing fluid as in the Tuttle experiments, [52] and all autoclaves were heated within tube furnaces at a constant pressure (35 $\pm$ 5 MPa). Temperatures were monitored via thermocouple temperature probes inserted into the autoclave ends ($\pm$ 2°C uncertainty). Experiments were conducted for 72 h at 350 - 353 °C, with autoclave heatup and cool down phases of an additional $\sim$ 2 h, and $<$ 0.3 h, respectively. To examine the effect of reduced temperature, a set of PDD and PDD-DEDS capsules was reacted at 253 $\pm$ 2°C (35 $\pm$ 5 MPa, Expt. 6) for 1001 h in a tube furnace.

Upon completion, capsules were cooled to room temperature under pressure to prevent rupture, removed, then rinsed in dichloromethane (DCM) to remove con-

taminants and immediately weighed to confirm no leakage occurred. Only those maintaining integrity within ± 1 mg difference were extracted for quantification of remnant PDD and liquid or gaseous products.

Capsules for liquid phase product analyses and quantification of PDD were extracted using similar procedures to YANG et al., 2012 and SHIPP et al., 2013. Briefly, capsules were first frozen in liquid N_2 to minimize losses of semi-volatile organic liquids, sliced open with a DCM-rinsed scalpel and quickly immersed in glass vials (with PTFE-silicone caps) containing pre-weighed chilled DCM (~ 22 g, CHROMOSOLV® HPLC grade, Sigma-Aldrich) and an added internal standard (naphthalene, analytical standard grade). Vials were shaken vigorously to promote thawing and mixing of the capsule contents into the DCM. In the cases of DEDS-containing experiments, activated Cu granules were added to aliquots of extracts to remove elemental sulfur (BLUMER, 1957) prior to mass spectrometry analyses. Arc welded capsules of unreacted initial PDD and PDD-DEDS mixtures were extracted to identify any contaminants either present in the organic liquids prior to reaction at conditions, or introduced as a result of loading and welding. Selected capsules at each experimental condition were weighed after solvent extraction for any solid phase carbonaceous products (e.g. char/pyrobitumen) that may have formed and adhered to the gold walls during reaction, yielding excess capsule wall mass.

5.2.3 Gas analysis

For the identification and quantification of gaseous products, the gold capsules were placed in glass serum vials (Bellco Glass Inc.) sealed with butyl rubber stoppers, which were then evacuated. Stoppers were previous boiled in NaOH solution, after Reeves [53] and Oremland [54], to reduce hydrocarbon contamination. Capsules were opened by piercing with a steel needle syringe containing N_2 gas to flush and homogenize gaseous products into the serum vial headspace. H_2 and CO were then determined as soon as possible after piercing (within ~ 15 min) to prevent substantial loss of H2 via diffusion through the stoppers. Loop injections of headspace gas were made into a gas chromatograph (GC) equipped with a molecular sieve column and thermal conductivity detector. Hydrocarbons (C_1 - C_6 n-alkanes and n-alkenes, methylpropane (iso-butane)) and CO_2 were subsequently determined for each of these capsules by loop injections of headspace gas onto a gas chromatograph (GC) equipped with HayeSep Q (divinylbenzene) packed column and serially connected thermal conductivity and flame ionization detectors. Commercial H2, CO, C_1 - C_6 hydrocarbon and CO_2 gas standards were similarly injected via loops to each GC column for

calibration. Absolute moles of gases evolved from each capsule were calculated by correcting for sequential losses by injection, and normalized to the average volume of the serum vials. The volume of the capsule and liquid contents once vented was minor (<2 % of the total serum vial volume) and therefore not explicitly corrected for. For semi-volatiles, there is likely some loss to the small organic liquid phase that is not accounted for, and the values presented should be taken as minimums.

5.3 Chemical kinetic modeling

The construction of detailed kinetic mechanisms for free radical networks by hand requires not only a pre-existing knowledge of the relevant chemistry and rate kinetics, but a systematic accounting of all the reactions that can take place, which can be tedious and error-prone. Although such mechanisms provide new insights on petroleum stability, [22, 55] they are difficult to extend to systems involving new chemical species. Recent computational advances have made it possible to better predict in detail the chemical kinetics for pyrolysis of complex organic molecules at high temperatures (e.g. during combustion). Such predictions are based on *ab initio* calculations or published data for key reaction steps and reactive intermediates, and use automated mechanism construction software to build and handle the complex networks of organic free radical reactions involved. Detailed models can now be constructed automatically by the RMG, which relies on an expansive and consistently updated database of thermodynamic and kinetic values sourced from either published literature or theory. RMG can generate kinetic predictions consistent with experimental pyrolysis of hydrocarbons, oxygenates and OS compounds at high temperatures. [56–58]

It is also capable of considering effects of inorganic components such as H_2S and H_2O, yielding kinetics of OS and alkylbenzene reactivity at >400°C in the presence of supercritical water that are consistent with experimental observations. [59, 60] Although widely applied in combustion and pyrolysis research, RMG has yet to be explored as a predictive tool for understanding lower temperature organic geochemical phenomena associated with petroleum formation and destruction. It has, however, predictive capability over a large range of temperature and pressure conditions, with inherently more realistic representations of organic reaction networks than traditional simulations currently in use in the petroleum industry. The approach differs strongly from traditional models for petroleum formation and decomposition, which tend to rely heavily on simplified Arrhenius parameters derived empirically at very high temperatures (up to 500°C), often lumping families of compounds together and assuming

simple first order kinetics. The veracity of this approach is regarded by many as highly questionable, [11, 28, 61] as it fails to explain the apparent stability of hydrocarbons in many high-temperature high-pressure reservoirs.

Several challenges exist in modeling the PDD-DEDS low temperature system. First, kinetics derived from *ab initio* calculations have uncertainties about their nominal values, which become larger at low temperatures. A modified Arrhenius rate coefficient $k(T)$ is used for reaction kinetics:

$$k(T) = A\left(\frac{T}{1\text{ K}}\right)^n \exp\left(-\frac{E_a}{RT}\right) \tag{5.1}$$

where A is the pre-exponential factor, n is the temperature exponential factor, E_a is the activation energy, R is the gas constant, and T is the temperature. At low temperatures, an uncertainty in the activation energy $E_a = E_{a,true} \pm \delta E$ can lead to increasing error in the overall $k(T)$:

$$\exp\left(-\frac{E_a}{RT}\right) = \exp\left(-\frac{E_{a,true} \pm \delta E}{RT}\right) = \exp\left(-\frac{E_{a,true}}{RT}\right)\exp\left(\frac{\pm\delta E}{RT}\right) \tag{5.2}$$

For instance, a typical rate coefficient with an energy uncertain by a factor of 2 at 1000 K will have double that error at 500 K. Thus, predictions by kinetic models extrapolated to geologic conditions must be validated extensively by experimental data

5.3.1 Mechanism generation

RMG was used to generate a comprehensive model for PDD and DEDS decomposition. This was done by first generating a pyrolysis mechanism for each component individually.

The PDD model was generated in RMG-Py at 35 MPa and 350°C, with a tolerance of $\epsilon = 0.04$ and reaction termination time of 72 hours. The number of radical electrons was restricted to 1 to aid model convergence and reduce the number of edge species being generated. Reaction filtering was turned on in order to generate the model in a timely manner. A skeletal PDD was produced containing 81 reactions and 1427 reactions.

The DEDS model was previously generated in RMG-Java by Caleb Class at 24 MPa and temperatures of 250, 300, and 350°C, with an initial DEDS mole fraction of 1.0, a tolerance of $\epsilon = 0.5$, and termination DEDS conversion of 0.80. The number of sulfur atoms per species was restricted to 2 and the number of carbon atoms was restricted to 4 in the mechanism generation. The Leeds University sulfur mechanism

extension to the methane oxidation mechanism was used as a reaction library. [62] The final DEDS pyrolysis model contains 105 species (including inerts) and 3260 reactions. A DEDS kinetic library was constructed by Caleb based on this model.

Generation of a comprehenesive PDD-DEDS model is not trivial because PDD is a C_{18} compound. With so many reaction sites, there are numerous memory and convergence problems unless the error tolerance ϵ is set loosely. However, doing so often leads to incomplete reaction networks. The comprehensive PDD-DEDS model we constructed merges the PDD model and the DEDS kinetic library created by Caleb. Cross reactions between the models were then added and merged in. This resulted in a final model containing 167 species and 3793 reactions. Currently, the model lacks a number of recombination products and is missing pathways to several observed experimental species. Improvements to RMG's capability to model large compounds and estimate parameters for akylaromatic reactions are necessary to fully capture the details within the PDD reaction network.

Reaction simulations used to create the plots found in this chapter were carried out in CHEMKIN-PRO [63].

5.4 Results

PDD was pyrolyzed alone and in the presence of DEDS ($\sim$ 8.5 mol.%) at 250 and 350°C and 35 MPa under confined, single phase conditions the absence of added minerals or water for 1001 and 72 hours, respectively. The PDD conversion for the experiments are presented in Table 5.1. The detailed speciation results from the Reeves experiments at 350 °C are presented in Table 5.2. The speciation data for the 250 °C experiments are omitted due to the low conversion lending poorly to high confidence quantitative measurements.

Therefore, we concern our story primarily with the overall conversion of PDD at the different temperatures, and the speciation differences between the neat PDD and PDD-DEDS case at 350 °C.

A number of PDD isomers are found among the experimental products. These isomers are depicted in Figure 5.3.

5.4.1 Neat PDD pyrolysis

The RMG model for PDD pyrolysis is simulated at various temperatures and 35 MPa and compared against existing literature data in Figure 5.4, with a shaded error bar of 20%. The model underpredicts the conversion of PDD at shorter timescales and lower tmperatures. The experiments conducted by Behar [22], Savage [19], and

Table 5.1: PDD conversion in Lewan and Reeves experiments. Experiments with DEDS contain an initial mole fraction of DEDS$\approx$ 0.085.

	Lewan	Reeves	
Temperature	350	350	250
Reaction time (hr)	72	72	1001
PDD conversion (neat) (%)	30	41.6 $\pm$ 1.9	1.8 $\pm$ 0.7
PDD conversion (w/ DEDS) (%)	54	43 $\pm$ 2.7	9.8 $\pm$ 0.6

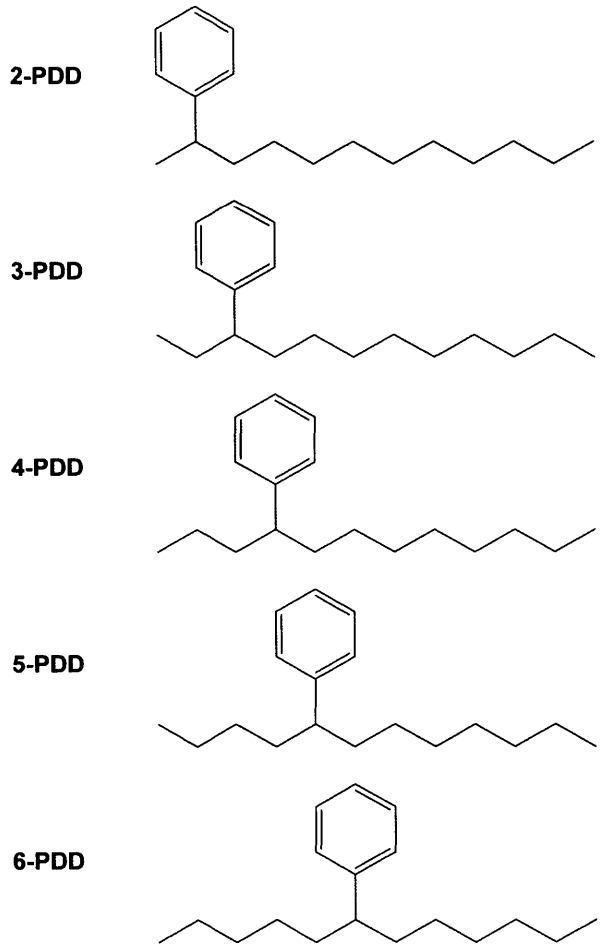


Figure 5.3: PDD isomers identified in experiment.

Lewan [21] are carried out at lower pressures than the Reeves experiments. However, there is general agreement between these datasets except for the Lewan experiments, which were conducted at atmospheric pressure, in contrast to the Behar and Reeves experiments conducted at 14 MPa and 35 MPa, respectively. Further experimental

Table 5.2: Experimental molar yield of product species after PDD neat pyrolysis and decomposition in the presence of DEDS at 350 °C after 72 hours of reaction time.

	PDD		PDD with DEDS	
$\chi_{PDD,initial}$	1.00		0.913	
$\chi_{DEDS,initial}$	0.00		.087	
Species	Molar Yield (%)	$\pm 2s$	Molar Yield (%)	$\pm 2s$
PDD	58.37	1.94	56.631	2.67
Toluene	15.625	0.41	16.835	1.553
n-Undecane	0.44	0.07	1.636	0.002
$\sum$ C_{11}-ene	0.89	0.05	2.42	0.271
Ethylbenzene	1.17	0.23	3.346	0.365
n-Decane	3.419	0.2	1.93	0.160
MeSH	0.00	0.00	0.0031	0.0011
H_2	0.42	0.09	0.126	0.005
CO_2	0.04	0.01	0.186	0.094
CH_4	0.28	0.11	1.33	0.23
Ethane	0.84	0.27	10.5	0.9
Ethene	0.005	0.002	0.0048	0.0001
Propane	0.77	0.24	1.202	0.14
Propene	0.08	0.02	0.034	0.004
n-Butane	0.65	0.16	1.025	0.11
i-Butane	0.0027	0.0016	0.021	0.003
$\sum$Butenes	0.06	0.01	0.0168	0.0019
n-Pentane	0.49	0.12	0.591	0.052
$\sum$Pentenes	0.02	0.01	0.0631	0.0061
n-Hexane	0.36	0.08	0.387	0.015
$\sum$Hexenes	0.003	0.002	0.08	0.003
n-Nonane	0.58	0.01	0.651	0.091
Propyl Benzene	0.25	0.01	0.705	0.069
Butyl Benzene	0.71	0.06	0.928	0.064
Pentyl Benzene	0.55	0.03	1.042	0.126
Hexyl Benzene	0.22	0.03	0.333	0.036
Heptyl Benzene	0.18	0.04	0.189	0.011
Octyl Benzene	0.39	0.01	0.243	0.023
Nonyl Benzene	0.47	0.02	0.398	0.0298
Decyl Benzene	0.24	0.06	0.124	0.032
Undecyl Benzene	0.26	0.03	0.244	0.048
2-PDD	1.914	0.03	2.554	0.056
3-PDD	0.14	0.01	0.659	0.083
4-PDD	0.81	0.02	0.69	0.037
5-PDD	0.598	0.01	0.407	0.027
6-PDD	0.13	0.01	0.208	0.0229

investigation may be necessary to determine the full extent of the effects of low pressure on PDD decomposition. The discrepancy between the model and the Reeves 250 °C data may be explained by a greater experimental error than estimated in Table 5.2. The measured conversion was very low (1.9%), and the initial and final measurements in the 250 °C experiment were taken 1001 hours apart, which could lead to greater systematic error. At the temperature and pressure conditions of the Reeves experiment, confined pyrolysis occurs in a single, liquid phase. We currently use RMG's gas phase kinetics to model this inside an ideal gas homogenous batch reactor. An investigation into the error caused by these approximations is necessary to distinguish the chemistry error from the physics error. In the future, using an equation of state to constrain the volume profile in CHEMKIN may be a partial solution to approximating the physics of the system.

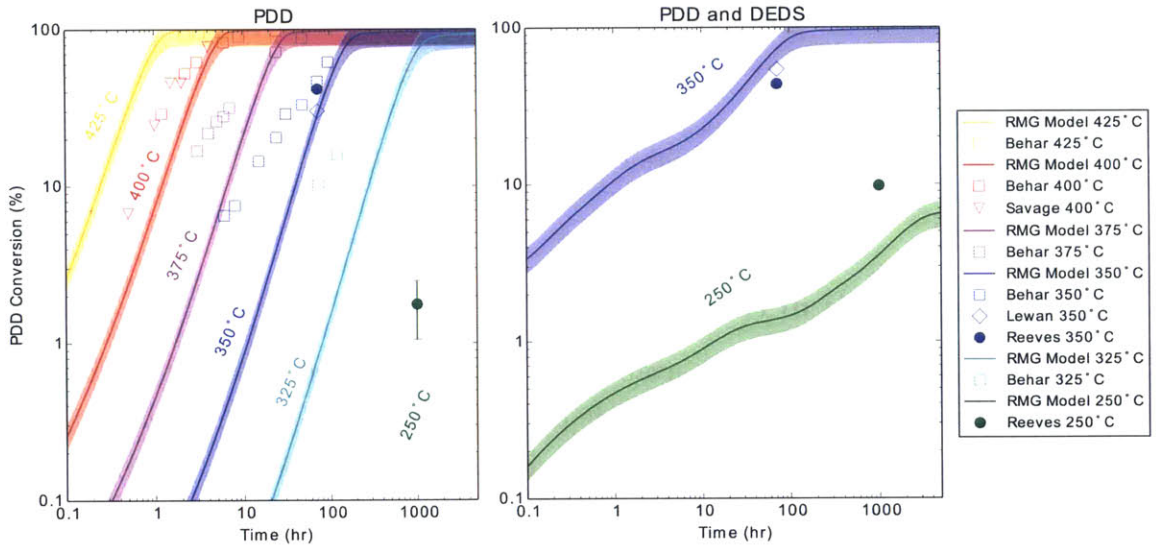


Figure 5.4: PDD conversion in neat pyrolysis and in the presence of DEDS with respect to temperature. The RMG model is simulated at P = 35 MPa and at different temperatures with an shaded error bar of ±20%. The model is plotted against against various experimental data by Lewan [21], Behar [22], Savage [19], and the new data by Reeves presented in this chapter.

Additional validation of the model was performed through comparisons with species profiles from the 400°C experiments of Savage and Klein [19] in Figures 5.5 to 5.7. The model slightly underpredicts the PDD conversion and the decane and ethylbenzene yields. Styrene is also overpredicted, while undecane is underpredicted.

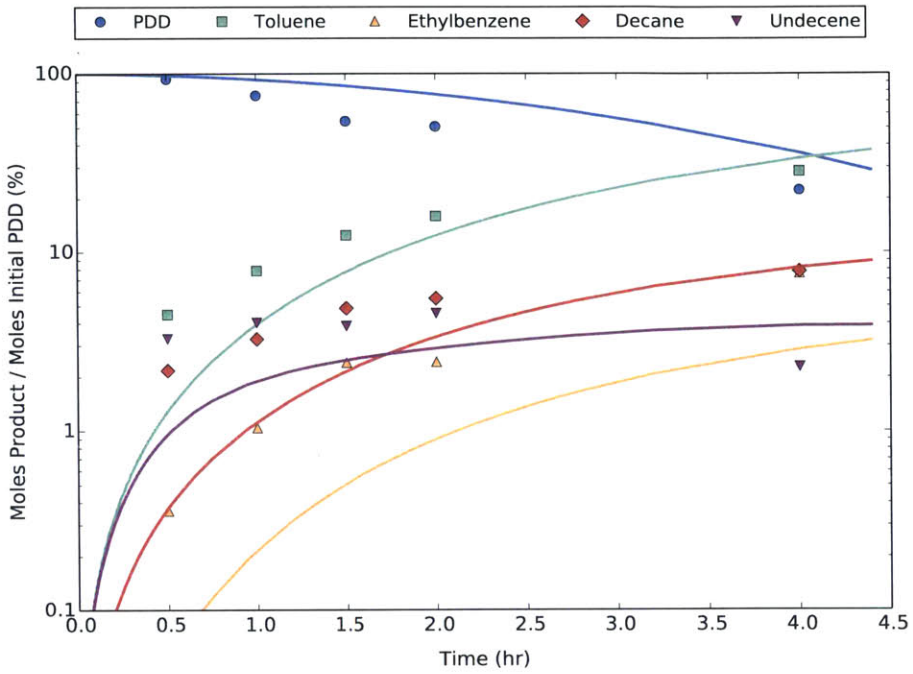


Figure 5.5: Simulated PDD neat pyrolysis molar yields of major species compared against Savage and Klein experiments conducted at 400°C.

5.4.2 PDD pyrolysis in the presence of DEDS

The amount of experimental data available for PDD decomposition in the presence of DEDS is scarce and consists of a single datapoint at 350 °C from Lewan's initial experiment carried out at atmospheric pressure, and our new experimental data at 350°C and 250°C. These data points are presented in Figure 5.4 and detailed in Table 5.1. At 72 hours and 350 °C, Lewan observed 54% conversion at the same initial mole fraction of DEDS as in our experiments, which showed only 43% conversion. Our results show that DEDS does not significantly accelerate anhydrous PDD decomposition at 350°C, in contrast to the Lewan [21] data. Though preliminary, our observations are consistent with other studies [64] showing a lack of enhancement of hydrocarbon breakdown with similar organic sulfides. However, there is more than a 5-fold increase in PDD conversion at 250°C when DEDS is present, yielding detectable quantities of decomposition products.

There are several major differences between products detected in neat PDD pyrolysis versus in the presence of DEDS. These differences are best highlighted using the FID response peaks for the two experiments at 350 °C, shown in Figure 5.8. This includes a change in the major alkane product distribution between C_{11} and C_{10}

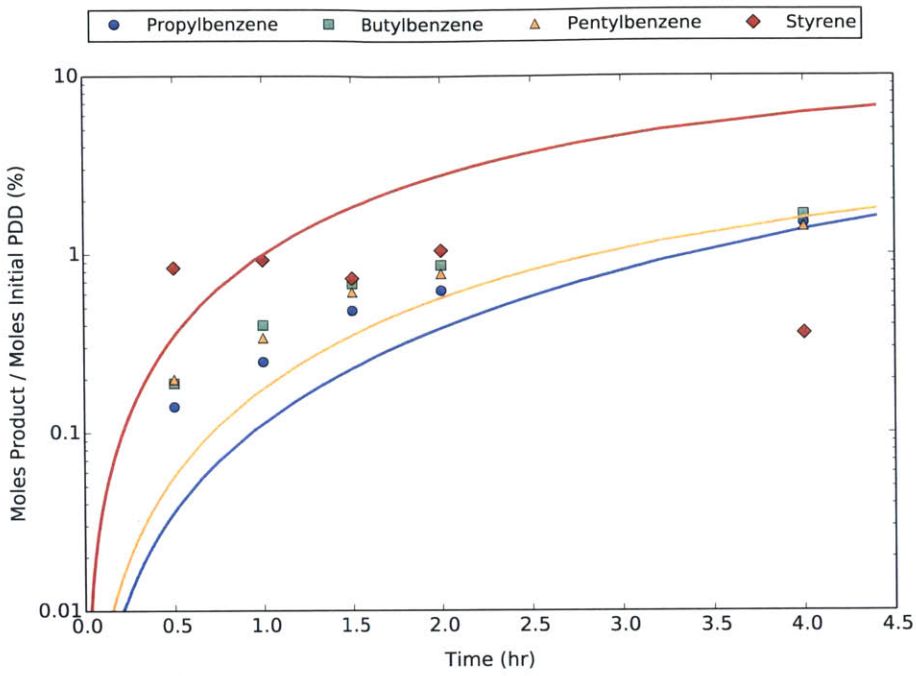


Figure 5.6: Simulated PDD neat pyrolysis molar yields of aromatic species compared against Savage and Klein experiments conducted at 400°C.

species, a difference in the PDD isomer selectivities, and an absence of high molecular weight recombination products when DEDS is present.

Given the lack of sound quantitative speciation data at 250 °C, we instead compare the experimental product selectivities at 350 °C against the model simulations. The major product selectivities in the Reeves experiments are compared against model predictions in Figure 5.9. The data was plotted against selectivity (moles product / moles PDD reacted) rather than product yield in order to separate the error in overall PDD conversion from the product selectivities. The major difference observed is the increased selectivity for the dominant alkane formed. When DEDS is present, higher concentrations of C_{11} is formed over C_{10}, whereas the reverse is true during PDD neat pyrolysis. This change is not captured well by the model: there is more decane formed in both cases, and a larger C_{10}/C_{11} ratio exists when DEDS is present. This could indicate a problem in the branching fraction kinetics of the PDD radical precursors to the C_{11} and C_{10} products.

The distribution of lighter alkanes in the PDD and PDD-DEDS experiments are compared against simulation results in Figure 5.10. There is a significant increase in ethane and methane generation when DEDS is present, which is reflected by the

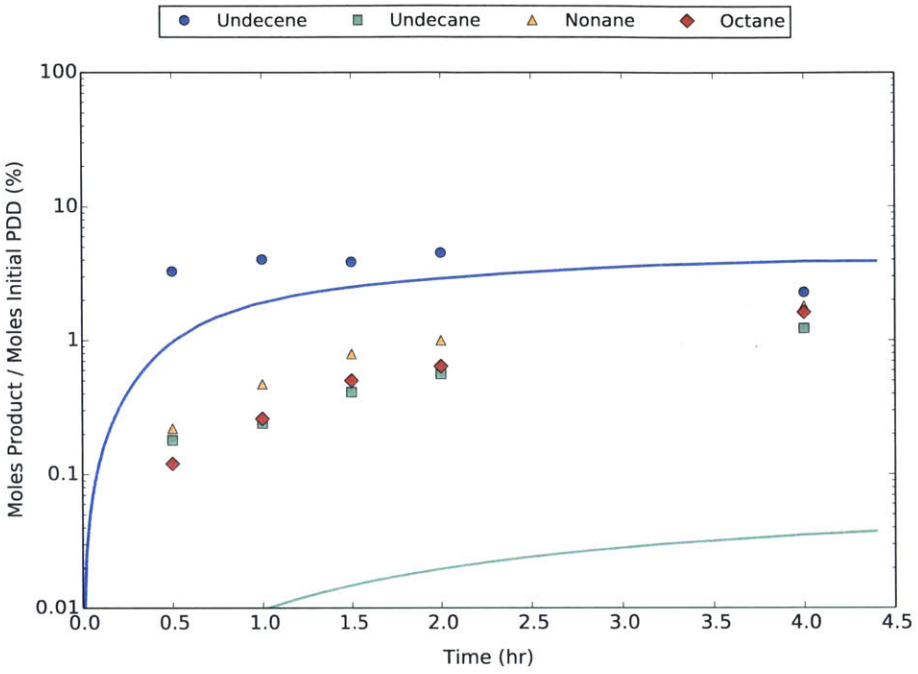


Figure 5.7: Simulated PDD neat pyrolysis molar yields of alkane species compared against Savage and Klein experiments conducted at 400°C.

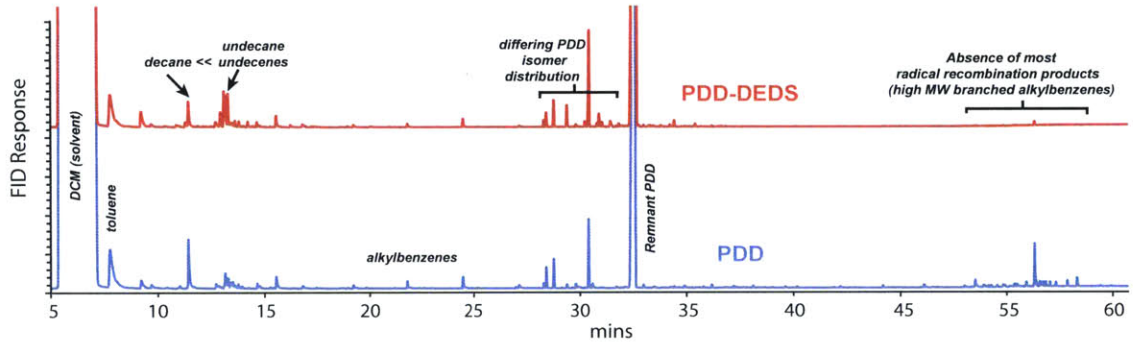


Figure 5.8: GC-FID traces of PDD-DEDS (top), and PDD (bottom) pyrolysis at 350 °C after 72 hours. Key differences between the product distributions are highlighted.

model, though it has an overprediction of the methane selectivity.

Another major difference is highlighted in Figure 5.11, where the selectivity of PDD isomers differs when DEDS is present. There is a larger selectivity towards 3-PDD as compared to other PDD isomers when DEDS is present. Unfortunately,

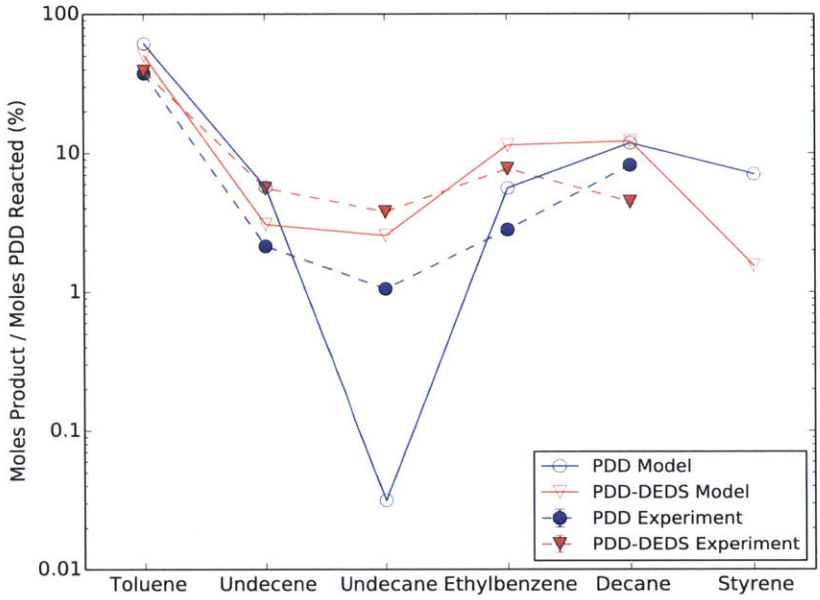


Figure 5.9: Simulated and experimental Reeves data of major species molar selectivities (moles produced / moles PDD reacted) at 350°C after 72 hours of reaction time for PDD neat pyrolysis (blue) and decomposition in the presence of DEDS (red).

these PDD isomers do not appear in the model. Efforts to include these paths are discussed in the next section.

5.5 Discussion

Flux analysis of the reaction network shows that shows that the major products predicted in PDD neat pyrolysis are toluene, undecene, ethylbenzene, and decane. Toluene and undecene arise from PDDrad3, the radical formed by hydrogen abstraction from the 3rd carbon to the phenyl group in PDD. Styrene, ethylbenzene, and decane form from PDDrad1, the radical formed by hydrogen abstraction from the 1st carbon to the phenyl group in PDD. This radical is resonantly stabilized by the presence of the aromatic ring, and its enthalpy is 12.55 kcal/mol more stable than all other PDD radicals formed from hydrogen abstraction on the aliphatic chain.

The Reeves experimental data shows that there was no styrene formed, and the Savage [19] experiments also show little styrene formation at 400 °C. This suggests that styrene formation from PDDrad1 should convert almost entirely to ethylbenzene. This pathway is found by RMG: styrene can participate in a reverse disproportion-

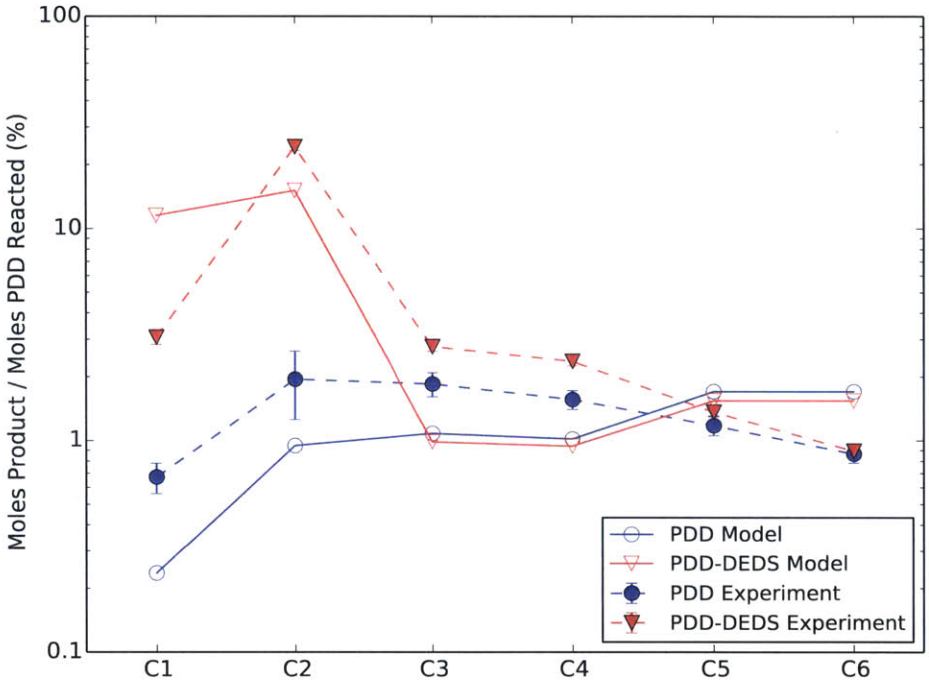


Figure 5.10: Simulated and experimental Reeves data of hydrogen and light alkane molar selectivities (moles produced / moles PDD reacted) at 350°C after 72 hours of reaction time for PDD neat pyrolysis (blue) and decomposition in the presence of DEDS (red).

ation reaction with PDD to form two resonantly stabilized radical species, leading to the formation of ethylbenzene; however, we still predict a substantial amount of styrene in the model.

Although the RMG model predicts the toluene selectivity well, it overpredicts the amount of undecene and underpredicts the undecane selectivity compared to both the Reeves experiments. In the RMG model, there is little further decomposition of undecene. Further reaction network analysis shows that undecene isomers are continually regenerated through hydrogen abstraction reactions with PDD radicals. However, radical alkyl chains should typically undergo *beta*-scissions, forming lighter species and contributing to the yields of small alkanes and alkenes. This problem may be due to lack of model convergence, since the overall concentrations of these lighter compounds detected in experiment are quite small, the model must be generated with a sufficiently tight error tolerance in order to capture these product pathways. The

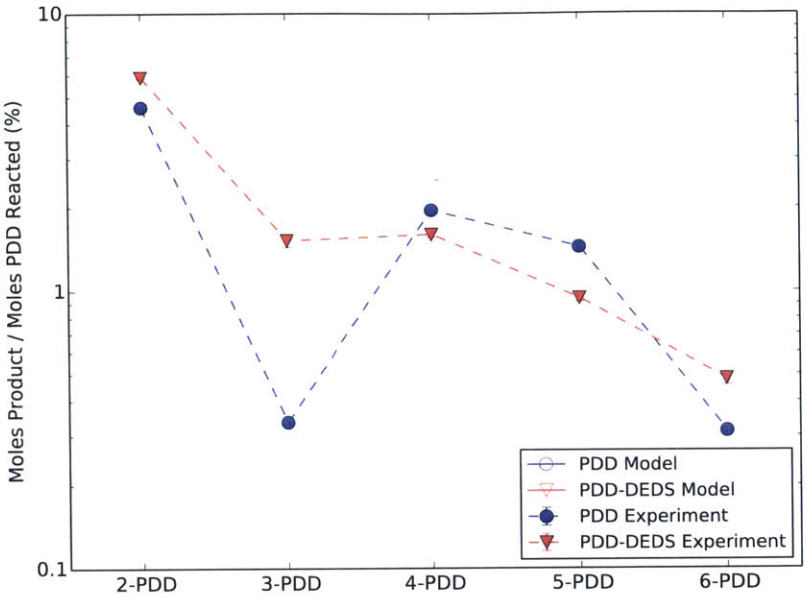


Figure 5.11: Simulated and experimental Reeves data of PDD isomer molar selectivities (moles produced / moles PDD reacted) at 350°C after 72 hours of reaction time for PDD neat pyrolysis (blue) and decomposition in the presence of DEDS (red).

inclusion of these reactions may also increase the overall convversion of PDD in the model. Some lighter hydrocarbons are produced in the model in low quantities from *beta*-scission reactions of the initial PDD radicals. We also expect some undecene to be consumed by radical addition reactions, forming larger hydrocarbon species; however, few of these appear in the model.

The undecane formation in the RMG model is primarily through the initiation reaction where PDD undergoes a C−C bond scission to form benzyl and undecyl radicals. The undecyl radical then abstracts a hydrogen from another species to form undecane. This initiation reaction is very slow. Once a radical pool is formed, abstraction reactions on the PDD molecule are much faster. Therefore the model should include another source of undecane production. One such reaction could be a reverse disproportionation reaction involving undecene to form the undecyl radical.

We compare our model to a similar mechanistic study of phenyldecane by Burklé-Vitzthum *et al.* [55] These authors propose a significant decomposition channel at low temperature through the molecular elimination "retro-ene" reaction which reacts phenyldecane directly to toluene and an alkene via a 4-centered transition state (see

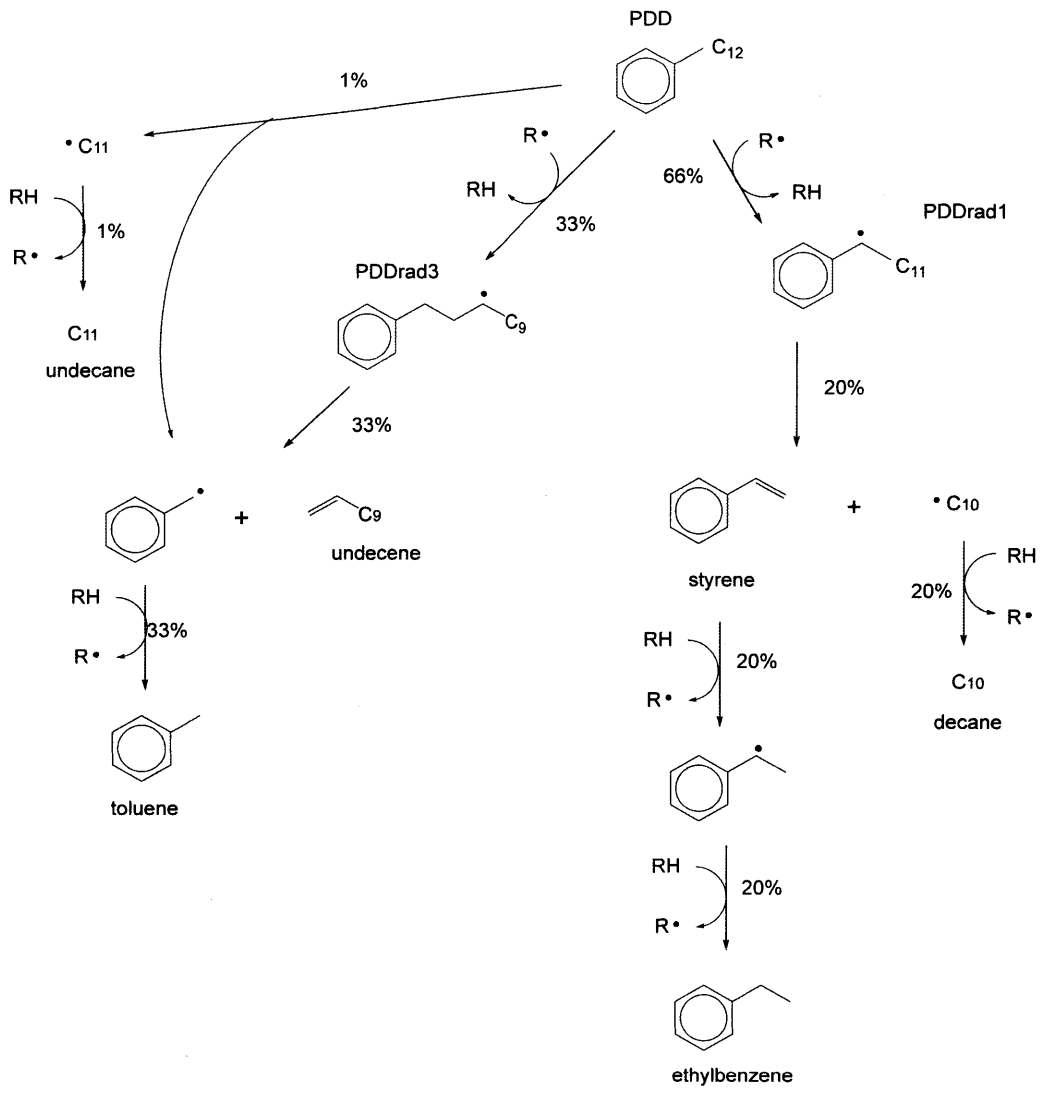


Figure 5.12: Fluxes of PDD decomposition at 350°C and 10 hours.

top reaction in Figure 5.13. The activation energy for this reaction was fitted to experiment and stated to be 54.4 kcal/mol. However, further investigation through CBS-QB3 calculations on propylbenzene by Class [65] gave an activation energy of over 100 kcal/mol with more than ten orders of magnitude decrease in the rate evaluated at 250 °C. A subsequent calculation by Liu [66] of a similar molecular elimination reaction via a lower energy 6-membered transition state gave a 72.1 kcal/mol activation energy which also results in a significantly slower direct elimination channel to toluene and alkene from the parent alkylbenzene. Thus, we use the 6-membered transitions state value rather than the value from the Burklé-Vitzthum mechanism.

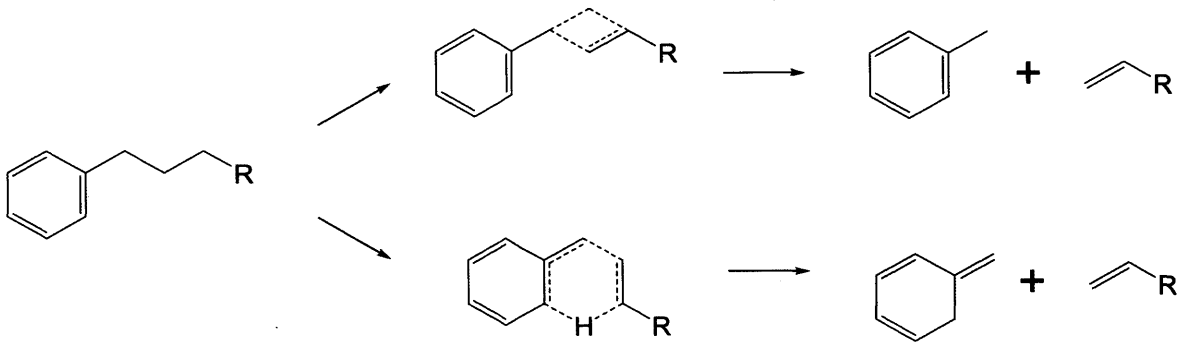


Figure 5.13: The Retro-ene molecular elimination reaction from alkylaromatic to toluene and alkene via a 4-membered transition state (top) and a 6-membered transition state(bottom). The 6-membered transition state is computed to be more favorable.

As shown in Figure 5.14, *ipso*-isomerizations of the PDD radicals generate the various carbon-shifted isomers depicted earlier in Figure 5.3. These reactions are analogous to the reactions previously published in the phenyldecane Burklé-Vitzthum mechanism. [55] RMG was able to find these reaction pathways; however, the fused cyclic radical intermediates have poorly estimated thermochemistry, often with up to 50 kcal/mol error. Therefore, the addition of these reactions will require the addition of the proper polycyclic ring corrections as well as review of the `Intra_R_Add_Exocyclic` and `Intra_R_Add_Endocyclic` family kinetics to ensure that these isomerizations are correctly estimated in the mechanism. This work is still in progress. The expected major product pathways resulting from β-scission and *ipso*-ipsomerization reactions of the initial PDD radicals is presented in Table 5.5.

Higher molecular weight products were detected in the Reeves experiments although their chemical structures could not be identified. We expect most of these products to form from the recombination of resonantly stabilized radicals (Figure 5.15), although only a few of these appear in the model due to model incompleteness. Further experimental probing using analytical chemistry techniques that can identify the structures of these recombination products will be useful.

A general product trend in neat pyrolysis is that unsaturated hydrocarbons are selectively formed over their saturated counterparts, i.e. there are higher concentrations of butane vs. butene, hexane vs. hexene, octane vs octene, etc. This result is corroborated in the Savage [19] data as well. One hypothesis is that at longer timescales, unsaturated hydrocarbons begin to cyclize via diels-alder type reactions,

Precursor	Reaction Type	Products
PDDrad1	β-scission	styrene + decane
PDDrad2	β-scission	propenylbenzene + nonane
	ipso-isomerization	2-PDD
PDDrad3	β-scission	toluene + undecene
		butenylbenzene + octane
	ipso-isomerization	3-PDD
PDDrad4	β-scission	ethylbenzene + decene
		pentenylbenzene + heptane
	ipso-isomerization	4-PDD
PDDrad5	β-scission	propylbenzene + nonene
		hexenylbenzene + hexane
	ipso-isomerization	5-PDD
PDDrad6	β-scission	butylbenzene + octene
		heptenylbenzene + pentane
	ipso-isomerization	6-PDD
PDDrad7	β-scission	pentylbenzene + heptene
		octenylbenzene + butane
	ipso-isomerization	5-PDD
PDDrad8	β-scission	hexylbenzene + hexene
		nonenylbenzene + propane
PDDrad9	β-scission	heptylbenzene + pentene
		decenylbenzene + ethane
PDDrad10	β-scission	octylbenzene + butene
		undecenylbenzene + methane
PDDrad11	β-scission	nonylbenzene + propene
PDDrad12	β-scission	decylbenzene + ethene

Table 5.3: The β-scission and *ipso*-isomerization reactions of the PDD initial radicals leading to products detected in experiment. PDDrad1 has a radical closest to the phenyl group, while PDDrad12 has a radical furthest away from the phenyl group.

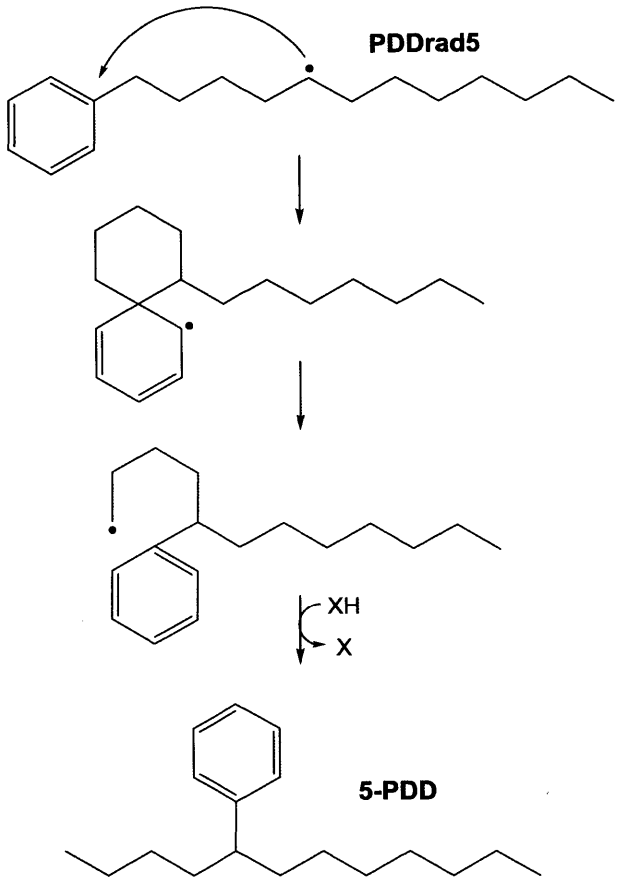


Figure 5.14: *Ipso*-isomerization reaction of PDD radical leading to the formation of a PDD isomer.

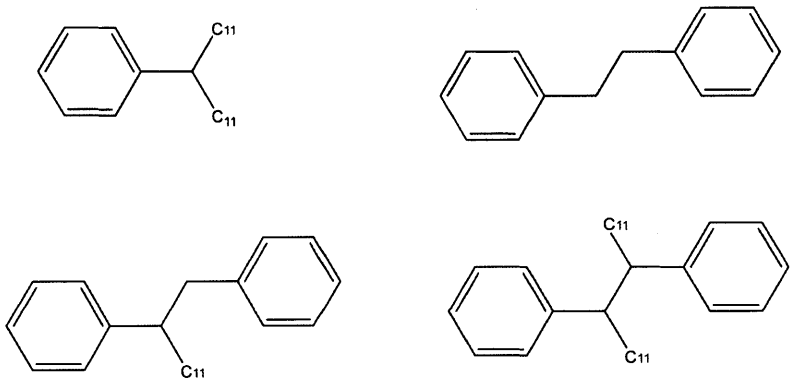


Figure 5.15: Recombination products in neat PDD pyrolysis.

forming polycyclic and polyaromatic ring structures leading to coke. However, RMG is currently not able to estimate these reactions due to our lack of elementary reaction

families for polyaromatic hydrocarbon (PAH) formation.

The model has some deficiencies in modeling neat PDD pyrolysis, namely poor prediction of the *ipso*-isomerization reactions and of the saturated vs. unsaturated alkyl chain compounds. However, it is capable of predicting most other major product selectivities, with the exception of undecane and styrene.

Next, we turn to the mechanistic investigation of the influence of DEDS on the decomposition of PDD. The Reeves experimental results at 350 °C puts the original Lewan effect under scrutiny; the conversion acceleration observed in stainless steel reactors may be either related to either low pressures or catalytic wall reactions. It is known that H_2S, a known product of DEDS decomposition, reacts with stainless steel to form pyrrhotite film on the reactor walls. [67]

The reaction network for PDD simulated with DEDS shows that DEDS decomposes completely to H_2S and ethane within an hour at 350°C prior to significant PDD conversion. DEDS-derived sulfanyl ($HS\cdot$) and ethyl radicals do, however, exert some effects on dominant hydrocarbon production pathways through competition with other radicals, which supports Lewan's hypothesis that sulfur participates in radical decomposition pathways. However, as DEDS decomposes, the $HS\cdot$ and other radicals recombine recombine with each other or other PDD-derived radicals, forming stable products. After some time, much of the sulfur is present as H_2S, which is relatively stable and is in equilibrium with the reaction system.

At 250°C, DEDS conversion is slower, and shown to participate in initial abstraction reactions on the parent PDD molecules. This is consistent with remnant DEDS only being detected experimentally in the products of the 250°C experiment. The timescale of available free radicals may therefore be important in controlling the pyrolysis of alkylaromatic compounds; the greater longevity of the sulfur-based free radicals at 250°C accelerates PDD decomposition.

A more in depth analysis of DEDS reacting alone in the presence of argon, shown in Figure 5.16, demonstrates that there is greater DEDS longevity with decreasing temperature. At 350°C, DEDS is consumed almost entirely after 10 hours, whereas less than 80% is converted at 250°C after 1000 hours.

Figure 5.17 shows the major reaction fluxes for DEDS at 350 °C after 60 seconds of reacting alone. Ethane and H_2S are the major products anhydrous pyrolysis, and agree with those observed experimentally when PDD is reacted in the presence of DEDS.

Due to the fast timescale of DEDS decomposition at 350 °C, the acceleration effect is not apparent at 72 hours reaction time. However, the model suggests that under

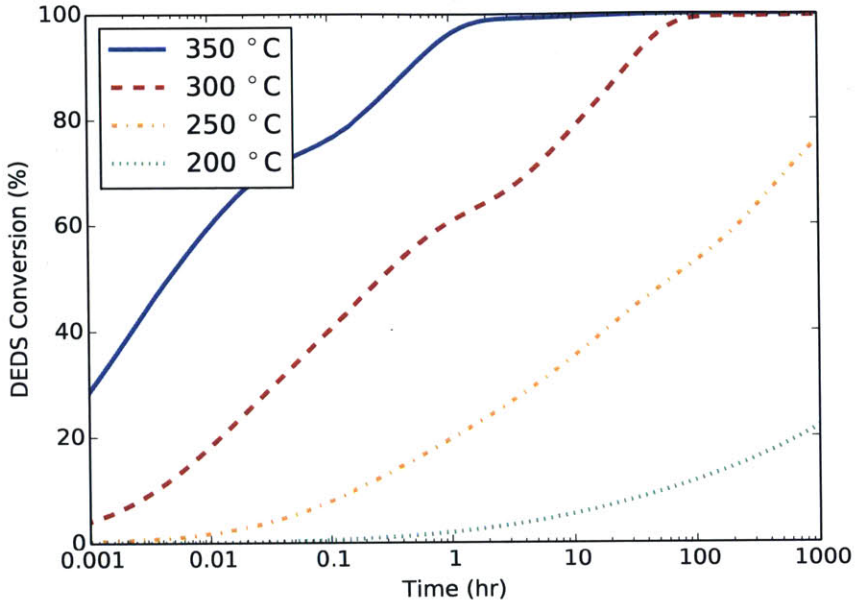


Figure 5.16: RMG model simulation of the decomposition of neat DEDS ($\chi = 0.0875$) in Argon as a function of temperature.

shorter reaction times, the sulfur radicals may have some effect on the decomposition. It would be useful to perform a series of short < 5 hour experiments at 350 °C and gather time-dependent data to further support our hypothesis that sulfur radical timescales matter.

The product selectivity differences noted in the C_{11}/C_{10} and PDD isomer distributions should be further investigated as it could be related to the reactivity of the DEDS-derived sulfanyl ($HS\cdot$) radical. In contrast, the predominant radical in PDD neat pyrolysis is benzyl. This difference is not yet predicted by the model; however, a larger branching fraction for PDDrad3/PDDrad1 when $HS\cdot$ radical attacks PDD could lead to a higher C_{11}/C_{10} ratio when DEDS is present as compared to PDD alone, since undecene results from the PDDrad3 *beta*-scission reaction, while decane forms from PDDrad1 (see Figure 5.12). Furthermore, this leads to a higher flux for *ipso*-isomerization reactions involving PDDrad3, analogous to the reaction depicted in Figure 5.14, resulting in the formation of 3-PDD product (also observed as a difference in Figure 5.11).

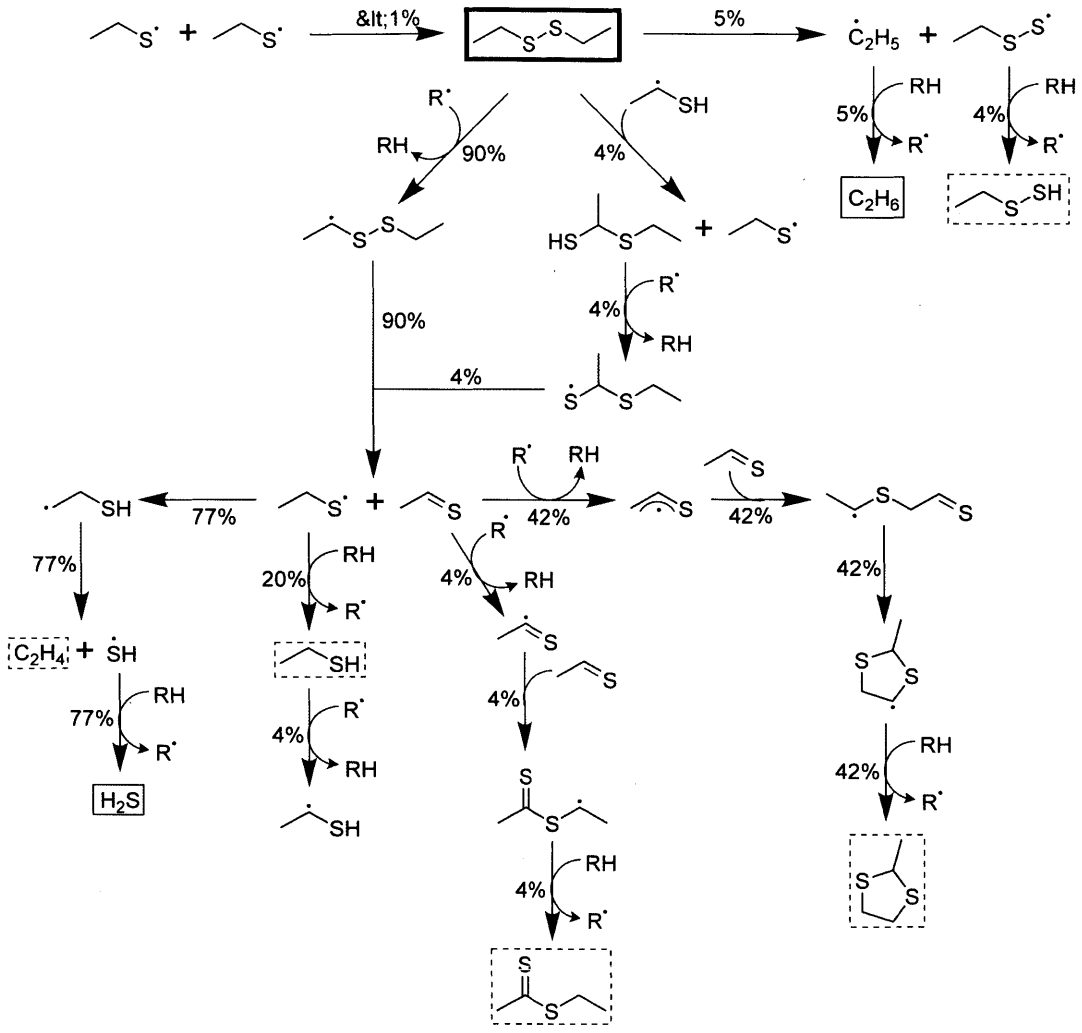


Figure 5.17: DEDS decomposition pathway at a temperature of 350 °C and at a reaction time of 60 seconds. Percentages are presented relative to total decomposition rate of DEDS at this time. Products in dashed boxes are predicted to be consumed later in the reaction timescale, while ethane and hydrogen sulfide are predicted to be stable products. Reproduced from Caleb Class's thesis [2].

5.6 Conclusion

Our combined experimental observations and reaction network simulations support the hypothesis that, under anhydrous conditions, organosulfur moieties could accelerate hydrocarbon free-radical reactions at the geologic temperatures associated with kerogen and oil maturation. Experimentally, we show that DEDS does not substantially accelerate PDD decomposition at 350°C, in contrast to previous investigations in fixed volume reactors, but does exhibit statistically significant acceleration effects

on decomposition at 250°C. The timescale of available free radicals are implicited in the acceleration of PDD decomposition at lower temperatures due to their longevity. These timescales may differ depending on the nature of the S moieties in question. Less reactive organic sulfur compounds such as thiols or thiophenes will likely have different effects on oil-to-gas composition. These effects will require further mechanistic investigation to model kerogen with diverse sulfur content in petroleum systems.

The model generated by RMG predicts the major products of PDD neat pyrolysis experiments found in literature and the new experiments performed by Reeves, but overpredicts the conversion of PDD, likely due to model convergence issues in RMG. There is also an overprediction of undecene and styrene and underprediction of undecane. These differences should be investigated by looking into the kinetics of the *beta*-scission and reverse disproportionation reactions involving the initial PDD radicals to verify that the branching ratios are correct. Polycyclic thermochemistry and *ipso*-isomerization reactions should also be included to account for the PDD isomers found in experiment.

Experimentally, we show that DEDS does not substantially accelerate PDD decomposition at 350°C (after 72h), in contrast to previous investigations in fixed volume reactors, but does exhibit statistically significant acceleration effects on decomposition at 250°C (after 1000h). The kinetic model shows that DEDS is predicted to react within an hour in the presence of PDD at 350°C, forming mostly ethane and hydrogen sulfide, whereas conversion of DEDS occurs more slowly at 250°C, allowing sustained interactions with PDD. This is consistent with remnant DEDS only being detected in the products of the 250°C experiment.

Reeves' flexible gold tube reactors allow for the independent control of reactant mixture pressures using water as the pressurizing fluid to carry out the experiment at high pressures far away from the 2-phase boundary for PDD, eliminatng vapor phase reactions unrealistic at geological conditions. However, further investigation is needed to understand the physics approximation error and whether RMG's gas phase kinetics are a reasonable approximation for the single-phase, high pressure system.

Future work should be directed to simulating sulfur effects at even lower temperatures relevant for petroleum reservoir modeling. The expected effect of disulfides at temperatures <250°C is to accelerate oil-to-gas decomposition, as HS· radicals are expected to be longer-lived at those temperatures. We hypothesize that the presence of other sulfur compounds that produce radicals more slowly than disulfides may also lead to a sustained radical pool that accelerates oil-to-gas decomposition in real petroleum systems.

Water has been hypothesized to decrease the rate of oil cracking by quenching radicals, forming oxygenated compounds in the process. [10] Practically all reservoirs are water-wet to varying degrees and contain organic sulfur, thus additional experiments were performed to investigate the effects of water in a combined PDD-DEDS-H2O system. The details of this followup study will not be a subject for this thesis.

We acknowledge financial support from ENI, a member of the MIT Energy Initiative, for the work done in this chapter.

5.7 References

[1] W. L. Orr. "Kerogen/asphaltene/sulfur relationships in sulfur-rich Monterey oils." *Org. Geochem.* 10 (1-3), 1986, pp. 499–516. DOI: 10.1016/0146-6380(86)90049-5.

[2] C. A. Class. "Predicting Organosulfur Chemistry in Fuel Sources." PhD thesis. Massachusetts Institute of Technology, 2015.

[3] T. Hoering. "Thermal reactions of kerogen with added water, heavy water and pure organic substances." *Org. Geochem.* 5 (4), 1984, pp. 267–278. DOI: 10.1016/0146-6380(84)90014-7.

[4] B. P. Tissot and D. H. Welte. *Petroleum Formation and Occurrence.* 2nd Ed. Springer-Verlag, 1984.

[5] Y. V. Kissin. "Catagenesis and composition of petroleum: Origin of n-alkanes and isoalkanes in petroleum crudes." *Geochim. Cosmochim. Ac.* 51 (9), 1987, pp. 2445–2457. DOI: 10.1016/0016-7037(87)90296-1.

[6] O. Weres, A. S. Newton, and L. Tsao. "Hydrous pyrolysis of alkanes, alkenes, alcohols and ethers." *Org. Geochem.* 12 (5), 1988, pp. 433–444.

[7] P. Ungerer. "State of the art of research in kinetic modelling of oil formation and expulsion." *Org. Geochem.* 16 (1–3), 1990, pp. 1–25. DOI: 10.1016/0146-6380(90)90022-R.

[8] K. J. Jackson, A. K. Burnham, R. L. Braun, and K. G. Knauss. "Temperature and pressure dependence of n-hexadecane cracking." *Org. Geochem.* 23 (10), 1995, pp. 941–953. DOI: 10.1016/0146-6380(95)00068-2.

[9] A. K. Burnham, H. R. Gregg, R. L. Ward, K. G. Knauss, S. A. Copenhaver, J. G. Reynolds, and R. Sanborn. "Decomposition kinetics and mechanism of n-hexadecane-1,2-13C2 and dodec-1-ene-1,2-13C2 doped in petroleum and n-hexadecane." *Geochim. Cosmochim. Ac.* 61 (17), 1997, pp. 3725–3737. DOI: 10.1016/S0016-7037(97)00182-8.

[10] M. D. Lewan. "Experiments on the role of water in petroleum formation." *Geochim. Cosmochim. Ac.* 61 (17), 1997, pp. 3691–3723. DOI: 10.1016/S0016-7037(97)00176-2.

[11] F. Dominé, D. Dessort, and O. Brévart. "Towards a new method of geochemical kinetic modelling: implications for the stability of crude oils." *Org. Geochem.* 28 (9-10), 1998, pp. 597–612. DOI: 10.1016/S0146-6380(98)00030-8.

[12] J. S. Seewald. "Organic-inorganic interactions in petroleum-producing sedimentary basins." *Nature* 426 (6964), 2003, pp. 327–333.

[13] G. Martin. "Pyrolysis of organosulfur compounds." *Supplement S: The chemistry of Sulphur-containing functional groups.* Ed. by S. Patai and Z. Rappoport. Wiley, 1993. Chap. 9, pp. 395–437.

[14] D. E. McKinney, F. Behar, and P. G. Hatcher. "Reaction kinetics and n-alkane product profiles from the thermal degradation of 13C-labeled n-C25 in two dissimilar oils as determined by SIM/GC/MS." *Org. Geochem.* 29 (1-3), 1998, pp. 119–136. DOI: 10.1016/S0146-6380(98)00071-0.

[15] R. Bounaceur, G. Scacchi, P.-M. Marquaire, F. Dominé, O. Brévart, D. Dessort, and B. Pradier. "Inhibiting Effect of Tetralin on the Pyrolytic Decomposition of Hexadecane. Comparison with Toluene." *Ind. Eng. Chem. Res.* 41 (19), 2002, pp. 4689–4701. DOI: 10.1021/ie0108853.

[16] V. Burklé-Vitzthum, R. Michels, G. Scacchi, P.-M. Marquaire, D. Dessort, B. Pradier, and O. Brevart. "Kinetic effect of alkylaromatics on the thermal stability of hydrocarbons under geological conditions." *Org. Geochem.* 35 (1), 2004, pp. 3–31. DOI: 10.1016/j.orggeochem.2003.06.001.

[17] F. Lannuzel, R. Bounaceur, R. Michels, G. Scacchi, and P.-M. Marquaire. "Reassessment of the kinetic influence of toluene on n-alkane pyrolysis." *Energy & Fuels* 24 (7), 2010, pp. 3817–3830. DOI: 10.1021/ef100253z.

[18] B. Blouri, F. Hamdan, and D. Herault. "Mild cracking of high-molecular-weight hydrocarbons." *Ind. Eng. Chem. Proc. DD* 24 (1), 1985, pp. 30–37. DOI: 10.1021/i200028a005.

[19] P. E. Savage and M. T. Klein. "Discrimination between molecular and free-radical models of 1-phenyldodecane pyrolysis." *Ind. Eng. Chem. Res.* 26 (2), 1987, pp. 374–376. DOI: 10.1021/ie00062a034.

[20] F. Billaud, P. Chaverot, M. Berthelin, and E. Freund. "Thermal decomposition of aromatics substituted by a long aliphatic chain." *Ind. Eng. Chem. Res.* 27 (8), 1988, pp. 1529–1536. DOI: 10.1021/ie00080a030.

[21] M. D. Lewan. "Sulphur-radical control on petroleum formation rates." *Nature* 391 (6663), 1998, pp. 164–166.

[22] F. Behar, F. Lorant, H. Budzinski, and E. Desavis. "Thermal Stability of Alkylaromatics in Natural Systems: Kinetics of Thermal Decomposition of Dodecylbenzene." *Energy & Fuels* 16 (4), 2002, pp. 831–841. DOI: 10.1021/ef010139a.

[23] P. E. Savage, G. E. Jacobs, and M. Javanmardian. "Autocatalysis and aryl-alkyl bond cleavage in 1-dodecylpyrene pyrolysis." *Ind. Eng. Chem. Res.* 28 (6), 1989, pp. 645–654. DOI: 10.1021/ie00090a001.

[24] G. W. Mushrush and R. N. Hazlett. "Pyrolysis of organic compounds containing long unbranched alkyl groups." *Ind. Eng. Chem. Fund.* 23 (3), 1984, pp. 288–294. DOI: 10.1021/i100015a004.

[25] T. Y. Ho, M. A. Rogers, H. V. Drushel, and C. B. Koons. "Evolution of sulfur compounds in crude oils." *AAPG Bulletin* 58 (11), 1974, pp. 2338–2348.

[26] E. Krein. "Organic sulfur in the geosphere: analysis, structures and chemical processes." *Supplement S: The chemistry of Sulphur-containing functional groups.* Ed. by S. Patai and Z. Rappoport. Wiley, 1993. Chap. 17, pp. 975–1032. DOI: 10.1002/9780470034408.ch17.

[27] J. Tomić, F. Behar, M. Vandenbroucke, and Y. Tang. "Artificial maturation of Monterey kerogen (Type II-S) in a closed system and comparison with Type II kerogen: implications on the fate of sulfur." *Org. Geochem.* 23 (7), 1995, pp. 647–660. DOI: 10.1016/0146-6380(95)00043-E.

[28] F. Dominé, R. Bounaceur, G. Scacchi, P.-M. Marquaire, D. Dessort, B. Pradier, and O. Brevart. "Up to what temperature is petroleum stable? New insights from a 5200 free radical reactions model." *Org. Geochem.* 33 (12), 2002, pp. 1487–1499. DOI: 10.1016/S0146-6380(02)00108-0.

[29] B. P. Tissot, R. Pelet, and P. Ungerer. "Thermal history of sedimentary basins, maturation indices, and kinetics of oil and gas generation." *AAPG Bulletin* 71 (12), 1987, pp. 1445–1466.

[30] J. S. Sinninghe Damsté, F. X. C. de las Heras, P. F. van Bergen, and J. W. de Leeuw. "Characterization of Tertiary Catalan lacustrine oil shales: Discovery of extremely organic sulphur-rich Type I kerogens." *Geochim. Cosmochim. Ac.* 57 (2), 1993, pp. 389–415. DOI: 10.1016/0016-7037(93)90439-4.

[31] H. Freund, C. C. Walters, S. R. Kelemen, M. Siskin, M. L. Gorbaty, D. J. Curry, and A. E. Bence. "Predicting oil and gas compositional yields via chemical structure-chemical yield modeling (CS-CYM): Part 1 - Concepts and implementation." *Org. Geochem.* 38 (2), 2007, pp. 288–305. DOI: 10.1016/j.orggeochem.2006.09.009.

[32] C. C. Walters, H. Freund, S. R. Kelemen, P. Peczak, and D. J. Curry. "Predicting oil and gas compositional yields via chemical structure-chemical yield modeling (CS-CYM): Part 2 - Application under laboratory and geologic conditions." *Org. Geochem.* 38 (2), 2007, pp. 306–322. DOI: 10.1016/j.orggeochem.2006.09.010.

[33] M. D. Lewan. "Laboratory simulation of petroleum formation: hydrous pyrolysis." *Org. Geochem.* New York: Plenum, 1983, pp. 419–442.

[34] D. A. Palmer and S. E. Drummond. "Thermal decarboxylation of acetate. Part I. The kinetics and mechanism of reaction in aqueous solution." *Geochim. Cosmochim. Ac.* 50 (5), 1986, pp. 813–823.

[35] G. D. Cody, N. Z. Boctor, J. A. Brandes, T. R. Filley, R. M. Hazen, and H. S. Yoder. "Assaying the catalytic potential of transition metal sulfides for abiotic carbon fixation." *Geochim. Cosmochim. Ac.* 68 (10), 2004, pp. 2185–2196. DOI: 10.1016/j.gca.2003.11.020.

[36] J. A. Shipp, I. R. Gould, E. L. Shock, L. B. Williams, and H. E. Hartnett. "Sphalerite is a geochemical catalyst for carbon-hydrogen bond activation." *Proc. Natl. Acad. Sci. USA* 111 (32), 2014, pp. 11642–5. DOI: 10.1073/pnas.1324222111.

[37] F. D. Mango, J. W. Hightower, and A. T. James. "Role of transition-metal catalysis in the formation of natural gas." *Nature* 368, 1994, pp. 536–538. DOI: 10.1038/368536a0.

[38] J. S. Seewald. "Evidence for metastable equilibrium between hydrocarbons under hydrothermal conditions." *Nature* 370 (6487), 1994, pp. 285–287. DOI: 10.1038/370285a0.

[39] J. S. Seewald. "Aqueous geochemistry of low molecular weight hydrocarbons at elevated temperatures and pressures: Constraints from mineral buffered laboratory experiments." *Geochim. Cosmochim. Ac.* 65 (10), 2001, pp. 1641–1664. DOI: 10.1016/S0016-7037(01)00544-0.

[40] J. S. Seewald. "Petroleum decomposition under redox-controlled aqueous conditions." *Geochim. Cosmochim. Ac.* 66, 2002, A695.

[41] R. Michels, P. Landis, R. P. Philp, and B. E. Torkelson. "Influence of Pressure and the Presence of Water on the Evolution of the Residual Kerogen during Confined, Hydrous, and High-Pressure Hydrous Pyrolysis of Woodford Shale." *Energy & Fuels* 9 (2), 1995, pp. 204–215. DOI: 10.1021/ef00050a002.

[42] R. A. McTavish. "The role of overpressure in the retardation of organic matter maturation." *J. Petrol. Geol.* 21 (2), 1998, pp. 153–186. DOI: 10.1111/j.1747-5457.1998.tb00652.x.

[43] J. Mi, S. Zhang, and K. He. "Experimental investigations about the effect of pressure on gas generation from coal." *Org. Geochem.* 74, 2014, pp. 116–122. DOI: 10.1016/j.orggeochem.2014.05.012.

[44] C. Zhonghong, M. Zunjing, Z. Ming, S. Xiaoyu, Z. Shouchun, and B. Youshu. "The effects of high pressure on oil-to-gas cracking during laboratory simulation experiments." *J. Petrol. Geol.* 37 (2), 2014, pp. 143–162.

[45] M. D. Lewan and T. E. Ruble. "Comparison of petroleum generation kinetics by isothermal hydrous and nonisothermal open-system pyrolysis." *Org. Geochem.* 33 (12), 2002, pp. 1457–1475. DOI: 10.1016/S0146-6380(02)00182-1.

[46] M. Monthioux, P. Landais, and J.-C. Monin. "Comparison between natural and artificial maturation series of humic coals from the Mahakam delta, Indonesia." *Org. Geochem.* 8 (4), 1985, pp. 275–292. DOI: 10.1016/0146-6380(85)90006-3.

[47] P. Landais, R. Michels, B. Poty, and M. Monthioux. "Pyrolysis of organic matter in cold-seal pressure autoclaves. Experimental approach and applications." *J. Anal. Appl. Pyro.* 16 (2), 1989, pp. 103–115. DOI: 10.1016/0165-2370(89)85010-7.

[48] P. Landais, R. Michels, and M. Elie. "Are time and temperature the only constraints to the simulation of organic matter maturation?" *Org. Geochem.* 22 (3-5), 1994, pp. 617–630. DOI: 10.1016/0146-6380(94)90128-7.

[49] V. P. Nguyen, V. Burkle-Vitzthum, P. M. Marquaire, and R. Michels. "Thermal reactions between alkanes and H2S or thiols at high pressure." *J. Anal. Appl. Pyro.* 103, 2013, pp. 307–319. DOI: 10.1016/j.jaap.2012.11.018.

[50] D. J. Seyfried W. and M. Berndt. "Rocking autoclaves for hydrothermal experiments II: The flexible reaction-cell system." *Hydrothermal Experimental Techniques.* Ed. by G. Ulmer and H. Barnes. New York: Wiley, 1987, pp. 216–239.

[51] C. Lazar, G. D. Cody, and J. M. Davis. "A kinetic pressure effect on the experimental abiotic reduction of aqueous CO2 to methane from 1 to 3.5 kbar at 300 ° C." *Geochim. Cosmochim. Ac.* 151, 2015, pp. 34–48. DOI: 10.1016/j.gca.2014.11.010.

[52] O. F. Tuttle. "Two pressure vessels for silicate-water studies." *GSA Bulletin* 60 (10), 1949, pp. 1727–1729.

[53] E. P. Reeves, J. S. Seewald, and S. P. Sylva. "Hydrogen isotope exchange between n-alkanes and water under hydrothermal conditions." *Geochim. Cosmochim. Ac.* 77, 2012, pp. 582–599. DOI: 10.1016/j.gca.2011.10.008.

[54] R. S. Oremland, L. G. Miller, and M. J. Whiticar. "Sources and flux of natural gases from Mono Lake, California." *Geochim. Cosmochim. Ac.* 51 (11), 1987, pp. 2915–2929.

[55] V. Burklé-Vitzthum, R. Michels, G. Scacchi, and P.-M. Marquaire. "Mechanistic Modeling of the Thermal Cracking of Decylbenzene. Application to the Prediction of Its Thermal Stability at Geological Temperatures." *Ind. Eng. Chem. Res.* 42 (23), 2003, pp. 5791–5808. DOI: 10.1021/ie030086f.

[56] K. M. Van Geem, M. F. Reyniers, G. B. Marin, J. Song, W. H. Green, and D. M. Matheu. "Automatic reaction network generation using RMG for steam cracking of n-hexane." *Aiche J.* 52 (2), 2006, pp. 718–730. DOI: 10.1002/Aic.10655.

[57] K. M. Van Geem, S. P. Pyl, G. B. Marin, M. R. Harper, and W. H. Green. "Accurate High-Temperature Reaction Networks for Alternative Fuels: Butanol Isomers." *Ind. Eng. Chem. Res.* 49 (21), 2010, pp. 10399–10420. DOI: 10.1021/ie1005349.

[58] A. G. Vandeputte, M. K. Sabbe, M.-F. Reyniers, and G. B. Marin. "Modeling the Gas-Phase Thermochemistry of Organosulfur Compounds." *Chem-Eur. J.* 17 (27), 2011, pp. 7656–7673. DOI: 10.1002/chem.201002422.

[59] Y. Kida, C. A. Class, A. J. Concepcion, M. T. Timko, and W. H. Green. "Combining experiment and theory to elucidate the role of supercritical water in sulfide decomposition." *Phys. Chem. Chem. Phys.* 16 (20), 2014, pp. 9220–9228. DOI: 10.1039/C4CP00711E.

[60] A. G. Carr, C. A. Class, L. Lai, Y. Kida, T. Monrose, and W. H. Green. "Supercritical Water Treatment of Crude Oil and Hexylbenzene: An Experimental and Mechanistic Study on Alkylbenzene Decomposition." *Energy & Fuels* 29 (8), 2015, pp. 5290–5302. DOI: 10.1021/acs.energyfuels.5b01040.

[61] J. G. Stainforth. "Practical kinetic modeling of petroleum generation and expulsion." *Mar. Petrol. Geol.* 26 (4), 2009, pp. 552–572. DOI: 10.1016/j.marpetgeo.2009.01.006.

[62] Leeds University. *Sulfur mechanism extension to the methane oxidation mechanism.* 2002.

[63] *CHEMKIN-PRO 15131.* Reaction Design: San Diego. 2013.

[64] A. Amrani, Q. Ma, W. S. Ahmad, Z. Aizenshtat, and Y. Tang. "Sulfur isotope fractionation during incorporation of sulfur nucleophiles into organic compounds." *Chem. Commun.* 2008, pp. 1356–1358. DOI: 10.1039/B717113G.

[65] Personal Communication. 2015.

[66] Personal Communication. 2016.

[67] M. D. Lewan. "Laboratory simulation of petroleum formation: hydrous pyrolysis." *Organic Geochemistry.* 1993, pp. 319–442.

6

Uncertainty propagation in automatic mechanism generation

Uncertainty analysis is the quantification of the relationship between the uncertainty, or variance, in a model's outputs with respect to the uncertainties in the input parameters. In chemical kinetics, these input parameters may include initial concentrations, temperature, pressure, rate coefficients, or thermodynamic quantities. In general, combustion systems are highly nonlinear; uncertainties in a few rate coefficients can become exaggerated while the rest become suppressed. [1, 2]

In RMG, uncertainty propagation is essential because rate coefficients are estimated and can have uncertainties up to several orders of magnitude, which can adversely affect both model generation and predictions. For most users, the model construction workflow is an iterative cycle: (1) generate model using RMG, (2) compare results against experimental data, (3) conduct sensitivity analysis for key observables to identify the most sensitive input parameters, (4) refine the most sensitive parameters by updating the RMG database, then regenerate the model via step (1) and repeat the cycle until a satisfactory model is generated. However, efforts to refine an input parameter are only worthwhile if the model is both sensitive to the parameter and if the uncertainty of the parameter is significant. Uncertainty analysis can help identify the parameters which most strongly contribute to the overall model error. In this chapter, local first order and global uncertainty analysis are implemented and performed for a toy phenyldodecane model generated by RMG. We also analyze the effect of correlated uncertainties arising from the use of kinetics rate rules and thermodynamic group additivity values.

6.1 Background

All models seek to describe the true behavior of some physical phenomena. In order to trust that a kinetic model is capturing the real pathways of a chemical reaction system, it must be able to predict multiple sets of observables measured in experiment. A comparison of error bars between the experimental data and model provides a reliability check for both. In Figure 6.1, two cases are shown. Case 1 shows a model that gives predictions that have overlapping error bars with the experimental data but systematically deviate from the data's nominal values, while case 2 shows a model that has no overlap with the data. In case 1, there is some confidence that the model parameters are reasonable, and refinement of the uncertain input parameters within their error bounds should be able achieve better predictions of the observable data. In case 2, refining the model input parameters within their error bounds will not produce better predictions, and we must understand the possible reasons for this discrepancy before proceeding further:

1. The error bars on the input parameters in the model are underestimated.
2. The error bars on the experimental data are underestimated.
3. The propagated error bounds on the model due to the uncertainties in the input parameters are underestimated.
4. The model structure is missing key features such as reaction pathways or species found in experiment.
5. The simulation is missing key approximations or using the incorrect assumptions to model the experimental conditions and physics.

The model generation workflow requires that these sources of error be investigated systematically in conjunction with the uncertainty propagation methods presented in this chapter, which primarily address (1) and (3): the assignment of error bars on input parameters and accurately propagating error bounds on the model outputs.

A kinetic model can be described as a set of ordinary differential equations of the form:

$$\begin{aligned} \frac{dy}{dt} &= f(y, t; \lambda) \\ y(t_0) &= y_0 \end{aligned} \tag{6.1}$$

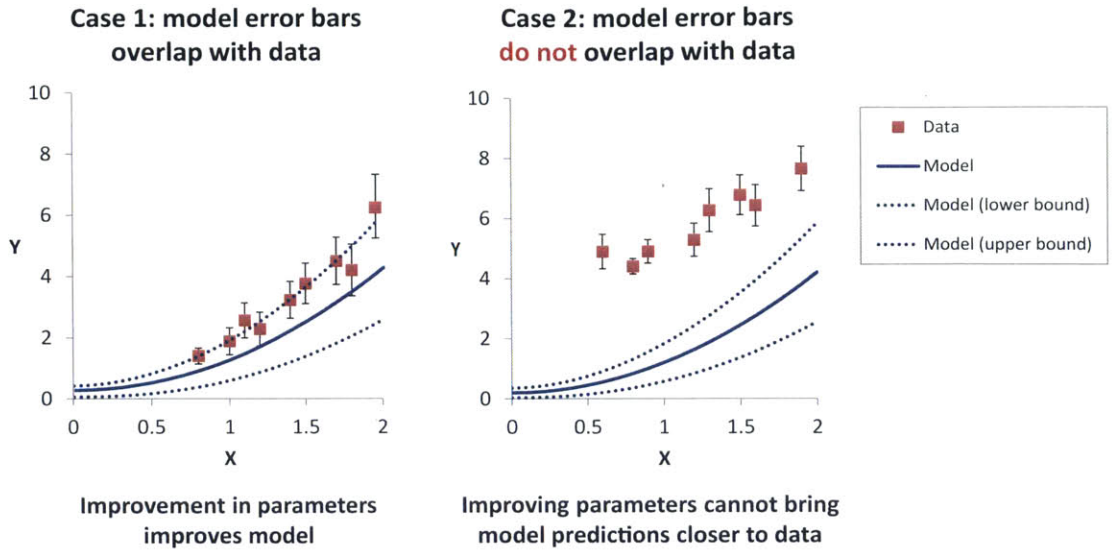


Figure 6.1: Checking model validity against experimental data: two cases. Case 1: model predictions and data overlap, though there is systematic deviation between the nominal experimental values and the model. Case 2: model predictions and data do not overlap.

where y is the solution vector, t is time, λ is the time-independent vector of input parameters, and y_0 is the initial value of y.

Uncertainty analysis is the study of the how the joint probability density function (pdf) of the input parameters λ, which can include temperature, pressure, rate coefficients, or thermodynamic quantities, affect the pdf of the output variables of interest.

The relationship between the output y and the input parameters can be described using a simple Taylor expansion about the nominal value of the input parameter λ_0:

$$y(t, \lambda_0 + \Delta\lambda) = y(t, \lambda_0) + \sum_j \left.\frac{\partial y}{\partial \lambda_j}\right|_{\lambda_0} \Delta\lambda_j + \frac{1}{2}\sum_j \sum_k \left.\frac{\partial^2 y}{\partial \lambda_j \partial \lambda_k}\right|_{\lambda_0} \Delta\lambda_j \Delta\lambda_k + \ldots \quad (6.2)$$

In local uncertainty analysis, the sensitivity coefficients are evaluated at nominal parameter values λ_0. This approximation can be extended to multiple orders. For instance, in highly nonlinear applications, second-order local sensitivities are used for greater accuracy. [3] More commonly, a "local first order" analysis is used, where the Taylor expansion is truncated to a first order approximation. For the rest of this chapter, local uncertainties will refer to the first order approximation, neglecting

covariance, which gives following expresion for D, the variance in the model output y:

$$D = \sigma^2(y) \approx \sum_j \left(\frac{\partial y}{\partial \lambda_j}\right)^2 \Bigg|_{\lambda_0} \sigma^2(\lambda_j) \tag{6.3}$$

where an individual input parameter's contribution to the uncertainty is the parameter variance $\sigma^2(\lambda_j)$ multiplied by the first order sensitivity coefficient of output y with respect to parameter λ_j:

$$D_j \approx \left(\frac{\partial y}{\partial \lambda_j}\right)^2 \Bigg|_{\lambda_0} \sigma^2(\lambda_j) \tag{6.4}$$

However, for kinetic input parameters $\lambda = k$, it is often more useful to rewrite y in terms of the transformed input parameter $\ln(k_j)$, under the assumption that rate coefficients have a multiplicative or logarithmic uncertainty distribution. [4]

$$D = \sigma^2(y) \approx \sum_j \left(\frac{\partial y}{\partial \ln k_j}\right)^2 \Bigg|_{k_0} \sigma^2(\ln k_j) \tag{6.5}$$

Therefore, the first order sensitivity index S_j, or contribution of uncertainty from an individual input kinetic parameter k_j on the output distribution of y is:

$$S_j = \frac{D_j}{D} \approx \frac{\left(\frac{\partial y}{\partial \ln k_j}\right)^2 \Big|_{k_0} \sigma^2(\ln k_j)}{\sigma^2(y)} \tag{6.6}$$

Approximating error with first order sensitivity coefficients at nominal parameter values may not be robust for handling large simultaneous uncertainties in model parameters. [1] In highly nonlinear systems, the behavior within the error bounds in an input parameter can vary greatly, and the linearity assumption in local uncertainty analysis does not hold. In other words, local uncertainty propagation will lead to incorrect output pdfs.

In RMG-generated models, rate coefficients are often uncertain to many orders of magnitude, hence first order sensitivity coefficients may poorly describe them. Furthermore, the assumption of independent random variables and no covariance does not hold for mechanisms generated by RMG, since an individual kinetic rate rule can be used to construct multiple rate coefficients in the same model. However, the simplicity and speed at which first order sensitivity coefficients can be computed make local uncertainty analysis a useful tool for gaining qualitative understanding. Using first order sensitivity indices, it is possible rank the input rate coefficients

in terms of their error contribution to the output uncertainties to determine which parameters should be refined.

Global uncertainty analysis is capable of handling large variations in input parameters because they simulate the system over the entire range of parameter uncertainties. A common method is stochastic Monte Carlo, which samples the parameter space assuming that their uncertainties can be represented using a known pdf about their nominal values. The output pdf is determined statistically after a large number of simulations probing the uncertain parameter space. Monte Carlo is simple to implement, but the computational costs are high and scale exponentially with the number of parameters. [5] To ameliorate the computational time, sampling techniques can be chosen carefully to minimize the number of simulations required. For example, Latin hypercube sampling or Sobol sequences can be used to generating accurate output distributions much faster than Monte Carlo methods alone. [1, 6]

Other global uncertainty methods rely on using basis set expansions to describe the output distributions post-sampling the parameter space. These methods also require sampling the uncertain parameter space and simulating the model, but typically converge faster than Monte Carlo. These methods include the Fourier Amplitute Sensitivity Test (FAST), High-dimensional Model Representations (HDMR), and Polynomial Chaos Expansions (PCE).

The Fourier Amplitude Sensitivity Test (FAST) method assumes that each input parameter is statistically independent with a given pdf. The parameters are perturbed simulataneously using sine functions with incommensurate frequencies to obtain Fourier coefficients that estimate the output pdf. The FAST method uses less sampling points than Monte Carlo methods. [5]

High-dimensional model representations (HDMR) [5] create a mapping between the input variables λ and the output y on the basis of their zeroth, first order, and higher order interactions of the form:

$$y = f(\lambda) = f_0 + \sum_{j}^{n} f_j(\lambda_j) + \sum_{1 \le i < j \le n} f_{i,j}(\lambda_j, \lambda_k) + ... + f_{12...n}(\lambda_1, \lambda_2, ..., \lambda_n) \quad (6.7)$$

where f_0 is the zeroth order effect, $f_j(\lambda_j)$ is the first-order term mapping the effect of the input λ_j on the output f, and so forth. The HDMR method is computationally efficient if the higher order interactions are negligible or converge quickly. Tomlin and Ziehn have used the HDMR method on a methane flame model study and demonstrated that a initial Morris Method screening step which discards non-

influential parameters can be used to reduce high dimensional input spaces and reduce computation time while maintaining accuracy. [2]

Polynomial chaos expansions (PCE) are becoming more widely used to track the evolution of probability density functions within combustion systems. [7, 8] Sheen and Wang have used PCEs in conjunction with experimental data to bound uncertainty in input parameters. [8] CHEMKIN-PRO [9] also contains global uncertainty capabilities using PCEs with parametric uncertainty analysis via the Deterministic Equivalent Modeling Method (DEMM). [10] This module allows the user to assign the PCE polynomial order and either a normal, lognormal, or uniform pdf for the input parameters. However, the current version of CHEMKIN-PRO's global analysis is cumbersome to use. All inputs and outputs must be viewed through the graphical user interface, only fixed order polynomial expansions can be used, and there is at least one unresolved bug associated with assigning the input parameter uncertainties.

The above methods use Bayesian inference to track global error propagation within nonlinear chemical models and require prior knowledge of the input parameters' probability distribution functions. A major drawback to these Bayesian methods is the computation time required relative to local uncertainty analysis. Frenklach and coworkers have devised an alternate method which uses bounded intervals to represent input uncertainties instead of probability distributions to cut back on computational costs. Rate parameters are then optimized on a feasible set constrained by the experimental data and their associated uncertainties. [11] This framework is the basis of the PrIMe online kinetics database [12] and the GRI-Mech3.0 dataset [13]. However, Bayesian methods are more widely used in the kinetics community.

6.2 Overview

This chapter describes the implementation of local and global uncertainty propagation methods. First, input parameter uncertainties were assigned to the rate coefficients and species thermochemistry estimated by RMG. Sensitivity analysis with respect to the nominal parameter values was implemented in RMG using DASPK 3.1 [14]. A framework for local uncertainty propagation was built to analyze output uncertainty dependence on either independent or correlated input parameters. Finally, a framework for global uncertainty propagation was implemented using PCEs constructed with the adaptive Smolyak pseudospectral approximation method. [15]

A toy model for phenyldodecane was used to compare the various methods. Local uncertainty analysis for independent and correlated coefficients was performed, fol-

lowed by a comparison of local vs. global methods for independent parameters. These results and their implications on the model generation workflow are then summarized.

6.3 Assigning uncertainties to RMG's databases

Rate coefficients and thermodynamic parmaters uncertainties must be assigned to track error propagation. This is difficult because most uncertainty information is often missing even from their original literature sources. Some uncertainties can be estimated for parameters derived from *ab initio* calculations based on the level of theory. Other rate coefficients are fitted using multiple quantum calculations or experimental data and have additional uncertainties associated with fitting. [16, 17]

Baulch *et al.* [4] assigns lognormal distributions to rate coefficients about their nominal value such that $k \in [k_{min}, k_{max}]$. In our analysis, we will instead be using log-uniform distribution about $\ln(k_0)$, an uncertainty factor f can be defined:

$$f = \log_{10}\left(\frac{k_0}{k_{min}}\right) = \log_{10}\left(\frac{k_{max}}{k_0}\right) = \frac{1}{\ln 10}\ln\left(\frac{k_{max}}{k_0}\right) \tag{6.8}$$

The standard deviation $\sigma \ln k$ of a uniform distribution about $ln(k_0)$ can be expressed in terms of f (note that this expression differs slightly from the definitions by Baulch *et al.* [4]):

$$\sigma \ln k = \sqrt{\frac{1}{12}}[\ln(k_{max}) - \ln(k_{min})] = \frac{2}{\sqrt{12}}\ln\left(\frac{k_{max}}{k_0}\right) = \frac{\ln 10}{\sqrt{3}}f \tag{6.9}$$

In most studies, the uncertainty in k is assumed to be the uncertainty associated with the preexponential factor A in the modified Arrhenius expression:

$$k = AT^n exp\left(\frac{-E_a}{RT}\right) \tag{6.10}$$

although it is unrealistic to characterize the uncertainties of E_a and n to be zero. Complete characterization of the uncertainties for A, E_a, and n requires a covariance matrix to describe their pdf, since Arrhenius coefficients are highly correlated and typically constrained through fitting to either experimental or theoretical data. Turanyi *et al.* have described a method in which elements of the covariance matrix can be determined through fitting to a set of temperature dependent rate uncertainty data. [18] This results in an uncertainty in k that contains temperature dependence.

6.3.1 Kinetics uncertainty

The RMG database contains kinetic data in the form of rate rules, training reactions, and library reactions. Although these entries are equipped to store uncertainty information, most contain no information. Therefore, a general assignment of uncertainties must be carried out.

For simplicity in our initial implementation, we will assign uncertainties in k using only a multiplicative uncertainty in the pre-exponential factor A, rather than complete uncertainty characterization for all the coefficients in the Arrhenius expression. Furthermore, we will consider the uncertainties to be uniform distributions about $\ln(k)$ in their uncertainty intervals $[\ln(k_{min}), \ln(k_{max})]$), rather than normal distributions in $\ln(k)$. This prevents problems during sampling for global analysis, arising from nonphysical extreme values at the tail ends of the normal distribution. It also supports the notion that the true rate coefficient could be any value within the error bound, rather than having a higher likelihood of being close to the nominal value.

Different sources of kinetics within an RMG-generated model have different expected uncertainties. A reaction that matches a training reaction or library reaction uses a rate calculated for that exact reaction, therefore the uncertainty matches that of the original rate. If there is no uncertainty associated with the entry, an uncertainty factor of $f = 0.5$ is assigned.

For rate coefficients estimated through the reaction families, a more thorough treatment is needed. The forward rate coefficient $k^{estimate}$ depends on the kinetics estimate k_m^{site} associated with the transition state of a particular reaction site m with degeneracy d_m:

$$k^{estimate} = d_m k_m^{site} \tag{6.11}$$

We can consider the uncertainty $\Delta \ln(k)$ between the true kinetics k^{true} and the estimated kinetics $k^{estimate}$ as equal to the uncertainty $\Delta \ln(k_m^{site})$ in the estimated reaction site kinetics (with no dependence on d_m):

$$\Delta \ln(k) = \left|\ln\left(k^{true}\right) - \ln\left(k^{estimate}\right)\right| = \Delta \ln\left(k_m^{site}\right) \tag{6.12}$$

In order to assign the uncertainty to $\Delta \ln(k_m^{site})$, we must first understand how k_m^{site} is calculated. In RMG, the kinetics k_m^{site} for a unique reaction site m is be estimated using the geometric average of N rate rule(s) with individual kinetics k_i^{rule}:

$$ln(k_m^{site}) = \frac{1}{N}\sum_{i=1}^{N} \ln\left(k_i^{rule}\right) \tag{6.13}$$

The rate rule(s) used to approximate k_m^{site} can be considered either an "exact" or "non-exact" match. In practice, an exact match occurs only in the case of a single rate rule ($N = 1$). A non-exact match can occur when the reaction site is described using the average of either a single ($N = 1$) or multiple ($N > 1$) rate rules.

We can think of the uncertainty $\Delta \ln(k_m^{site})$ in the estimated site kinetics to be broken up into two components, a residual error term Δ_m and an intrinsic error term associated with the error from the rate rules:

$$[\Delta \ln\left(k_m^{site}\right)]^2 = (\Delta_m)^2 + \sum_{rule\ i} \left(\frac{\partial \ln k_m^{site}}{\partial \ln k_i^{rule}}\right)^2 \left[\Delta \ln\left(k_i^{rule}\right)\right]^2 \tag{6.14}$$

The intrinsic second term describes the error associated with the rate rules' k_i^{rule} origin of calculation, i.e. experiment or theory, including error arising from any explicit dependence on species thermochemistry G_n. A single rate rule's kinetics k_i^{rule} can be described using the Evans-Polanyi modified Arrhenius expression:

$$\begin{aligned} k_i^{rule}(T) &= A_i \left(\frac{T}{1\ \mathrm{K}}\right)^{n_i} \exp\left(-\frac{E_{a,i}}{RT}\right) \\ E_{a,i} &= max(0,\ \alpha_i \Delta H_{rxn} + E_{0,i}) \end{aligned} \tag{6.15}$$

where A_i is the pre-exponential factor, n_i is the temperature exponential factor, $E_{a,i}$ is the activation energy, R is the gas constant, T is the temperature, ΔH_{rxn} is the enthalpy of reaction, and α_i and $E_{0,i}$ are constants. This expression simplifies to the modified Arrhenius expression when $\alpha_i = 0$.

To calculate the intrinsic error from the rate rules, first we derive $\ln\left(k_i^{rule}(T)\right)$:

$$\ln\left(k_i^{rule}(T)\right) = \ln(A_i) + n_i \ln(T) - \frac{max(0,\ \alpha_i \Delta H_{rxn} + E_{0,i})}{RT} \tag{6.16}$$

For simplicity in describing the uncertainty in a single rate rule, we do not consider the uncertainties associated with n_i, α_i, or $E_{0,i}$. Therefore the uncertainty in the rate rule depends only on the kinetics A-factor uncertainty $\Delta \ln A_i$ and the species thermochemistry uncertainty ΔG_n:

$$[\Delta \ln\left(k_i^{rule}(T)\right)]^2 = (\Delta \ln A_i)^2 + \left(\frac{-\alpha_i}{RT}\right)^2 \sum_{species\ n} \left(\frac{\partial \Delta H_{rxn}}{\partial G_n} \Delta G_n\right)^2 \tag{6.17}$$

Note that in RMG, rate rules are defined for the forward direction kinetics within a reaction family. Generated reactions are always defined in the same direction as the reaction family, therefore, there is no thermochemistry dependence in the forward

rate coefficients besides enthalpy dependence in the Evans-Polanyi form. A small number of rate rules are fitted from training reactions at the beginning of the RMG job. If the training reaction is defined in the reverse direction of the family's reaction template, the rate rule must be fitted with the reverse rate coefficient using the species thermochemistry. However, this thermochemistry dependence will not be considered here, as we attribute this uncertainty to be part of the general A-factor uncertainty $\Delta \ln A_i$ in the final rate rule.

For the preliminary analyses within this chapter, we ignore the thermochemistry dependence entirely in the rate rule uncertainty (the second term in Equation 6.17), since there are very few Evans-Polanyi rate rules in the database (the majority are modified Arrhenius expressions). However, future analyses should include this term, especially for low temperature systems where this uncertainty term becomes more relevant.

Therefore, the intrinsic uncertainty term can be simplified to:

$$\sum_{rule\ i} \left(\frac{\partial \ln k_m^{site}}{\partial \ln k_i^{rule}} \right)^2 \left[\Delta \ln\left(k_i^{rule}\right) \right]^2 = \frac{1}{N^2} \sum_{i=1}^{N} [\Delta \ln\left(k_i^{rule}\right)]^2 = \frac{1}{N^2} \sum_{i=1}^{N} (\Delta \ln A_i)^2 \quad (6.18)$$

The residual term Δm is the error associated with the assumption of chemical similarity through using the estimated reaction site kinetics k_m^{site} to describe the kinetics of a real reaction on the basis of matching functional groups. A set of similar reactions collapse to the same estimated reaction site kinetics, but their true rate coefficients will differ from each other. Given the implementation in RMG, we can think of the chemical similarity error term as being associated with both the error Δ_m^{family} from how specific the reaction sites are defined in a particular reaction family, as well as Δ_m^{match}, which differentiates between sites matched exactly or non-exactly by the rate rules:

$$(\Delta_m)^2 = (\Delta_m^{family})^2 + (\Delta_m^{match})^2 \quad (6.19)$$

The term Δ_m^{family} should differ for each reaction family based on the sparseness of the reaction family tree: a more densely populated tree with many assigned rate rules will contribute less chemical error. In the implementation in thie chapter, we assume the same error value Δ_m^{family} for all reaction families. Future implementations could use the leave-one-out cross validation statistic as a measure of this error. The term Δ_m^{match} should be large for non-exact rate rule matches for the reaction site and small for exact matches.

Table 6.1: Uncertainty factors for RMG kinetic parameters

Source of Kinetics	Uncertainty factor (f)
Reaction library	0.5
Traning reaction	0.5
Reaction family rate estimate	See Equation 6.21
Pressure dependent rate estimate	2

A value of $\Delta \ln\left(k_i^{rule}\right) = \Delta \ln A_i = 2.0$ is assigned to the uncertainty of each rate rule, $\Delta_m^{family} = 1.0$ for all families, and $\Delta_m^{match} = 2.0$ for non-exact matches and $\Delta_m^{match} = 0.0$ for exact rate rule to reaction site matches.

In summary, kinetics estimated by rate rules have uncertainties $\Delta ln(k)$ that can be broken down into both intrinsic and chemical similarity terms and expanded as follows:

$$\begin{aligned} \Delta \ln(k) = \Delta \ln\left(k_m^{site}\right) &= \sqrt{(\Delta_m)^2 + \sum_{rule\ i} \left(\frac{\partial \ln k_m^{site}}{\partial \ln k_i^{rule}}\right)^2 \left[\Delta \ln\left(k_i^{rule}\right)\right]^2} \\ &= \sqrt{(\Delta_m^{family})^2 + (\Delta_m^{match})^2 + \frac{1}{N^2}\sum_{i=1}^{N}\left(\Delta \ln k_i^{rule}\right)^2} \end{aligned} \tag{6.20}$$

Hence, the uncertainty factor f for kinetics estimated by rate rules is:

$$f = \frac{\sqrt{3}}{\ln 10}\Delta ln(k) = \frac{\sqrt{3}}{\ln 10}\sqrt{(\Delta_m^{family})^2 + (\Delta_m^{match})^2 + \frac{1}{N^2}\sum_{i=1}^{N}\left(\Delta \ln k_i^{rule}\right)^2} \tag{6.21}$$

For kinetics estimated through RMG's pressure-dependence module, we use simplified, lumped uncertainty for $\Delta \ln(k)$ with uncertainty factor $f = 2$.

The uncertainty factors for different kinetics sources are assigned in Table 6.1. Since most database rate coefficients do not have explicit uncertainties, these factors are used for the rest of the analyses in this chapter.

6.3.2 Thermochemistry uncertainty

In most uncertainty analyses, normal distributions are assumed for the uncertainties in the thermodynamic parameters ΔH, ΔS and C_p. [1, 19] In our implementation, we will using a uniform distribution across the uncertainty interval instead. Like the use of uniform distributions for kinetics, this prevents sampling the extreme values at the tails of a normal distribution.

In this chapter, we will be considering uncertainties ΔG_n for the free energy of an individual species n. We define the uncertainty distribution to be uniform about the nominal value G_0 within the interval $[G_{min}, G_{max}]$. The uncertainty factor f for thermodynamics can then be defined as:

$$f = G_{max} - G_0 = G_0 - G_{min} \tag{6.22}$$

The relationship between the uncertainty factor f and the standard deviation $\sigma(G) = \Delta G$ is:

$$\Delta G = \frac{1}{\sqrt{12}}(G_{max} - G_{min}) = \frac{2}{\sqrt{12}}(G_{max} - G_0) = \frac{f}{\sqrt{3}} \tag{6.23}$$

Thermodynamic parameters for each species are either taken from exact matches in Thermo Libraries or estimated through either Benson group additivity or on-the-fly QMTP (for cyclic molecules only). The ΔG_n for species thermochemistry estimated from libraries is the intrinsic error associated with the experimental or theoretical value. We assign an uncertainty of $\Delta G_n = 1.0$ kcal/mol for these values (note that this correponds to a ± 2.0 kcal/mol uncertainty interval). For thermochemistry estimated using the QMTP module, the error is associated with the level of theory used. Most QMTP calculations are carried out using semiempirical PM methods. We can consider their error to be approximately $\Delta G_n = 2.5$ kcal/mol. For Benson group additivity, the thermochemistry G_n for a species n is estimated using the sum of group values G_j^{group} corresponding to the degeneracy d_j of each Benson functional group j in the molecule:

$$G_n = \sum_{group\ j} d_j G_j^{group} \tag{6.24}$$

Similar to kinetics, the error we expect between the estimated thermochemistry and the true value comprises of the uncertainties in the themo groups as well as a residual error Δ_n from using the group additivity approximation to estimate the species thermochemistry:

$$(\Delta G_n)^2 = (\Delta_n)^2 + \sum_{group\ j} \left(\frac{\partial G_n}{\partial G_j^{group}} \right)^2 (\Delta G_j^{group})^2 \tag{6.25}$$

The residual error term Δn is expected to be small for most species, but likely high for certain classes of species such as cyclics, where the available groups are not specific enough. The certainty of each thermo group additivity value ΔG_j^{group} will

also be larger for groups fitted with less training data. The preliminary analyses in this chapter uses $\Delta n = 0$ and $\Delta G_j^{group} = 0.075$ kcal/mol for all Benson groups.

6.4 Sensitivity analysis

Previously, sensitivity analysis for RMG-generated models was conducted in CHEMKIN-PRO [9], a proprietary software with substantial license fees. In CHEMKIN-PRO, sensitivity coefficients for kinetic parameters can be obtained, but not for thermodynamic parameters, which can also have a great influence on model outputs under uncertainty.

Native sentivity analysis within RMG provides the advantage of local uncertainty propagation without the use of external software and also the possibility of using sensitivities as a metric for enlarging the model instead of fluxes. First order sensitivities can be calculated using a variety of methods, including finite differences, the decoupled direct method, Green's function method, or analytically integrated Magnus. [5] In addition, many commercial and open source post-processing codes are available such as SENKIN, KINALC, and the IDA solver in Sundials. [20] In the Java version of RMG, sensitivity analysis was implemented by coupling two Fortran codes: the automatic differentiation package DAEPACK [21] and the large scale differential algebraic equation solver DASPK 2.0 [22]. DAEPACK is the first stiff differential algebraic solver that uses sparse linear algebra for solving model systems. However, DAEPACK is only available commercially and cannot be freely distributed. We sought a solver with less restricted availability for the Python version of RMG. Assimulo, the Python wrapper for Sundials [20] was considered but discarded in favor of Petzold's fortran DASPK 3.1 [14] solver. DASPK 3.1 enables native sensitivity analysis and was the most compatible with the existing solvers in RMG, which relies on PyDAS, [23] a python wrapper for DASSL [24], an older solver written by Petzold. PyDAS was extended with a new submodule `pydas.daspk` containing a Python wrapper for DASPK 3.1. RMG's base solvers can be interchangeably configured to use either either the DASSL or DASPK 3.1 solver during compilation. This flexibility was given to users such that those without a DASPK 3.1 license (free-for-use for academic users only) can be able to fall back to the open-source DASSL solver. The RMG reaction system solver was optimized for speed using Cython [25]. Future work may find it worthwhile to implement automatic differentiation in RMG's solvers for speedup, as DASPK3.1 has some support capabilities for it when used in conjunction with with ADIFOR, [26] an automatic differentiation library written in fortran.

6.4.1 Mathematical description

Referring to the kinetic model of the form:

$$\begin{aligned} \frac{dy}{dt} &= f(y, t; \lambda) \\ y(t_0) &= y_0 \end{aligned} \tag{6.26}$$

where y is the solution vector, t is time, λ is the time-independent vector of input parameters, and y_0 is the initial value of y.

Define the first order sensitivity coefficient of output y (not to be confused with the first-order sensitivity index described earlier) with respect to parameter λ_j as s_j (lowercase):

$$s_j = \frac{\partial y}{\partial \lambda_j} \tag{6.27}$$

A set of differential equations for the sensitivity coefficients can be derived by differentiating equation 6.1 with respect to λ_j to obtain

$$\begin{aligned} \frac{ds_j}{dt} &= J(y, t; \lambda) s_j + f_j(y, t; \lambda) \\ s_j(t_0) &= \frac{\partial y_0}{\partial \lambda_j} \end{aligned} \tag{6.28}$$

where J is the Jacobian matrix whose terms are defined

$$J_{mn} = \frac{\partial f_m}{\partial y_n} \tag{6.29}$$

and f_j contains the explicit dependence of f on on the input parameter λ_j

$$f_j = \frac{\partial f}{\partial \lambda_j} \tag{6.30}$$

In RMG, we use a simple reactor system that is isothermal and isobaric. The variable y is cast as the moles of each species y_n and λ consists of both forward rate coefficients k_f and species thermochemistry $\lambda = G$. The functional form of f is:

$$\frac{\mathrm{d}y_n}{\mathrm{d}t} = f(y_n, t; \lambda) = V \sum_{rxn} \left(k_f \prod_{reactants} c_{reactant} - k_r \prod_{products} c_{product} \right) \tag{6.31}$$

where c is the species concentration, defined as $c = y_n/V$, and V is the volume, which is computed using the ideal gas law algebraically:

$$PV = RT \sum_i y_{n,i} \tag{6.32}$$

where P is the system pressure, R is the gas constant, and T is the system temperature. In our isothermal, isobaric system, V is proportional to the total moles $\sum_i y_{n,i}$.

Molar sensitivities $s_j = dy_n/d\lambda_j$ with respect to the forward rate coefficients and the species thermochemistry can both be derived with a single jacobian calculation and evaluation of the derivative f_j.

First, k_r is be substituted given the thermodynamic constraint

$$K_{eq} = \frac{k_f}{k_r} = \left(\frac{RT}{P^\circ}\right)^{-\Delta n} \exp\left(\frac{-\Delta G^\circ_{rxn}}{RT}\right) \tag{6.33}$$

where K_{eq} is the equilibrium constant of the reaction, T is the reaction temperature, R is the gas constant, P° is the standard pressure (1 bar), $\Delta G^\circ_{rxn}(T)$ is the standard reaction free energy, and Δn is the change in moles in the reaction.

This gives $\frac{\mathrm{d}y_n}{\mathrm{d}t}$ in terms of only the input parameters k_f and G:

$$\frac{\mathrm{d}y_n}{\mathrm{d}t} = V \sum_{rxn} \left(k_f \prod_{reactants} c_{reactant} - \frac{k_f}{\left(\frac{RT}{P^\circ}\right)^{-\Delta n} \exp\left(\frac{-\Delta G^\circ_{rxn}}{RT}\right)} \prod_{products} c_{product} \right) \tag{6.34}$$

In RMG, the residuals for dy_n/dt are defined using Equation 6.34 above, and the residuals for ds_j/dt are simulataneously computed through evaluating the jacobian analytically for the isothermal, isobaric system. These residuals are passed to DASPK 3.1 through the Python wrapper `PyDAS`, which integrates the system and computes the solution vector of species moles y_n and sensitivities s_j. Since the initial moles $y_{n,0}$ are independent of k_f and $G_{species}$, the initial values for the sensitivities $s_j(t_0) = 0$.

We are interested in deriving sensitivities with respect to concentrations c rather than moles y_n, for comparability with outputs from CHEMKIN [9]. In order to convert the molar sensitivities $\partial y_n/\partial \lambda_j$ outputted from RMG to concentration sensitivities $\partial c/\partial \lambda_j$, the following relationship is used

$$\begin{aligned} y_n &= cV \\ \frac{\partial y_n}{\partial \lambda_j} &= \frac{\partial c}{\partial \lambda_j} V + \frac{\partial V}{\partial \lambda_j} c \end{aligned} \tag{6.35}$$

Rearranging, the concentration sensitivities can be written in terms of the molar sensitivities and a volume sensitivity term $\frac{\partial V}{\partial \lambda_j}$:

$$\frac{\partial c}{\partial \lambda_j} = \frac{1}{V}\left(\frac{\partial y_n}{\partial \lambda_j} - c\frac{\partial V}{\partial \lambda_j}\right) \tag{6.36}$$

The volume sensitivity term $\partial V/\partial k_j$ can be derived from the ideal gas law constraint from Equation 6.32:

$$\frac{\partial V}{\partial \lambda_j} = \frac{\partial\left(\frac{RT}{P}\sum_i y_{n,i}\right)}{\partial \lambda_j} \tag{6.37}$$

In the case of the simple reactor at constant temperature and constant pressure, this term can be further simplified:

$$\frac{\partial V}{\partial \lambda_j} = \frac{RT}{P}\sum_i \frac{\partial y_{n,i}}{\partial \lambda_j} \tag{6.38}$$

6.4.2 Running sensitivity analysis within RMG

Sensitivity analysis can be performed within RMG with respect to either the forward kinetic rate parameters k_f or thermochemistry G of an individual species.

We are particularly interested in the normalized sensitivity of the concentration with respect to the forward rate coefficient k_m of reaction m:

$$\frac{\partial \ln c}{\partial \ln k_m} = \frac{k_m}{c}\left(\frac{\partial c}{\partial k_m}\right) \tag{6.39}$$

and the semi-normalized sensitivity with respect to the G_n of species n:

$$\frac{\partial \ln c}{\partial G_n} = \frac{1}{c}\left(\frac{\partial c}{\partial G_n}\right) \tag{6.40}$$

RMG-Py input files were modified to handle optional sensitivity analysis options. RMG prints time-dependent species sensitivities to a `.csv` file post model generation and allows user-specified thresholds for the printed sensitivities (the default threshold is to 1×10^{-3}) in order to limit excessive output file sizes, providing the same flexibility as CHEMKIN.

Both thermo and kinetic sensitivities can be automatically simulated within RMG, either in a stand-alone analysis for a pre-existing CHEMKIN format model, or at the end of an RMG job. The top 10 most sensitive rate coefficients and top 10 most sensitive species thermo are automatically plotted for the outputs of interest and saved to the `solver` folder.

6.5 Local uncertainty propagation

The first-order local uncertainty propagation of uncertainties when considering the forward rate coefficients k_m and species thermochemistry G_n as independent can be expressed as:

$$\begin{aligned}
[\Delta(\ln c)]^2 &= \sum_{rxn\ m} \left(\frac{\partial \ln c}{\partial \ln k_m}\right)^2 (\Delta \ln k_m)^2 + \sum_{species\ n} \left(\frac{\partial \ln c}{\partial G_n}\right)^2 (\Delta G_n)^2 \\
&= \sum_{rxn\ m} \left(\frac{\partial \ln c}{\partial \ln k_m}\right)^2 \left[(\Delta_m)^2 + \sum_{rule\ i} \left(\frac{\partial \ln k_m}{\partial \ln k_i^{rule}}\right)^2 (\Delta \ln k_i^{rule})^2\right] \\
&+ \sum_{species\ n} \left(\frac{\partial \ln c}{\partial G_n}\right)^2 \left[(\Delta_n)^2 + \sum_{group\ j} \left(\frac{\partial G_n}{\partial G_j^{group}}\right)^2 (\Delta G_j^{group})^2\right]
\end{aligned} \tag{6.41}$$

Here, the first order sensitivity indices, or the variance contribution from the kinetic parameter uncertainty $\Delta \ln k_m$ and the species free energy uncertainty ΔG_n on the species concentration variance are:

$$\begin{aligned}
S_m^{rxn} &= \frac{\left(\frac{\partial \ln c}{\partial \ln k_m}\right)^2 (\Delta \ln k_m)^2}{[\Delta(\ln c)]^2} \\
S_n^{species} &= \frac{\left(\frac{\partial \ln c}{\partial G_n}\right)^2 (\Delta G_n)^2}{[\Delta(\ln c)]^2}
\end{aligned} \tag{6.42}$$

6.5.1 Correlated uncertainty analysis

Uncertainty quantification for chemical kinetic models typically assumes independent kinetic and thermo parameters. However, there are implicit correlations between model parameters, even in hand constructed models. These correlations are usually difficult to track. In automatic reaction mechanism generation, however, thermodynamic and kinetic parameters are estimated using group-based contributions that are easily identified. Therefore, these parameters uncertainties are explicitly correlated, reducing the degrees of freedom and dimensionality of the model. Prager *et al.* conducted a uncertainty propagation study on a propane/ethane/air mixture using correlatated rate coefficients. [27] Their system contained correlated uncertainties for two rate rules used as rate coefficients for two reactions each. PCEs were used to construct the output ignition delay uncertainties. Their results show that the coupling of the reaction uncertainties produces an increase in output uncertainty compared to the case where the rate coefficient uncertainties are uncorrelated. Their analysis

showed that accounting for correlations in the kinetic model leads to higher fidelity uncertainty quantification as well as computational savings from reduced dimensionality. Wang and Sheen [28] also highlight two other viewpoints regarding rate-rule induced uncertainty during model development. One viewpoint suggests that rate parameters derived from a single rate rule or reaction class should be treated as decoupled from each other on the basis that the rate rule provides an initial guess for the real kinetic rate. The other viewpoint argues that rate parameters of a single reaction class should obey certain constraints due to chemical similarity. As described earlier, in our assignment of uncertainties to the rate parameters, we include both the error intrinsic to the rate rules as well as the residual error in approximating the real reaction.

The first order sensitivity indices with respect to the uncertainties in the rate rules Δk_i^{rule} and thermochemical groups ΔG_j^{group} can be derived by first rewriting the output uncertainty $\Delta(\ln c)$ in terms of the correlated rate rules and groups:

$$\begin{aligned}[\Delta(\ln c)]^2 = &\sum_{rule\ i} (\Delta \ln k_i^{rule})^2 \left[\sum_{rxn\ m} \left(\frac{\partial \ln c}{\partial \ln k_m}\right)\left(\frac{\partial \ln k_m}{\partial \ln k_i^{rule}}\right)\right]^2 \\ &+ \sum_{group\ j} (\Delta G_j^{group})^2 \left[\sum_{species\ n} \left(\frac{\partial \ln c}{\partial G_n}\right)\left(\frac{\partial G_n}{\partial G_j^{group}}\right)\right]^2 \\ &+ \sum_{rxn\ m} \left(\frac{\partial \ln c}{\partial \ln k_m}\right)^2 (\Delta_m)^2 + \sum_{species\ n} \left(\frac{\partial \ln c}{\partial G_n}\right)^2 (\Delta_n)^2\end{aligned} \tag{6.43}$$

The individual reaction kinetics k_m and species thermochemistry G_n are no longer independent and their sensitivity terms are lumped with respect to the rate rules and groups. Additional terms containing uncertainties Δ_m and Δ_n are still independent because they are residual errors.

In this representation, we see that the total uncertainty in the output concentration is smaller than when considering the parameters to be independent, due to the loss in degrees of freedom. We can now compute first-order sensitivity indices for the rate rules and thermochemical groups:

$$\begin{aligned}S_i^{rule} &= \frac{\left(\frac{\partial \ln c}{\partial \ln k_i^{rule}}\right)^2 (\Delta \ln k_i^{rule})^2}{[\Delta(\ln c)]^2} = \frac{\left[\sum_{rxn\ m}\left(\frac{\partial \ln c}{\partial \ln k_m}\right)\left(\frac{\partial \ln k_m}{\partial \ln k_i^{rule}}\right)\right]^2 (\Delta \ln k_i^{rule})^2}{[\Delta(\ln c)]^2} \\ S_j^{group} &= \frac{\left(\frac{\partial \ln c}{\partial G_j^{group}}\right)^2 (\Delta G_j^{group})^2}{[\Delta(\ln c)]^2} = \frac{\left[\sum_{species\ n}\left(\frac{\partial \ln c}{\partial G_n}\right)\left(\frac{\partial G_n}{\partial G_j^{group}}\right)\right]^2 (\Delta G_j^{group})^2}{[\Delta(\ln c)]^2}\end{aligned} \tag{6.44}$$

These metrics can identify the rate rules and thermochemical groups that are the most influential in a kinetic model. It is difficult to perceive when a single rate rule is being overextended and used to estimate highly dissimilar reactions. Therefore, by coupling group-based parameter uncertainties in our analysis, we are more likely to uncover large output uncertainties that can lead to human flagging of those rate rules, which will allow the user to identify cases where multiple rate rules should replace a single rule.

The residual errors should not be entirely ignored, since they describe the error from approximating chemical similarity. However, we will not be addressing these errors in the correlated case in this chapter.

6.6 Global uncertainty propagation

Global uncertainty propagation was implemented using a probabilistic sampling method and polynomial chaos expansions (PCE) to express the dependence of output variables on the uncertainties $\Delta \ln(k_m)$ in kinetics and ΔG_n in thermodynamic inputs. We will consider the uncertain parameters to be independent and compare the global sensitivity indices against the local first-order sensitivity indices.

Given that chemical kinetic models are often very large, it was desired to find a method that gives the user flexibility in allocating CPU time for global uncertainty analysis. We use the adaptive Smolyak pseudospectral approximation method described in Conrad and Marzouk's separate paper. [15] The adaptive method is well suited to chemical kinetic models because of the strong coupling expected between certain inputs and weak dependence on others. Conrad and Marzouk show that using the adaptive method leads to much faster convergence in the uncertainty propagation of a methane-air autognition system,with more than two orders of magnitude reduction in error using the same number of model evaluations as compared to Monte Carlo. This method is implemented in the MIT Uncertainty Quantification Library (MUQ) [29] and is suitable for our needs. MUQ is an open-source code written by Marzouk and his uncertainty quantification group at MIT. It is written in C++ with solvers from SUNDIALS. [20] Many of the classes in MUQ have Python wrappers, making it highly compatible for use with RMG.

We consider a set of canonical random variables ξ_i to correspond to the uncertain inputs $i = 1, ..., d$ for $d = M + N$, where M is the total number of uncertain kinetic parameters and N is the total number of uncertain thermodynamic parameters considered. The $\xi_i's$ are independent and uniformly distributed over the unit interval

[-1, 1]. We define the mapping of ξ_i to the log uniform uncertainty intervals in the corresponding kinetics rate coefficient k_i:

$$\begin{aligned} f_i = \log_{10}\left(\frac{k_{i,max}}{k_{i,0}}\right) &= \frac{\sqrt{3}}{\ln 10}\Delta \ln k_i \\ \log_{10}\left(\frac{k_i}{k_{i,0}}\right) &= \xi_i f_i \\ k_i &= 10^{\xi_i f_i} k_{i,0} \end{aligned} \tag{6.45}$$

where f_i is the uncertainty factor associated with kinetics of reaction i, and $k_{i,0}$ is the nominal value of the rate coefficient.

For the free energy G_i, we define the linear mapping of ξ_i with respect to the nominal value $G_{i,0}$ and the uncertainty factor f_i for thermodynamics:

$$\begin{aligned} f_i = G_{i,max} - G_{i,0} &= \sqrt{3}\Delta G_i \\ G_i &= \xi_i f_i + G_{i,0} \end{aligned} \tag{6.46}$$

The PCE expressing the dependence of the output concentration c on the uncertain inputs $\xi = \{\xi_1, \xi_2, ..., \xi_d\}$ can be written as an infinite spectral expansion in terms of a complete set of orthogonal basis functions Ψ_k:

$$c(\xi) = \sum_{k=0}^{\infty} \alpha_k \Psi_k(\xi) \tag{6.47}$$

In our case, Ψ_k are multi-dimensional Legendre polynomials defined as:

$$(n+1)\Psi_{n+1}(\xi) = (2n+1)\xi\Psi_n(\xi) - n\Psi_{n-1}(\xi) \tag{6.48}$$

Legendre polynomials are orthogonal with respect to the uniform distribution of ξ (note that Hermite polynomials are used in the case of Gauss-normal ξ distributions):

$$\langle \Psi_i \Psi_j \rangle = \int_{-1}^{1} \Psi_i(x)\Psi_j(x)dx = \frac{2}{2n+1}\delta_{ij} \tag{6.49}$$

The α_k coefficients in the PCE for the output concentration c can be computed using the Galerkin projection:

$$\alpha_k = \frac{\langle c\Psi_k \rangle}{\langle \Psi_k^2 \rangle} \tag{6.50}$$

Computationally, the PCE is truncated to a fixed number of terms up to $k = P$ as an approximation of the true dependence on the uncertainties:

$$c(\xi) \approx \sum_{k=0}^{P} \alpha_k \Psi_k(\xi) \tag{6.51}$$

Once the coefficients in the PCE are known, we can directly evaluate the moments, i.e. expected value and variance of the output variable. Key metrics for evaluating the influence of the uncertain parameters are the global main effect (first order) and the total-effect sensitivity indices. [6] To compute these, we first express the total variance on the output concentration c in terms of the first and higher order effects of the uncertain inputs:

$$Var(c) = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + ... + V_{12...d} \tag{6.52}$$

V_i is the contribution to output variance of c measuring the effect of varying an uncertain paramter λ_i alone, averaged over the variations in the other input parameters:

$$\begin{aligned} V_i &= Var_{\lambda_i}(E_{\lambda_{\sim i}}(c|\lambda_i)) \\ S_i &= \frac{V_i}{Var(c)} \end{aligned} \tag{6.53}$$

where $\lambda_{\sim_i}$ indicates the set of all λ except λ_i. V_{ij} and higher order terms describe the variance attributed to interaction effects between uncertain inputs. Using the first order variance contribution V_i, the main effect, or first order, sensitivity index can be computed (and is directly comparable to the first order sensitivity indices in local uncertainty analysis):

$$S_i = \frac{V_i}{Var(c)} \tag{6.54}$$

Computationally, obtaining a total effect sensitivity index can be more accurate when a small number of samples are taken. The total effect sensitivity index S_{Ti} measures the contribution of varying λ_i, including all variance contributed by its interactions with other uncertain inputs λ:

$$S_{Ti} = \frac{E_{\lambda_{\sim i}}(Var_{\lambda_i}(c|\lambda_{\sim i}))}{Var(c)} \tag{6.55}$$

It must be noted that the total-effect sensitivity indices typically do not sum to 1, because it double counts interaction effects between λ_i and λ_j.

6.6.1 Implementation

We use the nonintrusive stochastic collocation method, adaptive Smolyak pseudospectral approximations, to compute the approximate α_k values. The schematic for generating global variance statistics is visualized in Figure 6.2. First, the inputs ξ_i are sampled within their probability distributions and mapped to the appropriate kinetics k_i and thermodynamic G_i values. Then a sequence of independent simulations are performed in a Cantera class reactor to give the output mole fractions for species of interest. The α_k values are then computed by MUQ. In the adaptive Smolyak method, the sampling of the input uncertainties is conducted using a sparse grid, and the uncertainty space is sampled adaptively to minimize global approximation error to efficiently compute α_k.

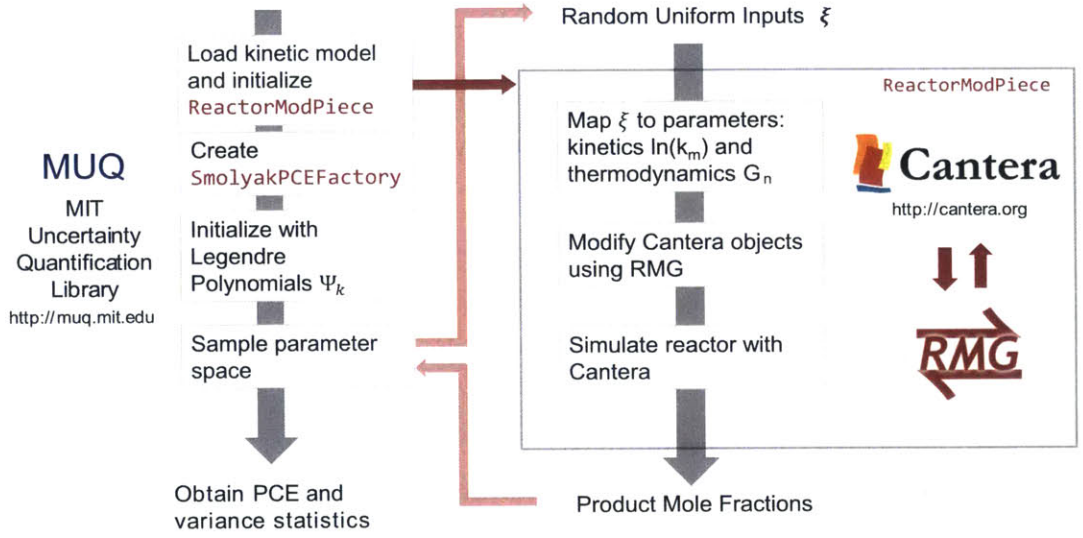


Figure 6.2: Schematic of the global uncertainty analysis workflow involving three key packages: MUQ, Cantera, and RMG.

The Cantera simulation of the reactor system is fully embedded in a constructed `ReactorModPiece` class, which inherits from the `libmuqmodeling.ModPiece` module. The class is initialized with a `rmgpy.tools.CanteraModel.Cantera` object containing the reactor conditions and an initialized RMG model, a list of desired output species mole fractions, and a list of the uncertain kinetics and thermo parameters to be used. This flexibility allows the user to select only the most influential parameters to perform the computation intensive global analysis. The `ReactorModPiece` class contains an `EvaluateImpl` function which takes an argument `ins`, a vector of the random uniform uncertain inputs $\xi = \{\xi_1, \xi_2, ..., \xi_d\}$ and returns the output mole

fractions of the user-specified output species from the Cantera simulation. These uncertain inputs are mapped to their appropriate values using Equations 6.45 and 6.46 and updated in the `Cantera` object prior to simulation.

To perform the adaptive PCE construction, a `libmuqApproximation.SmolyakPCEFactory` object is created outside of the `ReactorModPiece` class. This object is initialized with the set of legendre polynomials (`libmuqUtilities.LegendrePolynomials1DRecursive` objects) based on the number of uncertain inputs with Gauss-Patterson quadrature (`libmuqApproximation.GaussPattersonQuadrature1D` objects).

The MUQ library provides several options for constructing PCEs: (1) with fixed order P, (2) until a heuristic global error tolerance criteria is met, or (3) with a user-specified wall clock time. These options are given to the user for performing global uncertainty propagation on the kinetic model. After the PCE construction is completed, the user can evaluate the output at a given input test point, and the output variance, as well as the global main effect and total effect sensitivity indices can be computed directly from the PCEs.

This module can be used to run uncertainty analysis for a multitude of different targets and is specifically meant to be used *a posteriori* after local uncertainty analysis to reduce the number of input parameters investigated. Computation time is a large concern for global uncertainty methods, especially given the size of RMG-generated models, which often contain hundreds of species and tens of thousands of reactions. Therefore, using a small subset of uncertain parameters is recommended for global analysis.

A future workflow might include an automatic screening of parameters through local uncertainty analysis, such as screening via the Morris Method used by Tomlin *et al.* , to eliminate noninfluential inputs prior to global analysis. [2]

6.7 Case study using a toy model

The toy model used to compare uncertainty propagation methods is an intermediate, truncated phenyldodecane decomposition model used in developing the final mechanism in Chapter 5. This model was generated at 350 °C and 350 bar using a phenyldodecane mole fraction of 1.0 and a reaction time of 72 hours, with a limited number of reaction families (`H_Abstraction`, `R_Recombination`, `R_Addition_MultipleBond`, and `Disproportionation`). No reaction libraries were used during generation, so the reactions are all based on rate rule estimates. A single thermo library was used, such that some of the species thermochemistry was assigned from this library with trusted

values. The model contains 81 species and 1427 reactions. The undecene product is overpredicted by RMG in comparison to experimental and literature values, so uncertainty propagation was performed for the output concentration of undecene.

Although the model contains 1427 reactions, the overall number of rate rules accounted for is two orders of magnitude smaller, consisting of 14 rate rules from `H_Abstraction`, 6 rate rules from `R_Addition_MultipleBond`, 7 rate rules from `R_Recombination`, and 13 rate rules from `Disproportionation`, giving a total of only 40 rate rules used to estimate the kinetics in the entire model!

Uncertainties of $\Delta G_n = 1.0$ kcal/mol were assigned to the species thermochemistry taken from the thermo library. Uncertainties for thermochemistry estimated via Benson group additivity were assigned their uncertainties based on the contributing thermo groups. The model contains 81 species; however, only 18 distinct group additivity values are used to estimate their thermochemistry, along with 17 species with thermochemistry values assigned with thermo library values. Out of these group additivity values, 3 groups used placeholder, zero-value corrections: the ring group `Ring`, the group `R`, and the gauche correction `CsOsCdSs`, which likely indicates the lack of specific data for the corresponding functional group on the species.

6.7.1 Sensitivity analysis

RMG's native sensitivity analysis module was used to identify the most sensitive rate coefficients and species thermochemistry. The top 5 normalized sensitivities with respect to kinetics at 72 hours of reaction time are plotted in Figure 6.3. The top 5 semi-normalized sensitivities with respect to species thermochemistry are plotted in Figure 6.4.

6.7.2 Local uncertainty analysis

Using the assigned uncertainties for the rate coefficients and thermochemistry, local uncertainty propagation was performed considering the parameter uncertainties to be independent. The total variance in the undecene concentration was $[\Delta ln(c)]^2 = 1.47$. The first order sensitivity indices, or contribution of variance from each parameter was computed. The top 5 reactions contributing uncertainty to the undecene yield are shown in Figure 6.5. The top 5 species thermochemistry contributing to the uncertainty are shown in Figure 6.6.

Comparing against sensitivity analysis alone, local uncertainty analysis provides much more information. After accounting for the parameter uncertainties, we see that the order of the most influential parameters shifts between sensitivity vs. local uncertainty analysis. This weights highly uncertain and highly sensitive parameters,

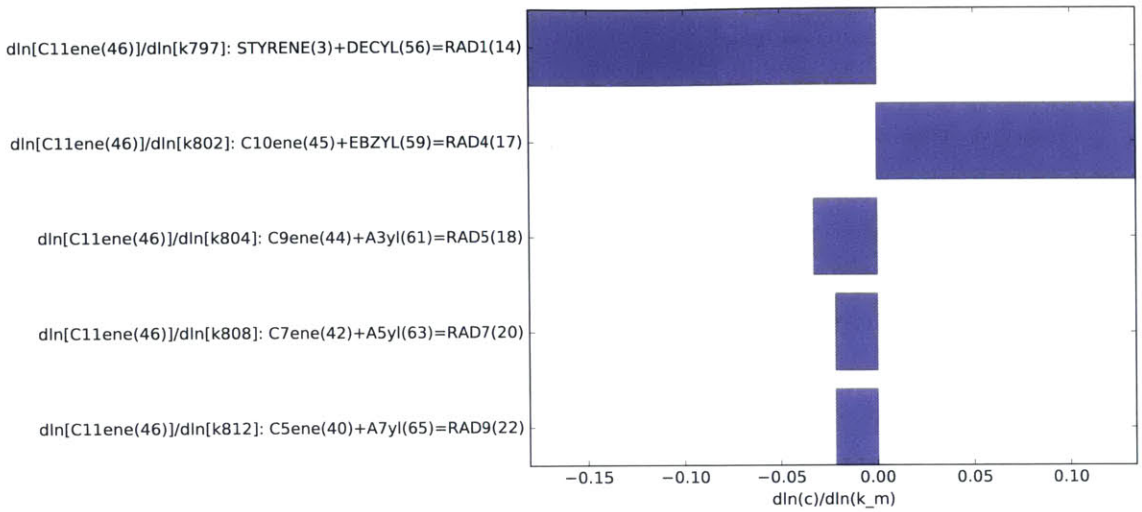


Figure 6.3: Top 5 most sensitive reaction kinetics affecting the concentration of undecene at 72 hours of reaction time.

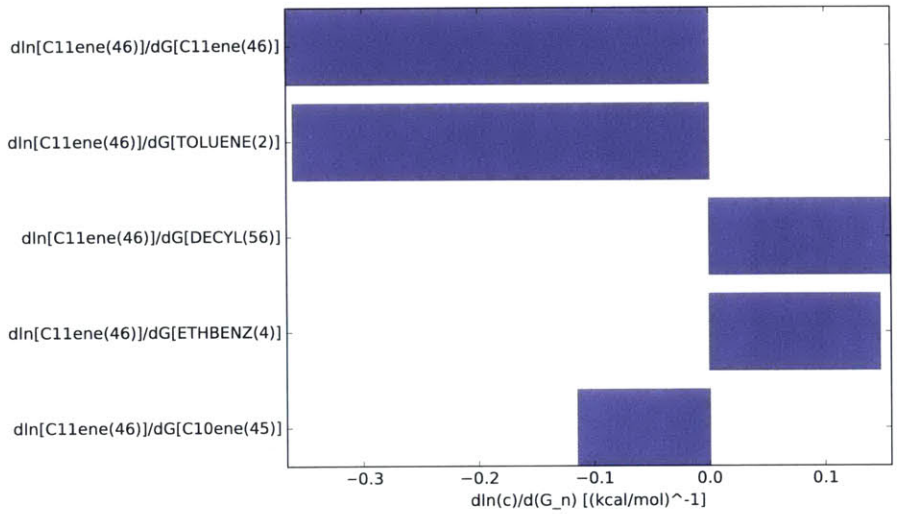


Figure 6.4: Top 5 most sensitive species thermochemistry affecting the concentration of undecene at 72 hours of reaction time.

letting the modeler quickly identify which parameter need to be improved first. In addition, the output concentration variance can provide an error bound on the model itself, which is useful in comparisons against experimental data. Finally, we see that when the uncertainties of both kinetics and thermo are accounted for, thermochemistry uncertainties dominate and affect the undecene yield more than kinetics in this toy model. In such cases, modelers should pay attention to refining thermochemical

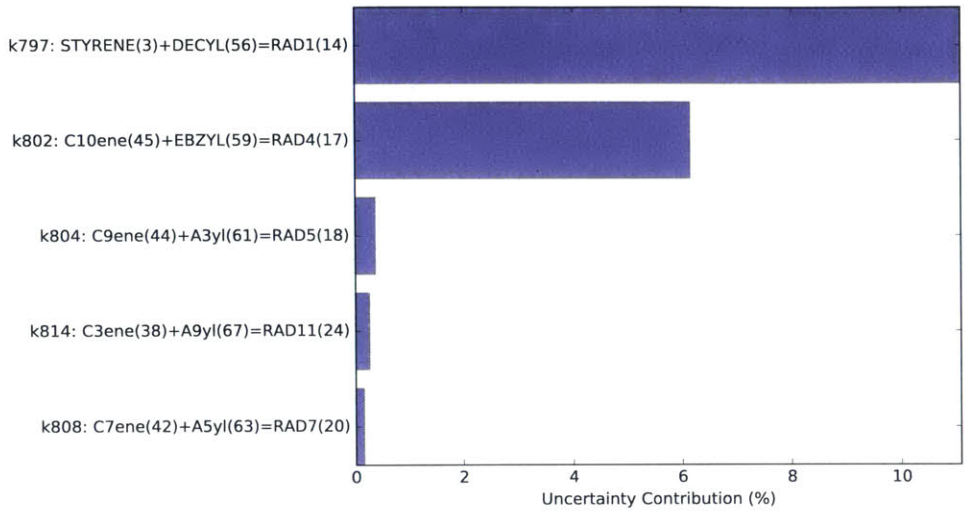


Figure 6.5: Top 5 reaction kinetics contributing to local uncertainty in the undecene concentration at 72 hours of reaction time.

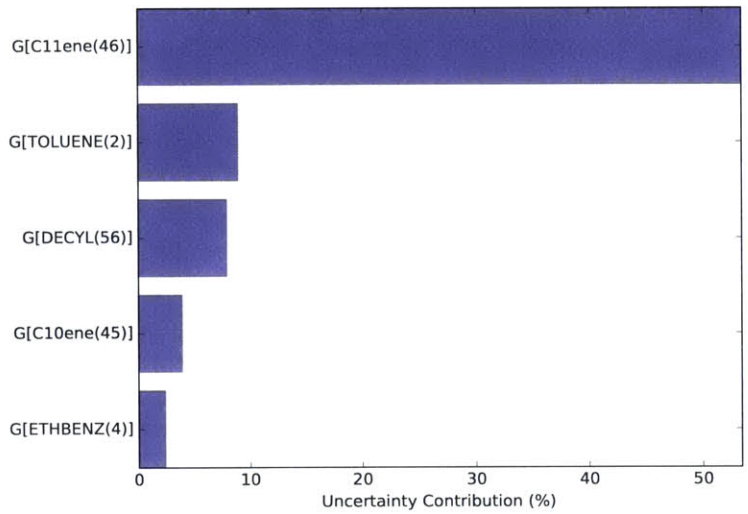


Figure 6.6: Top 5 species thermochemistry contributing to local uncertainty in the undecene concentration at 72 hours of reaction time.

parameters in addition to refining rate coefficients.

6.7.3 Correlated local uncertainty analysis

We now consider the case of correlations between the kinetics and thermodynamics, based on their construction through rate rules and thermo groups, respectively. The model contains 81 species and 1427 reactions, but these parameters collapse to only

40 independent rate rules and 35 independent thermochemical parameters (18 thermo groups and 17 thermo library values). There are, however, a full set of independent residual errors Δ_m in the kinetics and Δ_n in the thermochemistry. In this study, $\Delta_n = 0$, and Δ_m is set using the values and definitions from Section 6.3.1. The correlations drastically reduce the degrees of freedom in the model and also reduce the amount of expected variance $[\Delta ln(c)]^2 = 0.47$ in the output variable through error cancellation, as compared to considering the parameters to be fully independent $[\Delta ln(c)]^2 = 1.47$.

The largest first order sensitivity indices, or uncertainty contributions, for the rate rules are plotted in Figure 6.7 and for the thermochemical groups in Figure 6.8.

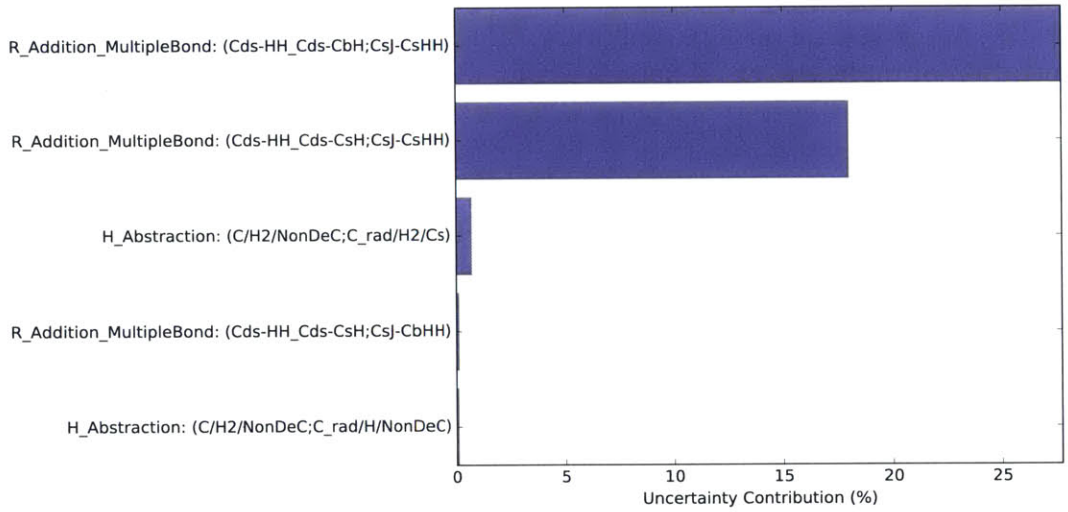


Figure 6.7: Top 5 rate rules contributing to uncertainty in the concentration of undecene at 72 hours of reaction time.

The most influential reaction kinetics contributed just over 10% of the total variance in the undecene concentration in the independent parameter consideration, while in the case of correlated parameters, we see that two of the `R_Addition_MultipleBond` rate rules dominate. This shows that even with cancellation effects, these two rate rules have very strong effects on the output concentration. We also see that although thermochemistry contributes greatly to uncertainty in the independent parameter consideration, they have less influence in the correlated case due to cancellation effects between the thermo groups. The results of the correlated analysis lead to very different conclusions from that of local analysis and provide a more accurate representation of the uncertainty propagation in group-based kinetic models.

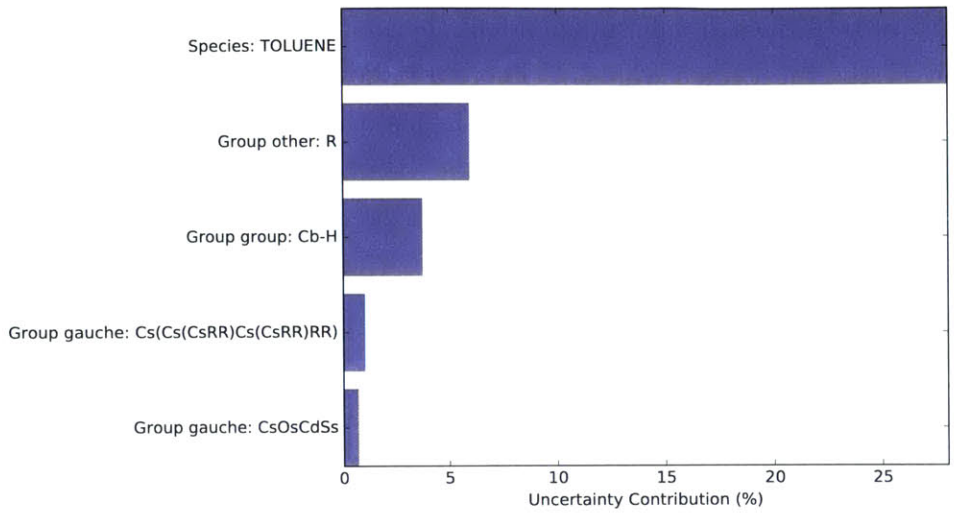


Figure 6.8: Top 5 thermo groups or library values contributing to uncertainty in the concentration of undecene at 72 hours of reaction time.

6.7.4 Global uncertainty analysis

Global uncertainty analysis was performed using the toy model and the uncertainty module utilizing functions from Cantera, RMG, and MUQ. Here, we consider the kinetics and thermodynamic parameters to be independent. Only the most influential parameters were used in the global analysis in order for the computation to be completed within a reasonable time and accuracy. The top 2 most influential rate coefficients and top 3 most influential species thermochemistry on the concentration of undecene were selected for the analysis. The parameters were sampled as a set of uniform distributions within their error bounds with the same input uncertainties as used in local uncertainty analysis.

The PCEs were constructed using Legendre polynomials with a Gauss-Patterson quadrature. The adaptive Smolyak pseudospectral method was used to construct the set of PCEs for describing the output concentration of undecene. A wall time of 30 minutes was chosen to perform the global UQ analysis (more accurate PCEs can be constructed by running the `SmolyakPCEFactory` for a longer time or with tolerance criteria.)

Global uncertainties for phenyldodecane and undecene were computed after propagation of uncertainties in the 5 parameters selected earlier. The results are shown in Table 6.2 and compared to the values derived from local uncertainty analysis when considering the same 5 parameter uncertainties.

The relative variance in the mole fraction x can be compared to $\Delta ln(c)$ used in

Table 6.2: Global vs. local output uncertainty for phenyldodecane and undecene in toy model

Species	Global Mean Mole Fraction	Local Nominal Mole Fraction	Global $[\Delta ln(c)]^2$	Local $[\Delta ln(c)]^2$
PDD(1)	0.187	0.146	0.58	1.64
C11ene(46)	0.141	0.108	0.53	1.28

local uncertainty analysis as they are equivalent when total concentration is fixed ($c_{total} = P/RT$ in a constant temperature and pressure system).

$$\begin{aligned} x &= \frac{c}{c_{total}} = \frac{cRT}{P} \\ \frac{\Delta x}{x_0} &= \frac{\Delta \frac{cRT}{P}}{\frac{c_0 RT}{P}} = \frac{\Delta c}{c_0} = \Delta ln(c) \end{aligned} \tag{6.56}$$

We see that there are drastic differences in the results from global uncertainty propagation as compared to local uncertainty analysis. The variance computed by local analysis is much higher than in the global case due to the nonlinearities in the system and the coupling of reactions. A linear increase in a rate coefficient will not change the output concentration proportionally due to the dependence on other parameters and species intermediates' concentrations. We also see that after sampling from the entire uncertainty space, the mean mole fraction computed in the global analysis differs significantly from the nominal mean fractions in local uncertainty analysis. These again point to the high degree of nonlinearity in the system and the inaccuracies of local uncertainty propagation.

The global first order sensitivity indices S_i and total effect indices $S_{T,i}$ were computed automatically for the phenyldodecane and undecene species with respect to the 5 parameters selected. The results are presented and compared to the first order sensitivity indices computed using local analysis in Table 6.3 and 6.4

Similar to local uncertainty analysis for independent parameters, global analysis shows that the influence of species thermochemistry on the undecene concentration is much greater than that of the rate coefficients. The thermodynamic parameters' influence is greater in the global case than in the local case. In addition, the global first order sensitivity indices vary widely from their corresponding values in the local case. In addition, the global first order sensitivity indices sum up to less than 1, indicating coupling in the system.

Table 6.3: Global reaction sensitivity indices for toy model: phenyldodecane and undecene

Output Species	Reaction Kinetics	Global $S_{T,i}$ (%)	Global S_i (%)	Local S_i (%)
PDD(1)	STYRENE(3)+DECYL(56)=RAD1(14)	39.4	25.1	43.3
PDD(1)	C10ene(45)+EBZYL(59)=RAD4(17)	29.3	18.6	22.4
C11ene(46)	STYRENE(3)+DECYL(56)=RAD1(14)	3.5	1.5	12.7
C11ene(46)	C10ene(45)+EBZYL(59)=RAD4(17)	2.3	0.9	7.0

Table 6.4: Global thermochemistry sensitivity indices for toy model: phenyldodecane and undecene

Output Species	Species Thermochemistry	Global $S_{T,i}$ (%)	Global S_i (%)	Local S_i (%)
PDD(1)	C11ene(46)	24.7	15.1	2.3
PDD(1)	TOLUENE(2)	4.3	2.7	0.7
PDD(1)	DECYL(56)	28.2	16.1	31.1
C11ene(46)	C11ene(46)	86.3	82.4	61.0
C11ene(46)	TOLUENE(2)	12.0	9.1	10.2
C11ene(46)	DECYL(56)	2.5	0.9	9.1

6.8 Conclusion

Sensitivity analysis is the most commonly used tool for improving and refining chemical kinetic mechanisms. However, this tool relies on human judgment regarding the the uncertainty of individual rates. We have assigned uncertainties to RMG's reaction rate coefficients and species thermochemistry and implemented native sensitivity analysis in RMG. This enables us to perform local uncertainty propagation for chemical kinetic mechanisms considering both input parameters as independent as well as correlated through the kinetics rate rules and thermochemical groups.

Using local uncertainty propagation, modelers are easily able to visualize and isolate the most influential and uncertain input parameters affecting model output targets. Using a toy model for phenyldodecane, it was found that 81 species thermochemistry parameters are estimated using just 35 independent thermodynamic parameters, while 1427 reaction rate coefficients collapse to only 40 rate rule parameters. We show that considering the group-based uncertainties to be correlated drastically reduces the degrees of freedom in the model and substantially reduces the uncertainty in the outputs. We find that uncertainty contributions from the ther-

modynamic inputs are much smaller in the correlated case than in the independent case due to cancellation terms between reactants and products containing the same thermo groups. This shows the importance of propagating correlated rather than independent uncertainties in an automatically generated kinetic mechanism.

Finally, global uncertainty analysis was implemented through coupling RMG classes with MUQ, [29] the MIT Uncertainty Quantification library. Adaptive Smolyak pseudospectral approximation methods were used to sample the uncertain kinetic and thermodynamic input space and generate output uncertainty distributions for a toy model of phenyldodecane considering the input uncertainties to be independent. Preliminary results show that nonlinearities exist and that there can be a substantial difference in output error between global vs local first-order uncertainty propagation. Future work should focus on implementing group-based global uncertainty propagation to provide a truly accurate assessment of model error.

In summary, we show that uncertainty analysis is necessary for understanding the effects of correlated uncertainties in automatically generated models. These tools should effectively replace sensitivity analysis in the model construction workflow.

6.9 References

[1] J. Zador, I. G. Zsely, and T. Turanyi. "Local and global uncertainty analysis of complex chemical kinetic systems." *Reliab. Eng. Syst. Safe.* 91 (10-11), 2006, pp. 1232–1240. DOI: 10.1016/j.ress.2005.11.020.

[2] T. Ziehn and A. S. Tomlin. "A global sensitivity study of sulfur chemistry in a premixed methane flame model using HDMR." *Int. J. Chem. Kin.* 40 (11), 2008, pp. 742–753. DOI: 10.1002/kin.20367.

[3] L. Vuilleumier, R. A. Harley, and N. J. Brown. "First- and Second-Order Sensitivity Analysis of a Photochemically Reactive System (a Green's Function Approach)." *Env. Sci. Tech.* 31 (4), 1997, pp. 1206–1217. DOI: 10.1021/es960727g.

[4] D. L. Baulch, C. J. Cobos, R. A. Cox, C. Esser, P. Frank, T. Just, J. A. Kerr, M. J. Pilling, J. Troe, R. W. Walker, and J. Warnatz. "Evaluated Kinetic Data for Combustion modeling." *J. Phys. Chem. Ref. Data* 21, 1992, pp. 411–734. DOI: 10.1063/1.555908.

[5] H. Rabitz, M. Kramer, and D. Dacol. "Sensitivity Analysis in Chemical Kinetics." *Ann. Rev. Phys. Chem.* 34, 1983, pp. 419–461. DOI: 10.1146/annurev.pc.34.100183.002223.

[6] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Comp. Phys. Comm.* 181 (2), 2010, pp. 259–270. DOI: 10.1016/j.cpc.2009.09.018.

[7] H. N. Najm, B. J. Debusschere, Y. M. Marzouk, S. Widmer, and O. P. Le Maître. "Uncertainty quantification in chemical systems." *Int. J. Numer. Meth. Eng.* 80 (6-7), 2009, pp. 789–814. DOI: 0.1002/nme.2551.

[8] D. A. Sheen and H. Wang. "The method of uncertainty quantification and minimization using polynomial chaos expansions." *Combust. Flame* 158 (12), 2011, pp. 2358–2374. DOI: 10.1016/j.combustflame.2011.05.010.

[9] *CHEMKIN-PRO 15131.* Reaction Design: San Diego. 2013.

[10] C. Wang. "Parametric Uncertainty Analysis for Complex Engineering Systems." PhD thesis. Massachusetts Institute of Technology, 1999.

[11] T. Russi, A. Packard, and M. Frenklach. "Uncertainty quantification: Making predictions of complex reaction systems reliable." *Chem. Phys. Lett.* 499 (1-3), 2010, pp. 1–8. DOI: 10.1016/j.cplett.2010.09.009.

[12] *PrIMe Data Model.* 2011. URL: http://primekinetics.org/data_model.

[13] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, R. K. H. C. T. Bowman, S. Song, J. W. C. Gardiner, V. V. Lissianski, and Z. Qin. *GRI-Mech 3.0.* 2011. URL: http://www.me.berkeley.edu/gri_mech/.

[14] L. Petzold and S. Li. *DASPK3.1.* URL: http://www.cs.ucsb.edu/~cse/software.html.

[15] P. R. Conrad and Y. M. Marzouk. "Adaptive Smolyak Pseudospectral Approximations." *SIAM J. Sci. Comp.* 35 (6), 2013, A2643–A2670. DOI: 10.1137/120890715.

[16] H.-H. Carstensen, A. M. Dean, and O. Deutschmann. "Rate constants for the H abstraction from alkanes (R-H) by radicals: A systematic study on the impact of R and R'." *Proc. Combust. Inst* 31 (1), 2007, pp. 149–157. DOI: 10.1016/j.proci.2006.08.091.

[17] S. M. Villano, L. K. Huynh, H.-H. Carstensen, and A. M. Dean. "High-Pressure Rate Rules for Alkyl + O2 Reactions. 2. The Isomerization, Cyclic Ether Formation, and *beta*-Scission Reactions of Hydroperoxy Alkyl Radicals." *J. Phys. Chem. A* 116 (21), 2012, pp. 5068–5089. DOI: 10.1021/jp3023887.

[18] T. Nagy and T. Turanyi. "Uncertainty of Arrhenius Parameters." *Int. J. Chem. Kin.* 43 (7), 2011, pp. 359–378. DOI: 10.1002/kin.20551.

[19] S. Petway. "Uncertainty Analaysis in Automatic Reaciton Mechanism Generation: Neopentyl + O2." MA thesis. Massachusetts Institute of Technology, 2003.

[20] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. "SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers." *ACM Trans. Math. Software* 31 (3), 2005, pp. 363–396. DOI: 10.1145/1089014.1089020. URL: http://computation.llnl.gov/casc/sundials/.

[21] J. Tolsma and P. I. Barton. *DAEPACK*. 2000. URL: https://yoric.mit.edu/daepack.

[22] P. N. Brown, A. C. Hindmarsh, and L. R. Petzold. *DASPK 2.0*. 1995.

[23] J. W. Allen and C. W. Gao. *PyDAS: A Python wrapper for the DASSL, DASPK, and DASKR differential algebraic system solvers*. 2016. URL: https://github.com/ReactionMechanismGenerator/PyDAS.

[24] L. R. Petzold. *Differential Algebraic System Solver (DASSL)*. 1982. URL: http://www.cs.ucsb.edu/~cse/software.html.

[25] R. Bradshaw, S. Behnel, D. S. Seljebotn, G. Ewing, et al. *The Cython compiler*. URL: http://cython.org/.

[26] Argonne National Laboratory and Rice University. *ADIFOR 2.0: Automatic Differentiation of Fortran*. 1998. URL: http://www.mcs.anl.gov/research/projects/adifor/.

[27] J. Prager, H. N. Najm, K. Sargsyan, C. Safta, and W. J. Pitz. "Uncertainty quantification of reaction mechanisms accounting for correlations introduced by rate rules and fitted Arrhenius parameters." *Combust. Flame* 160 (9), 2013, pp. 1583–1593. DOI: 10.1016/j.combustflame.2013.01.008.

[28] H. Wang and D. Sheen. "Combustion kinetic model uncertainty quantification, propagation and minimization." *Prog. Energy Combust. Sci.* 47, 2015, pp. 1–31. DOI: 10.1016/j.pecs.2014.10.002.

[29] P. R. Conrad, M. D. Parno, A. D. Davis, and Y. M. Marzouk. *MIT Uncertainty Quantification Library (MUQ)*. 2016. URL: http://muq.mit.edu/.

7

Recommendations for future work

My thesis has contributed several methodologies for constructing high fidelity kinetic models automatically. These contributions are available to the kinetics community through the RMG software package. To put these methodologies into practice, I have highlighted three combustion related processes modeled by RMG, highlighting our software's capability to model complex physical phenomena. This chapter discusses a number of challenges that remain in achieving the vision of predictive kinetics.

7.1 Improving the RMG database

We have reviewed the major pitfalls in RMG's thermochemistry and kinetics estimation schemes. These errors largely stem from an incomplete database and is a general problem for building computer-generated models. As with all group-based estimation schemes, we rely on the assumption of a densely populated database. In reality, the RMG database is sparsely populated and also biased. The database is filled by the users, most of whom are Green Group members building models for specific species. Subsequent users of RMG find a database populated with values that are relevant for classes of molecules previously modeled but generally incomplete for new classes of molecules. With uncertainty propagation in Chapter 6, it is easier to identify rate rules that need refinement. However, the process of adding new data must be streamlined for ease of use and transparency. Doing so will accelerate model development for new classes of molecules.

7.1.1 Infrastructure for an open access database

Currently, the RMG database can be visualized and searched on the RMG interactive website. The website also displays sources for each entry, i.e. the literature it was extracted from. The next step for improving user assessibility is to create new online infrastructure for adding new entries to the database. A wealth of theoretical and

experimental kinetic rates can already be found in literature but are difficult to add to RMG. Even with the addition of new scripts for translating reaction libraries to training reactions that fit kinetics rate rules, it is difficult and time consuming for most users to transcribe tables of rates into the Python database format. Therefore, new scripts should be added to facilitate the translation of real reaction data. The RMG website was previously capable of allowing registered users to add training reactions, which automatically made commits to Github. However, there has been insufficient upkeep and maintainance of this workflow. We must take steps to reduce the tedium and error associated with adding new entries and build better architecture for an open access database. To improve the predictive capability within RMG, the database must grow while maintaining transparency for the quantum chemists, experimentalists, and theorists working with our software.

7.1.2 An automated workflow for adding data from quantum chemistry calculations

To add all the rate rules would be a Herculean task. – Joshua W. Allen

For molecules with unexplored chemistries (i.e. cineole), associated thermochemistry and rate rules are difficult to find in literature. In this case, the fastest way to build the model is to compute new parameter values through quantum chemical calculations. However, adding these to a single library is not useful for extrapolating these values to other similar classes of molecules. We would prefer to add data to the thermodynamic groups and kinetics rate rules. RMG should borrow ideas from machine learning and automatically fit rate rules and groups, keeping a set of training data which retains original information. This will give the estimation schemes and group definitions the flexibility to change and be retrained at will.

For thermochemistry, on-the-fly computations can be performed for cyclic compounds through the QMTP interface. [1] However, these computations take substantially more time than group additivity and are performed using semi-empirical methods, which have higher error bars than *ab initio* methods. In addition, these calculations cannot be performed for non-cyclics in the current workflow. Ideally, we want to derive a robust set of thermodynamic group additivity values based on high fidelity quantum calculations. Currently, RMG has nearly all the capabilities to perform *ab initio* calculations through the `rmgpy.qm` module, as well as the capability to interpret those values through `rmgpy.cantherm`. By extending the functionality to perform *ab initio* calculations on-the-fly, and using them to fit new GAVs (such as for polycyclics) in the thermodynamic database, we can systematically improve

estimates for a much broader set of molecules. Furthermore, parallelization of the quantum calculations using the Python `scoop` package can lead to greater speed gains.

For kinetics, automatic transition state searches within RMG have been in development by Bhoorasingh and West. [2] Transition states geometry predictions for hydrogen abstraction reactions have been successfully constructed through 3D distance geometries generated using a group-additive method. Other automatic transition search software independent of RMG have also surfaced, such as the KinBot code by Zador [3] and the discovery of elementary chemical reactions using freezing string and berny optimization methods by Suleimanov. [4] The growing availibility of such codes gives us the opportunity to automate the generation of new training reactions from quantum calculations to fill up the kinetics database. It is up to future RMG developers to harness these new cheminformatics tools to fully utilize the raw computing power of supercomputers such as NERSC for running calculations.

7.2 Aromatic chemistry and recognition of PAH pathways

Aromatic chemistry is a key area for improvement in RMG. Recently, benzene, indene and naphthalene pathways from cyclopentadiene have been calculated by Vandeputte and Merchant [5] and have been added as training reactions in RMG. However, several problems prevent RMG from handling aromatics correctly, namely aromaticity representation and representation, and aromatic-specific kinetic parameters.

RMG is currently incapable of recognizing aromatics other than 6-membered benzene rings. Aromaticity perception is handled through RDKit [6]; however, conversion to aromatic atomtypes `Cb` is only done for 6-membered benzene rings. This means that rings containing heteroatoms such as furan or thiophene cannot be recognized in RMG. Consequently, representatations of naphthalene and higher order PAH's are problematic and can only be performed in their kekulized (single-double bond) form. An additional complication arises from resonance isomer generation for aromatics, which becomes unwieldy for higher order PAHs and radicals for which there exist many kekulized forms.

Finally, reactions are assigned through the reaction families, which are often acting on the kekulized isomers constituting an aromatic species, rather than its aromatic form. Therefore the problem for assigning rate coefficients is two-fold: kekulized forms do not represent aromatics accurately because they do not describe delocalized electronic structures, and the rate rules assigned to these kekulized forms are also generally derived from non-aromatic kinetic rates.

To mitigate these issues going forward, the representation of aromatics in RMG must be completely revamped. Generation of kekulized forms should be prevented, and new, isolated aromatic reaction families should be created separately from the original reaction families, or the aromatic rate rules should be segregated in their original families from those describing non-aromatic rates.

7.3 Lumped kinetic models

Detailed chemical kinetics are a powerful tool for understanding pyrolysis and oxidation chemistry. However, as molecules get larger, as seen in modeling PDD in Chapter 5, the number of species and reactions begin to grow rapidly due to the large number of reaction sites. These can lead to both CPU speed and memory issues in RMG, which can be partially remedied using methods such as reaction filtering, pruning or parallelization, but only to a certain degree. For molecules larger than PDD or mixtures (i.e. heavy crude, kerogen, etc.), knowledge of the precise chemical structure may be unknown, and lumping is necessary. A scheme for lumping species together based on their reactivity or chemical structure similarity is needed to proceed in modeling such chemistries. [7] For truly complex molecules and mixtures, post-processing model reduction methods are insufficient, and lumping on-the-fly is required to generate kinetic models for these chemistries and aid analytical chemistry analysis.

7.4 Robust model generation under uncertainty

In the rate-based enlarging algorithm, the nominal values of the input parameters (initial species concentration, thermodynamics, rate coefficients) determine which species enter the model core and which pathways are explored. Thus uncertainty in RMG's input parameters influences not only the output variables, but also the model structure itself. Error in a rate parameter used during the iterative model generation process may lead to "truncation error": exclusion of an important core species and its subsequent daughter reactions, outside the bounds of the model's tolerance specifications.

Formally, the rate-based termination rule in RMG can be expressed as

$$R_i(t, \boldsymbol{\theta}_0) < \epsilon R_{char}(t, \boldsymbol{\theta}_0) \qquad \text{for all edge species i} \tag{7.1}$$

where t is time, $\boldsymbol{\theta}_0$ is the vector of input parameters at their nominal values, $R_i(t)$ is the formation of edge species i, and $R_{char}(t, \boldsymbol{\theta}_0)$ is characteristic from all core species j, $R_{char}(t, \boldsymbol{\theta}_0) = \sqrt{\sum_j R_j^2(t, \boldsymbol{\theta}_0)}$.

Since the termination criteria applies only for the nominal parameter values, the criteria may not be satisfied over the whole range of parameter uncertainties. In order for the model to be valid over the entire parameter range, the following criteria must be satisfied:

$$R_i(t, \boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}) < \epsilon R_{char}(t, \boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}) \qquad \text{for all edge species i} \tag{7.2}$$

Jing Song performed some preliminary work describing an algorithm for generating robust models which satisfy this criteria for a user specified range of concentrations. [8] In Song's method, the constraint in Equation 7.2 defines a characteristic polyhedron for $\Delta\boldsymbol{\theta}$ in parameter space when the terms in the inequality are expanded about the nominal parameter point $\boldsymbol{\theta}_0$ using a first order Taylor approximation. The characteristic polyhedron can then be bounded with a unique hyper-rectangle Γ for the parameter range using feasibility analysis. [8]

$$\Gamma = \{\Delta\boldsymbol{\theta} | -d\Delta\boldsymbol{\theta}^- < \Delta\boldsymbol{\theta} < d\Delta\boldsymbol{\theta}^+\} \tag{7.3}$$

The algorithm describes a tighter practical tolerance ϵ_p to grow the model such that the original tolerance criteria ϵ is satisfied within the entirety of the uncertain parameter space. This could be implemented in RMG in order to generate robust models under uncertainty by tightening tolerances automatically based on the magnitude of the input uncertainties.

Alternatively, first order sensitivity coefficients can be computed on-the-fly through the PyDAS [9] backend during model generation, and a heuristic method for exploring uncertain and sensitive reaction pathways could be employed.

7.5 References

[1] G. R. Magoon and W. H. Green. "Design and implementation of a next-generation software interface for on-the-fly quantum and force field calculations in automated reaction mechanism generation." *Comp. & Chem. Eng.* 52, 2013, pp. 35–45. DOI: 10.1016/j.compchemeng.2012.11.009.

[2] P. L. Bhoorasingh and R. H. West. "Transition state geometry prediction using molecular group contributions." *Phys. Chem. Chem. Phys.* 17, 2015, pp. 32173–32182. DOI: 10.1039/C5CP04706D.

[3] J. Zador and H. N. Najm. *Automated exploration of the mechanism of elementary reactions.* Tech. rep. SAND2012-8095. Sandia National Laboratories, 2012.

[4] Y. V. Suleimanov and W. H. Green. "Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods." *Journal of Chemical Theory and Computation* 11 (9), 2015, pp. 4248–4259. DOI: 10.1021/acs.jctc.5b00407.

[5] S. S. Merchant. "Molecules to Engines: Combustion Chemistry of Alcohols and Their Application to Advanced eEngines." PhD thesis. Massachusetts Institute of Technology, 2015.

[6] G. Landrum. *RDKit.* URL: http://rdkit.org.

[7] E. Ranzi, M. Dente, A. Goldaniga, G. Bozzano, and T. Faravelli. "Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures." *Prog. Energy Combust. Sci.* 27 (1), 2001, pp. 99–139. DOI: 10.1016/S0360-1285(00)00013-7.

[8] J. Song, G. Stephanopoulos, and W. H. Green. "Valid parameter range analyses for chemical reaction kinetic models." *Chem. Eng. Sci.* 57 (21), 2002, pp. 4475–4491. DOI: 10.1016/S0009-2509(02)00416-5.

[9] J. W. Allen and C. W. Gao. *PyDAS: A Python wrapper for the DASSL, DASPK, and DASKR differential algebraic system solvers.* 2016. URL: https://github.com/ReactionMechanismGenerator/PyDAS.