Predictive Chemical Kinetics: Enabling Automatic Mechanism Generation and Evaluation

by

Joshua W. Allen

M. S. Chemical Engineering Practice
Massachusetts Institute of Technology, 2009
B. Chemical Engineering
University of Minnesota, 2007
B. S. Chemistry
University of Minnesota, 2007

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 2013

Author .....
Department of Chemical Engineering
May 1, 2013

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

Accepted by...
.....
Patrick S. Doyle
Professor of Chemical Engineering
Chairman, Committee for Graduate Students

Predictive Chemical Kinetics: Enabling Automatic Mechanism Generation and Evaluation

by

Joshua W. Allen

Submitted to the Department of Chemical Engineering
on May 1, 2013 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical
Engineering

Abstract

The use of petroleum-based fuels for transportation accounted for more than 25% of the total energy consumed in 2012, both in the United States and throughout the world. The finite nature of world oil reserves and the effects of burning petroleum-based fuels on the world's climate have motivated efforts to develop alternative, renewable fuels. A major category of alternative fuels is biofuels, which potentially include a wide variety of hydrocarbons, alcohols, aldehydes, ketones, ethers, esters, *etc.* To select the best species for use as fuel, we need to know if it burns cleanly, controllably, and efficiently. This is especially important when considering novel engine technologies, which are often very sensitive to fuel chemistry. The large number of candidate fuels and the high expense of experimental engine tests motivates the use of predictive theoretical methods to help quickly identify the most promising candidates.

This thesis presents several contributions in the areas of predictive chemical kinetics and automatic mechanism generation, particularly in the area of reaction kinetics. First, the accuracy of several methods of automatic, high-throughput estimation of reaction rates are evaluated by comparison to a test set obtained from the NIST Chemical Kinetics Database. The methods considered, including the classic Evans-Polanyi correlation, the "rate rules" method currently used in the RMG software, and a new method based on group contribution theory, are shown to not yet obtain the order-of-magnitude accuracy desired for automatic mechanism generation. Second, a method of very accurate computation of bimolecular reaction rates using ring polymer molecular dynamics (RPMD) is presented. RPMD rate theory enables the incorporation of quantum effects (zero-point energy and tunnelling) in reaction kinetics using classical molecular dynamics trajectories in an extended phase space. A general-purpose

software package named RPMDrate was developed for conducting such calculations, and the accuracy of this method was demonstrated by investigating the kinetics and kinetic isotope effect of the reaction $OH + CH_4 \longrightarrow CH_3 + H_2O$. Third, a general framework for incorporating pressure dependence in thermal unimolecular reactions, which require an inert third body to provide or remove the energy needed for reaction via bimolecular collisions, was developed. Within this framework, several methods of reducing the full, master equation-based model to a set of phenomenological rate coefficients $k(T, P)$ are compared using the chemically-activated reaction of acetyl radical with oxygen as a case study, and recommendations are made as to when each method should be used. This also resulted in a general-purpose code for calculating pressure-dependent kinetics, which was applied to developing an *ab initio* model of the reaction of the Criegee biradical CH_2OO with small carbonyls that reproduces recent experimental results.

Finally, the ideas and techniques of estimating reaction kinetics are brought together for the development of a detailed kinetics model of the oxidation of diisopropyl ketone (DIPK), a candidate biofuel representative of species produced from cellulosic biomass conversion using endophytic fungi. The model is evaluated against three experiments covering a range of temperatures, pressures, and oxygen concentrations to show its strengths and weaknesses. Our ability to automatically generate this model and systematically improve its parameters without fitting to the experimental results demonstrates the validity and usefulness of the predictive chemical kinetics paradigm. These contributions are available as part of the Reaction Mechanism Generator (RMG) software package.

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

Acknowledgments

If I have seen further it is by standing on the shoulders of giants.

– Sir Isaac Newton

Science in the twenty-first century is no longer possible to conduct in a vacuum. Although it may not be possible in this limited space to fully express my gratitude to the many folks who have made my thesis research possible, I shall neverless endeavor to make the attempt.

Thank you to my advisor, Prof. William Green. You knew when to let me explore one of my crazy ideas, and when to give me a psychological kick in the rear to get things done. Your feedback has been nothing but helpful. There was never a day where I doubted that I had chosen the correct person to advise me on my graduate career.

Thank you to my thesis committee, Profs. Paul Barton and Bruce Tidor. Your thoughts and suggestions have greatly improved my research and the thesis that resulted.

Thank you to all of the RMG developers I have had the pleasure of working with, including Richard, Sandeep, Michael, Greg, Amrit, Shamel, Connie, Nathan, and Nick. Your intelligence, dedication, and fortitude have made RMG a successful piece of software, a feat which is especially impressive considering that we are not formally trained as software engineers. Thanks in particular for always being willing to answer my questions, and for putting up with my crazy ideas and opinions regarding software development.

Thank you to my project collaborators in the greater scientific community. Thanks in particular go to those involved with the fungi biofuel codevelopment project, especially Drs. Craig Taatjes and Adam Scheer.

Thank you to the King Abdullah University of Science and Technology (KAUST) and the Combustion Energy Frontier Research Center (CEFRC) for providing the primary funding for my graduate research.

Thank you to all of the Green group members. It is said that a research group reflects the personality of the supervisor. Whether or not this is true, I have yet to meet another set of people so consistently friendly, helpful, genuine, and simply fun to be around. You all made coming to the office every day a pleasure, especially on days when the work was a struggle.

Thank you to Barb, Alison, and Gwen for all your wonderful support with the administrative aspects of my graduate life. With your warm enthusiasm and attention to detail, I know the Green group is in good hands.

Thank you to all of the great friends that I met in the MIT Department of Chemical Engineering. Whether competing on the athletic field, preparing for the holiday skits, or simply hanging out at a TG, you made me feel welcome and helped me get a break from my research. I was always impressed how, in every activity, you were eager to include everyone who wanted to participate, regardless of ability, and always ready with a word of encouragement.

Finally, thank you to my family. You have been nothing but supportive of me as I pursued my dream of getting a Ph.D.

Contents

List of Figures

List of Tables

1

INTRODUCTION

The use of petroleum-based fuels for transportation accounted for more than 25% of the total energy consumed in 2012, both in the United States and throughout the world [1, 2]. Although there have been a range of estimates of the available world oil reserves [3], what is not in doubt is that these reserves are finite. Since 2005, oil production has remained relatively unchanged at about 74 million barrels per day [4]; at this rate of consumption, even generous estimates of the world's oil reserves would be completely depleted within fifty years. However, the world energy consumption has increased at about 1.2% per year since 1970, and that rate is projected to continue through at least 2035, driven by increasing standard of living of populous developing countries, such as China and India [2]. The rising demand in the face of inelastic supply has caused the price of oil to fluctuate dramatically since 2005 [4]. This problem of supply motivates the search for alternatives to petroleum-based fuels.

The environmental impact of this petroleum consumption on the world climate is also a major concern. The executive summary of the 2012 *World Energy Outlook* states that, using current technologies, "No more than one-third of proven reserves of fossil fuels can be consumed prior to 2050 if the world is to achieve the 2 °C goal" of limiting global temperature rise [2]. Clearly, as we develop our alternative fuel infrastructure to affordably and sustainably satisfy the rising demand, we must also seek to minimize the environmental effects. This will require a significant, multifaceted effort [5, 6].

Efforts to develop alternatives to petroleum-based fuels fall into several categories. There are several ways to utilize other fossil fuels, including coal-to-liquids, gas-to-liquids, compressed natural gas, and liquefied natural gas. Many of these technologies may be cheaper than petroleum-based fuels today, and many also burn more cleanly; however, their basis in fossil fuels means that their supply is also finite. Another approach is to design vehicles to run on electricity, which enables the decoupling

of the energy generation problem from the energy storage problem. However, if we still utilize fossil fuels to generate the electricity, then the actual improvements in fuel efficiency and emissions are largely negated. Furthermore, current electric vehicles are hampered by reduced range, lengthy recharges, and missing infrastructure. Similar challenges exist for using hydrogen for energy storage instead of electricity.

The last major category of alternative fuels focuses on renewable resources, of which the leading category are biofuels. The first generation of biofuels, mainly ethanol and biodiesel, are produced from food stocks such as grains, vegetable oils, and animal fats, which couples food and fuel prices. There is also not enough farmland available to produce enough ethanol and biodiesel to serve as a significant substitute for petroleum-based fuels [7]. For these reasons, most current research focuses on conversion of cellulosic biomass to biofuel, either biologically using various microorganisms or thermochemically using traditional chemical processing methods. These methods yield a wide variety of chemical species, including hydrocarbons, alcohols, aldehydes, ketones, ethers, esters, *etc.*, that might be viable as fuels. To select the best species for use as a fuel, we need to know if it burns cleanly, controllably, and efficiently. This is especially important when considering novel engine technologies – such as those based on compression ignition – which are often very sensitive to fuel chemistry [8, 9].

1.1 Predictive chemical kinetics

An experimental test of one fuel in one engine is quite expensive, often costing tens to hundreds of thousands of dollars and requiring a significant amount of that fuel be produced. Given the large pool of candidate fuels and multiple engine technologies to consider, this is clearly not a problem we can solve by brute force. We could eliminate most of these experiments if we had a means of *predicting* the performance of fuel candidates, so that we could quickly focus our attention on the most promising systems. Ideally, an engineer would be able to simply sit down at a computer, input some information about the fuel, engine, and operating conditions of interest, and be presented with a quantitative prediction of the fuel and engine performance, including uncertainty information.

Figure 1.1 shows how predictive chemical kinetics can be envisioned as a cycle of several steps. First, we need a catalog of the currently-available chemistry information. This database is then used to develop a mechanism that quantitatively describes the underlying chemical processes by drawing on the gathered and curated chemistry

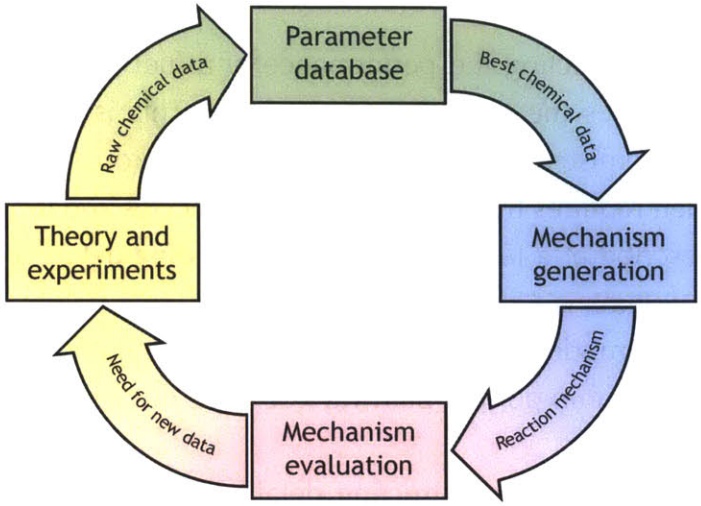


Figure 1.1: The predictive chemical kinetics cycle. A database of available chemical information is used to develop a model of a chemical process, such as combustion or pyrolysis. By comparing the predictions of this model to experimental measurements, we can systematically identify what chemical information would be most valuable to obtain in order to improve the model. This information can be determined via theoretical calculations or direct experimental measurements, which then increase our available chemical information used for developing future models.

knowledge. Comparing the predictions of this mechanism to experimental measurements generally yields some areas of agreement, but also other areas of disagreement. Using techniques such as rate-of-production and sensitivity analyses, we can systematically identify which chemical information would be most valuable to obtain in order to improve the model predictions. This information can be determined via theoretical calculations or direct experimental measurements, which then increase our available chemical information used for developing improved models for both this fuel and other fuels. By iterating around this cycle, we can efficiently converge on mechanisms that provide true fundamental insight into the underlying chemistry.

It is perhaps tempting to also use the experimental data we are using to evaluate the generated mechanism to fit the model parameters. Since the chemistry of many systems is complex, this *postdictive* fitting is typically very underspecified. Thus, although fitting might improve the performance of the current model at the current conditions of interest, the information gained is much less likely to extrapolate to similar fuels or other conditions.

1.2 Automatic mechanism generation

An important step in predicting the performance of a fuel is the development of a mechanism that describes the underlying chemistry. This is particularly challenging for combustion chemistry, which typically proceeds through a free-radical chain mechanism that often requires hundreds of chemical species and thousands of chemical reactions to describe at a fundamental level. Building detailed kinetics models by hand for these systems is tedious at best due to the high degree of bookkeeping involved. At worst, the model is subject to the whims and biases of the builder, with the potential to include unimportant pathways or leave out less obviously important pathways.

All of the things that make constructing detailed kinetics mechanisms by hand challenging are things that computers handle very well. This has encouraged the development of *automatic* mechanism generation: the use of computer software to build detailed kinetics mechanisms. To accomplish this, computers must be "taught" how to represent chemical species, determine which reactions can occur, estimate the thermodynamic and kinetic parameters for those species and reactions, decide which of the generated species and reactions are important, and decide when the model is sufficiently developed to accurately represent the underlying chemical phenomena.

1.3 Thesis overview

This thesis presents several contributions in the areas of predictive chemical kinetics and automatic mechanism generation. The contributions cover a range of topics, but can be broadly gathered under the heading of estimating the kinetics of chemical reactions. This is an important part of mechanism generation so that quantitative predictions can be made; as we will see, it is especially important for our automatic mechanism generation techniques since the estimated kinetics are used to discriminate the important species and decide when the mechanism is complete. Estimating chemical reaction rates is complicated by the tradeoff between the speed at which a rate can be estimated and the accuracy of the resulting estimate. The kinetics of thermal unimolecular reactions is further complicated by the coupling of reaction events with nonreactive bimolecular collisions, the latter of which provide and remove the energy necessary for unimolecular reactions to proceed. The chapters of this thesis will discuss how we have addressed these challenges, and how we are using our solutions to automatically build detailed kinetics models that help us understand the combustion of real alternative fuel candidates.

Chapter 2 discusses software frameworks for automatic mechanism generation, focusing on the software package RMG developed at MIT and used throughout this thesis. A new implementation of RMG using the Python programming language is presented. Particular effort was made in developing a logical, modular codebase, in developing a large suite of unit tests of individual features, and in selectively optimizing only performance-critical sections of code. A brief comparison of the new Python version of RMG to the previous Java implementation of RMG is shown.

Chapter 3 presents an evaluation of several methods of automatically and rapidly estimating chemical reaction rate coefficients by analogy, including a new group additivity-based method. A general framework for implementing these methods is discussed, including an emphasis on separating the training set of kinetic data from the hierarchy of functional groups used to establish analogous relationships. The uncertainty of each method is assessed using a test set of rate coefficient data obtained from the NIST Chemical Kinetics Database. This gives us some idea of the overall uncertainty in a kinetics estimate, and how far we are away from the accuracy desired for automatic mechanism generation.

Chapter 4 discusses how to calculate bimolecular reaction rates very accurately, as is often required for the sensitive reactions of a kinetics model. In particular, the chapter focuses on the method of ring polymer molecular dynamics (RPMD), which enables the incorporation of quantum mechanical effects in bimolecular reaction rates. The performance of RPMD is compared against several other methods using the reaction $OH + CH_4 \longrightarrow CH_3 + H_2O$, which has significant quantum character, as a case study. As part of this comparison, a general-purpose code for computing bimolecular reaction rates using RPMD rate theory, named `RPMDrate`, is presented.

Chapter 5 presents a framework for automatic estimation of pressure-dependent rate coefficients for thermal unimolecular reactions. These reactions require an inert third body to provide or remove the energy needed for reaction via bimolecular collisions. A master equation-based model of this competition between collisions and reactions is too detailed for automatic mechanism generation, so we wish to reduce the model to a set of phenomenological rate coefficients $k(T, P)$. Three methods of accomplishing this reduction are compared using the chemically-activated reaction of acetyl radical with oxygen as a case study. The strengths and weaknesses of each method are discussed in terms of accuracy, speed, and robustness, and guidelines for when each method should be used are given.

Chapter 6 demonstrates the importance of pressure-dependent kinetics under atmospheric (low-temperature) conditions via an investigation of the reaction of the

Criegee intermediate CH_2OO with small carbonyl compounds: formaldehyde, acetaldehyde, and acetone. By combining high-level quantum chemistry calculations with the master equation methods of the previous chapter, we are able to correctly predict several recent experimental observations. This chapter also discusses how pressure-dependence functionality was added to the `CanTherm` software package, and how both were integrated into the new Python version of RMG.

Chapter 7 brings together the ideas from the earlier chapters to present an automatically generated detailed kinetics model that describes the oxidation of diisopropyl ketone (DIPK), a candidate biofuel representative of species produced from cellulosic biomass conversion using endophytic fungi. The model is evaluated against three experiments covering a range of temperatures, pressures, and oxygen concentrations to show its strengths and weaknesses. Several systematically identified improvements to individual thermodynamic and kinetic parameters that improve the model are discussed, which illustrates how the cycle of Figure 1.1 occurs in practice. Our ability to automatically generate this model demonstrates the validity of the predictive chemical kinetics paradigm.

Finally, Chapter 8 discusses several ideas for future work in order to fully achieve the vision of predictive chemical kinetics. These ideas include general thoughts on the RMG software, as well as more specific next steps for the kinetics estimation, RPMD, and pressure dependence projects.

1.4 References

[1] U.S. Energy Information Administration. *Annual Energy Outlook 2012*. 2012.

[2] International Energy Agency. *World Energy Outlook 2012*. 2012.

[3] N. A. Owen, O. R. Inderwildi, and D. A. King. "The Status of Conventional World Oil Reserves – Hype or Cause for Concern?" *Energy Policy* 38 (8), 2010, pp. 4743–4749. DOI: 10.1016/j.enpol.2010.02.026.

[4] J. Murray and D. King. "Climate Policy: Oil's Tipping Point has Passed." *Nature* 481 (7382), 2012, pp. 433–435. DOI: 10.1038/481433a.

[5] S. Pacala and R. Socolow. "Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies." *Science* 305 (5686), 2004, pp. 968–972. DOI: 10.1126/science.1100103.

[6] S. J. Davis, L. Cao, K. Caldeira, and M. I. Hoffert. "Rethinking Wedges." *Environ. Res. Lett.* 8 (1), 2013, p. 011001. DOI: 10.1088/1748-9326/8/1/011001.

[7] G. P. Robertson, V. H. Dale, O. C. Doering, S. P. Hamburg, J. M. Melillo, M. M. Wander, W. J. Parton, P. R. Adler, J. N. Barney, R. M. Cruse, C. S. Duke, P. M. Fearnside, R. F. Follett, H. K. Gibbs, J. Goldemberg, D. J. Mladenoff, D. Ojima, M. W. Palmer, A. Sharpley, L. Wallace, K. C. Weathers, J. A. Wiens, and W. W. Wilhelm. "Sustainable Biofuels Redux." *Science* 322 (5898), 2008, pp. 49–50. DOI: 10.1126/science.1161525.

[8] 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.

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

2

Designing a modular software framework for automatic mechanism generation

Many industrially relevant chemical processes, including fuel combustion and pyrolysis, proceed through complex, nonlinear free-radical chemistry involving many kinetically-significant intermediates. Often the chemical kinetics is strongly coupled to thermodynamics, fluid dynamics, and/or heat and mass transport. Historically, the chemistry of these processes would be dramatically simplified in order to generate tractable reactor models. While this often works for interpolating over a small range of conditions, it does not provide accurate extrapolations to new operating conditions or reactor designs since it does not correctly represent critical features of the underlying physics and chemistry. For this reason, large-scale detailed kinetics models are now commonly used to model these complex processes at a fundamental level.

Building detailed kinetics models – which regularly comprise hundreds of species and thousands of reactions – by hand is a tedious and error-prone process. It is incumbent on the model builder to carefully keep track of all considered species and reactions, to systematically and comprehensively consider all of the chemistry that may contribute, and to impartially decide which chemistry is important. Some of these challenges can be mitigated by constructing models using a hierarchy of submechanisms, starting with small molecules and adding additional layers for larger molecules, validating each layer before moving onto the next. Many hand-curated models built in this manner are available from groups at Lawrence Livermore National Laboratory [1], NUI Galway [2], and other places.

Fortunately, these things that make constructing detailed kinetics mechanisms by humans challenging are things that computers are well-equipped to handle. The objective is thus to teach the computer how to do chemistry, so that it can "automatically" generate the reaction mechanisms. This chapter is concerned with the development of software for this purpose using the principles of good software engineering. After

a brief discussion of the principles of automatic mechanism generation and currently-available codes, a new version of the Reaction Mechanism Generator (RMG) software package written in the Python programming language will be presented.

2.1 Background

There are four principal features required for automatic reaction mechanism generation [3]:

- A means of representing chemical species uniquely and unambiguously;
- A means of determining which reactions a set of species can undergo, including which products are produced;
- A means of estimating the necessary thermodynamic, kinetic, and transport parameters for the species and reactions; and
- A means of deciding which species and reactions to include in the model, and which to exclude.

There have been several efforts to develop software for automatic mechanism generation. A brief history and review of these codes is given in the following paragraphs, and was adapted in part from Warth *et al.* [4] and a review by Pierucci and Ranzi [5].

The first generation of mechanism generation codes, including GRACE [6] and KING [7], used matrices to represent species and matrix transformations to represent matrices. This combinatorial approach allowed for the enumeration of the entire set of possible reactions without prior knowledge of chemical reaction classes, but also likely produced many nonsensical reactions. KING added the ability to specify some chemical constraints to limit the combinatorial explosion of products.

The second generation of mechanism generation codes moved to a functional-group based methodology. In this approach, reaction classes are defined as rules that manipulate matching functional groups to convert molecules from reactants to products. An early implementation of this rule-based approach was structure-oriented lumping [8, 9], which represents each species as a vector containing the numbers of each functional group in the molecule. Most current implementations now use chemical graph theory to represent molecules and functional group substructures using the graph abstract data type, with vertices representing atoms and edges representing bonds. Several mechanism generation codes that implement this technique appeared

in the 1990s, including MOLEC [10, 11], NETGEN [12–14], MAMOX [15], REACTION [16], and EXGAS [4]. More recent codes in this family include KUCRS [17, 18].

This switch to working with functional groups also enabled the ability to incorporate the automatic estimation of chemical parameters via analogy to other molecules for which such data was available. Many codes utilize an implementation of the Benson group contribution method [19, 20] for estimating thermodynamic parameters. Several standalone modules implementing the Benson method have been developed, including THERM [21], RADICALC [22], CHETAH [23], and THERGAS [24]. Kinetic parameters are also usually estimated using structure-reactivity relationships, such as Evans-Polanyi. Methods for estimating kinetic parameters will be the subject of Chapter 3.

Various methods of terminating the mechanism construction process have been implemented in different codes. For example, EXGAS constructs a reaction mechanism by successively adding generations of new species and reactions around an initial mechanism composed of validated C_0-C_2 small molecule chemistry; the termination criterion is specified simply as the desired number of generations. By contrast, Susnow *et al.* introduced a method of discriminating between important and unimportant species and reactions using species fluxes [25].

In addition to the codes mentioned previously, we are also aware of proprietary codes developed within several companies, including Dow and ExxonMobil, as well as an effort at the Laboratory for Chemical Technology at the University of Gent. These codes are not available to academic researchers.

2.2 RMG-Java

The remainder of this chapter focuses on the software package RMG (**R**eaction **M**echanism **G**enerator) developed at the Massachusetts Institute of Technology. RMG is a descendant of XMG, the ExxonMobil Mechanism Generator [26–28]. The objectives in the original development of RMG were threefold: (1) to use the graph representation of species and functional groups common to other contemporary mechanism generation codes; (2) to apply the software design concept of object-oriented programming to create a robust, modular codebase; and (3) to create a free, open-source code for automatic mechanism generation that anyone can use, modify, and redistribute [29].

Like many other "second-generation" automatic mechanism generation codes, RMG uses a functional-group based methodology to work with species and reactions.

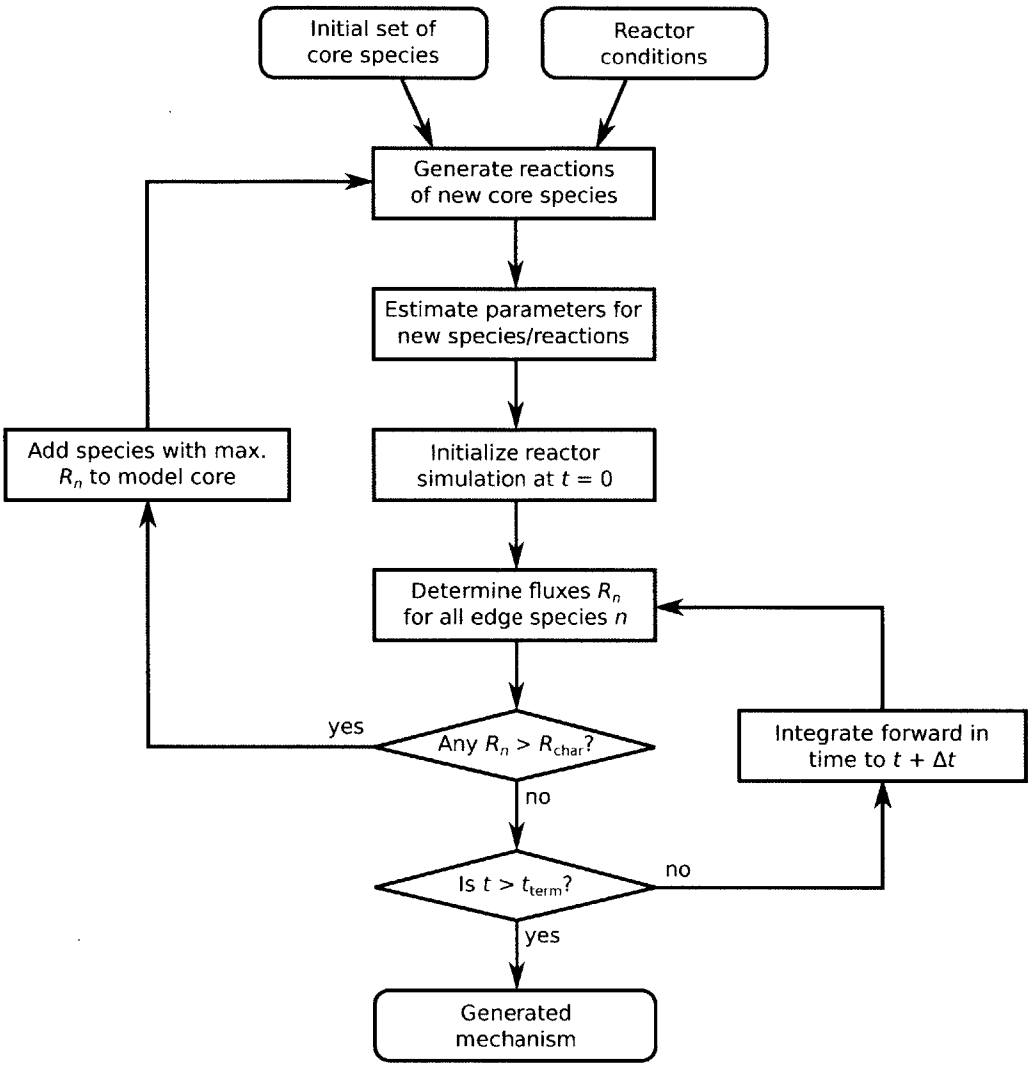


Figure 2.1: Flowchart of the automatic mechanism generation algorithm as implemented in RMG. The rate-based algorithm of Susnow *et al.* [25] is used to determine which species to incorporate into the final model.

Chemical species and functional groups are represented using chemical graph theory. Graph isomorphism allows for the comparing of chemical graphs to determine if they represent the same species, as well as searching species for various functional groups that identify reaction sites. A novel feature of RMG is the methods by which parameter estimation for thermodynamics and kinetics are performed; these incorporate a hierarchy of functional groups to enable using a more general analogy for the species or reaction when data for the more specific analogy is not yet available. RMG then uses the rate-based screening algorithm of Susnow *et al.* [25] to determine which species to incorporate into the final model.

A flowchart depicting the model generation algorithm as implemented in RMG is shown in Figure 2.1. The user supplies a starting pool of species, the operating conditions of interest (temperature, pressure, and initial mole fractions), and a criterion for terminating the reactor simulation (time and/or conversion). RMG places the initial species inside the model "core" to indicate that these species are important and therefore belong in the generated model. RMG then uses its reaction families to elucidate all of the reactions that these species can undergo. Many of these reactions will produce products that are not yet in the model core; these are placed on the model "edge" for the moment to indicate that we have not yet determined if these are important species. The parameters for the new species and reactions are estimated by analogy using the chemical database that comes with RMG. A simulation of the reactor is then initiated using the updated model. As the simulation marches forward in time, the fluxes to all species on the model edge are monitored. If one of these fluxes becomes significant relative to the species fluxes in the model core – as determined using a user-specified tolerance – the simulation is interrupted and that species is promoted from the edge to the core. New species and reactions are then generated by finding the reactions of the new core species with all of the other core species. This process iterates until no edge species have significant flux over the entire simulation before the termination time or conversion is achieved. At that point the model core is determined to contain all of the significant chemistry, and therefore the core species and reactions represent the final model.

The original implementation of RMG was done as part of the Ph.D. thesis of Jing Song in 2004 using the Java programming language [29]. Many other students have contributed since then [30–39], producing several major and minor releases which added many new features, including functionality for estimating pressure-dependent rate coefficients (see Chapter 5) [40], generating liquid-phase models that incorporate solvent effects [41], and applying automatic quantum chemistry calculations for improved estimates of the thermochemistry of cyclic molecules [42]. Several detailed kinetics mechanisms have been published that were automatically generated using RMG, including models for n-hexane [43], 1,3-hexadiene [44], methylformate [45], the isomers of butanol [46–49], and the jet fuel JP-10 [50].

The Java implementation of RMG has clearly been quite successful. However, as we continue to push the envelope in applying automatic mechanism generation to more exotic systems, several shortcomings of the RMG implementation have been encountered. First, we have discovered situations where RMG is limited by either execution speed or available memory. For example, building the JP-10 mechanism often

required considering more than a million candidate species for inclusion in the model; storing all of the structure and parameter information for these species currently requires tens of gigabytes of memory. Furthermore, since RMG is currently designed to only run on a single processor, generating a model of that size can take several days, especially when using more CPU-intensive options like automatic quantum chemistry calculations for cyclic molecule thermochemistry. The memory requirement per processor is heavily skewed; a better implementation would at least take advantage of the multiple processors available on modern systems, especially on computer clusters with large amounts of memory.

Second, RMG-Java has also lost the modularity originally envisioned by Song as new developers have come along and new features have been added. The original architecture as designed by Song grouped code into four main "packages":

- `jing.chemUtil` for implementing chemical graph theory to represent species and functional groups,
- `jing.chem` for manipulating species and functional groups,
- `jing.rxn` for working with chemical reactions, and
- `jing.rxnSys` for the automatic mechanism generation algorithms.

Many years later, RMG 4.0 still utilizes the same four main packages, the latter three of which have grown very large. Many new features which likely deserved their own packages were instead simply placed *ad hoc* into one of these packages based on which was closest to the intended feature. The loss of modularity means that adding new features or modifying existing features is much more likely to interfere with or break other, seemingly unrelated features. For example, my initial implementation of pressure dependence functionality discussed in Chapter 5 would often cause the feature of saving files from which the job could be restarted to break, and fixing the latter would then cause my pressure dependence implementation to break. A better design would disentangle these features and provide a means to test that each feature is working separately, so that we have more confidence that they will work together.

Third, some limitations of the choice of Java as the programming language have also been encountered. A particular example is our need to occasionally conduct a complex mathematical calculation, such as an optimization for fitting parameter values or an integration of a differential algebraic system of equations. Java is not very well suited for these CPU-intensive tasks, which require very fast access to arrays of

data. For this reason, several small Fortran utilities have been written that supplement the main Java codebase. These utilities do not benefit from the object-oriented design principles, require additional code for passing information from Java to Fortran and back, and reduce the portability of the RMG software.

Lastly, RMG-Java also suffers from a few troubling design paradigms. In RMG-Java, the program is designed to always give you a model. If an unexpected situation is encountered during the building of the model, RMG-Java will often just silently continue as if the issue had not appeared. Sometimes you might get a brief warning message buried deep in the log file as the only record that an issue had even occurred. Users would sometimes run jobs for a day or more, only to find that there was a critical error that occurred in the first few minutes of the job that was not handled in the proper manner, making the resulting model worthless. Much stronger error checking and handling was needed to ensure that RMG is working as the user expected.

2.3 RMG-Py

The Python programming language has seen increased adoption for scientific computing in recent years. Its easy-to-read syntax makes it much easier to reread and understand existing code, which is often just as important as writing that code in the first place. Python also comes with a large standard library of functionality – and makes it very easy to install third-party libraries – which significantly reduces the amount of new code that must be written. These features combine to dramatically accelerate the pace of development when compared to other languages, which makes Python a strong choice for scripting and rapid prototyping. Python is also commonly used as a server-side language in web development.

Much as how RMG-Java was originally written to introduce a cleaner, modular software design from XMG, so too is RMG-Py intended to refine the design of RMG-Java. RMG-Py was developed using several software paradigms, some of which are significant departures from those used to develop RMG-Java. These will be discussed in the next sections.

2.3.1 Organizing and modularizing the codebase

The value of modularization is that in enables us to take a very complex task – such as developing a code for automatic reaction mechanism generation – and break it up into several smaller tasks that are much more tractable. In RMG-Py we further subdivide the codebase into distinct components based on functionality. In Python vernacular, a *module* refers to an individual Python file, while a folder containing several Python

Table 2.1: An overview of the main RMG-Py packages and modules

Module	Description
`rmgpy.cantherm`	Chemical properties from quantum chemistry calculations
`rmgpy.chemkin`	Reading and writing models in CHEMKIN format
`rmgpy.constants`	Physical constants
`rmgpy.data`	Working with the RMG database
`rmgpy.kinetics`	Reaction kinetics models
`rmgpy.molecule`	Molecular representations using chemical graph theory
`rmgpy.pdep`	Pressure-dependent kinetics from master equation models
`rmgpy.quantity`	Physical quantities
`rmgpy.reaction`	Chemical reactions
`rmgpy.rmg`	Automatic reaction mechanism generation
`rmgpy.solver`	Modeling reaction systems
`rmgpy.species`	Chemical species
`rmgpy.statmech`	Statistical mechanics models of molecular degrees of freedom
`rmgpy.thermo`	Species thermodynamics models

files is called a *package*. Each module can contain multiple variable, function, and class definitions, giving us complete freedom to collect closely-related functionality as we see fit. (Java forced the convention of having exactly one class per file, which significantly reduced our organizing flexibility.)

A brief overview of the main RMG-Py components (packages and modules) is given in Table 2.1. All components are placed inside the `rmgpy` top-level package to ensure no collisions with components from other Python codes. Each component is designed to have a well-defined purpose that does not overlap with other components, although it can depend on other components. Functionality is much more important than size when dividing the codebase into components; for example, the `rmgpy.constants` module simply defines a few relevant physical constants, while the `rmgpy.data` package contains several large submodules and subpackages due to the intricacies of working with the RMG database.

A more thorough discussion of the RMG-Py codebase is presented as an appendix in Section 2.6.

2.3.2 Unit testing to encourage correctness and modularity

The primary method of ensuring that the RMG-Java code was working as intended was to simply generate a few small mechanisms and observe that the jobs completed successfully and produced chemically reasonable mechanisms. This top-down testing

approach makes it very difficult to isolate problems in individual components. It would be much better to first validate the function of each component separately before trying to test the system as a whole.

This bottom-up testing concept, called *unit testing*, gained popularity as part of a software development paradigm called extreme programming [51]. Unit testing provides a number of advantages to the software developer:

- Unit tests validate the correctness of small snippets of code independently. This usually leads to tests that are very quick to execute, which accelerates debugging when errors are encountered.

- Unit tests encourage the developer to think about how the component would be used as it is developed. This can help clarify the component design. In particular, the desire to test small pieces of functionality independently can help encourage modularity in the developed component.

- Unit tests offer another form of documentation, as they provide examples of how the component is used, and what the expected results should be.

- Unit tests can be executed after every change to the software. This helps to immediately identify changes that break functionality, so they can be addressed right away instead of manifesting into a bug at some point in the future. This is especially useful when refactoring code to be functionally identical but easier to understand.

Individual unit tests should be designed to check a single unit of functionality in a single situation. Tests should be written for both cases where the code should succeed and cases where it should fail. Multiple tests that validate a single component should be grouped into test suites. Ideally, enough tests would be written so that all of the functionality of the component is tested in all possible usage situations. Although this does require some additional coding investment during development, it leads to greatly increased confidence in the correctness and robustness of the individual components. The developer can then proceed to integration testing to ensure that the components interact as intended.

At the time of this writing, RMG-Py defined 677 distinct unit tests. The tests themselves are placed in Python modules that live in the same directory as the modules they are testing to encourage cocurrent development of the module and the associated tests.

2.3.3 Error handling

RMG-Py was consciously designed to be far more wary of errors than RMG-Java. This is especially important in automatic mechanism generation due to the computational expense required, so that we are not wasting time and resources waiting for nonsensical mechanisms to be created. When an error occurs, RMG-Py will tend to halt execution and print a (hopefully detailed) message describing the error and the immediate circumstances. This strictness has helped us identify bugs in RMG-Py that were hidden in RMG-Java, such as subtle errors in the construction of the hierarchy of functional groups for some reaction families.

As RMG-Py has matured, we have come across some errors that we have determined should not cause the mechanism generation to stop. Even in these situations, we print a clear warning message describing the situation encountered and the action taken. An example of this sort of error occurs when a rate coefficient estimate for an endothermic reaction is made with Arrhenius parameters such that the activation energy is below the enthalpy of activation, which implies that the reverse reaction has a negative activation energy. In these cases, we have found that raising the activation energy to match the enthalpy of reaction usually gives a better estimate of the reaction kinetics; nonetheless, we still print a warning message indicating that we have adjusted the activation energy to match the enthalpy of reaction.

2.3.4 Selectively accelerating execution with Cython

The readability, flexibility, and rapid pace of development that make writing Python code attractive come with a catch: Python often executes slower (and requires more memory) than other programming languages. The slow execution speed is often a result of one of two factors. First, Python is a dynamic, interpreted language, which means that it must interpret the meaning and evaluate the validity of each statement as it is encountered. Second, Python generally does not provide efficient indexing into arrays of numerical data, which makes iterating over large arrays of data slow.

The reference implementation of Python is written in the C programming language. This enables the opportunity to write C code that, when compiled, can be interacted with as if Python code, but executes much faster and uses less memory. However, writing significant amounts of C code would be a significant step backward, given that RMG-Java was written in part to move away from C. Instead, we turn to Cython, a superset of the Python language that allows for compilation directly to C code [52]. The key insights of Cython are the addition of explicit type declarations to the Python syntax, which allow for avoiding the expensive interpretations,

and the providing of direct access to C basic types and functions. Compiled Cython code with explicit type declarations often runs orders of magnitude faster than the equivalent pure Python code, and generally only requires some minor tweaks to the already-written Python code. Since Cython is (mostly) a superset of Python, you only need to invest the time to annotate the performance-critical bottlenecks in the code; the remaining code will run as before.

RMG-Py has annotated many of its lower-level modules using Cython in order to increase its execution speed. As we will show later, RMG-Py remains somewhat slower than RMG-Java in generating the same mechanism, but is far closer in performance than a pure Python implementation would be.

2.3.5 Integrating documentation

Another significant strength of Python is its ability to integrate documentation directly into the source code. Each module, class, function, and variable can be associated with a string that documents that component. (Java has similar functionality via the Javadoc *de facto* commenting standard, but it is not currently used significantly in the RMG-Java codebase.) Embedding the documentation in the code itself makes it easy for the developer to write documentation as the code is developed. This documentation can also be extracted and placed in standalone documentation using the Sphinx documentation generator, which can produce output in HTML, PDF (via LaTeX), and other formats. Perhaps not surprisingly, the documentation of the Python language is generally detailed and useful. We have adopted the same techniques in RMG-Py.

2.3.6 Developing in public with distributed version control

Both RMG-Java and RMG-Py are free and open source software. This means that anyone is free to download, use, modify, and redistribute the software as they wish. This is particularly important to enable others to reproduce the results of our software, a key component of scientific research [53]. We do not do this only for our official versioned releases, however; we also do our development in the open, which makes it easy for users to keep up with – and openly peer review – our changes. We have adopted the git version control system to manage and document changes to the code. Git is a distributed version control system, which means that everyone has a copy of the entire project history, and is free to make changes to the project independently of anyone else. Git provides a sophisticated system for merging in changes made by multiple authors, which makes collaboration much easier.

2.4 Comparing RMG-Py to RMG-Java

In order to assess the performance of RMG-Py compared to the more established RMG-Java, a model of diisopropyl ketone oxidation was constructed in both programs using the same settings. (A more rigorous form of this model is discussed in Chapter 7.) Figure 2.2 shows the CPU time, number of core reactions, numbers of edge species and reactions, and memory used plotted versus the model size (as indicated with the number of core species) for both programs. The models constructed by the two programs are nearly the same in size and composition, though they are not completely identical due to minor differences in the implementations of the kinetics estimation algorithm. Both programs also consider nearly the same number of species and reactions on the model edge. However, RMG-Py clearly requires about four times more CPU time and twice as much memory as RMG-Java to construct this model.

Python comes with an easy-to-use CPU profiler that allows us to systematically identify the performance bottlenecks in the code. Selective optimization of only these performance-critical components using Cython has resulted in a speedup of more than a full order of magnitude from the equivalent pure Python implementation would be. Several algorithms in RMG-Py has also been designed in anticipation of parallelization, which may also be contributing a bit to the speed difference. Ultimately, the performance of RMG-Py is close enough to that of RMG-Java that parallelization should easily be able to make up the difference, even when using only a few additional processors.

The increased memory use in RMG-Py is partly due to the presence of new features that RMG-Java does not have. For example, RMG-Py was designed to be able to load, manipulate, and save the entirety of the RMG database. To accomplish this, RMG-Py was written so that it stored the units and uncertainty information associated with each chemical parameter, not just the parameter value itself. This explains why RMG-Py uses around 250 MB in the early part of the mechanism generation, while RMG-Java only uses around 60 MB. Since the database is only loaded once, this additional memory use is less of a concern than the rate of memory use as the mechanism grows. Even here, though, RMG-Py consumes memory about twice as fast as RMG-Java. This is something that may need further optimization in order to reduce the number of systems where an RMG-Py job runs out of memory before a valid mechanism is reached.

Finally, it is important to remember that use of Python provides additional advantages outside of automatic mechanism generation. Use of Python has made it much

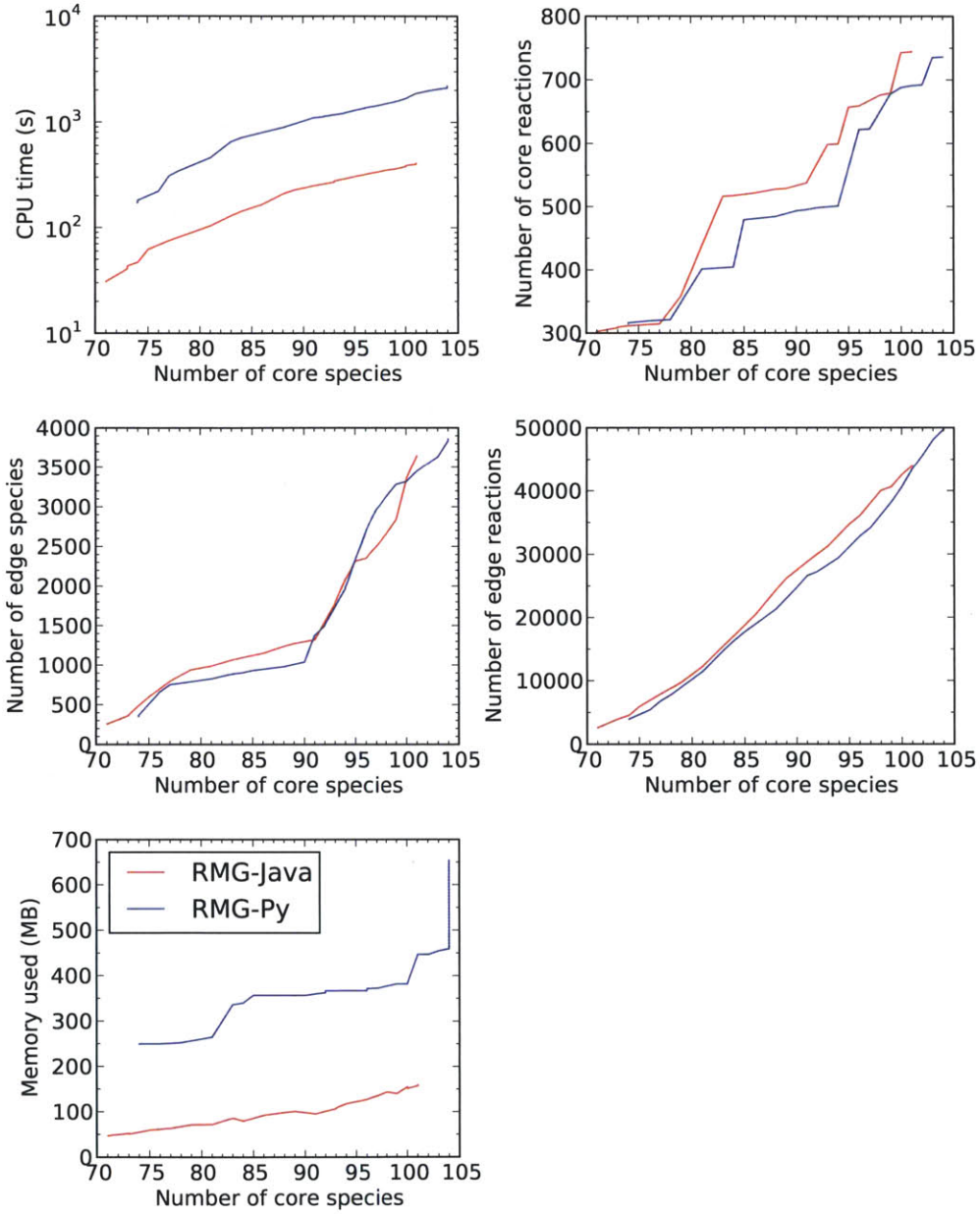


Figure 2.2: Comparison of execution statistics between RMG-Java 4.0 (red) and RMG-Py (blue) for a diisopropyl ketone (DIPK) oxidation model. Both RMG-Java and RMG-Py construct similar model sizes, but RMG-Py runs about four times slower and uses about twice the memory of RMG-Java.

easier to write scripts that utilize parts of the RMG codebase for other tasks. For instance, it has enabled the integration of the CanTherm software package for calculating properties from quantum chemistry calculations, which shares a fair amount of code with RMG-Py. (This integration is discussed in Chapter 6.) We have also made several features of RMG-Py – such as browsing and searching the RMG database – accessible via the internet by integrating with a Python-based web server.

2.5 Conclusion

This chapter has presented a new version of the Reaction Mechanism Generator software package written in the Python programming language. RMG-Py is implemented using a new modularization design that improves upon that of RMG-Java by creating smaller modules and packages grouped by more specific functionality. RMG-Py has also been developed using the software principles of unit testing, strong error handling, integrated documentation, and distributed version control. Although much effort has been made at selectively identifying and optimizing performance-critical components, RMG-Py currently runs somewhat slower and requires a bit more memory than RMG-Java. We expect that further optimization (including parallelization) will reduce or eliminate these differences in performance, and that the advanced features that RMG-Py enables are worth this temporary trade-off.

2.6 Appendix: A brief tour of RMG-Py

This section will briefly present the modular organization of RMG-Py, highlighting the key features of each component. These presentation will be necessarily incomplete; more details are available in the RMG-Py documentation, accessible at `http://greengroup.github.com/RMG-Py/`.

2.6.1 `rmgpy.cantherm` – CanTherm

RMG-Py integrates the CanTherm utility originally developed by Sandeep Sharma and Michael Harper for computing thermodynamic and kinetic properties from quantum chemistry calculations [54]. A number of improvements to the CanTherm code were made during this integration, including adding functionality for pressure-dependent master equation calculations using quantum chemistry calculations as input parameters. Further discussion of the updated CanTherm will be deferred until Chapter 6.

2.6.2 `rmgpy.data` – Working with the RMG database

The `rmgpy.data` package contains functionality for working with the RMG database of chemical information. Both the new, more flexible format used by RMG-Py and the original format used by RMG-Java can be loaded and saved; this allows us to backport improvements made to the RMG-Py database back to RMG-Java if desired.

2.6.2.1 General database organization

Many database components are divided conceptually into three components: a *dictionary* associating string labels with species or functional group structures, an optional *tree* that places these structures in a hierarchy from most generic to most specific, and a *library* that associates each structure with a set of parameter values. These three components are represented together using the `Database` class for the hierarchy itself and the `Entry` class to store the string label, structure, and parameter values. The `Database` class stores all of the entries by label, as well as the top-level entries in the hierarchy. Individual databases can be marked as recommended if they contain values that we trust and generally recommend to use during RMG jobs. Both brief and verbose descriptions of the database can also be stored.

Each `Entry` object also stores references to its immediate parent and children in the hierarchy if applicable, as well as information about the parameter values, particularly its source. Comments about each entry are divided into a short description (generally a brief, one-line summary) and a long description (which can be as verbose as desired, including multiple paragraphs). The source reference is stored separately from the short and long descriptions to enable easier formatting and cataloging of the reference information.

Each database type is generally divided into at least two components: a set of libraries containing collections of curated parameter values, and some sort of hierarchy of functional groups and associated values for estimating parameters by analogy. Some components also define one or more depositories containing an uncurated record of all available information.

2.6.2.2 Thermodynamics database

The RMG thermodynamics database, represented in total by the `ThermoDatabase` class, is divided into depositories, libraries, and group values via the `ThermoDepository`, `ThermoLibrary`, and `ThermoGroups` classes. Entries in the thermodynamics database can contain any of the heat capacity models defined in the `rmgpy.thermo` package as parameter values. This is an improvement over the RMG-Java database, which

mandates use of the Benson group additivity format, and therefore the `ThermoData` class.

2.6.2.3 *Kinetics database*

The RMG kinetics database, represented in total by the `KineticsDatabase` class, is divided into depositories, libraries, and families via the `KineticsDepository`, `KineticsLibrary`, and `KineticsFamily` classes. Entries in the kinetics database can contain any of the kinetics models defined in the `rmgpy.kinetics` package as parameter values. Each family has multiple ways to estimate the kinetics using its hierarchy of functional groups, including the `KineticsRules` class for the "rate rules" method originally developed for RMG-Java, and the `KineticsGroups` class for using a kinetics group additivity method. Further discussion of kinetics estimation methods can be found in Chapter 3.

Reactions generated from the database are created using one of several subclasses of the general `Reaction` class defined in the `rmgpy.reaction` module. A `DepositoryReaction` comes from a depository, a `LibraryReaction` from a library, and a `TemplateReaction` from a family. Each of these classes allows for storing additional metadata about the source of the reaction. Depository and library entries can store the source component and the actual entry that the reaction corresponds to. Template reactions can store the family the reaction belongs to, as well as the most specific functional groups that define the reaction.

Although significant effort was made to minimize the amount of hardcoding of reaction families required, there were a few places where it was unavoidable, and these places are clearly marked. One example of this involved the radical-radical association reaction family, which double-counts all of the reactions in the forward direction since both functional groups Y· are identical. The reaction-path degeneracies must be divided by two in order to obtain the correct rate coefficient values. A second example involved families such as hydrogen abstraction and intra-H migration, which are the same in both the forward and reverse direction. In these families, one must swap a few of the labels on the product molecules in order to obtain products that also match the reaction family template. We could probably add new reaction recipe steps for this in order to eliminate the hardcoded atom label swapping.

2.6.2.4 *Statistical mechanics database*

The RMG frequency database, represented in total by the `StatmechDatabase` class, is divided into depositories, libraries, and group values via the `StatmechDepository`, `StatmechLibrary`, and `StatmechGroups` classes. Entries in the depositories and libraries can define any of the molecular degrees of freedom defined in the `rmgpy.statmech`

module, which are encapsulated in `Conformer` objects. The functional group entries are associated with a set of vibrational frequency ranges and the numbers of vibrations in each range for each occurrence of that functional group. Each collection of frequency lower bound, frequency upper bound, and number of occurrences are stored in memory as `GroupFrequencies` objects.

2.6.2.5 Overall database

The `RMGDatabase` class represents the entirety of the RMG database, including thermodynamics, kinetics, and statistical mechanics sections. The `RMGDatabase` object also provides a place to store functional groups that are forbidden in any generated molecular structures. (Ideally there would be no explicitly forbidden functionalities, and we would rely on thermodynamic and kinetic parameter estimation to discourage inclusion of these unusual species. In practice, however, these functionalities are often so unusual that our parameter estimates are likely to have significant error.)

2.6.3 `rmgpy.kinetics` – Kinetics models

The `rmgpy.kinetics` package provides representations of various models used to represent the kinetics of elementary reactions. The most commonly used for pressure-independent kinetics is the `Arrhenius` class, which implements the modified Arrhenius expression. If a single Arrhenius expression is not enough to accurately fit the kinetics, the `MultiArrhenius` class provides support for using a sum of Arrhenius expressions. A discrete set of $k(T)$ values can also be stored directly using the `KineticsData` class; this may be better than fitting an `Arrhenius` model, which introduces some fitting error.

Several classes that represent traditional models of pressure-dependent falloff are available. The `ThirdBody` class represents kinetics that are always in the low-pressure limit by multiplying an Arrhenius expression with the collider concentration. The `Lindemann` and `Troe` classes model the transition from the high-pressure to the low-pressure limit, the latter with additional parameters that modify the transition region. Two more general models of pressure-dependent kinetics are also available. The `PDepArrhenius` class uses a set of Arrhenius expressions at several pressures, and interpolates between them on a logarithmic scale to compute the $k(T, P)$ value. (This corresponds to the `PLOG` format in a Chemkin file.) A `MultiPDepArrhenius` class provides sum-of-`PDepArrhenius` support in the same way that the `MultiArrhenius` class does for `Arrhenius`. The `Chebyshev` class is also available for fitting of Chebyshev polynomials over inverse temperature and logarithmic pressure axes.

The rmgpy.kinetics package also provides two simple models of quantum mechanical tunneling through a reaction barrier. The Wigner class only requires the imaginary frequency of the transition state, but is also not very accurate. The Eckart class also requires the barrier heights of the reactants and products, but is significantly more accurate.

2.6.4 rmgpy.molecule – Molecular representations

The rmgpy.molecule package provides functionality for working with species and functional group representations as chemical graphs. The graph representation itself is defined by the Vertex, Edge, and Graph classes. Each Vertex object stores a dictionary containing the set of neighboring vertices and the connecting Edge objects to each. This dictionary-of-dictionaries format is very fast for editing the graph structure (*i.e.* adding or removing vertices and edges), and also provides reasonable traversal speeds.

An important task that is done often in RMG jobs is the determination of whether or not two graphs represent the same species or functional group, called a graph isomorphism. RMG-Py implements the efficient VF2 algorithm [55, 56] for this purpose. Some variables used only during isomorphism are cached on the Vertex objects for quick access. The connectivity values originally proposed by Morgan [57] were also used to accelerate the isomorphism by representing the graph connectivity near each vertex using a series of integer values. For each vertex, the connectivity value c_{k+1} is the sum of the c_k values of the neighboring vertices; each vertex has an implicit value of one for c_0, so that c_1 is simply the number of directly-connected neighbors to that vertex. The VF2 algorithm also works well for subgraph isomorphism, which allows us to efficiently identify the location of functional group matches in a molecule.

The Element class enables representation of chemical elements and their properties: atomic number, name, mass, and symbol. Only one instance of this class should be used to represent each element; these are automatically defined in the rmgpy.molecule.elements module for all 118 elements. The getElement() function returns the Element object corresponding to a given atomic number or symbol.

The AtomType class generalizes the concept of a chemical element by also including information about the local bond structure around that element: the number of double, triple, aromatic, and carbonyl bonds. This is useful because it makes the specification of functional groups more flexible, as well as accelerating the graph isomorphism evaluations. Only one instance of this class should be used to represent each atom type, and these should be defined in their entirety in the rmgpy.molecule.atomtype module. To define a new atom type, you must specify its unique name, the atom types

that are more generic and more specific than it, and the atom type(s) formed when reaction recipe steps are applied (e.g. bond forming and breaking). You must also add to the `getAtomType()` function so that the atom type can be identified for an atom in a given molecule.

Molecules are represented as chemical graphs using the `Atom`, `Bond`, and `Molecule` classes, which derive from the `Vertex`, `Edge`, and `Graph` classes and add semantic information for atoms and bonds. Atoms are currently represented by the element, number of radical electrons, spin multiplicity, formal charge, and an optional label for tagging purposes. The atom type is calculated automatically and cached; its use in `Atom` objects is completely hidden from the user. Bonds are represented by their bond order: single, double, triple, or aromatic. Besides the atoms and bonds, `Molecule` objects also store a molecular fingerprint that is used to accelerate comparison to other molecules by avoiding the full isomorphism algorithm in the majority of cases. Currently we simply use the molecular formula as the fingerprint, but it is straightforward to include other structure information in the fingerprint if it is quick enough to calculate.

Functional groups are represented as chemical graphs using the `GroupAtom`, `GroupBond`, and `Group` classes, which derive from the `Vertex`, `Edge`, and `Graph` classes and add semantic information for atoms and bonds in functional groups. `GroupAtom` objects are similar to `Atom` objects, except they use atom types instead of elements and allow for specifying multiple matching atom types, numbers of free electrons, spin multiplicities, and formal charges. Similarly, `GroupBond` objects allow for specifying multiple matching bond orders. `Group` objects also have a form of molecular fingerprint that often avoids the full isomorphism algorithm by first checking that the molecule has at least as many carbon, oxygen, and sulfur atoms and radical electrons as the subgraph. This especially helps in the forbidden structures checking, which often has to deal with unusual functionalities, e.g. O–O–O–O.

There are several ways to specify a chemical graph without constructing it manually from `Atom` and `Bond` (or `GroupAtom` and `GroupBond`) objects. Molecules can be constructed from the InChI [58] and SMILES [59–61] line notations, as well as from Chemical Markup Language (CML), an XML-based format [62]. Conversions to and from these formats are done using the OpenBabel software library [63]. Both molecules and functional groups can be constructed from the adjacency list format developed for RMG-Java. Adjacency lists in particular provide the ability to specify labels to tag individual atoms with, which we use to mark the central atoms in functional groups.

Algorithms for estimating the external + internal symmetry number from the graph representation of a molecule are available in the `rmgpy.molecule.symmetry` module. Separate functions are used to calculate atom-centered, bond-centered, axis-centered, and ring-centered contributions to the overall symmetry. A variety of tests of each function separately and the overall symmetry number are run; out of these, a few reflect either known limitations of the current algorithms or cases where we're not sure what the correct result should be.

A custom algorithm for rendering molecules via their skeletal formulae was implemented in the `rmgpy.molecule.draw` module. This was deemed necessary since existing drawing routines could not draw free radicals, which are common in our applications. The algorithm can also draw charged atoms if requested.

2.6.5 `rmgpy.pdep` – Pressure-dependent reactions

The `rmgpy.pdep` package provides functionality for estimating the pressure-dependent kinetics of thermal unimolecular reactions by various methods of reducing a master equation model. The theory of pressure-dependent reactions is the subject of Chapter 5, and some implementation details of this package can be found in Chapter 6.

2.6.6 `rmgpy.quantity` – Physical quantities

The `rmgpy.quantity` module provides functionality for storing and manipulating physical quantities, *i.e.* numbers with associated units and/or uncertainties. The main classes are `ScalarQuantity` and `ArrayQuantity`, which can store scalar and array (vector, matrix, *etc.*) quantities, respectively. Instead of worrying about calling `ScalarQuantity` or `ArrayQuantity`, however, you can simply call the `Quantity` function, which will return an object of the appropriate class based on the values provided.

In both classes, the value is stored internally in SI units to enable fast computations using those quantities. The idea is that we do very little reading and writing of quantities, and much more computation of values using quantities, so it's more valuable to have the computations occur quickly than the reading and writing. The `value` attribute gives the value of the quantity in the given units, whereas the `value_si` attribute gives the value of the quantity in the corresponding SI units. The stored units are available in the `units` attribute, and the uncertainty type (additive or multiplicative) and value are available in the `uncertaintyType` and `uncertainty` attributes, respectively.

A number of functions have also been provided that construct a physical quantity of a given dimensionality, such as mass, length, concentration, *etc.* This allows for validation that the provided units are in fact consistent with the intended dimensionality. This also allows for handling of a few special cases, particularly frequencies (which

are stored in cm^{-1} and rate coefficients (which have different units depending on the reaction order).

2.6.7 `rmgpy.reaction` – Chemical reactions

The `rmgpy.species` module provides the `Reaction` class for representing a general chemical reaction. Each reaction can be identified by a unique string label and/or integer index. The reactants and products of the reaction are stored as lists of `Species` objects in most cases, although some parts of the database code use `Molecule` or `Group` objects. The pairings of reactants to products are also often stored, particularly to aid in flux diagram generation (so that we don't have to resort to an atom balance). Reaction objects can be associated with a transition state, a kinetics model (pressure-independent or pressure-dependent), and a reaction-path degeneracy. Note that the convention in RMG-Py is to store the reaction-path degeneracy in the appropriate kinetics model parameter(s), so that the kinetics model always reflects the true rate coefficient, not the per-site rate. Reactions can also be marked as reversible or irreversible, or as duplicates of other reactions. The latter is particularly useful when loading or saving files in the Chemkin format.

This module also provides the `ReactionModel` class for representing a collection of chemical reactions and their associated species. The stoichiometry matrix and reaction rates for the model can be evaluated with the available methods.

2.6.8 `rmgpy.rmg` – Reaction mechanism generation

The `rmgpy.rmg` package implements the automatic mechanism generation algorithm. An individual mechanism generation calculation is represented using the `RMG` class. This provides a place to collect all of the options and functions that control the generated mechanism.

The `CoreEdgeReactionModel` class represents the mechanism being generated. It consists of separate `ReactionModel` objects representing the core (the set of species and reactions known to be important) and the edge (the set of species and reactions that might be important). An alternative means of directly accessing the created species, reactions, and pressure-dependent reaction networks are also defined on this class:

- The species dictionary associates molecular formulae with lists of all of the `Species` objects that match that formula. This makes it easy to quickly access the small subset of the defined species that might be isomorphic with a newly-proposed species.

- The reaction dictionary associates reaction families/libraries, reactants, and products with lists of all of the `Reaction` objects that have the same properties, using a nested sequence of dictionaries. This makes it easy to quickly access the small subset of the defined reactions with the same reactants, products, and source as a newly-proposed reaction.

- The network dictionary associates a set of one or two "source" species with lists of all of the pressure-dependent reaction networks (as `Network` objects) that contain that source configuration. This makes it easy to quickly access the small subset of the defined networks that a new path reaction may belong to.

A separate `Species` class that inherits the class of the same name from the `rmgpy.species` module is defined specifically for use in RMG jobs. It contains attributes and methods that are only useful in this context, such as methods for estimating various species properties from the database.

2.6.9 `rmgpy.solver` – Modeling reaction systems

The `rmgpy.solver` package provides functionality for simulating reaction systems, a critical part of the rate-based model enlargement algorithm. All reaction systems are implemented as descendants of the `ReactionSystem` class, which defines an interface for initializing the reaction system, conducting the simulation, determining if a species flux exceeds the characteristic flux, and determining if the simulation has reached its termination point. The termination point can be specified using one or more `TerminationConversion` objects to indicate a target conversion, and/or a single `TerminationTime` object to indicate a target time.

As in RMG-Java, the only reaction system currently implemented is the isothermal, isothermic batch reactor. This system is particularly efficient to simulate because the reaction rate coefficients and equilibrium constants need only be computed once, at the start of the simulation. The `SimpleReactor` class implements the functionality for evaluting the residual for the simple batch reactor. Integrating the governing equations of the reaction system is done using the DASSL differential algebraic equation solver. A Python wrapper to DASSL was developed as a separate project, titled `PyDAS` [64].

2.6.10 `rmgpy.species` – Chemical species

The `rmgpy.species` module defines the `Species` class for representing a single chemical species. Each species stores the chemical graphs that define each resonance isomer as a list of `Molecule` objects. The corresponding molecular conformation, if

known, is stored as a `Conformer` object. Each species can also be associated with a unique string label and integer index, a thermodynamics model (that derives from `HeatCapacityModel`), and a variety of collision parameters, including Lennard-Jones parameters, molecular weight, dipole moment, polarizability, rotational relaxation collision number, and a collisional energy transfer model. Convenience methods are available for calculating the values of the thermodynamic state functions and statistical mechanical properties.

This module also contains the `TransitionState` class for storing properties of a transition state for a reaction. This includes the molecular conformation, imaginary frequency, reaction-path degeneracy, and desired tunneling model.

2.6.11 `rmgpy.statmech` – Statistical mechanics models

The `rmgpy.statmech` package provides representations of various stastical mechanics models of molecular degrees of freedom. These classes are implemented as children of a base class `Mode`, which defines a common interface for computing the partition function, heat capacity, enthalpy, entropy, and sum and density of states of that mode. These properties can be computed either quantum mechanically or classically. Where appropriate, functions for computing the quantum mechanical energy levels and degeneracies of individual modes are provided.

The only currently-available model for translational motion is the `IdealGasTranslation` class, which implements the three-dimensional particle-in-a-box model for an ideal gas. The only attribute is the mass of the ideal gas molecules. Note that the heat capacity returned by this class is $\frac{5}{2}R$ since it is at constant pressure.

Several models of rotational motion are available. The `KRotor`, `LinearRotor`, and `SphericalTopRotor` classes represent one-, two-, and three-dimensional rotational motions governed by a single moment of inertia. The `NonlinearRotor` class is also available for rotational motion of a nonlinear molecule with distinct moments of inertia. All classes also allow for specifying the symmetry number of the rotational motion.

Vibrations are represented using the `HarmonicOscillator` class for harmonic vibrations. A single instance of this class is generally used to store all of the vibrational degrees of freedom of a molecule, although this use is not required.

One-dimensional hindered internal rotations (torsions) are represented using the `HinderedRotor` class. This class allows for storing the hindered rotor potential function as either a simple cosine expression

$$V(\phi) = \frac{V_0}{2}\left[1 - \cos n\phi\right] \tag{2.1}$$

or as a Fourier series

$$V(\phi) = \sum_k \left[A_k \cos k\phi + B_k \sin k\phi\right] \tag{2.2}$$

In either case, the reduced moment of inertia and symmetry number of the torsion must also be given. An optional semiclassical correction is also available for hindered rotors.

The `rmgpy.statmech.schrodinger` module contains functionality for evaluating quantum mechanical properties using the energy levels and degeneracies of an individual mode. This module also contains implementations of direct count methods for convoluting harmonic and anharmonic degrees of freedom when computing the sum and density of states.

The `Conformer` class represents a single molecular configuration, as generally defined by the three-dimensional geometry of the atoms (indicated by atomic number and mass). Each conformer can be associated with a ground-state energy (including zero-point energy), the list of molecular degrees of freedom, the spin multiplicity, and the number of optical isomers. Functions for determining the moments of inertia from the geometry are provided.

2.6.12 `rmgpy.thermo` – Thermodynamics models

The `rmgpy.thermo` package provides representations of various models used to represent the thermodynamics of individual species. These models differ in the functional form used to represent the constant-pressure heat capacity $C_\mathrm{p}(T)$. The `ThermoData` class stores the values of the heat capacity at several temperatures, and uses linear interpolation to evaluate the heat capacity between these data points. The `Wilhoit` class uses the Wilhoit polynomial

$$C_\mathrm{p}(T) = C_\mathrm{p}(0) + \left[C_\mathrm{p}(\infty) - C_\mathrm{p}(0)\right] y^2 \left[1 + (y-1)\sum_{i=0}^{3} a_i y^i\right] \tag{2.3}$$

as the heat capacity model This form is particularly advantageous because it guarantees that the correct limits of heat capacity at zero and infinite temperature are

obtained; however, it is less efficient to evaluate. The NASA class uses one or more NASA polynomials to fit the heat capacity:

$$\frac{C_\mathrm{p}(T)}{R} = a_{-2}T^{-2} + a_{-1}T^{-1} + a_0 + a_1T + a_2T^2 + a_3T^3 + a_4T^4 \tag{2.4}$$

The NASA polynomial is very fast to evaluate, but has no physical interpretation; often multiple polynomials are required to accurately model a wide temperature range. The `ThermoData`, `Wilhoit`, and NASA classes all inherit from a common base class `HeatCapacityModel`, which stores common attributes and defines a common interface for computing the heat capacity, enthalpy, entropy, and Gibbs free energy. This allows species to interchangeably use any of the available classes.

Converting between the various thermodynamics models is something that occurs frequently in RMG jobs. For example, we often construct a thermodynamics model from group additivity in the `ThermoData` format, but wish to save it in the NASA format for use in chemical kinetics modeling software, such as Chemkin and Cantera. Particular unit tests were designed to test the conversions between each format.

2.7 References

[1] URL: https://www-pls.llnl.gov/?url=science_and_technology-chemistry-combustion-mechanisms.

[2] URL: http://c3.nuigalway.ie/mechanisms.html.

[3] 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*. Ed. by M. J. Pilling. Vol. 35. Comprehensive Chemical Kinetics. Elsevier, 1997, pp. 293–437. DOI: 10.1016/S0069-8040(97)80019-2.

[4] V. Warth, F. Battin-Leclerc, R. Fournet, P. A. Glaude, G. M. Côme, and G. Scacchi. "Computer Based Generation of Reaction Mechanisms for Gas-Phase Oxidation." *Comp. & Chem.* 24 (5), 2000, pp. 541–560. DOI: 10.1016/S0097-8485(99)00092-3.

[5] S. Pierucci and E. Ranzi. "A Review of Features in Current Automatic Generation Software for Hydrocarbon Oxidation Mechanisms." *Comp. & Chem. Eng.* 32 (4-5), 2008, pp. 805–826. DOI: 10.1016/j.compchemeng.2007.03.005.

[6] Y. Yoneda. "A Computer Program Package for the Analysis, Creation, and Estimation of Generalized Reactions – GRACE. I. Generation of Elementary Reaction Network in Radical Reactions – A/GRACE(I)." *Bull. Chem. Soc. Japan* 52 (1), 1979, pp. 8–14. DOI: 10.1246/bcsj.52.8.

[7] F. P. Di Maio and P. G. Lignola. "KING, a KInetic Network Generator." *Chem. Eng. Sci.* 47 (9-11), 1992, pp. 2713–2718. DOI: 10.1016/0009-2509(92)87118-A.

[8] R. J. Quann and S. B. Jaffe. "Structure-Oriented Lumping: Describing the Chemistry of Complex Hydrocarbon Mixtures." *Ind. Eng. Chem. Res.* 31 (11), 1992, pp. 2483–2497. DOI: 10.1021/ie00011a013.

[9] R. J. Quann and S. B. Jaffe. "Building Useful Models of Complex Reaction Systems in Petroleum Refining." *Chem. Eng. Sci.* 51 (10), 1996, pp. 1615–1635. DOI: 10.1016/0009-2509(96)00023-1.

[10] C. Chevalier, J. Warnatz, and H. Melenk. "Automatic Generation of Reaction Mechanisms for the Description of the Oxidation of Higher Hydrocarbons." *Ber. Bunsenges. Phys. Chem.* 94, 1990, pp. 1362–1367.

[11] C. Chevalier, W. Pitz, J. Warnatz, C. Westbrook, and H. Melenk. "Hydrocarbon Ignition: Automatic Generation of Reaction Mechanisms and Applications to Modeling of Engine Knock." *Symposium (International) on Combustion* 24 (1), 1992, pp. 93–101. DOI: 10.1016/S0082-0784(06)80016-0.

[12] L. J. Broadbelt, S. M. Stark, and M. T. Klein. "Computer Generated Pyrolysis Modeling: On-the-Fly Generation of Species, Reactions, and Rates." *Ind. Eng. Chem. Res.* 33 (4), 1994, pp. 790–799. DOI: 10.1021/ie00028a003.

[13] L. J. Broadbelt, S. M. Stark, and M. T. Klein. "Termination of Computer-Generated Reaction Mechanisms: Species Rank-Based Convergence Criterion." *Ind. Eng. Chem. Res.* 34 (8), 1995, pp. 2566–2573. DOI: 10.1021/ie00047a003.

[14] L. J. Broadbelt, S. M. Stark, and M. T. Klein. "Computer Generated Reaction Modelling: Decomposition and Encoding Algorithms for Determining Species Uniqueness." *Comp. & Chem. Eng.* 20 (2), 1996, pp. 113–129. DOI: 10.1016/0098-1354(94)00009-D.

[15] E. Ranzi, T. Faravelli, P. Gaffuri, and A. Sogaro. "Low-Temperature Combustion: Automatic Generation of Primary Oxidation Reactions and Lumping Procedures." *Combust. Flame* 102 (1-2), 1995, pp. 179–192. DOI: 10.1016/0010-2180(94)00253-0.

[16] E. S. Blurock. "Reaction: System for Modeling Chemical Reactions." *J. Chem. Inf. Comput. Sci.* 35 (3), 1995, pp. 607–616. DOI: 10.1021/ci00025a032.

[17] T. Ogura, Y. Nagumo, A. Miyoshi, and M. Koshi. "Chemical Kinetic Mechanism for High Temperature Oxidation of Butane Isomers." *Energy & Fuels* 21 (1), 2007, pp. 130–135. DOI: 10.1021/ef0603151.

[18] T. Ogura, Y. Sakai, A. Miyoshi, M. Koshi, and P. Dagaut. "Modeling of the Oxidation of Primary Reference Fuel in the Presence of Oxygenated Octane Improvers: Ethyl Tert-Butyl Ether and Ethanol." *Energy & Fuels* 21 (6), 2007, pp. 3233–3239. DOI: 10.1021/ef700321e.

[19] 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.

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

[21] E. R. Ritter and J. W. Bozzelli. "THERM: Thermodynamic Property Estimation for Gas Phase Radicals and Molecules." *Int. J. Chem. Kin.* 23 (9), 1991, pp. 767–778. DOI: 10.1002/kin.550230903.

[22] E. R. Ritter and J. W. Bozzelli. "RADICALC: A Computer Code to Estimate Entropies and Heat Capacities for Radical Species and Transition State Structures." *Chem. Phys. Processes Combust.* 52, 1993, pp. 459–462.

[23] J. R. Downey, D. J. Frurip, M. S. LaBarge, A. N. Syverud, N. K. Grant, M. D. Marks, B. K. Harrison, W. H. Seaton, D. N. Treweek, and T. B. Selover. *The ASTM Program for Chemical Thermodynamic and Energy Release Evaluation: CHETAH*. National Institute of Standards and Technology, 1994.

[24] C. Muller, V. Michel, G. Scacchi, and G. M. Come. "THERGAS: A Computer Program for the Evaluation of Thermochemical Data of Molecules and Free Radicals in the Gas Phase." *J. Chim. Phys.* 92 (5), 1995, pp. 1154–1178.

[25] 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.

[26] J. M. Grenda, A. M. Dean, P. Peczak, and W. H. Green. "Recent Advances in Computational Chemical Kinetics Mechanism Generation Techniques." *Twenty-seventh Symposium on Combustion*. 1998.

[27] J. M. Grenda, J. W. Bozzelli, and A. M. Dean. "Automated Methods of Treating Chemically Activated Reactions in Kinetics Mechanism Generation." *Eighth International Conference on Numerical Combustion*. 2000.

[28] 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.

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

[30] D. M. Matheu. "Integrated Pressure-Dependence in Automated Mechanism Generation: A New Tool for Building Gas-Phase Kinetic Models." PhD thesis. Massachusetts Institute of Technology, 2003.

[31] C. D. Wijaya. "Developing Fundamentally Based Models for Autoignition." PhD thesis. Massachusetts Institute of Technology, 2003.

[32] P. E. Yelvington. "Design of a Viable Homogenous-Charge Compression-Ignition (HCCI) Engine: A Computation Study with Detailed Chemical Kinetics." PhD thesis. Massachusetts Institute of Technology, 2005.

[33] J. Yu. "Estimation Method for the Thermochemical Properties of Polycyclic Aromatic Molecules." PhD thesis. Massachusetts Institute of Technology, 2005.

[34] S. V. Petway. "Uncertainty Analysis in Automatic Reaction Mechanism Generation: Neopentyl + 02." MA thesis. Massachusetts Institute of Technology, 2006.

[35] R. W. Ashcraft. "Ab Initio Modeling of Complex Aqueous and Gaseous Systems Containing Nitrogen." PhD thesis. Massachusetts Institute of Technology, 2008.

[36] S. Sharma. "Predictive Modeling of Combustion Processes." PhD thesis. Massachusetts Institute of Technology, 2009.

[37] C. F. Goldsmith. "Predicting Combustion Properties of Hydrocarbon Fuel Mixtures." PhD thesis. Massachusetts Institute of Technology, 2010.

[38] M. R. Harper. "Automated Reaction Mechanism Generation: Data Collaboration, Heteroatom Implementation, and Model Validation." PhD thesis. Massachusetts Institute of Technology, 2011.

[39] G. R. Magoon. "Automated Reaction Mechanism Generation: Improving Accuracy and Broadening Scope." PhD thesis. Massachusetts Institute of Technology, 2012.

[40] J. W. Allen, C. F. Goldsmith, and W. H. Green. "Automatic Estimation of Pressure-Dependent Rate Coefficients." *Phys. Chem. Chem. Phys.* 14, 2012, pp. 1131–1155. DOI: 10.1039/c1cp22765c.

[41] R. H. W. Amrit Jalan and W. H. Green. "An Extensible Framework for Capturing Solvent Effects in Computer Generated Kinetic Models." *J. Phys. Chem. B*, 2013 (submitted).

[42] G. R. Magoon and W. H. Green. "Design and Implementation of a Next-Generation Software System for On-the-Fly Quantum and Force Field Calculations in Automated Reaction Mechanism Generation." *Comp. & Chem. Eng.* 2013 (submitted).

[43] K. M. V. 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.

[44] 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.

[45] R. H. West, C. F. Goldsmith, M. R. Harper, W. H. Green, L. Catoire, and N. Chaumeix. "Kinetic Modeling of Methyl Formate Oxidation." *7th US National Technical Meeting of the Combustion Institute*. 2011.

[46] M. R. Harper, K. M. V. 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.

[47] N. Hansen, M. R. Harper, and W. H. Green. "High-Temperature Oxidation Chemistry of n-Butanol – Experiments in Low-Pressure Premixed Flames and Detailed Kinetic Modeling." *Phys. Chem. Chem. Phys.* 13, 2011, pp. 20262–20274. DOI: 10.1039/C1CP21663E.

[48] S. S. Merchant, E. F. Zanoelo, R. L. Speth, M. R. Harper, K. M. V. Geem, and W. H. Green. "Combustion and Pyrolysis of iso-Butanol: Experimental and Chemical Kinetic Modeling Study." *Combust. Flame*, 2013 (submitted).

[49] N. Hansen, S. S. Merchant, M. R. Harper, and W. H. Green. "The Predictive Capability of an Automatically Generated Combustion Chemistry Mechanism: Chemical Structures of Premixed iso-Butanol Flames." *Combust. Flame*, 2013 (submitted).

[50] G. R. Magoon, J. Aguilera-Iparraguirre, W. H. Green, J. J. Lutz, P. Piecuch, H.-W. 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.

[51] K. Beck. *Extreme Programming Explained: Embrace Change*. Second edition. Addison-Wesley, 2004.

[52] S. Behnel, R. Bradshaw, C. Citro, L. D. 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.

[53] D. C. Ince, L. Hatton, and J. Graham-Cumming. "The Case for Open Computer Programs." *Nature* 482, 2012, pp. 485–488. DOI: 10.1038/nature10836.

[54] S. Sharma, M. R. Harper, and W. H. Green. *CanTherm*. 2010. URL: https://github.com/GreenGroup/CanTherm.

[55] P. Foggia, C. Sansone, and M. Vento. "A Performance Comparison of Five Algorithms for Graph Isomorphism." *in Proceedings of the 3rd IAPR TC-15 Workshop on Graph-based Representations in Pattern Recognition*. 2001, pp. 188–199.

[56] 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.

[57] H. L. Morgan. "The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service." *Journal of Chemical Documentation* 5 (2), 1965, pp. 107–113. DOI: 10.1021/c160017a018.

[58] URL: http://www.iupac.org/home/publications/e-resources/inchi.html.

[59] D. Weininger. "SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules." *J. Chem. Inf. Comput. Sci.* 28 (1), 1988, pp. 31–36. DOI: 10.1021/ci00057a005.

[60] D. Weininger, A. Weininger, and J. L. Weininger. "SMILES. 2. Algorithm for Generation of Unique SMILES Notation." *J. Chem. Inf. Comput. Sci.* 29 (2), 1989, pp. 97–101. DOI: 10.1021/ci00062a008.

[61] D. Weininger. "SMILES. 3. DEPICT. Graphical Depiction of Chemical Structures." *J. Chem. Inf. Comput. Sci.* 30 (3), 1990, pp. 237–243. DOI: 10.1021/ci00067a005.

[62] P. Murray-Rust, H. S. Rzepa, and M. Wright. "Development of Chemical Markup Language (CML) as a System for Handling Complex Chemical Content." *New J. Chem.* 25, 2001, pp. 618–634. DOI: 10.1039/B008780G.

[63] N. O'Boyle, M. Banck, C. James, C. Morley, T. Vandermeersch, and G. Hutchison. "Open Babel: An Open Chemical Toolbox." *J. Cheminformatics* 3 (1), 2011, p. 33. DOI: 10.1186/1758-2946-3-33.

[64] URL: http://github.com/jwallen/PyDAS.

3

Evaluating the accuracy of automatic kinetics estimation methods

One of the primary challenges in developing detailed kinetics models, either by hand or automatically, is the estimation of the physical parameters associated with the chemical species and reactions in the model. There are usually many parameters to estimate for a given model; for example, modern combustion mechanisms routinely contain hundreds of species and thousands of reactions, and automatic mechanism generation codes can easily consider orders of magnitudes more of each when building a mechanism. Today, our automatic mechanism generation code, RMG [1], is often estimating $\sim 10^6$ reaction rate coefficients in the course of building a model. In the near future, it is expected that a parallelized version of RMG will routinely be considering $\sim 10^8$ reactions while constructing models. Clearly we need methods of estimating chemical parameters that are reasonably accurate, but also very quick to apply, since we will be repeating them often while constructing the mechanism. Accuracy is especially important in automatic mechanism generation codes that utilize a rate-based method to discriminate between important and unimportant species when building the model [2], as the parameters directly affect the model composition. If our parameter estimates are too far off, we may miss important reaction pathways or include pathways that are actually not important.

Estimating kinetics parameters is particularly challenging due to the paucity of available kinetic data from the literature. There are only $\sim 10^4$ reactions whose rate coefficients have been directly measured experimentally over some range of conditions (which may or may not be similar to the conditions you are interested in). Significant progress has been made in applying *a priori* quantum chemistry calculations to the calculation of reaction rates, but these calculations often have significant – and poorly known – uncertainties [3], and at present these calculations are far too slow to bridge the gap between the $\sim 10^6$ to $\sim 10^8$ rate coefficients needed and the $\sim 10^4$

rate coefficients that are known. This means that we are often required to extrapolate kinetics estimates for novel functionalities at unstudied temperatures and pressures. Fortunately, the number of rate-controlling reactions that significantly affect the model predictions is generally quite small, even for large mechanisms. For the other, non-sensitive parameters, a more approximate estimate of the kinetics is often sufficient to establish that the reaction is either too slow to occur on the relevant timescales or so fast that it reaches thermodynamic equilibrium on the relevant timescales. In either of these cases, the precise values of the kinetics parameters are much less important.

It is important to know the uncertainty of an estimated parameter, e.g. so that we can identify if discrepancies between model predictions and experimental observations are due to poor parameter estimates or instead to missing chemistry. The uncertainty of thermodynamics parameters, as estimated using Benson group contribution theory [4–7], is generally thought to be around 2-5 kcal/mol for most species, and higher for more complex polycyclics [8]. However, the uncertainty of kinetics parameter estimation methods is not as well-known.

In this chapter, we shed some light on the current accuracy of high-throughput kinetics estimation methods by comparing their predictions against a test set of reactions obtained from the NIST Chemical Kinetics Database [9]. Three kinetics estimation methods are discussed: the classic Evans-Polanyi method, the "rate rules" method used in RMG, and a new method based on kinetics group additivity [10–19]. A common framework for implementing both of these methods is presented that emphasizes the separation of the training set from the functional groups that define the kinetics estimation method. This framework also places the two methods on an equal footing for the purposes of comparing their overall accuracy. The accuracy of each method is evaluated for three reaction families common in gas-phase combustion chemistry: hydrogen abstractions, radical additions to multiple bond (the reverse of β-scissions), and radical-radical recombinations.

3.1 Background

3.1.1 The Evans-Polanyi method

Several methods of estimating high-pressure limit rate coefficients for gas-phase reactions have been developed previously. Among the earliest was the Evans-Polanyi method for correlating activation energies E_a to enthalpies of reaction ΔH_{rxn} using the linear relationship [20, 21]

$$E_a = \alpha \Delta H_{rxn} + E_0 \tag{3.1}$$

with parameters α and E_0. Usually the preexponential factor is assumed to be the same for all similar reactions. Several variations of the above expression also exist [22]. Perhaps the most obvious of these variations – and the one used in this work – is to simply require that the activation energy be nonnegative (for exothermic reactions) and be at least as large as the enthalpy of reaction (for endothermic reactions), *i.e.*

$$E_a = \begin{cases} 0 & \Delta H_{rxn} < -E_0/\alpha \\ \alpha \Delta H_{rxn} + E_0 & -E_0/\alpha < \Delta H_{rxn} < E_0/\alpha \\ \Delta H_{rxn} & \Delta H_{rxn} > E_0/\alpha \end{cases} \tag{3.2}$$

The choice $\alpha = \frac{1}{2}$ makes this formula work for both forward and reverse reactions, satisfying the expected relation; this is probably useful for reaction families such as hydrogen abstraction where the forward and reverse reactions can be identical.

3.1.2 The rate rules method

The success of the Benson group contribution theory for estimating thermodynamic properties has encouraged the development of methods that utilize the transition state structure to estimate the kinetics. The rate coefficient of a reaction is largely determined by the atoms in the region around its transition state. This region, containing several polyvalent atoms, can be called a "supergroup" [10, 23]. The supergroup can be decomposed into component groups. For example, in hydrogen abstraction reactions X–H + Y· ⟶ X· + Y–H, the component groups would be the abstracting group (Y) and the group from which a hydrogen is abstracted (X). Once one defines N_X "X" groups and N_Y "Y" groups, there are $N_X N_Y$ X–H–Y· supergroups which each will have a different rate coefficient $k(T)$. Different methods exist for associating kinetic parameter values with supergroups.

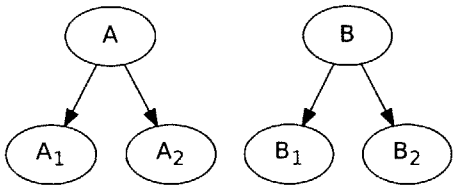
In the so-called "rate rules" method used in the RMG software package, a database of rules that associate an entire supergroup X–H–Y· with a set of rate parameters is constructed. These functional groups are placed in a hierarchical tree such that the children of a given group define mutually exclusive functionalities that are more specific than the parent. To estimate the kinetics of a reaction, the most specific functional groups that describe that reaction are determined by descending the tree as far as possible. These functional groups are then used to identify the best matching

rate rule. When an exact rate rule is not available, the rates of rules with the most similar functional groups, as determined by the degree of separation in the hierarchy, are averaged together. This approach has the advantage of being extensible, and as additional rate rules are added, related rate estimates are improved. However, the trees grow exponentially, so a very large number of rate rules are needed to provide reliably accurate estimates for all possible reactions. At present the trees are only sparsely populated with rate estimates.

The rate rules used in this work were originally constructed from high-level quantum chemistry calculations of Sumathi *et al.* [10–12, 23, 24], Wijaya *et al.* [25, 26], and Saeys *et al.* [17] Some rules were also created from rate estimates from the *n*-heptane oxidation mechanism of Curran *et al.* [27] Additional rules have since been added from the high-level quantum chemistry calculations of Harper *et al.* [28] and from Vandeputte [29]. Averaged rules are also constructed for supergroups higher in the hierarchy by averaging together rules from the more specific supergroups below them. Otherwise, the algorithm will tend to fall all the way up the tree to the rule associated with the most general supergroup, which has a generic rule that is less likely to be accurate for the specific reaction of interest.

Although this method performs reasonably well – as we will demonstrate later in this chapter – there are a few important shortcomings. First, the method as currently implemented requires that the functional groups be assigned manually during rule creation. This is not always straightforward, especially for reactant molecules with multiple nearby functional groups. Currently the rule creator must decide immediately which functionality is most important when constructing the rule. The information about the original reactant species is discarded, making it difficult to change the chosen functional group hierarchy.

Second, the steps to take when estimating a rate for a reaction that does not match one of the existing rate rules are not obvious. Consider the simple hierarchy of functional groups for a bimolecular reaction. The two primary groups are labeled A and B:



Let us say we are trying to estimate the kinetics of a reaction that matches functional group pair (A_2,B_2), but that the database only has a rate rule for the (A_1,B_1) pair. Since we have a hierarchy of functional groups, we can fall up to a more general functional group in order to obtain a kinetics estimate. However, we don't necessarily want to make *both* groups more generic, as the pairs (A,B_2) and (A_2,B) are a better description of the functional groups than the more general (A,B). The current implementation in RMG is to average together all of the available rules that are a minimum number of functional group generalization steps (often called a "distance" for brevity) from the most specific groups. In the worst case, we would fall all the way up to the most general groups, where we have placed a very general rate rule so that a kinetics estimate can always be generated.

If we only use the explicitly-stored rate rules, the kinetics estimation algorithm will usually fall all the way up to the most general functional groups unless it happens to match an exact rule, which is uncommon. The rule for the most general functional groups is going to have a very large uncertainty. To remedy this, we use our hierarchy of functional groups to fill in estimated rate rules between the most general groups and the exact rate rules by averaging together the exact rules. In our example, we could construct a rule for (A,B) by averaging together the rules for (A_1,B_1), (A_1,B_2), (A_2,B_1), and (A_2,B_2). Unusually, the current RMG implementation does not fill in averaged rate rules for (A,B_1), (A,B_2), (A_1,B), or (A_2,B) in this manner. These two sources of averaging – caused by the complexity of having to work with multiple functional group trees – make it very difficult to clearly document the method by which the averaged rate rule was constructed, and make it difficult to know the uncertainty of the estimate.

3.1.3 Kinetics group additivity method

More recently, several efforts have been made to generate a group contribution method for kinetics that mimics that of the Benson method for thermodynamics. In these approaches, the effect on the rate expression from each of the component groups is assumed to be independent and additive. By assuming the effects of X and Y are independent, one obtains $N_X + N_Y$ distinct $k(T)$ values, many fewer than the $N_X N_Y$ distinct values required by the rate rules method. Many fewer estimation rules are therefore required to fill in these smaller trees. Separating the contributions from each functional group also makes it much easier to understand how to fall up the functional group tree if there is no group additivity value for any of the most specific functional groups, since we can move up each tree independently. However, the accuracy of the separability approximation is unknown.

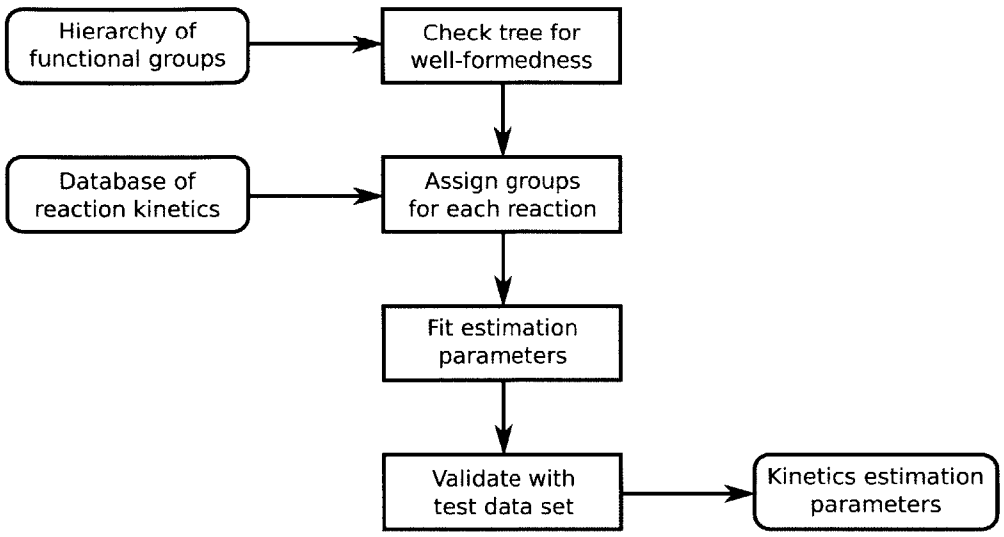


Figure 3.1: A general framework for developing kinetics estimation methods. After validating the input data and functional groups for correctness (see text for details), the reactions are classified by functionality. This can then be used to determine the parameters for the desired method, possibly using an optimization program. If a separate set of test data is available, it can be used to ensure the fit was done correctly and get an idea of the uncertainty in the kinetics estimates. Particularly important is the separation of the database of training data from the hierarchy of functional groups, which allows for easy recalculating of the method parameters when a change to either input is made.

Sumathi *et al.* developed a set of Benson-like group additivity parameters that enable the estimation of the thermodynamic properties of the transition state [10–12]. Independent efforts by Willems and Froment [13, 14], Truong *et al.* [15, 16], Saeys *et al.* [17], and Sabbe *et al.* [18, 19] have produced methods that utilize group contribution theory to apply corrections to a reference reaction of the same class. Willems and Froment, Saeys *et al.*, and Sabbe *et al.* represent their corrections using Arrhenius parameters (preexponential and activation energy), while Truong *et al.* use factors for tunneling, reaction symmetry, partition function, and potential energy. In this work we use an adaptation of the method of Saeys *et al.*and Sabbe *et al.* where we fit modified Arrhenius parameters A, n, and E_a instead of just A and E_a.

3.2 A general kinetics estimation framework

Our implementations of kinetics estimation methods using the structure of the transition state are based in the framework shown in Figure 3.1. There are two inputs:

a hierarchy of functional groups used to classify reactions and a database containing kinetic information for our method to train against. A key contribution of this work was the recognition that these two inputs ought to be kept independent. This way, the kinetics estimation method can easily be retuned as a result to a change in either the tree of functional groups or the training data set. The training of the estimation method can be fully automated, which lowers the opportunity for human errors.

The hierarchy of functional groups is first checked for validity. A valid tree requires that (1) all children of a given parent define functional groups that are more specific than the parent functional group and (2) all siblings define functional groups that are mutually exclusive to one another. Note, however, that the siblings do not need to be mutually *exhaustive*; that is, they do not need to enumerate all of the possible substructures of the common parent that they belong to. Requiring otherwise would defeat the primary purpose of using the hierarchy in the first place: extensibility.

Once the functional group tree is known to be well-formed, it can be used to classify the functional groups of each reaction in the training set. This step is exactly the same as when the estimation algorithm is applied to estimate a particular reaction, so nothing new is required. The functional groups and associated kinetics for each reaction are then used to train the kinetics estimation method. Clearly this step will be different depending on the choice of method. For the rate rule method, we would simply output the functional groups and corresponding parameters as the rules themselves, perhaps in addition to manually-defined rules for the more generic functional groups nearer to the top of the tree. A group additivity-based method would require constructing and solving an optimization problem to fit the group additivity values for each functional group.

The final step is to evaluate the newly-trained kinetics estimation method using a test set of reactions that were not used to train the method. This provides an idea of the uncertainty in the method (and a means to check if the fitted parameters are reasonable). In the remainder of this chapter, we will do exactly this for the Evans-Polanyi, rate rules, and group additivity methods of estimating kinetics.

3.3 Methodology

The original test set for evaluting the accuracy of the kinetics estimation methods was generated using the entries in the PrIMe data warehouse [30]. However, this test set was quickly revealed to have a few shortcomings. First, many of the entries were found to have been transcribed incorrectly from the literature, often with incorrect units for

the Arrhenius preexponential – causing errors as large as an Avogadro number! – or the activation energy. These were identifed as clear outliers when plotted on parity plots against the RMG-estimated rate coefficients. (Incidentally, this demonstrates that this methodology can also be used as a sanity check on the parameters of a test set such as PrIMe.) Second, and more importantly, the PrIMe data was missing the information about the valid temperature range of each kinetics entry. As many of the entries come from direct experimental measurements over a limited range of temperatures, it would be inappropriate to use these values outside of the experimental range in performing the evaluation. The PrIMe entries provided no means of doing so.

For these reasons, a test set of reactions was created from the NIST Chemical Kinetics Database [9]. Entries in the NIST database store the reactants and products (as CAS numbers), the Arrhenius parameters, the valid temperature range, and details about the literature source; however, they do not store the reaction family or the transition state supergroup structure. To determine these for each reaction, we used RMG to generate all possible reactions involving the reaction's reactants that lead to the reaction's products. If this resulted in a single unique reaction, the entry was kept and assigned that reaction's family and supergroup. Only reactions involving C/H/O atoms with kinetics that represent the high-pressure limit were kept for the purposes of this comparison. Many reactions have multiple entries in the NIST database, each with different values for the kinetics parameters due to the differing methods used to determine them (experiment, theory, and/or review). We decided to keep all of these entries instead of simply selecting the one we thought was the most accurate.

Figure 3.2 shows an illustration of the kinetics estimation procedures used in this work, using as an example the hydrogen abstraction reaction $CH_3COCH_3 + CH_3 \longrightarrow CH_4 + CH_3COCH_2$. This reaction is identified by searching the reactants for the functional groups that match the most general "root" template; for hydrogen abstraction, we search for hydrogen donor X–H and acceptor Y· functionalities. Each general functional group is the top node in a tree of functional groups of increasing specificity, which can be descended to uniquely identify the functional groups that best describe the reaction. For the example reaction, the hydrogen donor group X–H is identified to by a carbon atom with only single bonds (labeled Cs_H in the figure), a primary carbon atom (C_pri), and ultimately a primary carbon atom adjacent to a carbonyl group (C/H3/CO). Note that the most specific functional group does not completely describe the reactant molecule, as it does not say anything about what functionality occupies the nonreactive carbonyl bond. The hydrogen acceptor group Y· is determined to be a carbon radical with only single bonds (Cs_rad) and a methyl

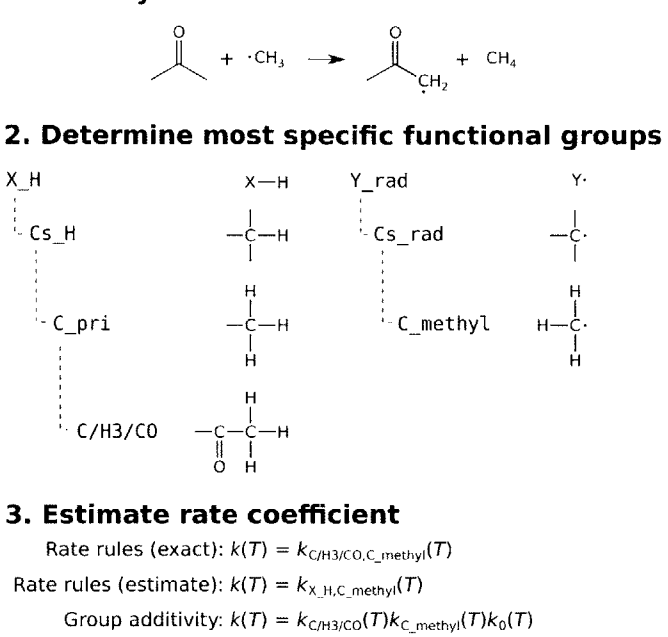


Figure 3.2: An illustration of the kinetics estimation procedures used in this work. A reaction is identified by the participating functional groups; in this hydrogen abstraction example, the groups are the hydrogen donor X—H and acceptor Y·. Each general functional group is the top node in a tree of functional groups of increasing specificity, which can be descended to find the functional groups that best describe the reaction. Each kinetics estimation method uses these specific groups to construct the reaction rate coefficient.

radical (C_methyl). In this case the radical is small enough that the functional group describes it completely.

The rate rules and group additivity methods use these specific groups to construct the reaction rate coefficient. In the rate rules method, the database is checked to see if there is an exact rule for the most specific groups (C/H3/CO, C_methyl); if so, this rule is used for the rate coefficient. If not, the algorithm will check for the existence of rules involving more general pairs of functional groups, such as (C_pri, C_methyl) or (Cs_H, Cs_rad). Some of these more general combinations have been filled in by averaging together rules from the more specific child nodes. If multiple rules are found for more general functional group pairs, the rules "closest" to the most specific group – as determined by the number of generalizations required to move from the most specific group to the more general group where a rule exists – are averaged together. A general rate rule is always provided for the root template, which ensures that a rate estimate can always be made.

The group additivity method again starts with the most specific groups (C/H3/CO, C_methyl), but this time it checks the database for the contributions of each group separately to the overall rate coefficient. If no numerical value exists for one of the groups, the algorithm ascends the tree until it finds a more general group with a value and uses that. Since the contributions of each group are (assumed to be) independent, each group value can be looked up independently.

The rate rule method was used as currently implemented in RMG 4.0 [1], using all of the rate rules provided in that release. The group additivity method parameters were determined by training against the same rate rules, in order to ensure the two methods use the same training set. The Evans-Polanyi method parameters could not be trained against the existing rate rules, as without knowing the actual reactants and products involved, it was not possible to determine the enthalpy of reaction for each rule. Instead, the subset of the NIST kinetics database that matched one of our exact rate rules was used as a training set. This way, we could still use the same test set as with the other methods: the subset of the NIST database that can be classified as a type of reaction known to RMG, but where RMG is missing the corresponding rate rule.

3.4 Results and discussion

The three reaction families with the highest number of entries in the NIST database involving only C/H/O atoms were hydrogen abstractions (with 1318 entries), radical additions to multiple bonds (475), and radical-radical recombinations (519). Note that the latter two reaction families can have pressure-dependent rate coefficients. These reactions are common in gas-phase pyrolysis and combustion mechanisms. We therefore focused our attention on evaluating the accuracy of kinetics estimation methods for these three families.

3.4.1 Evans-Polanyi plots

Evans-Polanyi plots for the hydrogen abstraction and radical addition to multiple bond families are shown in Figure 3.3. The points represent the entire NIST data set, including both entries that matched exact rate rules and those that did not. The solid lines represent the Evans-Polanyi parameters fitted to only the exact matches, while the corresponding dashed lines indicate the 95% confidence intervals. The piecewise linearity of the Evans-Polanyi model reflects the constraints that the activation energy be nonnegative and be at least as large as the enthalpy of reaction, as shown in Equation (3.2). In both families, the fitted parameters can be used to predict the activation

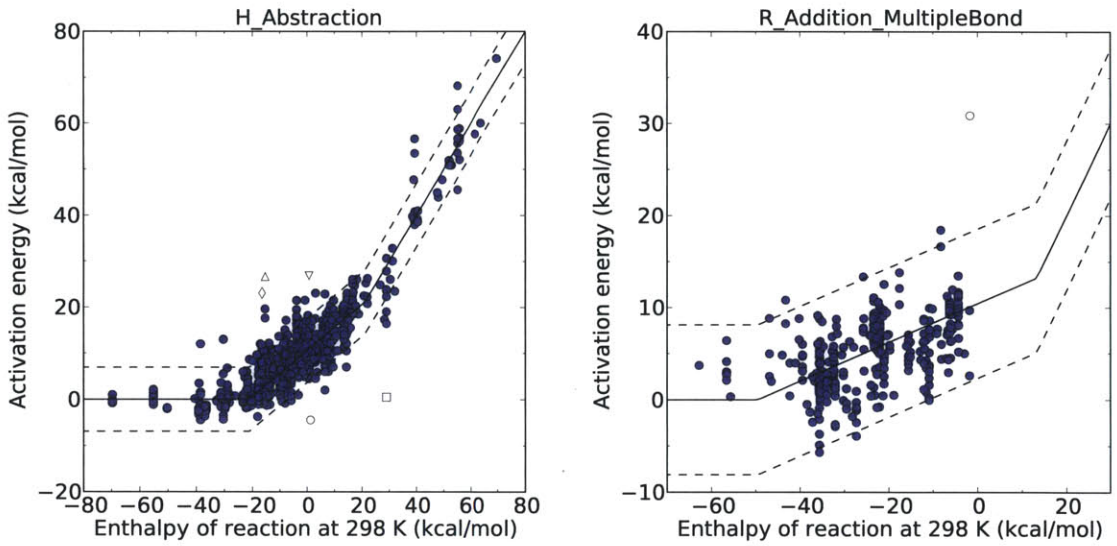
energy to within about 8-10 kcal/mol, though there are several significant outliers. The Evans-Polanyi correlation is not useful for the radical-radical recombination family, as these are generally barrierless reactions.

In fitting the Evans-Polanyi parameters for hydrogen abstraction reactions, we have constrained α to a value of exactly 0.5, with the fitted E_0 having a value of 10.6 kcal/mol. If we also allow α to be optimized, the results are α = 0.447 and E_0 = 10.2 kcal/mol. Both fits give a 95% confidence limit range of 6.9 kcal/mol, with the α-optimized version slightly below that value and the other slightly above. We will therefore contiune to use the constrained α = 0.5 fit in this work, though selecting the other fit gives virtually identical results.

Several outliers on the Evans-Polanyi plot in Figure 3.3(a) for hydrogen abstraction reactions are worthy of comment as illustrative examples. The open circle point corresponds to the reaction $C_2H_5OH + H \longrightarrow H_2 + C_2H_5O$ as fit to $k(T)$ values calculated by Park *et al.* [31]. They performed the fit over the range 300-1000 K, which they demonstrated to have very large curvature due to tunneling, and the resulting fit has an n value of 10.58 and an E_a value of -4.5 kcal/mol in order to capture this tunneling effect. Park *et al.* also performed a separate fit over the range 1000-3000 K – included as a separate entry in the NIST database – where the tunneling is much less important; the resulting E_a of 8.8 kcal/mol is much closer to the Evans-Polanyi estimate of 10.7 kcal/mol.

Two of the outliers are the result of fitting small mechanisms to experimental concentration profiles from thermal decomposition experiments. The square point corresponds to the reaction $C_2H_2 + H \longrightarrow H_2 + C_2H$ as fitted to data from the thermal decomposition of acetylene by Tanzawa and Gardiner [32]. The triangle up point corresponds to the reaction i-$C_4H_8 + CH_3 \longrightarrow CH_4 +$ i-C_4H_7 as fitted to data from the thermal decomposition of neopentane by Bradley and West [33]. As these results from from fitting many rate coefficients rather than isolating a single reaction, the fit may not be representative of the underlying fundamental kinetics, and probably does not extrapolate. Note that the fitting can cause the error to occur in either direction, as the former gives a much lower activation energy than expected from the Evans-Polanyi correlation, while the latter gives a much higher value than expected.

An additional pair of outliers occur due to fitting of Arrhenius parameters to high-temperature experimental data. The diamond point corresponds to the reaction $HCHO + CH_3 \longrightarrow CH_4 + CHO$, whose literature value of 23.1 kcal/mol results from a fit to experimental measurements in the range 1170-1630 K by Choudhury *et al.* [34]. The triangle down point corresponds to the reaction $CH_4 + H \longrightarrow H_2 + CH_3$, whose ac-



(a) (b)

Figure 3.3: Evans-Polanyi plots for the (a) hydrogen abstraction and (b) radical addition to multiple bond families using the NIST kinetics database. The fitted Evans-Polanyi curves are also shown, along with the 95% confidence intervals. Both methods can generally estimate the activation energy to within around 10 kcal/mol.

tivation energy of 26.8 kcal/mol was obtained by fitting to methane-oxygen-air flame data over the range 1325-1700 K by Biordi *et al.* [35]. In both cases the observed $k(T)$ values are very similar to those in many other experiments, as seen in the Baulch *et al.* review [36]. The recommended Arrhenius expressions from Baulch *et al.*, which incorporate these and many other experiments, give activation energies of 4.3 and 9.6 kcal/mol, respectively, which are much closer to the predicted activation energies of 2.8 and 10.5 kcal/mol.

There is also one notable outlier on the Evans-Polanyi plot for radical addition to multiple bond reactions, as indicated in Figure 3.3(b) using an open circle marker. This outlier represents the reaction $H + CO_2 \longrightarrow HOCO$, as estimated by Larson *et al.* using an RRKM/master equation model [37]. The authors indicate that they fitted the barrier heights in order to match some available experimental data obtained at low pressures; their adjusted barriers are much larger than those computed more recently using high-level quantum chemistry calculations [38]. Accordingly, the activation energy reported by Larson *et al.* for the high-pressure limit is about 20 kcal/mol larger than that predicted by our Evans-Polanyi fit. This could perhaps justify removal of this entry from the test set, but we have left it in for now as an example of how this analysis can serve as a validation check.

Table 3.1: Fitted top-level rate rules for each reaction family.

Family	A[a]	n	α	E_0
Hydrogen abstraction				
Original	1.0×10^{5}	0.0	0.0	10.0
Updated	4.2×10^{1}	3.21	0.5	10.6
Updated	1.9×10^{12}	0.0	0.5	10.6
Updated	4.3×10^{5}	2.0	0.5	10.6
Radical addition to multiple bond				
Original	1.0×10^{13}	0.0	0.0	0.5
Updated	3.9×10^{9}	0.778	0.207	10.4
Radical-radical recombination				
Original	1.0×10^{13}	0.0	0.0	0.0
Updated	3.4×10^{13}	-0.134	0.0	0.016

[a] The units for A are $cm^3\ mol^{-1}\ s^{-1}$, with E_0 in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp[-(\alpha\Delta H_{rxn} + E_0)/RT]$.

The fitted Evans-Polanyi parameters were used to update the top-level rate rules for the root (most generic) nodes in the three reaction families considered in this study. The Arrhenius preexponential and temperature exponent factors were also updated; this was done by averaging the A and n values of all of the rate rules in the RMG database. The updated Arrhenius and Evans-Polanyi parameters are given in Table 3.1, along with the original parameters for comparison. The updated generic rules for the hydrogen abstraction and radical addition to multiple bond families are significantly different than the original rules, while those for the radical-radical recombination family are within about a factor of two of the original rule. The updated values were used when evaluating the performance of the rate rules method in the next sections, in order to compare the methods on an equal footing.

3.4.2 Kinetics estimation methods

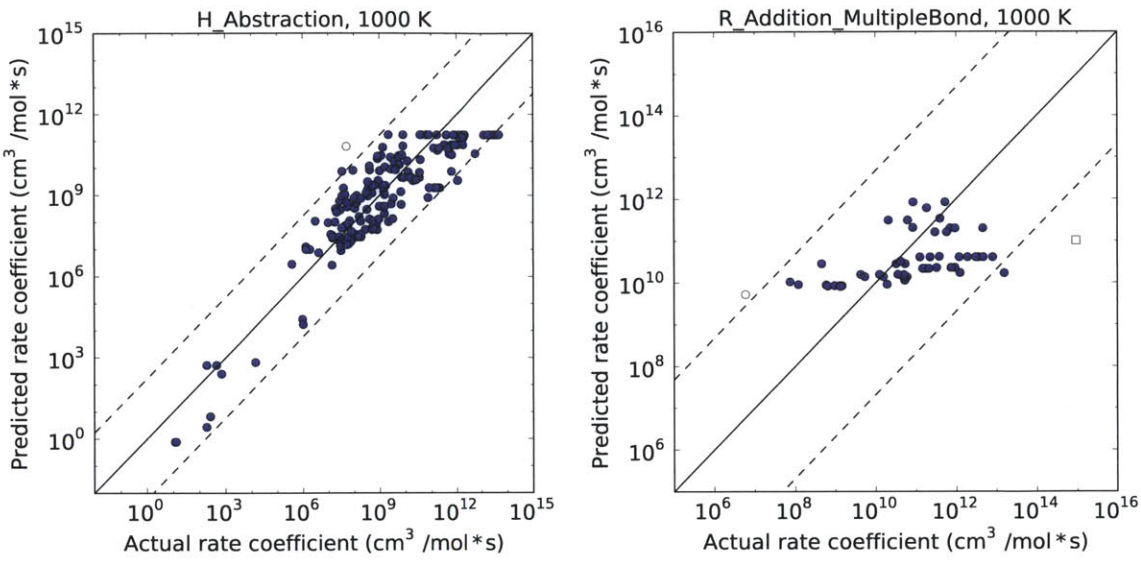
A summary of the observed uncertainties of each kinetics estimation method for each family at several temperatures is shown in Table 3.2. The numbers in parentheses indicate the number of entries in the test set at that temperature, which changes since we are honoring the valid temperature ranges included with each entry. The uncertainty of each method increases significantly as temperature is decreased, as small errors in the activation energy get magnified at lower temperatures.

Table 3.2: Summary of evaluation of automatic kinetics estimation methods

Family	Accuracy (No. of reactions)[a,b] 500 K	1000 K	1500 K	2000 K
Hydrogen abstraction				
Exact rules	$10^{1.6}$ (409)	$10^{1.1}$ (372)	$10^{1.1}$ (303)	$10^{1.1}$ (249)
Estimated rules	$10^{2.4}$ (278)	$10^{1.6}$ (206)	$10^{1.6}$ (179)	$10^{1.7}$ (160)
Group additivity	$10^{4.5}$ (278)	$10^{2.4}$ (206)	$10^{2.3}$ (179)	$10^{2.3}$ (160)
Evans-Polanyi	$10^{3.6}$ (278)	$10^{2.2}$ (206)	$10^{1.9}$ (179)	$10^{1.9}$ (160)
Hydrogen abstraction (n = 0)				
Exact rules	$10^{1.6}$ (409)	$10^{1.1}$ (372)	$10^{1.1}$ (303)	$10^{1.1}$ (249)
Estimated rules	$10^{3.2}$ (278)	$10^{2.0}$ (206)	$10^{1.8}$ (179)	$10^{1.7}$ (160)
Group additivity	$10^{4.7}$ (278)	$10^{2.6}$ (206)	$10^{2.2}$ (179)	$10^{1.9}$ (160)
Evans-Polanyi	$10^{4.5}$ (278)	$10^{2.6}$ (206)	$10^{2.2}$ (179)	$10^{1.9}$ (160)
Hydrogen abstraction (n = 2)				
Exact rules	$10^{1.6}$ (409)	$10^{1.1}$ (372)	$10^{1.1}$ (303)	$10^{1.1}$ (249)
Estimated rules	$10^{2.6}$ (278)	$10^{1.7}$ (206)	$10^{1.6}$ (179)	$10^{1.7}$ (160)
Group additivity	$10^{4.6}$ (278)	$10^{2.5}$ (206)	$10^{2.3}$ (179)	$10^{2.1}$ (160)
Evans-Polanyi	$10^{3.5}$ (278)	$10^{2.2}$ (206)	$10^{1.9}$ (179)	$10^{1.9}$ (160)
Radical addition to multiple bond				
Exact rules	$10^{1.5}$ (155)	$10^{1.5}$ (122)	$10^{1.4}$ (100)	$10^{1.4}$ (84)
Estimated rules	$10^{5.2}$ (66)	$10^{3.4}$ (56)	$10^{2.5}$ (53)	$10^{2.3}$ (50)
Group additivity	$10^{8.6}$ (66)	$10^{5.0}$ (56)	$10^{3.8}$ (53)	$10^{3.2}$ (50)
Evans-Polanyi	$10^{4.1}$ (66)	$10^{2.7}$ (56)	$10^{2.1}$ (53)	$10^{2.1}$ (50)
Radical-radical recombination				
Exact rules	$10^{1.5}$ (38)	$10^{1.0}$ (47)	$10^{1.1}$ (35)	$10^{1.7}$ (26)
Estimated rules	$10^{3.1}$ (91)	$10^{2.8}$ (158)	$10^{2.4}$ (139)	$10^{2.5}$ (101)
Group additivity	$10^{2.5}$ (91)	$10^{2.4}$ (158)	$10^{1.9}$ (139)	$10^{1.9}$ (101)
Evans-Polanyi	$10^{3.0}$ (91)	$10^{2.7}$ (158)	$10^{2.2}$ (139)	$10^{2.3}$ (101)

[a] The accuracy was computed as the 95% confidence interval.

[b] The number of reactions tested differs at each temperature since only reactions from the test set valid at that temperature were kept.



(a) (b)

Figure 3.4: Parity plot comparing the RMG-estimated rate coefficients at 1000 K using the Evans-Polanyi method for (a) hydrogen abstraction and (b) radical addition to multiple bond reactions to the actual values, as taken from the NIST kinetics database. The dashed lines represent the 95% confidence limit.

3.4.2.1 Evans-Polanyi method

In general, the transition state structure affects the reaction kinetics through both entropic (preexponential) and enthalpic (activation energy) effects. The Evans-Polanyi method only adjusts the activation energy, meaning that it can only account for enthalpic effects, and even then is only a very basic linear energy relationship. Nonetheless, its simplicity makes it a good baseline for comparing the accuracy of the more advanced methodologies.

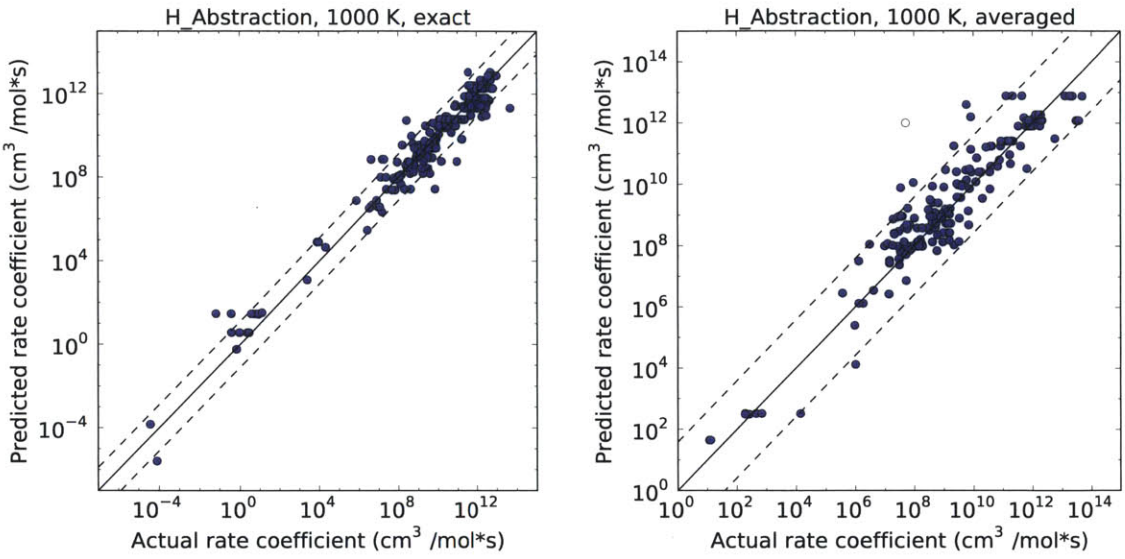
Parity plots comparing the predicted and actual values of the rate coefficient for hydrogen abstraction and radical addition to multiple bond reactions at 1000 K using the Evans-Polanyi method are shown in Figure 3.4. The dashed lines indicate the 95% confidence limit, which we have used as a simple indicator of the overall uncertainty of the method. For hydrogen abstraction reactions, Figure 3.4(a) shows that the uncertainty of a Evans-Polanyi prediction is $10^{2.2}$, more than two orders of magnitude. From the Evans-Polanyi plots, the uncertainty in an activation energy estimated from the Evans-Polanyi expression is about 8 kcal/mol; at 1000 K this corresponds to an uncertainty of around $10^{1.7}$. The difference is likely due to uncertainty in the preexponential factor.

Some of the scatter is due to disagreement in various estimates of the kinetics in

the entries from the NIST database. For example, the hydrogen abstraction reaction $HO_2 + OH \longrightarrow H_2O + O_2$ has 13 entries in the NIST database, with kinetics estimates that span nearly three orders of magnitude, from 6×10^{10} to 4×10^{13} $\mathrm{cm^3/mol \cdot s}$. Judging which of these sources is the "correct" one was beyond the scope of this analysis. Note that the Evans-Polanyi estimate of 2×10^{11} $\mathrm{cm^3/mol \cdot s}$ is safely within the literature range.

An outlier on the Evans-Polanyi method parity plot for hydrogen abstraction, indicated with the open circle, corresponds to the reaction $C_3H_6 + CH_3O \longrightarrow CH_3OH +$ a-C_3H_5. The literature value for this rate comes from Tsang [39], who assigned Arrhenius parameters by analogy to the reaction $C_3H_8 + CH_3O \longrightarrow CH_3OH +$ i-C_3H_7. Thus, the literature value is itself actually an estimate of the kinetics, rather than an *ab initio* calculation or direct experimental measurement. We estimate a bond dissociation energy at 1000 K of 90.7 kcal/mol for the allylic hydrogen in propene and 99.6 kcal/mol for the secondary hydrogen in propane [40], which suggests the relative stabilities of the transition states in the two reactions are significantly different. Adjusting the Tsang estimate to use a more stable transition state would increase the value of the corresponding rate coefficient, which would reduce the deviation from the Evans-Polanyi predicted value. Note that the disagreement between the Tsang estimate and the Evans-Polanyi estimate is outside the order-of-magnitude uncertainty asserted by Tsang for his estimate.

The Evans-Polanyi predictions for the radical addition to multiple bond family at 1000 K, shown in Figure 3.4(b), show a range of predictions of only two orders of magnitude, but with the actual rate coefficients spanning a range of more than eight orders of magnitude. Overall, the uncertainty for this family of $10^{2.7}$ is somewhat larger than that of hydrogen abstraction. Two outliers are marked in Figure 3.4(b). The open circle corresponds to the $H + CO_2 \longrightarrow HOCO$ reaction discussed earlier. A second outlier, indicated by an open square, represents the reaction $HCHO + H \longrightarrow CH_2OH$ as investigated by Greenhill *et al.* using RRKM/master equation calculations [41]. Their reported high-pressure limit rate expression is in the reverse direction, which we have converted to the forward direction using thermochemistry obtained from high-level quantum chemistry calculations [40]. Their reported activation energy of 29.6 kcal/mol is very close to our computed enthalpy of reaction of 29.5 kcal/mol, which would indicate that the forward reaction is nearly barrierless. However, their rate is significantly faster than the rate used in several published mechanisms, including GRI-Mech 3.0 [42]. We suspect that Greenhill's preexponential factor is too high, perhaps due to an incorrect enthalpy of reaction.



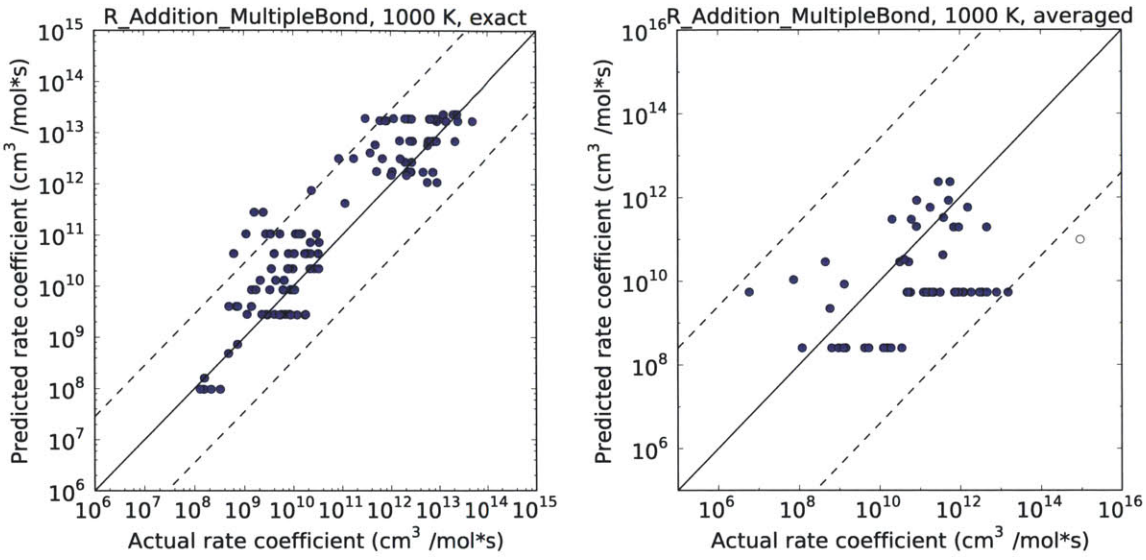
(a) (b)

Figure 3.5: Parity plots comparing the RMG-estimated rate coefficients at 1000 K using the rate rules method for hydrogen abstraction reactions to the actual values, as taken from the NIST kinetics database. The dashed lines represent the 95% confidence limit. (a) The reactions that matched an exact rate rule are generally estimated to within about an order of magnitude. (b) The reactions that required an averaged rate rule have a much larger uncertainty, about two orders of magnitude.

3.4.2.2 Rate rules method

The rules that are utilized in the rate rules method can be divided into two categories: the "exact" rules that were explicitly created from a literature value and the "averaged" rules that combine multiple exact rules to help fill in gaps in the tree of supergroups. To evaluate the rate rules method, we therefore separate the entries in the NIST database that match an exact rule from those that match an averaged rule. Not surprisingly, the estimated constructed from exact rules are far more accurate overall than those made from averaged rules. The uncertainty in the averaged rules is much more indicative of the overall uncertainty of the method.

For hydrogen abstraction, the parity plots in Figure 3.5 show that the exact rules are nearly order-of-magnitude accurate, while the uncertainty from an averaged rule is $10^{1.6}$, significantly larger. This is still a noticeable improvement over the Evans-Polanyi method (Figure 3.4(a)). The estimated rate coefficients on the averaged rules plot that fall outside the 95% confidence interval are mostly constructed using rules created for very generic supergroups; these rules are intended to cover a wider range of structures, and therefore may be constructed using values for functionalities very

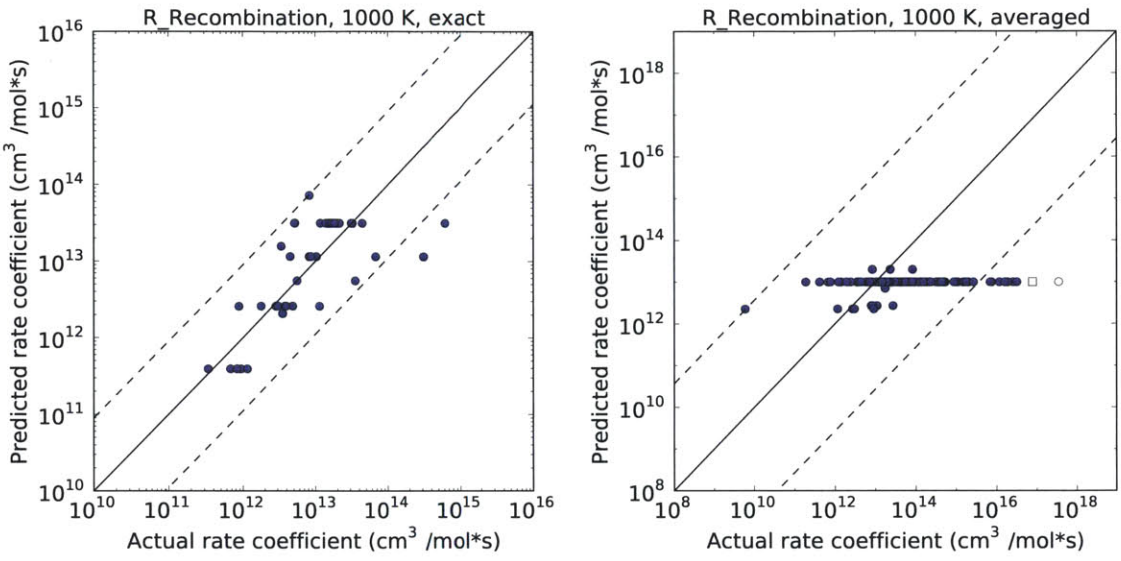


(a) (b)

Figure 3.6: Parity plots comparing the RMG-estimated rate coefficients at 1000 K using the rate rules method for radical addition to multiple bond reactions to the actual values, as taken from the NIST kinetics database. The dashed lines represent the 95% confidence limit. (a) The reactions that matched an exact rate rule are generally estimated to between one and two orders of magnitude. (b) The reactions that required an averaged rate rule have a much larger uncertainty, about three orders of magnitude. The horizontal streaks are due to many missing data in the rate estimation tree, leading to many different reactions being assigned the same "generic" rate coefficent.

different than those of the reaction of interest. The outlier indicated by the open circle in Figure 3.5(b) corresponds to the $C_3H_6 + CH_3O \longrightarrow CH_3OH +$ a-C_3H_5 entry from Tsang discussed earlier.

Interestingly, while the uncertainty of the rate rules method for the radical addition to multiple bond family (Figure 3.6) for exact rules is better than that for the Evans-Polanyi method, the uncertainty for averaged rules is actually worse. The rate rules method is able to make estimates that span four orders of magnitude, a wider range of values than via the Evans-Polanyi method; however, this has simply made the scatter worse. Also notice two "rows" of points at 6×10^5 and 1×10^7 $\mathrm{cm}^3/\mathrm{mol} \cdot \mathrm{s}$, neither of which is centered on the parity line; these correspond to entries estimated by falling up the tree of supergroups to a more general supergroup, with an averaged rule that is clearly not reflective of the test dataset. Both of these observations suggest that the coverage of the existing rate rules in this family is inadequate. Indeed, there are only 132 exact rules in this family, while hydrogen abstraction has 982. The outlier, indicated by the open circle, corresponds to the $HCHO + H \longrightarrow CH_2OH$ entry from



(a) (b)

Figure 3.7: Parity plots comparing the RMG-estimated rate coefficients at 1000 K using the rate rules method for radical-radical recombination reactions to the actual values, as taken from the NIST kinetics database. The dashed lines represent the 95% confidence limit.(a) The reactions that matched an exact rate rule are generally estimated within one order of magnitude. (b) The reactions that required an averaged rate rule have a much larger uncertainty, more than two orders of magnitude.

Greenhill *et al.* discussed earlier.

Radical-radical recombination reactions are very difficult to calculate theoretically due to their inherent barrierless nature, which often requires variational calculations. Having an accurate method of estimating the kinetics of these reactions would be especially valuable if it can reduce the number of these tricky and expensive calculations. The rate rules method is again able to very nearly achieve order-of-magnitude accuracy for radical-radical recombination reactions (Figure 3.7) that match an exact rule. However, the accuracy is only three orders of magnitude when averaged rules are required. Nearly all of the averaged estimates use the top-level rate rule (since RMG knows only 63 rules for this family) of 1×10^{13} $\mathrm{cm}^3/\mathrm{mol} \cdot \mathrm{s}$, while the test set spans a range from 10^{11} to 10^{18} $\mathrm{cm}^3/\mathrm{mol} \cdot \mathrm{s}$.

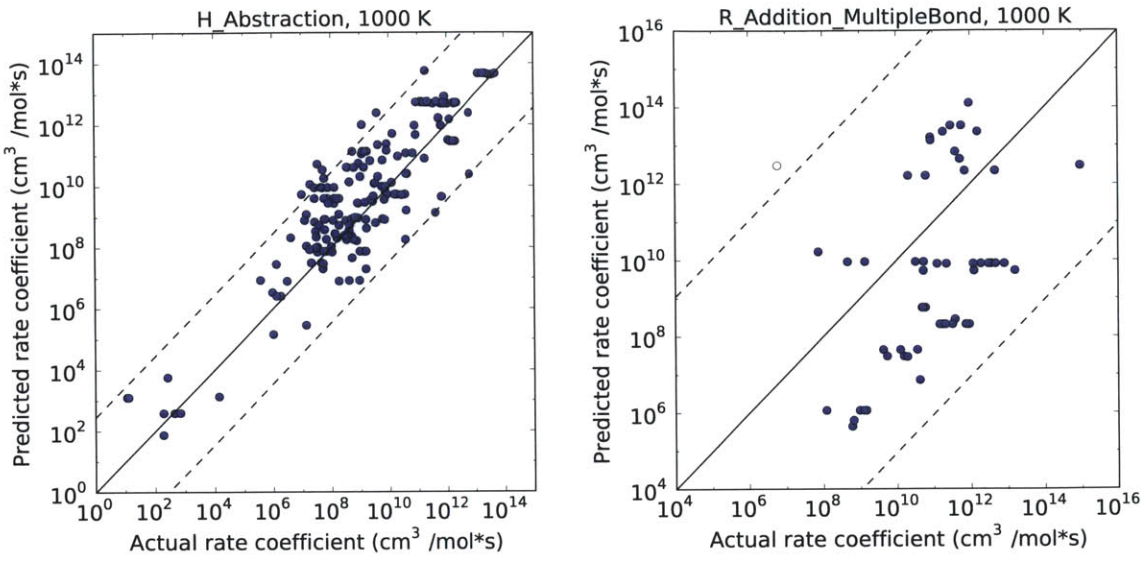
Many of the test set values are above the collision theory limit, which is generally considered an upper bound to the kinetics. This is likely because many of these reactions are reported in the literature in the bond-breaking direction, and have activation energies that are not consistent with the enthalpy of reaction, *i.e.* many of the literature values are incorrect. The two outliers on the radical-radical recombination

parity plot in Figure 3.7(b) are examples of this. The open circle represents the reaction $H + C_6H_5O \longrightarrow C_6H_5OH$, with literature value from Lovell *et al.* via Zhu and Bozzelli given in the reverse (dissociation) direction [43, 44]. The dissociation activation energy of 86.9 kcal/mol was estimated based on phenol pyrolysis experments over 1064-1162 K, a relatively narrow temperature range. A Benson group additivity estimate of the enthalpy of reaction gives 104.1 kcal/mol, which is significantly larger than the activation energy. Increasing the activation energy to 104.1 kcal/mol would cause the predicted rate coefficient to decrease by almost four orders of magnitude at 1000 K, which would move it within an order of magnitude of the rate rule estimate. The second outlier, represented by the open square, corresponds to the reaction $C_2H_5 + C_6H_5O \longrightarrow C_6H_5OC_2H_5$, with kinetics determined by phenyl ethyl ether pyrolysis experiments over the range 950-1220 K and very low pressures by Colussi *et al.* [45]. As with the previous outlier, the literature activation energy, given in the dissociation direction, is 60.4 kcal/mol, much lower than the group additivity-estimated enthalpy of reaction of 83.6 kcal/mol. Increasing the activation energy of both outliers would again move the literature result much closer to the rate rule estimate.

3.4.2.3 Kinetics group additivity method

Parity plots demonstrating the accuracy of our implementation of the kinetics group additivity method for hydrogen abstraction and radical addition to multiple bond at 1000 K are shown in 3.8. In both families, the uncertainty in the group additivity method is not only worse than that of the rate rules method (for exact or averaged rules), it is actually also worse than that of the Evans-Polanyi method! For the radical addition to multiple bond family, the group additivity method is able to predict rate coefficient values that span the largest range, but this again has seemed to simply increase the scatter in the predictions. The outlier in Figure 3.8(b) is the $H + CO_2 \longrightarrow HOCO$ entry from Larson *et al.* discussed earlier.

The smaller size of the training set for radical addition to multiple bond reactions suggests that we cannot draw firm conclusions about the accuracy of the group additivity approximation for that family from this analysis. However, the hydrogen abstraction family has a very large training set, and represents a family for which we expect the group additivity approximation to be reasonable. We have not yet attempted to adjust the hierarchy of supergroups in order to more closely gather groups with similar group values; this may help improve the accuracy of the method. The group additivity method also has other advantages, such as a clearer method of estimating rates for reactions that do not match an exact rule, as an estimate can be constructed for each



(a) (b)

Figure 3.8: Parity plot comparing the RMG-estimated rate coefficients at 1000 K using the group additivity method for (a) hydrogen abstraction and (b) radical addition to multiple bond reactions to the actual values, as taken from the NIST kinetics database. The dashed lines represent the 95% confidence limit. In both cases, the group additivity method is less accurate than both the Evans-Polanyi and rate rules methods.

functionality in the subgroup independently. The group additivity method is also the only method that is able to make predictions that span the entire range of literature values for the radical addition to multiple bond family. For these reasons, we believe that the group additivity method remains worthy of further effort. However, it is not currently advantageous from an accuracy standpoint to use the group additivity method when building detailed kinetics models.

3.5 Conclusion

This chapter presented a general framework for implementing methods of obtaining rapid estimates of high-pressure limit rate coefficients from a training set of literature data. The new framework emphasized the decoupling of the training data set from the hierarchy of functional groups. Within that framework, three methods of estimating the kinetics were evaluated to determine their uncertainty using the hydrogen abstraction, radical addition to multiple bond, and radical-radical recombination reaction families.

None of the kinetics estimation methods are currently able to achieve the order-of-magnitude accuracy desired for detailed kinetics models. The rate rules method is the closest to achieving this, but only for hydrogen abstraction, for which we have a very large training set. Although the group additivity method makes it much easier to document how a kinetics estimate was arrived at, it also appears to need further refinement in order to obtain similar accuracy as the rate rules method. Despite this large uncertainty in kinetics estimates, large detailed kinetics models are being successfully developed, both by hand and by computer. However, these models often require refinement of nearly all of the sensitive reactions in order to achieve reasonable accuracy in their predictions. It is important that we capture the information learned in these refinements in order to allow future models to benefit as well. This will require a coordinated effort by the kinetics community to build a large database of training and test data and to refine the kinetics estimation rules and procedures.

No matter which method is used, it is clear that significant effort must be made in order to achieve order-of-magnitude accuracy in kinetics estimation. This will likely require a large increase in the size of the training set of kinetics data, which itself will require a coordinated effort by the kinetics community. As the training data set grows, we may also be able to obtain some increase in accuracy by redesigning our hierarchy of functional groups to better reflect similarities in the underlying kinetics. Separating the functional group tree from the training data set makes this straightforward to achieve.

3.6 References

[1] J. W. Allen, R. W. Ashcraft, G. J. Beran, B. A. Buesser, C. A. Class, C. Gao, C. F. Goldsmith, M. R. Harper, A. Jalan, M. Keceli, G. R. Magoon, D. M. Matheu, S. S. Merchant, J. D. Mo, S. Petway, S. Ruman, S. Sharma, K. M. Van Geem, J. Song, Y. Suleymanov, J. Wen, R. H. West, A. Wong, H.-W. Wong, N. W.-W. Yee, P. E. Yelvington, J. Yu, and W. H. Green. *RMG (Reaction Mechanism Generator) version 4.0*. 2013. URL: `http://rmg.sourceforge.net/`.

[2] 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.

[3] W. H. Green. "Building and Solving Accurate Combustion Chemistry Simulations." *J. Combust. Soc. Japan* 50 (151), 2008, pp. 19–28.

[4] 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.

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

[6] N. Cohen and S. W. Benson. "Estimation of Heats of Formation of Organic Compounds by Additivity Methods." *Chem. Rev.* 93 (7), 1993, pp. 2419–2438. DOI: 10.1021/cr00023a005.

[7] 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.

[8] G. R. Magoon and W. H. Green. "Design and Implementation of a Next-Generation Software System for On-the-Fly Quantum and Force Field Calculations in Automated Reaction Mechanism Generation." *Comp. & Chem. Eng.* 2013 (submitted).

[9] J. A. Manion, R. E. Huie, R. D. Levin, D. R. Burgess Jr., V. L. Orkin, W. Tsang, W. S. McGivern, J. W. Hudgens, V. D. Knyazev, D. B. Atkinson, E. Chai, A. M. Tereza, C.-Y. Lin, T. C. Allison, W. G. Mallard, F. Westley, J. T. Herron, R. F. Hampson, and D. H. Frizzell. *NIST Chemical Kinetics Database and NIST Standard Reference Database 17 and Version 7.0 (Web Version) and Release 1.4.3, Data version 2008.12*. 2008. URL: http://kinetics.nist.gov/.

[10] R. Sumathi, H.-H. Carstensen, and W. H. Green. "Reaction Rate Prediction via Group Additivity Part 1: H Abstraction from Alkanes by H and CH_3." *J. Phys. Chem. A* 105 (28), 2001, pp. 6910–6925. DOI: 10.1021/jp010697q.

[11] R. Sumathi, H.-H. Carstensen, and W. H. Green. "Reaction Rate Prediction via Group Additivity, Part 2: H-Abstraction from Alkenes, Alkynes, Alcohols, Aldehydes, and Acids by H Atoms." *J. Phys. Chem. A* 105 (39), 2001, pp. 8969–8984. DOI: 10.1021/jp011827y.

[12] R. Sumathi, H.-H. Carstensen, and W. H. Green. "Reaction Rate Predictions Via Group Additivity. Part 3: Effect of Substituents with CH_2 as the Mediator." *J. Phys. Chem. A* 106 (22), 2002, pp. 5474–5489. DOI: 10.1021/jp013957c.

[13] P. A. Willems and G. F. Froment. "Kinetic Modeling of the Thermal Cracking of Hydrocarbons. 1. Calculation of Frequency Factors." *Ind. Eng. Chem. Res.* 27(11), 1988, pp. 1959–1966. DOI: 10.1021/ie00083a001.

[14] P. A. Willems and G. F. Froment. "Kinetic Modeling of the Thermal Cracking of Hydrocarbons. 2. Calculation of Activation Energies." *Ind. Eng. Chem. Res.* 27(11), 1988, pp. 1966–1971. DOI: 10.1021/ie00083a002.

[15] T. N. Truong. "Reaction Class Transition State Theory: Hydrogen Abstraction Reactions by Hydrogen Atoms as Test Cases." *J. Chem. Phys* 113(12), 2000, pp. 4957–4964. DOI: 10.1063/1.1287839.

[16] S. Zhang and T. N. Truong. "Kinetics of Hydrogen Abstraction Reaction Class H + H−C(sp3): First-Principles Predictions Using the Reaction Class Transition State Theory." *J. Phys. Chem. A* 107(8), 2003, pp. 1138–1147. DOI: 10.1021/jp021265y.

[17] M. Saeys, M.-F. Reyniers, G. B. Marin, V. V. Speybroeck, and M. Waroquier. "Ab Initio Group Contribution Method for Activation Energies for Radical Additions." *AIChE J.* 50(2), 2004, pp. 426–444. DOI: 10.1002/aic.10038.

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

[19] M. K. Sabbe, M.-F. Reyniers, M. Waroquier, and G. B. Marin. "Hydrogen Radical Additions to Unsaturated Hydrocarbons and the Reverse β-Scission Reactions: Modeling of Activation Energies and Pre-Exponential Factors." *Chem. Phys. Chem.* 11(1), 2010, pp. 195–210. DOI: 10.1002/cphc.200900509.

[20] M. G. Evans and M. Polanyi. "Further Considerations on the Thermodynamics of Chemical Equilibria and Reaction Rates." *Trans. Faraday Soc.* 32, 1936, pp. 1333–1360. DOI: 10.1039/TF9363201333.

[21] M. G. Evans and M. Polanyi. "Inertia and Driving Force of Chemical Reactions." *Trans. Faraday Soc.* 34, 1938, pp. 11–24. DOI: 10.1039/TF9383400011.

[22] P. Blowers and R. Masel. "Engineering Approximations for Activation Energies in Hydrogen Transfer Reactions." *AIChE J.* 46(10), 2000, pp. 2041–2052. DOI: 10.1002/aic.690461015.

[23] R. Sumathi and W. H. Green. "A Priori Rate Constants for Kinetic Modeling." *Theor. Chem. Acc.* 108, 2002, pp. 187–213. DOI: 10.1007/s00214-002-0368-4.

[24] R. Sumathi and W. H. Green. "Oxygenate, Oxyalkyl and Alkoxycarbonyl Thermochemistry and Rates for Hydrogen Abstraction from Oxygenates." *Phys. Chem. Chem. Phys.* 5, 2003, pp. 3402–3417. DOI: 10.1039/B307050F.

[25] C. D. Wijaya, R. Sumathi, and W. H. Green. "Thermodynamic Properties and Kinetic Parameters for Cyclic Ether Formation from Hydroperoxyalkyl Radicals." *J. Phys. Chem. A* 107(24), 2003, pp. 4908–4920. DOI: 10.1021/jp027471n.

[26] C. D. Wijaya. "Developing Fundamentally Based Models for Autoignition." PhD thesis. Massachusetts Institute of Technology, 2003.

[27] H. J. Curran, P. Gaffuri, W. J. Pitz, and C. K. Westbrook. "A Comprehensive Modeling Study of n-Heptane Oxidation." *Combust. Flame* 114(1-2), 1998, pp. 149–177. DOI: 10.1016/S0010-2180(97)00282-4.

[28] M. R. Harper, S. S. Merchant, W. E. Boettcher, and W. H. Green. "Hydrogen-Abstraction Rate Coefficients for Hydrocarbons and Oxygenated Hydrocarbons." *J. Phys. Chem. Ref. Data*, 2013 (submitted).

[29] A. Vandeputte. "The Thermochemistry and Decomposition Mechanism of Organosulfur and Organophosphorus Compounds." PhD thesis. University of Gent, 2012.

[30] URL: http://primekinetics.org/.

[31] J. Park, Z. F. Xu, and M. C. Lin. "Thermal Decomposition of Ethanol. II. A Computational Study of the Kinetics and Mechanism for the $H+C_2H_5OH$ Reaction." *J. Chem. Phys* 118(22), 2003, pp. 9990–9996. DOI: 10.1063/1.1573182.

[32] T. Tanzawa and W. C. G. Jr. "Thermal Decomposition of Acetylene." *Symposium (International) on Combustion* 17(1), 1979, pp. 563–573. DOI: 10.1016/S0082-0784(79)80057-0.

[33] J. N. Bradley and K. O. West. "Single-Pulse Shock Tube Studies of Hydrocarbon Pyrolysis. Part 5.-Pyrolysis of Neopentane." 72, 1976, pp. 8–19. DOI: 10.1039/F19767200008.

[34] T. K. Choudhury, W. A. Sanders, and M. C. Lin. "A Shock Tube and Modeling Study of the Methyl + Formaldehyde Reaction at High Temperatures." *J. Phys. Chem.* 93(13), 1989, pp. 5143–5147. DOI: 10.1021/j100350a024.

[35] J. C. Biordi, C. P. Lazzara, and J. F. Papp. "Molecular Beam Mass Spectrometry Applied to Determining the Kinetics of Reactions in Flames II. A Critique of Rate Coefficient Determinations." *Combust. Flame* 26, 1976, pp. 57–76. DOI: 10.1016/0010-2180(76)90057-2.

[36] D. L. Baulch, C. T. Bowman, C. J. Cobos, R. A. Cox, T. Just, J. A. Kerr, M. J. Pilling, D. Stocker, J. Troe, W. Tsang, R. W. Walker, and J. Warnatz. "Evaluated Kinetic Data for Combustion Modeling: Supplement II." *J. Phys. Chem. Ref. Data* 34 (3), 2005, pp. 757–1397. DOI: 10.1063/1.1748524.

[37] C. W. Larson, P. H. Stewart, and D. M. Golden. "Pressure and Temperature Dependence of Reactions Proceeding via a Bound Complex. An Approach for Combustion and Atmospheric Chemistry Modelers. Application to HO + CO $\longrightarrow$ [HOCO] $\longrightarrow$ H + CO_2." *Int. J. Chem. Kin.* 20 (1), 1988, pp. 27–40. DOI: 10.1002/kin.550200105.

[38] J. S. Francisco, J. T. Muckerman, and H.-G. Yu. "HOCO Radical Chemistry." *Acc. Chem. Res.* 43 (12), 2010, pp. 1519–1526. DOI: 10.1021/ar100087v.

[39] W. Tsang. "Chemical Kinetic Data Base for Combustion Chemistry Part V. Propene." *J. Phys. Chem. Ref. Data* 20 (2), 1991, pp. 221–273. DOI: 10.1063/1.555880.

[40] C. F. Goldsmith, G. R. Magoon, and W. H. Green. "Database of Small Molecule Thermochemistry for Combustion." *J. Phys. Chem. A* 116 (36), 2012, pp. 9033–9057. DOI: 10.1021/jp303819e.

[41] P. G. Greenhill, B. V. O'Grady, and R. G. Gilbert. "Theoretical Prediction of CH_3O and CH_2OH Gas-Phase Decomposition Rate Coefficients." *Aus. J. Chem.* 39 (12), 1986, pp. 1929–1942. DOI: 10.1071/CH9861929.

[42] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, C. T. Bowman, R. K. Hanson, S. Song, J. William C. Gardiner, V. V. Lissianski, and Z. Qin. *GRI-Mech 3.0*. 1999. URL: http://www.me.berkeley.edu/gri_mech/.

[43] A. B. Lovell, K. Brezinsky, and I. Glassman. "The Gas Phase Pyrolysis of Phenol." *Int. J. Chem. Kin.* 21 (7), 1989, pp. 547–560. DOI: 10.1002/kin.550210706.

[44] L. Zhu and J. W. Bozzelli. "Kinetics and Thermochemistry for the Gas-Phase Keto-Enol Tautomerism of Phenol $\rightleftharpoons$ 2,4-Cyclohexadienone." *J. Phys. Chem. A* 107 (19), 2003, pp. 3696–3703. DOI: 10.1021/jp0212545.

[45] A. J. Colussi, F. Zabel, and S. W. Benson. "The Very Low-Pressure Pyrolysis of Phenyl Ethyl Ether, Phenyl Allyl Ether, and Benzyl Methyl Ether and the Enthalpy of Formation of the Phenoxy Radical." *Int. J. Chem. Kin.* 9 (2), 1977, pp. 161–178. DOI: 10.1002/kin.550090202.

4

Calculating quantum effects in bimolecular rate coefficients using ring polymer molecular dynamics

Modern detailed kinetics models often require hundreds of chemical species and thousands of reactions in order to quantitatively describe complex chemical phenomena over a wide range of conditions. (Chapter 7 describes such a model for the oxidation of diisopropyl ketone.) As was discussed in Chapter 2, we must estimate the thermodynamic and kinetic parameters for each species and reaction when constructing the model. Fortunately, only a small subset of the model parameters have a significant effect on any particular experimental observable; the other parameters can in principle be estimated approximately using methods such as those described in Chapter 3. For the remaining sensitive parameters, however, it is often necessary to obtain very accurate values in order for the model prediction to be quantitatively correct.

Many classes of chemical reactions exhibit significant quantum mechanical character at low temperatures due to the effects of zero-point vibrational energy and/or tunneling through an activation barrier. A rigorous determination of these effects via solution of the quantum reactive scattering Schrödinger equation is currently feasible only for systems containing six or fewer atoms [1]. For this reason, several theoretical methods have been developed in an effort to approximately incorporate these important quantum effects in the calculation of chemical reaction rate coefficients for larger systems. Many previous methods are refinements of transition state theory. Modern implementations of both conventional and variational transition state theory incorporate zero-point energy by utilizing quantum mechanical partition functions for at least the internal degrees of freedom, and there are several ways of including a semiclassical tunneling correction, both one-dimensional and multidimensional [2]. There

has also been significant effort toward developing a fully quantum transition state theory (QTST) [3–7], such as the quantum instanton (QI) theory of Miller *et al.* [8, 9], which improves on an earlier semiclassical transition state theory [10] by treating the Boltzmann operator fully quantum mechanically.

Recently, an alternative method of incorporating quantum effects in chemical reaction rates has been developed based on ring polymer molecular dynamics (RPMD) [11, 12]. RPMD exploits the isomorphism between the statistical properties of the quantum system and those of a classical fictitious ring polymer consisting of many copies of the original system connected by harmonic springs [13]. This isomorphism enables the inclusion of quantum effects via classical molecular dynamics simulations in an extended phase space, an approach with several desirable features for calculating bimolecular gas-phase rate coefficients. First, the RPMD rate coefficient becomes exact in the high temperature limit, where the ring polymer collapses to a single bead. Second, the RPMD rate coefficient has a well-defined short-time limit that acts as an upper bound on the RPMD rate. When the transition state dividing surface is defined in terms of the centroid of the ring polymer, this short-time limit coincides with a well-known (centroid density) version of quantum transition state theory (QTST). Finally, and most importantly, the RPMD rate coefficient is rigorously independent of the choice of the transition state dividing surface used to compute it, a feature that distinguishes it from transition state theory-based methods.

Previous applications of RPMD rate theory have demonstrated that it provides systematic and consistent performance across a wide range of system dimensionalities, including one-dimensional Eckart barriers [11, 12], atom-diatom reactions [14–16], and small polyatomic reactions [17–19]. In all systems considered so far, the RPMD rate coefficient captures almost perfectly the zero-point energy effect, and is within a factor of 2-3 of the exact rate at low temperatures in the deep tunneling regime, as determined by comparison to rigorous quantum mechanical results available for these systems. Furthermore, RPMD rate theory has been shown to consistently underestimate the rates for symmetric reactions and overestimate them for asymmetric reactions [20, 21]. The systematic and consistent performance of RPMD rate theory contrasts with the performance of transition state theory-based methods, which depend on the choice of the transition state dividing surface. Proper identification of this dividing surface becomes increasingly difficult as the dimensionality of the problem increases due to the multidimensional nature of tunneling at low temperatures and high amount of recrossings at high temperatures; as a result, QTST methods are often less accurate in higher dimensionalities. Furthermore, more elaborate quantum im-

plementations of transition state theory methods do not necessarily provide improved accuracy when compared to less sophisticated methods, and therefore are not guaranteed to provide a predictable level of error when applied to higher-dimensionality systems. RPMD rate theory has been shown to give better agreement with accurate quantum mechanical (multiconfigurational time-dependent Hartree) [22] calculations than transition state theory methods for the six-atom reactions $H + CH_4 \longrightarrow CH_3 + H_2$ and $O + CH_4 \longrightarrow CH_3 + OH$ [17, 19]. The predictable level of accuracy, which occurs because of the rigorous independence of the method from the choice of transition state dividing surface, is one of the most important and attractive features of RPMD rate theory.

The successful application of RPMD to small systems motivated us to develop a new general-purpose software package for using RPMD to accurately compute bimolecular reaction rates of arbitrary polyatomic systems. This chapter describes the development of a new general-purpose software package for using ring polymer molecular dynamics simulations to accurately compute bimolecular reaction rates. This chapter describes the development of this open source software package, `RPMDrate` [21], and the use of the new software to study the kinetics of the hydrogen abstraction reaction $OH + CH_4 \longrightarrow CH_3 + H_2O$, which has significant quantum mechanical character. This represents the first time that RPMD rate theory has been applied to a seven-atom system, for which no exact quantum mechanical results are currently available. In particular, this work focuses on the kinetic isotope effect (KIE) of this reaction relative to the reaction $OH + CD_4 \longrightarrow CD_3 + HDO$. KIE is widely used in experiments to detect quantum mechanical effects; as a ratio of rate coefficients, the KIE depends weakly on the height of the classical reaction barrier, and therefore provides a useful means for comparison to experiment even when the PES is less accurate. As we will see, the choice of a hydrogen abstraction and a deuterium abstraction reaction will be useful in illustrating the predictable level of accuracy of RPMD rate theory.

4.1 Bimolecular reaction rate theory

4.1.1 Quantum rate theory

Consider a generic bimolecular reaction of the form $A + B \longrightarrow$ products. The quantum Hamiltonian operator $\hat{H}$ describing the system in terms of momenta $\hat{p}_i$ and positions $\hat{q}_i$ is

$$\hat{H} = \sum_{i=1}^{f} \frac{\hat{p}_i^2}{2m_i} + \hat{V}(\mathbf{q}) \tag{4.1}$$

where the i subscript indexes over atoms, m_i is the mass of atom i, $V(\mathbf{q})$ is the potential energy, and f is the number of atoms in the combined A + B system. To compute the bimolecular rate coefficient, a surface $s(\mathbf{q})$ is constructed that divides the momentum-position phase space $(\mathbf{p}, \mathbf{q})$ into reactants and products such that the products are in $s > 0$. The quantum mechanical thermal rate coefficient $k(T)$ for conversion of reactants to products can be expressed in terms of the correlation of the thermal flux through this dividing surface at time $t = 0$ with the projection onto the states of the products in the long time limit [23–25]. This is expressed mathematically as

$$k(T) = \frac{1}{Q_{\text{reac}}(T)} \lim_{t \to \infty} c_{fs}(t; s) \tag{4.2}$$

where $Q_{\text{reac}}(T)$ is the partition function of the reactants and $c_{fs}(t; s)$ is the cross-correlation function between position and flux for our chosen dividing surface, given by

$$c_{fs}(t; s) = \text{tr}\left[e^{-\beta\hat{H}/2}\hat{F}e^{-\beta\hat{H}/2}e^{+i\hat{H}t/h}\hat{h}e^{-i\hat{H}t/h}\right] \tag{4.3}$$

Above, $\beta \equiv (k_{\text{B}}T)^{-1}$ is reciprocal temperature, $\hat{h}$ is the Heaviside step function operator for selecting the product states,

$$\hat{h} \equiv \hat{h}(s) = \begin{cases} 1 & s > 0 \\ 0 & s < 0 \end{cases} \tag{4.4}$$

and $\hat{F}$ is the flux operator, given by

$$\hat{F} = \frac{i}{\hbar}\left[\hat{H}, \hat{h}\right] \tag{4.5}$$

Although the correlation function $c_{fs}^{(n)}(t; s)$ is dependent on the choice of dividing surface, its value in the long time limit is not, which makes the bimolecular rate coefficient $k(T)$ rigorously independent of the choice of dividing surface $s(\mathbf{q})$ [12].

4.1.2 Classical rate theory

The formal expression for the classical rate coefficient looks very similar to the quantum mechanical formulation, except that traces are converted to phase space averages and operators become functions of classical position and momenta. We will use the superscript (1) to denote the classical limit, as this will more clearly relate it to the subsequent discussion of RPMD rate theory. The classical rate coefficient is given by

$$k^{(1)}(T) = \frac{1}{Q^{(1)}_{\mathrm{reac}}(T)} \lim_{t\to\infty} c^{(1)}_{fs}(t;s) \tag{4.6}$$

The classical correlation function $c^{(1)}_{fs}(t;s)$ is given by a phase space average instead of a trace,

$$c^{(1)}_{fs}(t;s) = \frac{1}{(2\pi\hbar)^{3f}} \int\int e^{-\beta H(\mathbf{p}_0,\mathbf{q}_0)} F(\mathbf{p}_0,\mathbf{q}_0) h[s(\mathbf{q}_t)]\, d^{3f}\mathbf{q}_0\, d^{3f}\mathbf{p}_0 \tag{4.7}$$

where $H^{(1)}(\mathbf{p},\mathbf{q})$ is the classical Hamiltonian,

$$H^{(1)}(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{f} \frac{|\mathbf{p}_i|^2}{2m_i} + V(\mathbf{q}) \tag{4.8}$$

$F(\mathbf{p}_0,\mathbf{q}_0)$ is the classical flux through the dividing surface at time $t = 0$,

$$F(\mathbf{p}_0,\mathbf{q}_0) = \left.\frac{dh[s(\mathbf{q})]}{dt}\right|_{t=0} = \delta[s(\mathbf{q}_0)] v_s(\mathbf{p}_0,\mathbf{q}_0) \tag{4.9}$$

$v_s(\mathbf{p}_0,\mathbf{q}_0)$ is the classical velocity through the dividing surface at time $t = 0$,

$$v_s(\mathbf{p},\mathbf{q}) = \frac{ds(\mathbf{q})}{dt} = \sum_{i=1}^{f} \frac{\partial s(\mathbf{q})}{\partial q_i} \frac{p_i}{m_i} \tag{4.10}$$

and $h[s(\mathbf{q}_t)]$ counts the phase space trajectories that are in the product space at time t, as determined by the classical Newtonian equations of motion.

4.1.3 Ring polymer molecular dynamics (RPMD) rate theory

In ring polymer molecular dynamics theory, the phase space is extended by making n copies of the classical system, and connecting the copies of each atom, called "beads", by a ring of harmonic springs. The RPMD Hamiltonian $H^{(n)}(\mathbf{p},\mathbf{q})$ is given by

$$H^{(n)}(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{f}\sum_{j=1}^{n}\frac{|\mathbf{p}_{i,j}|^2}{2m_i} + \sum_{i=1}^{f}\sum_{j=1}^{n}\frac{1}{2}m_i\omega_n^2|\mathbf{q}_{i,j}-\mathbf{q}_{i,j-1}|^2 + \sum_{j=1}^{n}V(\mathbf{q}_j) \tag{4.11}$$

where the j subscript indexes over beads, $\omega_n \equiv (\beta\hbar/n)^{-1}$ is the frequency of the springs, and n is the total number of beads. (Because each bead is a ring of atoms, we use $\mathbf{q}_{i,0} \equiv \mathbf{q}_{i,n}$ when evaluating the Hamiltonian.) Since we have made n copies of the system in our extended phase space, we must use a reciprocal temperature of $\beta_n \equiv \beta/n$ to compensate. Clearly, setting $n = 1$ recovers the original classical Hamiltonian exactly, ensuring RPMD theory is consistent with the classical result in the limit of a single bead.

The RPMD rate coefficient is given by

$$k^{(n)}(T) = \frac{1}{Q_{\text{reac}}^{(n)}(T)}\lim_{t\to\infty}c_{fs}^{(n)}(t;s) \tag{4.12}$$

where $Q_{\text{reac}}^{(n)}(T)$ is the partition function of the reactants and $c_{fs}^{(n)}(t;s)$ is the position-flux cross-correlation function, given by a phase space average

$$c_{fs}^{(n)}(t;s) = \frac{1}{(2\pi\hbar)^{3fn}}\int\int e^{-\beta_n H_n(\mathbf{p}_0,\mathbf{q}_0)}F(\bar{\mathbf{p}}_0,\bar{\mathbf{q}}_0)h[s(\bar{\mathbf{q}}_t)]\,d^{3fn}\mathbf{q}_0\,d^{3fn}\mathbf{p}_0 \tag{4.13}$$

Above, $F(\bar{\mathbf{p}}_0,\bar{\mathbf{q}}_0)$ is the classical flux through the dividing surface at time $t = 0$ as determined using the centroid of the ring polymer beads (as indicated with an overbar),

$$F(\bar{\mathbf{p}}_0,\bar{\mathbf{q}}_0) = \left.\frac{dh[s(\bar{\mathbf{q}})]}{dt}\right|_{t=0} = \delta[s(\bar{\mathbf{q}}_0)]v_s(\bar{\mathbf{p}}_0,\bar{\mathbf{q}}_0) \tag{4.14}$$

$v_s(\bar{\mathbf{p}}_0,\bar{\mathbf{q}}_0)$ is the classical velocity of the ring polymer centroids through the dividing surface at time $t = 0$,

$$v_s(\bar{\mathbf{p}},\bar{\mathbf{q}}) = \frac{ds(\bar{\mathbf{q}})}{dt} = \sum_{i=1}^{f}\frac{\partial s(\bar{\mathbf{q}})}{\partial\bar{q}_i}\frac{\bar{p}_i}{m_i} \tag{4.15}$$

and $h[s(\bar{\mathbf{q}}_t)]$ counts the phase space trajectories that are in the product space at time t. The centroid values for the position and momentum of each bead are defined by simple averages:

$$\bar{q}_i \equiv \frac{1}{n}\sum_{j=1}^{n} q_{i,j} \qquad\qquad \bar{p}_i \equiv \frac{1}{n}\sum_{j=1}^{n} p_{i,j} \tag{4.16}$$

RPMD rate theory has several advantages that make it attractive as a general-purpose rate theory:

- RPMD rate theory has a well-defined short-time limit that provides an upper bound on the RPMD rate. When the dividing surface is defined in terms of the centroid of the ring polymer, this short-time limit coincides with a well-known centroid density [3–5] version of quantum transition state theory [12].

- RPMD rate theory gives the exact quantum mechanical result for a parabolic barrier at all temperatures at which a rate coefficient can be defined.

- The RPMD rate coefficient is rigorously independent of the choice of the transition state dividing surface [12]. This is especially valuable in situations where the optimum dividing surface is difficult to determine, such as multidimensional reactions.

- The RPMD rate coefficient becomes exact in the high temperature (classical) limit, where the ring polymer collapses to a single bead [11]. At all temperatures, the RPMD rate coefficient with one bead ($n = 1$) is identical to the classical rate coefficient.

The construction of the ring polymer beads incorporates an artificial uncertainty into the position of each atom in the reacting system. This is what enables a purely classical molecular dynamics simuation to capture the quantum effects of tunneling through a reaction barrier and zero-point energy. Earlier work has demonstrated that RPMD rate theory captures almost perfectly the zero-point energy effect [15], and also captures the tunneling effect to within a factor of two to three, even in the deep tunneling regime [14, 17].

4.1.4 Bennett-Chandler factorization

Unlike the quantum correlation function, the classical and RPMD versions are discontinuous at $t = 0$ and have a nonzero limit at t approaches zero from above. This implies that a well-defined transition state theory rate coefficient $k^{\mathrm{TST}}(T; s)$ exists:

$$k^{\mathrm{TST}}(T;s) = \frac{1}{Q^{(n)}_{\mathrm{reac}}(T)} \lim_{t\to 0^+} c^{(n)}_{fs}(t;s) \tag{4.17}$$

The assumption of no recrossing to reactants used in transition state theory allows us to replace the time-dependent step function $h[s(\bar{\mathbf{q}}_t)]$ with a step function in the initial velocity $h[v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0)]$. This allows the evaluation of the momentum integrals, giving

$$\lim_{t\to 0^+} c^{(n)}_{fs}(t;s) = \prod_{i=1}^{f} \left(\frac{m_i}{2\pi\beta\hbar^2}\right)^{3/2} \int e^{-\beta V^{(n)}(\mathbf{q})} \delta[s(\bar{\mathbf{q}})] f_s(\bar{\mathbf{q}})\, d^{3fn}\mathbf{q} \tag{4.18}$$

where

$$f_s(\bar{\mathbf{q}}) = \left\{ \sum_{i=1}^{f} \frac{1}{2\pi\beta m_i} \left[\frac{\partial s(\bar{\mathbf{q}})}{\partial \bar{q}_i}\right]^2 \right\}^{1/2} \tag{4.19}$$

The deviation of the QTST rate coefficient from the full quantum rate coefficient (due to recrossing) is defined as the transmission coefficient $\kappa^{(n)}(T;s)$, given by [26–28]

$$\kappa^{(n)}(T;s) \equiv \frac{k^{(n)}(T)}{k^{\mathrm{TST}}(T;s)} = \frac{c^{(n)}_{fs}(t\to\infty;s)}{c^{(n)}_{fs}(t\to 0^+;s)} \tag{4.20}$$

Unlike the full rate coefficient, which is independent of the choice of dividing surface, the TST rate coefficient is exponentially sensitive to the choice of dividing surface. For this reason, the efficient evaluation of the transmission coefficient requires the dividing surface be placed very close to the reaction barrier.

The Bennett-Chandler factorization [29, 30] provides a means of efficiently evaluating the TST rate coefficient. Instead of a dividing surface near the reaction barrier, a very simple dividing surface $s_0(\mathbf{q})$ is chosen in the asymptotic reactant valley,

$$s_0(\mathbf{q}) \equiv R_\infty - |\mathbf{R}| \tag{4.21}$$

where $\mathbf{R}$ is the vector connecting the centers of mass of the two reactant molecules, and R_∞ is some distance above which the interaction between the reactant molecules is negligible. The TST rate coefficient $k^{\mathrm{TST}}(T;s_0)$ can now be evaluated to give

$$k^{\mathrm{TST}}(T;s_0) = 4\pi R_\infty^2 \left(\frac{1}{2\pi\beta\mu}\right)^{1/2} \tag{4.22}$$

where μ is the reduced mass of the two reactant molecules. Physically, the first factor is the surface area of a sphere of radius R_∞, and the second is the thermally-averaged speed of the reactant molecules entering this sphere. The deviation between the TST rate coefficient evaluated at the dividing surface near the reaction barrier – hereafter referred to as $s_1(\mathbf{q})$ – and that evaluated above is defined as the factor $p^{(n)}(T; s_0, s_1)$,

$$p^{(n)}(T; s_1, s_0) \equiv \frac{k^{\mathrm{TST}}(T; s_1)}{k^{\mathrm{TST}}(T; s_0)} = \frac{c_{fs}^{(n)}(t \to 0^+; s_1)}{c_{fs}^{(n)}(t \to 0^+; s_0)} \tag{4.23}$$

The full bimolecular rate coefficient is now given by

$$k^{(n)}(T) = \kappa^{(n)}(T; s_1) p^{(n)}(T; s_1, s_0) k^{\mathrm{TST}}(T; s_0) \tag{4.24}$$

Since $k^{(n)}(T)$ is rigorously independent of the choice of dividing surface, we can instead use the expression

$$k^{(n)}(T) = \kappa^{(n)}(T; s_{\xi^\ddagger}) p^{(n)}(T; s_{\xi^\ddagger}, s_0) k^{\mathrm{TST}}(T; s_0) \tag{4.25}$$

where $\xi^\ddagger$ is the value of the reaction coordinate that maximizes the value of $p^{(n)}(T; s_\xi, s_0)$, and $s_{\xi^\ddagger}(\mathbf{q})$ is the corresponding dividing surface. This is done to maximize the value of the transmission coefficient $\kappa^{(n)}(T; s_\xi)$, and therefore minimize the number of trajectories required to accurately determine this value. For symmetric reactions $\xi^\ddagger = 1$ by definition; however, for asymmetric reactions the value of $\xi^\ddagger$ can deviate slightly from unity, depending on the accuracy of the defined transition state $s_1(\mathbf{q})$.

4.2 RPMDrate: A general-purpose software package for applying RPMD rate theory

The viability of RPMD as a method for accurately computing bimolecular rate coefficients had previously been demonstrated [14, 15, 17]. These works used standalone software packages developed specifically for the particular systems of interest in those works. The next logical step was to generalize from these specialized implementations to develop a general-purpose implementation that could easily be used on a variety of bimolecular reactions with minimal additional coding. The first major version of the resulting software package, named `RPMDrate`, was released in summer 2012 [21].

This section describes the computational strategies and relevant implementation details used in `RPMDrate` to calculate bimolecular rate coefficients using RPMD rate theory.

4.2.1 Generalized definition of the transition state dividing surface

Many bimolecular reactions can be described by the breaking and forming of pairs of bonds, where exactly one of the atoms in the breaking bond is also involved in a corresponding forming bond. (For example, all hydrogen abstraction reactions can be described in this manner.) In the general case there are multiple pairs of bonds being broken and formed in a single reaction event. Many reactions have multiple equivalent transition states; for example, the reaction $OH + CH_4 \longrightarrow CH_3 + H_2O$, has four equivalent transition states, since each of the four methane hydrogens can be abstracted with equal probability. In the general case there are N_{channels} equivalent transition states to consider.

The general definition of the transition state dividing surface $s_1(\mathbf{q})$ involving N_{bonds} forming/breaking bond pairs in each of N_{channels} equivalent transition states is

$$s_1(\mathbf{q}) = \min\left[s_{1,1}(\mathbf{q}), \ldots, s_{1,N_{\text{bonds}}}(\mathbf{q})\right] \tag{4.26}$$

Above, $s_{1,k}(\mathbf{q})$ describes a particular breaking/forming bond pair k. Since all bond pairs must be in the product state for the reaction to have occurred, we use the minimum in the above expression to select the pair closest to the reactants, as when that pair reaches $s = 0$, all others must be in $s > 0$. The value of each $s_{1,k}(\mathbf{q})$ is given by

$$s_{1,k}(\mathbf{q}) = \max\left[s_{1,k}^{(1)}(\mathbf{q}), \ldots, s_{1,k}^{N_{\text{channels}}}(\mathbf{q})\right] \tag{4.27}$$

where $s_{1,k}^{(\ell)}(\mathbf{q})$ describes a particular breaking/forming bond pair k in configuration ℓ. Since any configuration can be in the product state for the reaction to have occurred, we use the maximum in the above expression to select the configuration closest to the products. The value of each $s_{1,k}^{(\ell)}(\mathbf{q})$ is given by

$$s_{1,k}^{(\ell)}(\mathbf{q}) = \left(\left|\mathbf{q}_{12,k}^{(\ell)}\right| - q_{12,k}^{\ddagger(\ell)}\right) - \left(\left|\mathbf{q}_{23,k}^{(\ell)}\right| - q_{23,k}^{\ddagger(\ell)}\right) \tag{4.28}$$

where $q_{12,k}^{\ddagger(\ell)}$ and $q_{23,k}^{\ddagger(\ell)}$ represent the lengths of the breaking and forming bonds, respectively, at the geometry that defines the transition state.

In RPMDrate, the user specifies $s_1(\mathbf{q})$ by giving the desired transition state geometry – the Cartesian coordinates of each atom at the transition state – and a list of the breaking and forming bond pairs for each configuration. RPMDrate automatically extracts the values of $q_{12,k}^{\ddagger(\ell)}$ and $q_{23,k}^{\ddagger(\ell)}$ from this information and constructs the expression for $s_1(\mathbf{q})$ and its analytical gradient and Hessian, which are required for the subsequent calculations.

Although the above definition may not be optimal for many bimolecular reactions, we are nonetheless guaranteed to obtain the correct RPMD rate coefficient, as its value is independent of the choice of dividing surface.

4.2.2 Transition state theory rate

The static factor $p^{(n)}(T; s_{\xi^\ddagger}, s_0)$ can be expressed in the form [14, 17, 31]

$$p^{(n)}(T; s_{\xi^\ddagger}, s_0) = \frac{P^{(n)}(\xi^\ddagger)}{P^{(n)}(0)} \tag{4.29}$$

where

$$P^{(n)}(\xi) = \frac{\int\int e^{-\beta_n H_n^s(\mathbf{p},\mathbf{q})} \delta[s(\bar{\mathbf{q}}) - \xi] \, d^{fn}\mathbf{q} \, d^{fn}\mathbf{p}}{\int\int e^{-\beta_n H_n^s(\mathbf{p},\mathbf{q})} \, d^{fn}\mathbf{q} \, d^{fn}\mathbf{p}} \tag{4.30}$$

represents the probability that a reaction coordinate $s(\bar{\mathbf{q}})$ that interpolates between $s_0(\bar{\mathbf{q}})$ and $s_1(\bar{\mathbf{q}})$ has the value of ξ. This probability is evaluated using a modified ring polymer Hamiltonian of the form

$$H_s^{(n)}(\mathbf{p}, \mathbf{q}) = H^{(n)}(\mathbf{p}, \mathbf{q}) - \frac{1}{\beta_n} \ln f_s(\bar{\mathbf{q}}) \tag{4.31}$$

with $f_s(\bar{\mathbf{q}})$ as defined in Equation (4.19). A reasonable choice for the interpolating reaction coordinate $s(\bar{\mathbf{q}})$ is

$$s(\bar{\mathbf{q}}) = \frac{s_0(\bar{\mathbf{q}})}{s_0(\bar{\mathbf{q}}) - s_1(\bar{\mathbf{q}})} \tag{4.32}$$

Using the above definition, $s(\bar{\mathbf{q}}) = 0$ represents the dividing surface near the reactants and $s(\bar{\mathbf{q}}) = 1$ represents the transition state dividing surface.

Another way of representing $p^{(n)}(T; s_{\xi^\ddagger}, s_0)$ is

$$p^{(n)}(T; s_{\xi^\ddagger}, s_0) = e^{-\beta\left[W(\xi^\ddagger) - W(0)\right]} \tag{4.33}$$

where

$$W(\xi) = -\frac{1}{\beta} \ln P^{(n)}(\xi) \tag{4.34}$$

is the centroid potential of mean force (free energy) associated with the probability distribution $P^{(n)}(\xi)$. The free energy difference $W(\xi^{\ddagger}) - W(0)$ can be efficiently evaulated using the technique of umbrella integration [32, 33]. In this method, the reaction coordinate is discretized into a series of N_{windows} windows. Molecular dynamics trajectories are simulated within each window using a parabolic umbrella potential

$$w_i(\bar{\mathbf{q}}) = \frac{1}{2} k_i \left[s(\bar{\mathbf{q}}) - \xi_i \right]^2 \tag{4.35}$$

with force constant k_i to determine the mean $\bar{\xi}_i^{\text{b}}$ and variance $(\sigma_i^{\text{b}})^2$ in each window. The umbrella potential has the advantage of greatly encouraging sampling of the high-energy windows near $s(\bar{\mathbf{q}}) = 1$, but it also introduces a bias into the computed trajectories (which is reflected by the superscript b).

The biased probability distribution $P_i^{(n)\text{b}}(\xi)$ in each window is approximated as a normal distribution using the determined mean and variance,

$$P_i^{(n)\text{b}}(\xi) = \frac{1}{\sigma_i^{\text{b}} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\xi - \bar{\xi}_i^{\text{b}}}{\sigma_i^{\text{b}}} \right)^2 \right] \tag{4.36}$$

This bias can be easily removed while computing the derivative of the potential of mean force $\partial W(\xi)/\partial \xi$ on a grid in the reaction coordinate ξ:

$$\frac{\partial W(\xi)}{\partial \xi} = \sum_{i=1}^{N_{\text{windows}}} \frac{N_i P_i^{(n)\text{b}}(\xi)}{\sum_{j=1}^{N_{\text{windows}}} N_j P_j^{(n)\text{b}}(\xi)} \left[\frac{1}{\beta} \frac{\xi - \bar{\xi}_i^{\text{b}}}{(\sigma_i^{\text{b}})^2} - k_i(\xi - \xi_i) \right] \tag{4.37}$$

Numerical integration of $\partial W(\xi)/\partial \xi$ over the ξ values gives the value of $W(\xi^{\ddagger}) - W(0)$.

4.2.3 Transmission coefficient

If the transition state dividing surface $s_{\xi^{\ddagger}}(\bar{\mathbf{q}})$ is chosen carefully such that the number of recrossings is negligible, then the value of the transmission coefficient $\kappa^{(n)}(T; s_{\xi^{\ddagger}})$ is unity. In practice, however, constructing such a dividing surface is complex and

tedious, so $\kappa^{(n)}(T; s_{\xi^\ddagger})$ may be less than one (since the TST rate coefficient is an upper bound on the true rate coefficient).

The transmission coefficient $\kappa^{(n)}(T; s_{\xi^\ddagger})$ can be written in the form

$$\kappa^{(n)}(T; s_{\xi^\ddagger}) = \lim_{t\to\infty} \frac{\left\langle v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0) h[s(\bar{\mathbf{q}}_t - \xi^\ddagger)] \delta[s(\bar{\mathbf{q}}_0 - \xi^\ddagger)] \right\rangle}{\left\langle v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0) h[v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0)] \delta[s(\bar{\mathbf{q}}_0 - \xi^\ddagger)] \right\rangle} \tag{4.38}$$

where the angle brackets represent a phase space average

$$\langle A \rangle = \frac{\int\int e^{-\beta_n H_n(\mathbf{p},\mathbf{q})} A(\mathbf{p},\mathbf{q})\, d^{fn}\mathbf{q}\, d^{fn}\mathbf{p}}{\int\int e^{-\beta_n H_n(\mathbf{p},\mathbf{q})}\, d^{3fn}\mathbf{q}\, d^{3fn}\mathbf{p}} \tag{4.39}$$

Physically, the above expression represents the fraction of trajectories that start at the transition state dividing surface with a positive velocity (*i.e.* moving from reactants to products) that end up in the product space in the long time limit. The delta functions in both the numerator and denominator indicate that the phase space averages can be evaulated using the constraint that the centroid of the ring polymer begins on the dividing surface. For trajectories where this constraint is enforced, the transmission coefficient expression becomes

$$\kappa^{(n)}(T; s_{\xi^\ddagger}) = \lim_{t\to\infty} \frac{\left\langle f_s(\bar{\mathbf{q}}_0)^{-1} v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0) h[s(\bar{\mathbf{q}}_t - \xi^\ddagger)] \right\rangle_s}{\left\langle f_s(\bar{\mathbf{q}}_0)^{-1} v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0) h[v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0)] \right\rangle_s} \tag{4.40}$$

where we have added a subscript s to the angle brackets to represent the constrained ensemble. The value $f_s(\bar{\mathbf{q}}_0)^{-1}$ represents a correction for this constraint.

Evaluating the value of $\kappa^{(n)}(T; s_{\xi^\ddagger})$ using molecular dynamics is performed as follows. A parent trajectory constrained to the transition state dividing surface using the RATTLE algorithm [34] is evolved in the presence of a thermostat to obtain a statistical sampling of transition state configurations. For each configuration obtained from the parent trajectory, a set of child trajectories are initiated with sampled initial momenta and evolved without the thermostat to obtain $s(\bar{\mathbf{q}}_t)$. For each child trajectory, the value of $f_s(\bar{\mathbf{q}}_0)^{-1} v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0)$ is counted at each time t if the trajectory is in the product space (for the numerator). For the denominator, this value is counted if at $t = 0$ the child trajectory has a velocity moving toward the product space (*i.e.* $v_s(\bar{\mathbf{p}}_0, \bar{\mathbf{q}}_0) > 0$). After many trajectories have been sampled, the ratio of the numerator and denominator at long time is taken to obtain the value of the transmission coefficient.

4.3 Application of RPMD rate theory to the kinetic isotope effect of $OH + CH_4 \longrightarrow CH_3 + H_2O$

The hydrogen abstraction reaction

$$OH + CH_4 \longrightarrow CH_3 + H_2O \qquad (R1)$$

is an important reaction in atmospheric chemistry as the main process by which methane is removed from the atmosphere [35], and in combustion chemistry as a propagation reaction of high-temperature methane combustion [36]. This reaction has significant quantum character due to both zero-point energy and tunneling effects; however, no exact quantum mechanical results are currently available for this seven-atom system. In this section we compare the results of applying RPMD rate theory to this reaction with other methods of incorporating quantum effects, including variational transition state theory with multidimensional tunneling and quantum instanton (QI) theory. In particular, this discussion focuses on exploring the kinetic isotope effect (KIE) of this reaction relative to

$$OH + CD_4 \longrightarrow CD_3 + HDO \qquad (R2)$$

in order to minimize the effect of inaccuracies in the analytical potential energy surface, and thereby focus on the performance of the various rate theories. The KIE of reactions R1 and R2 has been measured in three independent experiments, two of which agree very closely over the temperature range 300-400 K where their experiments overlap [37–39]. However, although there have been several theoretical calculations of the KIE, they generally are not consistent with one another or the experimental result [40–46].

4.3.1 Methodology

RPMD rate theory was used to calculate the rate coefficients of the reactions R1 and R2 at several temperatures over the range 200 to 1000 K. The RPMD calculations utilized 128 beads at all temperatures below 700 K and 64 beads at the higher temperatures. Energies and forces were determined using the global potential energy surface of Espinosa-García and Corchado [42]. The $^2\Pi_{1/2}$ excited state of OH, which is only 140 cm^{-1} above the ground state [47], was included in the electronic partition function. All RPMD calculations were performed using the `RPMDrate` software package.

The RPMD results were compared against a variety of previously published theoretical results using the same potential energy surface. The canonical unified statistical model with microcanonical optimized multidimensional tunneling (CUS/μOMT) results were taken from Espinosa-García and Corchado [42], while the quantum instanton (QI) results were taken from Wang and Zhou [46]. Two additional theoretical results obtained via direct dynamics are also included. Melissas and Truhlar applied canonical variational transition state theory with small curvature tunneling (CVT/SCT) at the MP-SAC2//MP2/adj-cc-pVTZ level [40, 41]. Masgrau and coworkers also utilized variational transition state theory with interpolated single-point energy corrections and small curvature tunneling (VTST-ISPE/SCT) using the MCCM-CCSD(T)-1*sc* multicoefficient correlation method [43–45].

Representative experimental results for the $OH + CH_4 \longrightarrow CH_3 + H_2O$ rate coefficient are taken from Vaghjiani and Ravishankara [48], Sharkey and Smith [49], and Bryukov *et al.* [50] The experimental kinetic isotope effects are taken from Gordon and Mulac [37], Dunlop and Tully [38], and Gierczak *et al.* [39]

4.3.2 Results and dicussion

Table 4.1 compares the values of the rate coefficient $k(T)$ for the reaction $OH + CH_4 \longrightarrow CH_3 + H_2O$ as computed using various theoretical methods. The same results are plotted in Figure 4.1. The CUS/μOMT results of Espinosa-García and Corchado show very good agreement with experimental data, even well into the deep tunneling regime (computed by us to be below 387 K, based on the imaginary frequency of 1703 cm^{-1} used by Espinosa-García and Corchado). This level of accuracy is initially surprising due to the many approximations utilized in the CUS/μOMT method, including a 1D hindered rotor approximation along the minimum energy path and harmonic approximation for all other modes, which are unlikely to accurately describe tunneling at low temperatures. However, we note that the CUS/μOMT method was applied by the same authors who constructed the analytical PES; the authors calibrated this PES so that the CUS/μOMT method reproduced the experimental results.

The quantum instanton (QI) method is expected to be more accurate than the CUS/μOMT method since it eliminates several of the latter's approximations. However, the results for reaction R1 are mixed; at 300 K the QI result is more accurate than the CUS/ÎijOMT result, but at 250 K the QI method is less accurate. At 200 K – for which no experimental data was available when the analytical PES was constructed – the QI rate is faster than the CUS/μOMT one by a factor of three, a much larger disagreement than at the other temperatures, suggesting that the CUS/μOMT

Table 4.1: Comparison of rate coefficients (in cm^3 $molecule^{-1}$ s^{-1}) for the $OH + CH_4 \longrightarrow CH_3 + H_2O$ reaction calculated using several theoretical methods.

T (K)	CVT/SCT[a]	VTST-ISPE/SCT[b]	CUS/μOMT[c]	QI[d]	cd-QTST[e]	RPMD[e]	Experiment[f]
200		1.51×10^{-16}	5.40×10^{-16}	1.53×10^{-15}	7.08×10^{-15}	3.04×10^{-15}	
250	2.64×10^{-15}		2.20×10^{-15}	3.33×10^{-15}	1.58×10^{-14}	8.18×10^{-15}	2.04×10^{-15}
298	5.48×10^{-15}	3.45×10^{-15}					6.30×10^{-15}
							6.81×10^{-15}
300	5.65×10^{-15}	3.61×10^{-15}	6.40×10^{-15}	7.65×10^{-15}	3.07×10^{-14}	1.75×10^{-14}	7.60×10^{-15}
400	2.30×10^{-14}	2.36×10^{-14}	3.10×10^{-14}	2.39×10^{-14}	9.33×10^{-14}	5.83×10^{-14}	3.48×10^{-14}
500	7.04×10^{-14}	7.31×10^{-14}	9.30×10^{-14}	8.08×10^{-14}	2.20×10^{-13}	1.37×10^{-13}	1.04×10^{-13}
600	1.71×10^{-13}	1.80×10^{-13}	2.10×10^{-13}	2.13×10^{-13}	4.55×10^{-13}	2.77×10^{-13}	2.35×10^{-13}
700		3.63×10^{-13}		4.21×10^{-13}	8.46×10^{-13}	4.86×10^{-13}	4.45×10^{-13}
800	6.41×10^{-13}		6.80×10^{-13}	8.31×10^{-13}	1.46×10^{-12}	7.24×10^{-13}	7.49×10^{-13}
1000	1.67×10^{-12}	1.53×10^{-12}	1.60×10^{-12}	1.80×10^{-12}	3.56×10^{-12}	1.37×10^{-12}	1.69×10^{-12}

[a] From Melissas and Truhlar [40, 41].
[b] From Masgrau and coworkers [43–45].
[c] From Espinosa-García and Corchado [42].
[d] From Wang and Zhou [46].
[e] From this work.
[f] From Vaghjiani and Ravishankara [48], Sharkey and Smith [49], and Bryukov *et al.* [50]

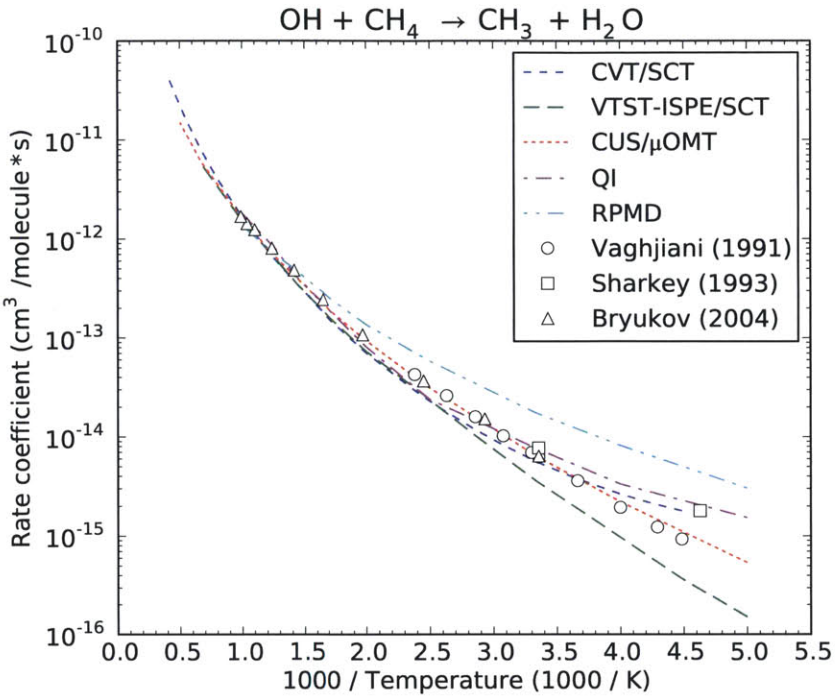


Figure 4.1: Plot of rate coefficients for the $OH + CH_4 \longrightarrow CH_3 + H_2O$ reaction calculated using several theoretical methods. In the deep tunneling regime (below 387 K) the transition state theory-based methods show significant disagreement. The RPMD result overpredicts the rate coefficient by a larger value than anticipated, likely due to less accurate regions of the potential energy surface.

method significantly underestimates the tunneling contribution, and therefore that the analytical PES is inaccurate.

The RPMD results of this work further confirm the inaccuracy of the PES. Table 4.1 shows the RPMD results with and without the transmission coefficient, with the latter labeled cd-QTST to reflect its coincidence with the centroid-density version of quantum transition state theory. At 1000 K the RPMD rate is 20-30% lower than both the other theoretical methods and the experimental results. Since RPMD rate theory is exact at high temperatures, the deviation of the RPMD result from experiment is likely due to the fitting of the analytical PES with the CUS/μOMT method, *i.e.* the true reaction barrier is a little lower than that on the fitted PES. The RPMD transmission coefficient – which can be calculated by dividing the cd-QTST value by the RPMD value – is very small (0.4) at 1000 K, consistent with the expected high-level of recrossing in heavy-light-heavy reactions.

Based on previous experience, we expected the RPMD rate coefficient to overes-

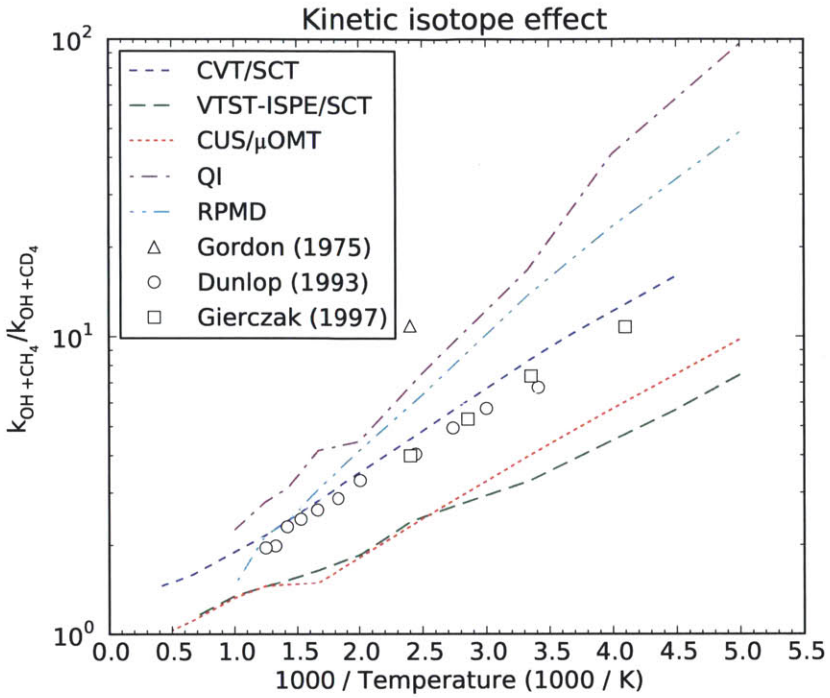


Figure 4.2: Plot of kinetic isotope effects for $OH + CH_4 \longrightarrow CH_3 + H_2O$ to $OH + CD_4 \longrightarrow CD_3 + HDO$ calculated using several theoretical methods. The predictions of the transition state theory-based methods again vary significantly, with some methods overpredicting and others underpredicting the KIE. The RPMD result overpredicts the KIE by the expected factor, and is also generally closer to the experimental values than the other methods.

timate the exact rate by about a factor of two to three in the deep tunneling regime (below 387 K). However, the results show a larger error than expected; for instance, the RPMD rate coefficient at 250 K is about four times larger than that observed experimentally. This discrepancy is also likely due to the fitting of the analytical PES, as the RPMD result at 250 K is only 2.5 times larger than the QI result. For this reason we focus our evaluation more on the predicted kinetic isotope effects, which are much less dependent on the accuracy of the potential energy surface, and should give a better indication of the relative performance of each method.

Table 4.2 and Figure 4.2 shows the values of the kinetic isotope effect for R1 to R2 at several temperatures as predicted by the various theoretical methods and measured experimentally. Because KIEs are less sensitive to the PES parameters, we have included two two additional theoretical calculations that utilized direct dynamics instead of the analytical PES to compute energies and forces: the CVT/SCT results of

Table 4.2: Comparison of kinetic isotope effects for $OH + CH_4 \longrightarrow CH_3 + H_2O$ to $OH + CD_4 \longrightarrow CD_3 + HDO$ calculated using several theoretical methods.

T (K)	CVT/SCT[a]	VTST-ISPE/SCT[b]	CUS/μOMT[c]	QI[d]	cd-QTST[e]	RPMD[e]	Experiment[f]
200		7.45	9.79	96.39	73.69	48.64	
250			5.73	41.11	34.15	23.47	10.19
300	8.27	3.27	3.97	16.81	19.15	13.78	6.96
400	4.82	2.43	2.45	7.54	9.00	6.38	4.31
500		1.85	1.81	4.45	5.36	4.14	3.30 (498 K)
600		1.64	1.49	4.16	3.90	3.08	
700		1.52		3.12	3.07	2.41	
800	2.16		1.45	2.80	2.64	2.16	1.96
1000		1.34	1.32	2.25	2.20	1.46	

[a] From Melissas and Truhlar [40, 41].
[b] From Masgrau and coworkers [43–45].
[c] From Espinosa-García and Corchado [42].
[d] From Wang and Zhou [46].
[e] From this work.
[f] From Gordon and Mulac [37], Dunlop and Tully [38], and Gierczak *et al.* [39]

Melissas and Truhlar and the VTST-ISPE/SCT results of Masgrau *et al.* The experimental KIE values are larger than two even at 1000 K, and increase as temperature decreases, showing a much smaller tunneling effect for R2 than R1.

The transition state theory-based methods give contradictory estimates of the KIE, with the CVT/SCT and QI methods overpredicting the KIE and the VTST-ISPE/SCT and CUS/μOMT methods underpredicting the KIE. Surprisingly, the oldest and least sophisticated CVT/SCT calculation is the most accurate, though this accuracy is likely a result of error cancellation. The CUS/μOMT method improves on CVT/SCT by accounting for some recrossing of the optimized dividing surface, while the VTST-ISPE/SCT method uses dual level direct dynamics to correct the minimum energy path using interpolated single-point energies. Both refinements cause the calculated KIEs to be much lower than those of the CVT/SCT method, but both are also further from the experimental data. Conversely, the QI method gives KIEs that are much larger than experiment, especially at very low temperatures, where the QI result is four times larger than experiment. This inconsistent, unpredictable performance of the TST-based methods is a direct result of their sensitivity to the choice of transition state dividing surface.

Our cd-QTST results are very similar to the QI results except at very low temperatures, where the cd-QTST result is slightly more accurate. By contrast, in previous investigations of 1D Eckart barrier and gas-phase atom-diatom reactions, the QI method often provided equal or more accurate results than those of CVT, cd-QTST, or RPMD rate theory [9, 31]. This system is the second demonstration that the accuracy of the QI method, which depends on the transition state dividing surface, decreases when the method is applied to larger polyatomic reactions where it is harder to draw an accurate dividing surface [17].

Since the experiments show that the deuterium abstraction reaction R2 has a smaller tunneling contribution than the corresponding hydrogen abstraction R1, we expect the RPMD rate coefficient for R2 to be computed very accurately. As a result, we anticipate that the KIE will be overpredicted by RPMD theory by about factor of two at low temperatures due to the anticipated overprediction of the rate of R1. Indeed, the results in Table 4.2 are consistent with our expectations. Within the methods that utilize the same potential energy surface, the RPMD result is consistently closer to the experimental data than the QI result at all temperatures, and is also significantly more accurate than the CUS/μOMT result at high temperatures, where the KIE is even less sensitive to the classical parameters of the PES. RPMD gives a more accurate KIE than CUS/μOMT even despite the tuning of the analytical PES to repro-

duce the experimental KIE data using the latter method [42]. Note that the inclusion of the transmission coefficient, which distinguishes RPMD from the otherwise-identical cd-QTST method, improves the accuracy significantly. The transmission coefficient correction is what makes RPMD theory independent of the choice of transition state dividing surface.

4.4 Conclusion

Ring polymer molecular dynamics reaction rate theory is the most accurate method to date of predicting the kinetic isotope effect of $OH + CD_4 \longrightarrow CD_3 + HDO$ to $OH + CH_4 \longrightarrow CH_3 + H_2O$. The consistent and predictable accuracy of RPMD rate theory is a direct result of its independence of the choice of transition state dividing surface. This feature is not shared by any of the approximate quantum transition state theory methods, and makes RPMD extremely attractive for further application to polyatomic chemical reactions when the exact quantum mechanical calculations are impossibly expensive.

The result of this work has been a general-purpose, easy-to-use software package named `RPMDrate` for calculating bimolecular rate coefficients using RPMD rate theory. In particular, we have made it easy to (1) define the transition state dividing surface in terms of the forming and breaking bonds, (2) connect either an existing potential energy surface or to direct dynamics calculations, (3) spread the calculation across multiple processors in order to accelerate the computation via parallelization, (4) allow for easy restarting of jobs if additional sampling trajectories are desired, and (5) included several optional validity checks on the trajectories and resulting values.

4.5 References

[1] S. C. Althorpe and D. C. Clary. "Quantum Reactive Scattering Calculations on Chemical Reactions." *Ann. Rev. Phys. Chem.* 54 (1), 2003, pp. 493–529. DOI: 10.1146/annurev.physchem.54.011002.103750.

[2] A. Fernández-Ramos, J. A. Miller, S. J. Klippenstein, and D. G. Truhlar. "Modeling the Kinetics of Bimolecular Reactions." *Chem. Rev.* 106 (11), 2006, pp. 4518–4584. DOI: 10.1021/cr050205w.

[3] M. J. Gillan. "Quantum Simulation of Hydrogen in Metals." *Phys. Rev. Lett* 58, 1987, pp. 563–566. DOI: 10.1103/PhysRevLett.58.563.

[4] M. J. Gillan. "Quantum-Classical Crossover of the Transition Rate in the Damped Double Well." *J. Phys. C: Solid State Phys.* 20, 1987, pp. 3621–3641. DOI: 10.1088/0022-3719/20/24/005.

[5] G. A. Voth, D. Chandler, and W. H. Miller. "Rigorous Formulation of Quantum Transition State Theory and its Dynamical Corrections." *J. Chem. Phys.* 91, 1989, pp. 7749–7760. DOI: 10.1063/1.457242.

[6] N. F. Hansen and H. C. Andersen. "A New Formulation of Quantum Transition State Theory for Adiabatic Rate Constants." *J. Chem. Phys.* 101, 1994, pp. 6032–6037. DOI: 10.1063/1.467318.

[7] N. F. Hansen and H. C. Andersen. "Properties of Quantum Transition State Theory and Its Corrections." *J. Phys. Chem.* 100 (4), 1996, pp. 1137–1143. DOI: 10.1021/jp951605y.

[8] W. H. Miller, Y. Zhao, M. Ceotto, and S. Yang. "Quantum Instanton Approximation for Thermal Rate Constants of Chemical Reactions." *J. Chem. Phys.* 119, 2003, pp. 1329–1342. DOI: 10.1063/1.1580110.

[9] C. Venkataraman and W. H. Miller. "The Quantum Instanton (QI) Model for Chemical Reaction Rates: The "Simplest" QI with One Dividing Surface." *J. Phys. Chem. A* 108 (15), 2004, pp. 3035–3039. DOI: 10.1021/jp031112q.

[10] W. H. Miller. "Semiclassical Limit of Quantum Mechanical Transition State Theory for Nonseparable Systems." *J. Chem. Phys.* 62 (5), 1975, pp. 1899–1906. DOI: 10.1063/1.430676.

[11] I. R. Craig and D. E. Manolopoulos. "Chemical Reaction Rates from Ring Polymer Molecular Dynamics." *J. Chem. Phys.* 122, 2005, p. 084106. DOI: 10.1063/1.1850093.

[12] I. R. Craig and D. E. Manolopoulos. "A Refined Ring Polymer Molecular Dynamics Theory of Chemical Reaction Rates." *J. Chem. Phys.* 123, 2005, p. 034102. DOI: 10.1063/1.1954769.

[13] D. Chandler and P. G. Wolynes. "Exploiting the Isomorphism Between Quantum Theory and Classical Statistical Mechanics of Polyatomic Fluids." *J. Chem. Phys.* 74 (7), 1981, pp. 4078–4095. DOI: 10.1063/1.441588.

[14] R. Collepardo-Guevara, Y. V. Suleimanov, and D. E. Manolopoulos. "Bimolecular Reaction Rates from Ring Polymer Molecular Dynamics." *J. Chem. Phys.* 130, 2009, p. 174713. DOI: 10.1063/1.3127145.

[15] R. P. de Tudela, F. J. Aoiz, Y. V. Suleimanov, and D. E. Manolopoulos. "Chemical Reaction Rates from Ring Polymer Molecular Dynamics: Zero Point Energy Conservation in $Mu + H_2 \longrightarrow MuH + H$." *J. Phys. Chem. Lett.* 3 (4), 2012, pp. 493–497. DOI: 10.1021/jz201702q.

[16] Y. V. Suleimanov, R. P. de Tudela, P. G. Jambrina, J. F. Castillo, V. Saez-Rabanos, D. E. Manolopoulos, and F. J. Aoiz. "A Ring Polymer Molecular Dynamics Study of the Isotopologues of the $H + H_2$ Reaction." *Phys. Chem. Chem. Phys.* 15, 2013, pp. 3655–3665. DOI: 10.1039/C2CP44364C.

[17] Y. V. Suleimanov, R. Collepardo-Guevara, and D. E. Manolopoulos. "Bimolecular Reaction Rates from Ring Polymer Molecular Dynamics: Application to $H + CH_4 \longrightarrow H_2 + CH_3$." *J. Chem. Phys.* 134, 2011, p. 044131. DOI: 10.1063/1.3533275.

[18] Y. Li, Y. V. Suleimanov, J. Li, W. H. Green, and H. Guo. "Rate Coefficients and Kinetic Isotope Effects of the $X + CH_4 \longrightarrow CH_3 + HX$ (X = H, D, Mu) Reactions from Ring Polymer Molecular Dynamics." *J. Chem. Phys.* 138 (9), 2013, p. 094307. DOI: 10.1063/1.4793394.

[19] Y. Li, Y. V. Suleimanov, M. Yang, W. H. Green, and H. Guo. "Ring Polymer Molecular Dynamics Calculations of Thermal Rate Constants for the $O(^3P) + CH_4 \longrightarrow OH+CH_3$ Reaction: Contributions of Quantum Effects." *J. Phys. Chem. Lett.* 4 (1), 2013, pp. 48–52. DOI: 10.1021/jz3019513.

[20] J. O. Richardson and S. C. Althorpe. "Ring-Polymer Molecular Dynamics Rate-Theory in the Deep-Tunneling Regime: Connection with Semiclassical Instanton Theory." *J. Chem. Phys.* 131 (21), 2009, p. 214106. DOI: 10.1063/1.3267318.

[21] Y. V. Suleimanov, J. W. Allen, and W. H. Green. "RPMDrate: Bimolecular Chemical Reaction Rates from Ring Polymer Molecular Dynamics." *Comp. Phys. Comm.* 184, 2013, pp. 833–840. DOI: 10.1016/j.cpc.2012.10.017.

[22] H.-D. Meyer, U. Manthe, and L. Cederbaum. "The Multi-Configurational Time-Dependent Hartree Approach." *Chem. Phys. Lett.* 165 (1), 1990, pp. 73–78. DOI: 10.1016/0009-2614(90)87014-I.

[23] T. Yamamoto. "Quantum Statistical Mechanical Theory of the Rate of Exchange Chemical Reactions in the Gas Phase." *J. Chem. Phys.* 33, 1960, pp. 281–289. DOI: 10.1063/1.1731099.

[24] W. H. Miller. "Quantum Mechanical Transition State Theory and a New Semiclassical Model for Reaction Rate Constants." *J. Chem. Phys.* 61, 1974, pp. 1823–1834. DOI: 10.1063/1.1682181.

[25] W. H. Miller, S. D. Schwartz, and J. W. Tromp. "Quantum Mechanical Rate Constants for Bimolecular Reactions." *J. Chem. Phys.* 79, 1983, pp. 4889–4899. DOI: 10.1063/1.445581.

[26] E. Wigner. *Z. Phys. Chem. Abt. B* 19, 1932, p. 203.

[27] H. Eyring. "The Activated Complex in Chemical Reactions." *J. Chem. Phys.* 3, 1935, pp. 107–115. DOI: 10.1063/1.1749604.

[28] J. C. Keck. "Variational Theory of Chemical Reaction Rates Applied to Three-Body Recombinations." *J. Chem. Phys.* 32, 1960, pp. 1035–1050. DOI: 10.1063/1.1730846.

[29] C. H. Bennett. "Molecular Dynamics and Transition State Theory: The Simulation of Infrequent Events." *Algorithms for Chemical Computations*. Ed. by R. E. Christoffersen. Vol. 46. ACS Symposium Series. Chap. 5, pp. 63–97. DOI: 10.1021/bk-1977-0046.ch004.

[30] D. Chandler. "Statistical Mechanics of Isomerization Dynamics in Liquids and the Transition State Approximation." *J. Chem. Phys.* 68, 1978, pp. 2959–2970. DOI: 10.1063/1.436049.

[31] T. Yamamoto and W. H. Miller. "On the Efficient Path Integral Evaluation of Thermal Rate Constants Within the Quantum Instanton Approximation." *J. Chem. Phys.* 120, 2004, pp. 3086–3099. DOI: 10.1063/1.1641005.

[32] J. Kästner and W. Thiel. "Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: Umbrella Integration." *J. Chem. Phys.* 123, 2005, p. 144104. DOI: 10.1063/1.2052648.

[33] J. Kästner and W. Thiel. "Analysis of the Statistical Error in Umbrella Sampling Simulations by Umbrella Integration." *J. Chem. Phys.* 124, 2006, p. 234106. DOI: 10.1063/1.2206775.

[34] H. C. Andersen. "Rattle: A "Velocity" Version of the SHAKE Algorithm for Molecular Dynamics Calculations." *J. Comput. Phys.* 52(1), 1983, pp. 24–34. DOI: 10.1016/0021-9991(83)90014-1.

[35] R. P. Wayne. *Chemistry of Atmospheres*. Third edition. Oxford University Press, 2000.

[36] S. M. Hwang, S.-O. Ryu, K. J. De Witt, and M. J. Rabinowitz. "Rate Coefficient Measurements of the Reaction $CH_3 + O_2 = CH_3O + O$." *J. Phys. Chem. A* 103 (30), 1999, pp. 5949–5958. DOI: 10.1021/jp990998o.

[37] S. Gordon and W. A. Mulac. *Proc. Symp. Chem. Kinet. Data Upper Lower Atmos.* 1, 1975, p. 289.

[38] J. R. Dunlop and F. P. Tully. "A Kinetic Study of Hydroxyl Radical Reactions with Methane and Perdeuterated Methane." *J. Phys. Chem.* 97 (43), 1993, pp. 11148–11150. DOI: 10.1021/j100145a003.

[39] T. Gierczak, R. K. Talukdar, S. C. Herndon, G. L. Vaghjiani, and A. R. Ravishankara. "Rate Coefficients for the Reactions of Hydroxyl Radicals with Methane and Deuterated Methanes." *J. Phys. Chem. A* 101 (17), 1997, pp. 3125–3134. DOI: 10.1021/jp963892r.

[40] V. S. Melissas and D. G. Truhlar. "Interpolated Variational Transition State Theory and Tunneling Calculations of the Rate Constant of the Reaction OH^+CH_4 at 223-2400 K." *J. Chem. Phys.* 99 (2), 1993, pp. 1013–1027. DOI: 10.1063/1.465401.

[41] V. S. Melissas and D. G. Truhlar. "Deuterium and Carbon-13 Kinetic Isotope Effects for the Reaction of OH with CH_4." *J. Chem. Phys.* 99 (5), 1993, pp. 3542–3552. DOI: 10.1063/1.466230.

[42] J. Espinosa-García and J. C. Corchado. "Potential Energy Surface for a Seven-Atom Reaction. Thermal Rate Constants and Kinetic Isotope Effects for CH_4 + OH." *J. Chem. Phys.* 112 (13), 2000, pp. 5731–5739. DOI: 10.1063/1.481148.

[43] L. Masgrau, A. Gonzalez-Lafont, and J. M. Lluch. "The Reactions CH_nD_{4-n} + OH $\longrightarrow$ P and CH_4 + OD $\longrightarrow$ CH_3 + HOD as a Test of Current Direct Dynamics Computational Methods to Determine Variational Transition-State Rate Constants. I." *J. Chem. Phys.* 114 (5), 2001, pp. 2154–2165. DOI: 10.1063/1.1335655.

[44] L. Masgrau, A. Gonzalez-Lafont, and J. M. Lluch. "The Reactions CH_nD_{4-n} + OH $\longrightarrow$ P and CH_4 + OD $\longrightarrow$ CH_3 + HOD as a Test of Current Direct Dynamics Multicoefficient Methods to Determine Variational Transition State Rate Constants. II." *J. Chem. Phys.* 115 (10), 2001, pp. 4515–4526. DOI: 10.1063/1.1389848.

[45] L. Masgrau, A. González-Lafont, and J. M. Lluch. "Test of Variational Transition State Theory with Multidimensional Tunneling Contributions Against Experimental Kinetic Isotope Effects for the $CH_nD_{4-n} + OH \longrightarrow P$ (n = 0, 4) Reactions." *Theor. Chem. Acc.* 108, 2002, pp. 38–40. DOI: 10.1007/s00214-002-0357-7.

[46] W. Wang and Y. Zhao. "Quantum Instanton Calculation of Rate Constant for $CH_4 + OH \longrightarrow CH_3 + H_2O$ Reaction: Torsional Anharmonicity and Kinetic Isotope Effect." *J. Chem. Phys.* 137(21), 2012, p. 214306. DOI: 10.1063/1.4768874.

[47] M. W. C. Jr., C. A. Davies, J. R. Downey, D. J. Frurip, R. A. McDonald, and A. N. Syverud. *JANAF Thermochemical Tables*. Vol. 14. Washington, D.C.: National Bureau of Standards, 1985.

[48] G. L. Vaghjiani and A. R. Ravishankara. "New Measurement of the Rate Coefficient for the Reaction of OH with Methane." *Nature* 350, 1991, pp. 406–409. DOI: 10.1038/350406a0.

[49] P. Sharkey and I. W. M. Smith. "Kinetics of Elementary Reactions at Low Temperatures: Rate Constants for the Reactions of OH with HCl ($298 \geq T/K \geq 138$), CH_4($298 \geq T/K \geq 178$) and C_2H_6($298 \geq T/K \geq 138$)." *J. Chem. Soc. Faraday Trans.* 89, 1993, pp. 631–637. DOI: 10.1039/FT9938900631.

[50] M. G. Bryukov, V. D. Knyazev, S. M. Lomnicki, C. A. McFerrin, and B. Dellinger. "Temperature-Dependent Kinetics of the Gas-Phase Reactions of OH with Cl_2, CH_4, and C_3H_8." *J. Phys. Chem. A* 108(47), 2004, pp. 10464–10472. DOI: 10.1021/jp047340h.

5

EVALUTING MASTER EQUATION METHODS FOR AUTOMATIC ESTIMATION OF PRESSURE-DEPENDENT RATE COEFFICIENTS

The conventional physical picture of a chemical reaction in the gas phase is that reaction events occur as the result of bimolecular collisions between reactant molecules. If the collision occurs with the reactant molecules in the necessary spatial orientation and with sufficient energy to overcome the intrinsic activation barrier, the reaction proceeds and the reactant molecules are converted to product molecules. This description is suitable for bimolecular reactions (of the form A + B $\longrightarrow$ products).

However, many families of chemical reactions are unimolecular (of the form A $\longrightarrow$ products). Thermal unimolecular reactions require an inert third body to provide or remove the energy needed for reaction via nonreactive collisions with the reactant molecule. The rate of these collisions depends on the concentration of the inert, which in turn is related to the pressure of the system. Therefore, under conditions where this collision rate is rate-limiting, the observed phenomenological rate coefficient $k(T, P)$ is a function of both temperature T and pressure P. An accurate estimate of the reaction rate for unimolecular reactions must account for this potential pressure dependence of the kinetics.

Accounting for the pressure-dependent kinetics of unimolecular reactions is an important step in developing detailed kinetics models of many industrially-relevant processes, including combustion and pyrolysis. This is especially important in our automatic reaction mechanism generation code, RMG, which uses a flux-based criterion for deciding which species are important to the model, as the kinetics directly influence the model composition. Since RMG often considers hundreds of thousands of unimolecular reactions in the course of constructing a reaction mechanism, we require a method that is fully automatable, successful over a wide range of conditions, reasonably accurate, and not too computationally intensive.

In this chapter, several methods of estimating phenomenological pressure- depen-

dent rate coefficients $k(T, P)$ will be evaluated to determine the appropriateness of each for automatic mechanism generation. The strengths and weaknesses of each method will be evaluated in terms of general accuracy, speed, and robustness. Particular scenarios in which certain methods work better than others will also be discussed. Finally, a set of criteria will be given to aid the user in choosing the best method for their intended application.

5.1 Unimolecular reaction rate theory

A graphical representation of a typical unimolecular system is shown in Figure 5.1. An activated species, labeled C* in the figure, can be formed either as the product of an association reaction (chemical activation) or via collisional excitation (thermal activation). 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. Often the reactive events occur more rapidly than collision events, and an excited molecule will traverse multiple reactive events – appearing to "skip" over intermediate species – before being collisionally stabilized. This suggests that there is a net reaction rate from each isomer and product to every other isomer and product set in the network, not just those directly adjacent. For example, this implies that there will be a phenomenological rate for the net reaction A + B $\longrightarrow$ D + E in the figure, despite there being no direct transition state pathway connecting them.

The discussion of unimolecular reaction rate theory in this section is based on several existing treatments [1–6]. Further details can be found in those works.

5.1.1 Collision models

The importance of bimolecular collisions in unimolecular reactions was first proposed independently by Lindemann [7] and Christiansen [8] in the early 1920s. It was soon recognized by Hinshelwood and others that a rigorous treatment of these processes required consideration of molecular energy levels [9]. Most current models of collisional energy transfer decompose the pseudo-first-order collision rate coefficient $k_{\text{coll}}(E, E')[\text{M}]$ into two terms:

$$k_{\text{coll}}(E, E')[\text{M}] = \omega \mathbb{P}(E, E') \tag{5.1}$$

In this formation, ω represents the total frequency of collisions and $\mathbb{P}(E, E')$ represents the probability that a single collision causes a transfer of energy such that the molecule

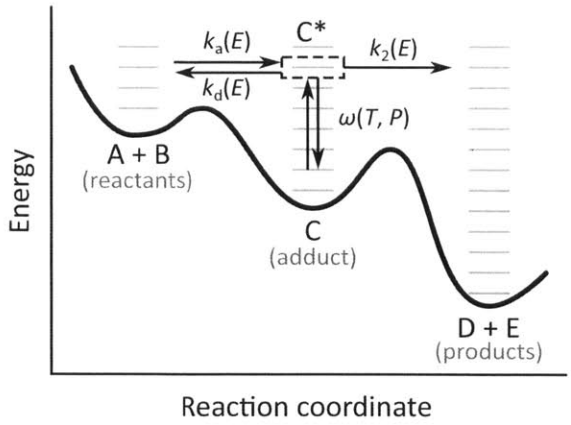


Figure 5.1: A typical unimolecular system. An activated species C* can be formed either from chemical activation (as the product of an association reaction) or thermal activation (via collisional excitation). Once activated, multiple isomerization and dissociation reactions may become competitive with 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.

moves from energy E' to energy E. A common model of the collision frequency ω is the Lennard-Jones model

$$\omega_{\mathrm{LJ}} = \Omega^{(2,2)*}(\tilde{T})\sqrt{\frac{8k_{\mathrm{B}}T}{\pi\mu}}\pi\sigma^2[\mathrm{M}] \tag{5.2}$$

where k_{B} is the Boltzmann constant, μ is the reduced mass of the colliders, σ and ϵ are the Lennard-Jones parameters, and $[\mathrm{M}]$ is the bath gas concentration. The factor $\Omega^{(2,2)*}(\tilde{T})$ represents a configurational correction, and is well-approximated by the expression

$$\Omega^{(2,2)*}(\tilde{T}) \approx 1.16145\tilde{T}^{-0.14874} + 0.52487e^{-0.77320\tilde{T}} + 2.16178e^{-2.43787*\tilde{T}} \tag{5.3}$$

where $\tilde{T} \equiv k_{\mathrm{B}}T/\epsilon$.

Models of the collisional energy transfer probability $\mathbb{P}(E, E')$ are subject to two constraints. The first is normalization of the probability:

$$\int_0^\infty \mathbb{P}(E, E')\, dE' = 1 \tag{5.4}$$

The second is detailed balance, which ensures that the correct Boltzmann distribution $b(E) \equiv \rho(E)e^{-E/k_{\mathrm{B}}T}/Q(T)$ is obtained:

$$\mathbb{P}(E, E')b(E') = \mathbb{P}(E', E)b(E) \tag{5.5}$$

The most common model for $\mathbb{P}(E, E')$, and the one used throughout this work, is the single exponential down model

$$\mathbb{P}(E, E') = C(E') \exp\left(\frac{E' - E}{\alpha}\right) \quad E < E' \tag{5.6}$$

The single exponential down model accurately reflects the intuition that a collision is much more likely to cause a small transfer of energy than a large one. This model is popular because it only contains one parameter: the average energy transferred in a deactivating collision $\alpha \equiv \langle \Delta E_{\mathrm{down}} \rangle$. Often we don't know enough about the details of the collisions to justify use of more complex models. In fact, many modeling efforts treat this as an adjustable parameter used to tune the model to the experimental measurements, although there have been some very recent efforts to compute this parameter from first principles [10–12].

5.1.2 Reaction models

Reaction events in unimolecular rate theory cause a change of configuration at constant total energy, and are therefore characterized by a microcanonical rate coefficient $k(E)$ that is a function of energy. The Rice-Ramsperger-Kassel-Marcus (RRKM) expression for the microcanonical rate coefficient $k(E)$ provides the microcanical analogue of canonical transition state theory [13–15]:

$$k(E) = \frac{N^{\ddagger}(E)}{h\rho(E)} \tag{5.7}$$

Above, $N^{\ddagger}(E)$ is the sum of states of the transition state, $\rho(E)$ is the density of states of the reactant, and h is the Planck constant. If reactant and transition state energies, geometries, and vibrational frequencies are available (e.g. from a quantum chemistry calculation), then RRKM theory gives excellent results. However, such information is often expensive to obtain and nontrivial to automate, since it requires the sum of states for the transition state. Furthermore, anharmonicity effects – such as hindered internal rotations – become increasingly important for larger polyatomic molecules, and

anharmonic corrections are often necessary to achieve rate coefficients at accuracies better than an order of magnitude.

An alternative approach for when such information is not available is to utilize the inverse Laplace transform to transform the canonical rate $k(T)$ in the high-pressure limit into the microcanonical rate coefficient $k(E)$. The relationship between these, recognized by Slater [16], is

$$k(E)\rho(E) = \mathcal{L}^{-1}\left[Q(\beta)k_\infty(\beta)\right] \tag{5.8}$$

where $\beta \equiv (k_\mathrm{B}T)^{-1}$ is the transform variable corresponding to the energy E. Exact formulas exist for simple Arrhenius kinetics [5] and modified Arrhenius kinetics $k_\infty(T) = AT^n \exp\left(-E_\mathrm{a}/RT\right)$ for $n > -\frac{1}{2}$ [17]. The inverse Laplace transform method assumes that the given $k(T)$ expression is valid over the temperature range from zero to infinity, and that the activation energy E_a is physically equivalent to the reaction barrier height E_0, both of which are nontrivial approximations. Note that, for simple Arrhenius kinetics ($n = 0$) the quantum Rice-Ramsperger-Kassel (QRRK) method is essentially the same as the inverse Laplace transform method.

The RRKM or inverse Laplace transform method is used to determine the microcanonical rate coefficient for the forward reaction only, as the reverse rate coefficient is determined from detailed balance. For an isomerization reaction $\mathrm{A} \longrightarrow \mathrm{B}$ the detailed balance equation is

$$k_\mathrm{f}(E)\rho_\mathrm{A}(E) = k_\mathrm{b}(E)\rho_\mathrm{B}(E) \tag{5.9}$$

and for a dissociation reaction $\mathrm{A} \longrightarrow \mathrm{B} + \mathrm{C}$ the equation is

$$k_\mathrm{f}(E)\rho_\mathrm{A}(E) = k_\mathrm{b}(E)\rho_\mathrm{BC}(E) \tag{5.10}$$

where $\rho_\mathrm{BC}(E)$ is the convolved vibrational-rotational density of states for species B and C, and also includes relative translational motion.

An alternative formulation for the detailed balance expressions incorporates the macroscopic equilibrium coefficient $K_\mathrm{eq}(T)$ and equilibrium distributions $b_i(E,T)$ at each temperature. For isomerization and dissociation, respectively, the alternative formulation is

$$k_{\mathrm{f}}(E)b_{\mathrm{A}}(E,T) = K_{\mathrm{eq}}(T)k_{\mathrm{b}}(E)b_{\mathrm{B}}(E,T) \tag{5.11}$$

$$k_{\mathrm{f}}(E)b_{\mathrm{A}}(E,T) = K_{\mathrm{eq}}(T)k_{\mathrm{b}}(E)b_{\mathrm{BC}}(E,T) \tag{5.12}$$

where $b_{\mathrm{BC}}(E,T)$ is the combined equilibrium distribution for species B and C. These two formulations are equivalent; however, there are multiple reasons to use Equations (5.11) and (5.12) instead of Equations (5.9) and (5.10):

- Only the density of states of the unimolecular isomers need be computed. This is a result of the assumption of thermalized bimolecular channels, which means that we only need to compute the product $k_{\mathrm{b}}(E)b_{\mathrm{BC}}(E,T)$, and not the individual values of $k_{\mathrm{b}}(E)$ and $b_{\mathrm{BC}}(E,T)$. (In Equation (5.21) the required products are the $\mathbf{F}_{im}\mathbf{b}_m$ terms.)

- Only the reactive vibrational-rotational modes need be included in the density of states. Missing modes will not affect the observed equilibrium because we are imposing the macroscopic equilibrium via $K_{\mathrm{eq}}(T)$. This is particularly important in automatic mechanism generation, where we do not yet have an efficient, accurate method of estimating the external rotational constants.

- Constants of proportionality in the density of states become unimportant, as they cancel when taking the ratio $\rho(E)/Q(\beta)$, e.g. when computing the equilibrium distribution $b(E,T)$. We will use this to include an arbitrary active K-rotor in the density of states expression.

Thus, our implementation of the master equation uses Equations (5.11) and (5.12).

5.1.3 Treatment of external rotations

In the above discussion of reaction models we treated the total energy E as the only relevant conserved quantity. In this section we briefly comment on another important conserved quantity: the total angular momentum J. The discussion is based on a symmetric top molecule, which many molecules can be approximated as. Further discussion is available in the literature [2, 4, 18, 19].

A symmetric top molecule has two of the three principal moments of inertia equivalent: $I_{\mathrm{a}} = I_{\mathrm{b}} \neq I_{\mathrm{c}}$. Such a molecule is described by a two-dimensional J-rotor and a one-dimensional K-rotor, such that the rotational energy $E_{\mathrm{r}}(J,K)$ is given by

$$E_{\mathrm{r}}(J,K) = \frac{\hbar^2}{2I_{\mathrm{a}}}J(J+1) + \frac{\hbar^2}{2}\left(\frac{1}{I_{\mathrm{c}}} - \frac{1}{I_{\mathrm{a}}}\right)K^2 \tag{5.13}$$

The formulation of RRKM theory changes depending on which rotational modes are treated as active – able to exchange energy with the reaction coordinate – and which are treated as adiabatic. If neither the J-rotor nor the K-rotor are active for reaction, the RRKM expression is

$$k(E,J,K) = \frac{N^{\ddagger}(E,J,K)}{h\rho(E,J,K)} = \frac{N^{\ddagger}_{\mathrm{v}}[E - E_0 - E^{\ddagger}_{\mathrm{r}}(J,K)]}{h\rho_{\mathrm{v}}[E - E_{\mathrm{r}}(J,K)]} \tag{5.14}$$

where $N^{\ddagger}_{\mathrm{v}}(\epsilon)$ and $\rho_{\mathrm{v}}(\epsilon)$ only contain the internal degrees of freedom (vibrations and torsions) and E_0 is the barrier height (the difference in ground-state energy between the transition state and the reactants).

More commonly there is some degree of rovibrational coupling due to Coriolis effects. This is generally handled by making the K-rotor active, which gives a new two-dimensional RRKM expression of the form

$$k(E,J) = \frac{N^{\ddagger}(E,J)}{h\rho(E,J)} = \frac{N^{\ddagger}_{\mathrm{vk}}[E - E_0 - E^{\ddagger}_{\mathrm{r}}(J)]}{h\rho_{\mathrm{vk}}[E - E_{\mathrm{r}}(J)]} \tag{5.15}$$

where

$$E_{\mathrm{r}}(J) = \frac{\hbar^2}{2I_{\mathrm{a}}} J(J+1) \tag{5.16}$$

and $N^{\ddagger}_{\mathrm{vk}}(\epsilon)$ and $\rho_{\mathrm{vk}}(\epsilon)$ contain the internal degrees of freedom as well as the K-rotor.

Rigorously, the total angular momentum should be conserved for a reaction, and so the J-rotor should be treated as adiabatic. At low and moderate temperatures, treatment of the J-rotor as adiabatic is needed to obtain very accurate estimates of $k(T,P)$. However, the resulting two-dimensional master equation is significantly more expensive to solve. To obtain a one-dimensional formulation, the J-rotor can also be made active, which gives an RRKM expression of

$$k(E) = \frac{N^{\ddagger}(E)}{h\rho(E)} = \frac{N^{\ddagger}_{\mathrm{vr}}(E - E_0)}{h\rho_{\mathrm{vr}}(E)} \tag{5.17}$$

where $N^{\ddagger}_{\mathrm{vr}}(\epsilon)$ and $\rho_{\mathrm{vr}}(\epsilon)$ contain the internal degrees of freedom as well as all rotational modes.

For the purposes of automatic mechanism generation we will limit ourselves to the one-dimensional formulation in which the J-rotor is treated as active. This master

equation is small enough to be solved quickly, and the result is generally accurate enough for all but the most sensitive reactions. It is also likely that, in the context of automatic mechanism generation, the error introduced by treating the J-rotor as active is smaller than other sources of error, such as uncertainty in the many estimated parameters.

5.1.4 The master equation

In the late 1950s master equation models of chemical systems began appearing [20–24], including an early linear integral-differential equation formulation by Widom [25]. Analytical solutions for a variety of simple models soon followed [26–28], as did the first numerical approaches [29]. Numerical methods – which are required for complex unimolecular reaction networks – became much more attractive in the 1970s with the appearance of new algorithms, including Gear's method for solving stiff systems of ordinary differential equations [30] and efficient algorithms for calculating the density of states [31–33]. In the 1990s computing power had increased to the point where it was practical to solve them numerically by discretizing the integrals over energy.

The master equation combines the collision and reaction models into a mathematical description of the unimolecular reaction network. The dependent variables describing the isomers are $p_i(E,t)$, the population distribution of isomer i over energy E at time t. From statistical mechanics we expect a Boltzmann distribution $b_i(E,T)$ to result at long times:

$$\lim_{t\to\infty} p_i(E,t) = x_{i\infty}\frac{\rho_i(E)e^{-\beta E}}{Q_i(\beta)} = x_{i\infty}b_i(E,T) \tag{5.18}$$

Above, $x_{i\infty}$ is the total population of isomer i at equilibrium, $\rho_i(E)$ and $Q_i(\beta)$ are the rovibrational density of states and corresponding partition function of isomer i, and $\beta \equiv (k_\mathrm{B}T)^{-1}$. Assuming that inelastic collisions are much more common than reactive collisions, we can treat the bimolecular reactants as thermalized, and represent the concentrations of reactant configuration n by $y_{n\mathrm{A}}(t)$ and $y_{n\mathrm{B}}(t)$, where A and B distinguish between the two reactants. Product channels are treated as irreversible sinks, and their concentration is therefore not explicitly modeled.

Neglecting the dependencies of the microcanonical rates and other quantities on the angular momentum quantum number J or any other quantum numbers besides the total energy E, and neglecting reactions which do not proceed through one of the isomers, the master equation can be written as

$$\frac{d}{dt}p_i(E,t) = \omega_i \int_0^\infty \mathbb{P}_i(E,E')p_i(E',t)\,dE' - \omega_i p_i(E,t) + \sum_{j\neq i}^{N_{\text{isom}}} k_{ij}(E)p_j(E,t) - \sum_{j\neq i}^{N_{\text{isom}}} k_{ji}(E)p_i(E,t) + \sum_{m=1}^{N_{\text{reac}}} y_{m\text{A}}(t)y_{m\text{B}}(t)f_{in}(E)b_n(E) - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} g_{ni}(E)p_i(E,t) \tag{5.19}$$

$$\frac{d}{dt}y_{m\text{A}}(t) = \frac{d}{dt}y_{m\text{B}}(t) = \sum_{i=1}^{N_{\text{isom}}} \int_0^\infty g_{mi}(E)p_i(E,t)\,dE - \sum_{i=1}^{N_{\text{isom}}} y_{m\text{A}}(t)y_{m\text{B}}(t)\int_0^\infty f_{im}(E)b_m(E)\,dE \tag{5.20}$$

A quick review of the terms in the master equation:

- ω_i is the collision frequency;
- $\mathbb{P}_i(E,E')$ is the probability of collisional transfer from energy E' to energy E;
- $k_{ij}(E)$, $f_{in}(E)$, and $g_{ni}(E)$ are the microcanonical rate coefficients for isomerization, association, and dissociation, respectively;
- $b_n(E)$ is the Boltzmann distribution for bimolecular reactant channel n;
- N_{isom}, N_{reac}, and N_{prod} are the numbers of isomers, bimolecular reactant channels, and bimolecular product channels, respectively;
- indices i and j refer to unimolecular isomers;
- index m refers to bimolecular reactant channels; and
- index n refers to bimolecular product channels.

In Equation (5.19), the first pair of terms correspond to collision, the second pair to isomerization, and the final pair to association/dissociation. Equation (5.19) applies to isomers, while Equation (5.20) applies to bimolecular reactants. Note that there are almost always additional reactions creating and destroying all the species, so Equations (5.19) and (5.20) are for an idealized situation.

Equations (5.19) and (5.20) are nonlinear, both due to the presence of the bimolecular reactant terms and because both ω_i and $\mathbb{P}_i(E,E')$ depend on the composition,

which is changing with time. The rate coefficients can be derived from considering the pseudo-first-order situation where $y_{m\mathrm{A}}(t) \ll y_{m\mathrm{B}}(t)$, and all $y(t)$ are negligible compared to the bath gas M. From these assumptions the changes in ω_i, $\mathbb{P}(E, E')$, and all $y_{m\mathrm{B}}$ can be neglected, which yields a linear equation system.

To extract the phenomenological rate coefficients numerically, it is helpful to discretize the energy E into N_{grains} grains $\{E_r\}$. This converts the linear integro-differential equations into a system of first-order ordinary differential equations with the form

$$\frac{d}{dt}\begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ y_{1\mathrm{A}} \\ y_{2\mathrm{A}} \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{K}_{12} & \dots & \mathbf{F}_{11}\mathbf{b}_1 y_{1\mathrm{B}} & \mathbf{F}_{12}\mathbf{b}_2 y_{2\mathrm{B}} & \dots \\ \mathbf{K}_{21} & \mathbf{M}_2 & \dots & \mathbf{F}_{21}\mathbf{b}_1 y_{1\mathrm{B}} & \mathbf{F}_{22}\mathbf{b}_2 y_{2\mathrm{B}} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ (\mathbf{g}_{11})^T & (\mathbf{g}_{12})^T & \dots & h_1 & 0 & \dots \\ (\mathbf{g}_{21})^T & (\mathbf{g}_{22})^T & \dots & 0 & h_2 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ y_{1\mathrm{A}} \\ y_{2\mathrm{A}} \\ \vdots \end{bmatrix} \tag{5.21}$$

where the elements of the vectors $\mathbf{p}_i$ are such that

$$(\mathbf{p}_i)_r = \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} p_i(E)\, dE \tag{5.22}$$

(We will use indices r and s throughout to indicate energy grain.) The diagonal matrices $\mathbf{K}_{ij}$ and $\mathbf{F}_{im}$ and the vectors $\mathbf{g}_{ni}$ contain the microcanonical rate coefficients for isomerization, association, and dissociation, respectively:

$$(\mathbf{K}_{ij})_{rs} = \begin{cases} \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} k_{ij}(E)\, dE & r = s \\ 0 & r \neq s \end{cases} \tag{5.23a}$$

$$(\mathbf{F}_{im})_{rs} = \begin{cases} \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} f_{im}(E)\, dE & r = s \\ 0 & r \neq s \end{cases} \tag{5.23b}$$

$$(\mathbf{g}_{ni})_r = \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} g_{ni}(E)\, dE \tag{5.23c}$$

The collisional energy transfer probabilities $(\mathbf{P}_i)_{rs}$ for isomer i are given by

$$(\mathbf{P}_i)_{rs} = \frac{1}{\Delta E_r \Delta E_s} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} \int_{E_s - \Delta E_s/2}^{E_s + \Delta E_s/2} \mathbb{P}_i(E, E')\, dE'\, dE \tag{5.24}$$

The matrices $\mathbf{M}_i$ represent the collisional energy transfer probabilities of isomer i minus the rates of reactive loss to other isomers and to reactants and products:

$$(\mathbf{M}_i)_{rs} = \begin{cases} \omega_i \left[(\mathbf{P}_i)_{rr} - 1\right] - \sum_{j \neq i}^{N_{\text{isom}}} (\mathbf{K}_{ij})_{rs} - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} (\mathbf{g}_{ni})_r & r = s \\ \omega_i (\mathbf{P}_i)_{rs} & r \neq s \end{cases} \tag{5.25}$$

The scalars h_m are simply the total rate coefficient for loss of reactant channel m due to chemical reactions:

$$h_m = -\sum_{i=1}^{N_{\text{isom}}} \sum_{r=1}^{N_{\text{grains}}} y_{m\text{B}} (\mathbf{F}_{im})_{rr} b_m(E_r) \tag{5.26}$$

5.2 Master equation reduction methods

The master equation contains more detail than is required for our detailed kinetics models, where we are generally only interested in the total population instead of the population distribution in energy. We would therefore like to transform the complex master equation matrix from Equations (5.19) and (5.20) into a small number of phenomenological rate coefficients $k(T, P)$. In this section we present the theoretical description of several master equation reduction methods. In Section 5.4 we will compare the methods using a case study to illustrate their strengths and weaknesses.

All of the methods discussed in this section share a common formalism in that they seek to express the population of each isomer $p_i(E, t)$ as a linear combination of the total populations $x_j(t)$ and $y_{m\text{A}}(t) y_{m\text{B}}$ of all of the unimolecular isomers and bimolecular reactant channels:

$$p_i(E, t) \approx \sum_{j=1}^{N_{\text{isom}}} u_{ij}(E) x_j(t) + \sum_{m=1}^{N_{\text{reac}}} v_{im}(E) y_{m\text{A}}(t) y_{m\text{B}} \tag{5.27}$$

The functions $u_{ij}(E)$ represent the portion of the population distribution of energy states of unimolecular isomer i that tracks the population of isomer j. Similarly, the functions $v_{im}(E)$ represent the population distribution of energy states of unimolecular isomer i that tracks the population of bimolecular reactant channel m. Note that, even for the best $u_{ij}(E)$ and $v_{im}(E)$, Equation (5.27) is an approximation to the true $p_i(E, t)$ given by the solution of the full master equation.

By inserting the above expression into the master equation and collecting terms, we can arrive at general expressions for the various phenomenological rate coefficients $k(T,P)$ in terms of only the $u_{ij}(E)$ and $v_{im}(E)$:

$$k_{ij}(T,P) \approx \sum_{\ell \neq i}^{N_{\text{isom}}} \int_0^\infty k_{i\ell}(E) u_{\ell j}(E)\, dE - \sum_{\ell \neq i}^{N_{\text{isom}}} \int_0^\infty k_{\ell i}(E) u_{ij}(E)\, dE - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} \int_0^\infty g_{ni}(E) u_{ij}(E)\, dE \tag{5.28a}$$

$$k_{im}(T,P) \approx \sum_{\ell \neq i}^{N_{\text{isom}}} \int_0^\infty k_{i\ell}(E) v_{\ell m}(E)\, dE + \int_0^\infty f_{im}(E) b_m(E)\, dE - \sum_{\ell \neq i}^{N_{\text{isom}}} \int_0^\infty k_{\ell i}(E) v_{im}(E)\, dE - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} \int_0^\infty g_{ni}(E) v_{im}(E)\, dE \tag{5.28b}$$

$$k_{nj}(T,P) \approx \sum_{\ell=1}^{N_{\text{isom}}} \int_0^\infty g_{n\ell}(E) u_{\ell j}(E)\, dE \tag{5.28c}$$

$$k_{nm}(T,P) \approx \sum_{\ell=1}^{N_{\text{isom}}} \int_0^\infty g_{n\ell}(E) v_{\ell m}(E)\, dE \tag{5.28d}$$

As before, the indices i and j represent unimolecular isomers of the initial adduct, m represents bimolecular reactants, and n represents bimolecular reactants and products. Thus, the rate coefficients above are for isomerization, association, dissociation, and bimolecular reactions $\mathrm{A}_m + \mathrm{B}_m \longrightarrow \mathrm{A}_n + \mathrm{B}_n$, respectively.

As the subsequent sections will show, the various master equation reduction methods provide different values of $u_{ij}(E)$ and $v_{im}(E)$, and therefore different $k(T,P)$ values. In addition, having the $u_{ij}(E)$ and $v_{im}(E)$ values enables the reconstruction of the approximate population distributions $p_i(E,t)$, which helps us further understand the performance of each method.

5.2.1 The modified strong collision (MSC) method

As described by Chang, Bozzelli, and Dean [17], the modified strong collision method utilizes a greatly simplified collision model that allows for a decoupling of the energy grains. In this model, collisional stabilization of a reactive isomer is treated as a single step process, ignoring the effects of collisional energy redistribution within the reactive energy space. An attempt to correct for the effect of collisional energy redistribution is made by modifying the collision frequency $\omega_i(T,P)$ with a collision

efficiency $\beta_i(T)$, following early work by Troe on the single-isomer "fall-off" case [34, 35]. By approximating the reactive populations as existing in pseudo-steady state, a matrix equation is formed at each energy through which $u_{ij}(E)$ and $v_{im}(E)$ expressions, and subsequently the $k(T,P)$ values, can be determined.

After applying the modified strong collision approximation, a population balance performed at a reactive energy E gives (in discretized form)

$$\begin{aligned}\frac{d}{dt}\begin{bmatrix} p_1(E,t) \\ p_2(E,t) \\ \vdots \end{bmatrix} &= \begin{bmatrix} m_1(E) & k_{12}(E) & \dots \\ k_{21}(E) & m_2(E) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}\begin{bmatrix} p_1(E,t) \\ p_2(E,t) \\ \vdots \end{bmatrix} \\ &+ \sum_{i=1}^{N_{\text{isom}}} x_i(t)\begin{bmatrix} \omega_i(T,P)\beta_i(T)b_i(E;T)\delta_{i1} \\ \omega_i(T,P)\beta_i(T)b_i(E;T)\delta_{i2} \\ \vdots \end{bmatrix} \\ &+ \sum_{n=1}^{N_{\text{reac}}} y_{n\text{A}}(t)y_{n\text{B}}\begin{bmatrix} f_{1n}(E)b_n(E;T) \\ f_{2n}(E)b_n(E;T) \\ \vdots \end{bmatrix}\end{aligned} \tag{5.29a}$$

$$\frac{d}{dt}\mathbf{p}(E,t) = \mathbf{L}(E)\mathbf{p}(E,t) + \sum_{i=1}^{N_{\text{isom}}} x_i(t)\mathbf{z}_i(E) + \sum_{m=1}^{N_{\text{reac}}} y_{m\text{A}}(t)y_{m\text{B}}\mathbf{w}_m(E) \tag{5.29b}$$

where $x_i(t)$ reflects the total population of the nonreactive energies of isomer i, δ_{ij} is the Kronecker delta, and

$$m_i(E) = -\omega_i(T,P)\beta_i(T) - \sum_{j\neq i}^{N_{\text{isom}}} k_{ji}(E) - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} g_{ni}(E) \tag{5.30}$$

is the total rate of loss from isomer i. Applying the pseudo-steady state approximation to the above gives

$$\mathbf{L}(E)\mathbf{p}(E) = -\sum_{i=1}^{N_{\text{isom}}} x_i(t)\mathbf{z}_i(E) - \sum_{n=1}^{N_{\text{reac}}} y_{n\text{A}}(t)y_{n\text{B}}\mathbf{w}_n(E) \tag{5.31}$$

for which the solution is

$$\begin{aligned}\mathbf{p}^{(i)}(E) &= -[\mathbf{L}(E)]^{-1}\mathbf{z}_i(E) \\ \mathbf{p}^{(n)}(E) &= -[\mathbf{L}(E)]^{-1}\mathbf{w}_n(E)\end{aligned} \tag{5.32a}$$

Above, $\mathbf{p}^{(i)}(E)$ is the pseudo-steady population of each isomer at reactive energy E resulting from the thermal activation of isomer i, and $\mathbf{p}^{(n)}(E)$ is the pseudo-steady population of each isomer at reactive energy E resulting from the chemical activation of reactant channel n. These can then be used to compute the $k(T,P)$ values:

$$k_{ij}T,P = \omega_i(T,P)\beta_i(T)\sum_{r=r_{0,i}}^{N_{\text{grains}}} p_i^{(j)}(E_r) \tag{5.33a}$$

$$k_{nj}(T,P) = \sum_{i=1}^{N_{\text{isom}}}\sum_{r=r_{0,i}}^{N_{\text{grains}}} g_{ni}(E)p_i^{(j)}(E_r) \tag{5.33b}$$

$$k_{im}(T,P) = \omega_i(T,P)\beta_i(T)\sum_{r=r_{0,i}}^{N_{\text{grains}}} p_i^{(m)}(E_r) \tag{5.33c}$$

$$k_{nm}(T,P) = \sum_{j=1}^{N_{\text{isom}}}\sum_{r=r_{0,j}}^{N_{\text{grains}}} g_{nj}(E_r)p_j^{(m)}(E_r) \tag{5.33d}$$

The indices above have the same meaning as in Equations (5.28): i and j represent unimolecular isomers, m represents bimolecular reactants, n represent s bimolecular reactants and products, and r represents energy grains.

The modified strong collision method can be expressed in terms of the notation presented in the introduction to this section, Equation (5.27):

$$u_{ij}(E_r) = \begin{cases} b_i(E_r)\delta_{ij} & E_r < E_{j,\text{crit}} \\ p_i^{(j)}(E_r) & E_r \geq E_{j,\text{crit}} \end{cases} \tag{5.34a}$$

$$v_{im}(E_r) = \begin{cases} 0 & E_r < E_{m,\text{crit}} \\ p_i^{(m)}(E_r) & E_r \geq E_{m,\text{crit}} \end{cases} \tag{5.34b}$$

where δ_{ij} is the Kronecker delta. The critical energy that divides the low-energy Boltzmann grains from the high-energy steady-state grains is determined as the first reactive energy grain, *i.e.* the lowest energy where any microcanonical rate coefficient $k(E)$ is nonzero. Inserting the above equations into Equations (5.28) gives identical $k(T,P)$ values as those in Equations (5.33).

5.2.2 The reservoir state (RS) method

An alternative to approximating collisional stabilization as a single-step process, proposed by Green and Bhatti [36], is to assume that, except for a depleted region near

the transition state energy, the low energy grains are Boltzmann distributed. Therefore, we partition the energy grains for each isomer; the low-energy grains form the *reservoir*, while the high-energy grains form the *active-state*. Green and Bhatti suggest the cutoff be placed a few $k_{\mathrm{B}}T$ below the lowest transition-state energy adjacent to that isomer; our work suggests that a better choice is simply to place the cutoff at the lowest adjacent transition-state energy, as the intervening grains are often better modeled as Boltzmann distributed than as pseudo-steady. (This choice only matters at temperatures where the equilibrium population of these grains is significant.) Here we neglect all reactions (including tunneling) from these reservoir grains; if tunneling below the transition state energy is important, one should set the reservoir cutoff a little lower.

As the development below implies, the mathematics of the reservoir state method parallels that of the modified strong collision method. The primary difference is that, for the reservoir state method, we will be working with all high-energy grains for all isomers at once, rather than being able to treat each energy grain independently. First, a Boltzmann distribution $\mathbf{b}_i^{\mathrm{r}}$ is imposed on the reservoir grains $\mathbf{p}_i^{\mathrm{r}}$ for each isomer i, *i.e.*

$$\mathbf{p}_i^{\mathrm{r}} = x_i(t)\mathbf{b}_i^{\mathrm{r}} \tag{5.35}$$

leaving a single time-dependent constant of proportionality $x_i(t)$ which is related to the total population of isomer i. Note that a significant fraction of the equilibrium population is in the activated grains at high temperatures, *i.e.* $\sum_{s=1}^{N_{\mathrm{res},i}}(\mathbf{b}_i^{\mathrm{r}})_s < 1$. The quantity $\mathbf{b}_i^{\mathrm{r}}$ is not renormalized, however, in order to ensure that the computed $k(T,P)$ values satisfy macroscopic equilibrium. With the reservoir approximation, the master equation for the active-state grains $\mathbf{p}_i^{\mathrm{a}}$ for each isomer is

$$\frac{d}{dt}\begin{bmatrix}\mathbf{p}_1^{\mathrm{a}}\\ \mathbf{p}_2^{\mathrm{a}}\\ \vdots\end{bmatrix} = \begin{bmatrix}\mathbf{M}_1^{\mathrm{aa}} & \mathbf{K}_{12}^{\mathrm{a}} & \cdots\\ \mathbf{K}_{21}^{\mathrm{a}} & \mathbf{M}_2^{\mathrm{aa}} & \cdots\\ \vdots & \vdots & \ddots\end{bmatrix}\begin{bmatrix}\mathbf{p}_1^{\mathrm{a}}\\ \mathbf{p}_2^{\mathrm{a}}\\ \vdots\end{bmatrix} + \sum_{j=1}^{N_{\mathrm{isom}}} x_j(t)\begin{bmatrix}\mathbf{M}_1^{\mathrm{ar}}\mathbf{b}_1^{\mathrm{r}}\delta_{1j}\\ \mathbf{M}_2^{\mathrm{ar}}\mathbf{b}_2^{\mathrm{r}}\delta_{2j}\\ \vdots\end{bmatrix} + \sum_{m=1}^{N_{\mathrm{reac}}} y_{m\mathrm{A}}(t)y_{m\mathrm{B}}\begin{bmatrix}\mathbf{F}_{1m}^{\mathrm{a}}\mathbf{b}_m^{\mathrm{a}}\\ \mathbf{F}_{2m}^{\mathrm{a}}\mathbf{b}_m^{\mathrm{a}}\\ \vdots\end{bmatrix} \tag{5.36a}$$

$$\frac{d}{dt}\mathbf{p}^{\mathrm{a}} = \mathbf{L}\mathbf{p}^{\mathrm{a}} + \sum_{j=1}^{N_{\mathrm{isom}}} x_j(t)\mathbf{z}_j + \sum_{m=1}^{N_{\mathrm{reac}}} y_{m\mathrm{A}}(t)y_{m\mathrm{B}}\mathbf{w}_m \tag{5.36b}$$

As with the modified strong collision method, the pseudo-steady state approximation is applied to the active-state grains to give a matrix equation

$$\mathbf{L}\mathbf{p}^{\mathrm{a}} = -\sum_{j=1}^{N_{\mathrm{isom}}} x_j(t)\mathbf{z}_j - \sum_{m=1}^{N_{\mathrm{reac}}} y_{m\mathrm{A}}(t)y_{m\mathrm{B}}\mathbf{w}_m \tag{5.37}$$

for which the solution is

$$\mathbf{p}^{\mathrm{a}(j)} = -\mathbf{L}^{-1}\mathbf{z}_j \tag{5.38a}$$

$$\mathbf{p}^{\mathrm{a}(m)} = -\mathbf{L}^{-1}\mathbf{w}_m \tag{5.38b}$$

where $\mathbf{p}^{(j)}$ is the pseudo-steady active-state population of each isomer resulting from the thermal activation of isomer j, and $\mathbf{p}^{(m)}$ is the pseudo-steady active-state population of each isomer resulting from the chemical activation of reactant channel m. These can then be used to compute the $k(T,P)$ values via

$$k_{ij}(T,P) = \sum_r \left(\mathbf{M}_i^{\mathrm{ra}}\mathbf{p}_i^{\mathrm{a}(j)}\right)_r \tag{5.39a}$$

$$k_{nj}(T,P) = \sum_{i=1}^{N_{\mathrm{isom}}} \mathbf{g}_{ni}\cdot\mathbf{p}_i^{\mathrm{a}(j)} \tag{5.39b}$$

$$k_{im}(T,P) = \sum_r \left(\mathbf{M}_i^{\mathrm{ra}}\mathbf{p}_i^{\mathrm{a}(m)}\right)_r \tag{5.39c}$$

$$k_{nm}(T,P) = \sum_{i=1}^{N_{\mathrm{isom}}} \mathbf{g}_{ni}\cdot\mathbf{p}_i^{\mathrm{a}(m)} \tag{5.39d}$$

where $\mathbf{p}_i^{\mathrm{a}(j)}$ and $\mathbf{p}_i^{\mathrm{a}(m)}$ contain the components of $\mathbf{p}^{\mathrm{a}(j)}$ and $\mathbf{p}^{\mathrm{a}(m)}$ relating to isomer i. Once again, indices i and j represent unimolecular isomers, index m represents bimolecular reactants, index n represents bimolecular reactants and products, and index r represents energy grains. Note that we have not needed the common pseudo-first-order assumption on bimolecular reactants (*i.e.* $y_{m\mathrm{A}}(t) \ll y_{m\mathrm{B}}$) to arrive at the above result. However, the steady-state approximations used in both the modified strong collision and reservoir state methods implicitly assume that both $y_{m\mathrm{A}}(t)$ and $y_{m\mathrm{B}}$ are small enough that $y_{m\mathrm{A}}(t)y_{m\mathrm{B}}\mathbf{F}_{im}\mathbf{b}_m$ is small.

The terms of the pseudo-steady state vector $\mathbf{p}_i^{\mathrm{a}}$ and the reservoir populations $\mathbf{b}_i^{\mathrm{r}}$ can be used to construct the $\mathbf{u}_{ij}$ and $\mathbf{v}_{im}$ vectors for use in Equation 5.27:

$$\mathbf{u}_{ij} = \begin{bmatrix} \delta_{ij}\mathbf{b}_j^{\mathrm{r}} \\ \mathbf{p}_i^{\mathrm{a}(j)} \end{bmatrix} \qquad \mathbf{v}_{im} = \begin{bmatrix} \mathbf{0} \\ \mathbf{p}_i^{\mathrm{a}(m)} \end{bmatrix} \tag{5.40}$$

Again, using the above expressions with Equations (5.28) gives identical $k(T,P)$ values as those from Equations (5.39).

Finally, a note on efficient solving of Equation (5.37). Collisional energy transfer models favor small transfers of energy over large ones. For the single exponential down model, $P(E,E') \to 0$ exponentially as $|E' - E| \to \infty$. This suggests that the probabilities are negligible for sufficiently large energy transfers, which in turn results in a matrix $\mathbf{P}_i$ that is strongly banded. We can take advantage of this bandedness by using an indexing scheme that iterates over energies as the outer loop and isomers as the inner loop. The equations in this manuscript utilize the opposite indexing scheme for ease of understanding, but our software implementation of these equations uses the more efficient indexing. A typical sparsity pattern for an active-state matrix $\mathbf{L}$ using the two indexing schemes is shown in Figure 5.2 for a three-isomer network. The grain-major indexing scheme, Figure 5.2b, yields a matrix with a much narrower bandwidth than the isomer-major indexing scheme, Figure 5.2a.

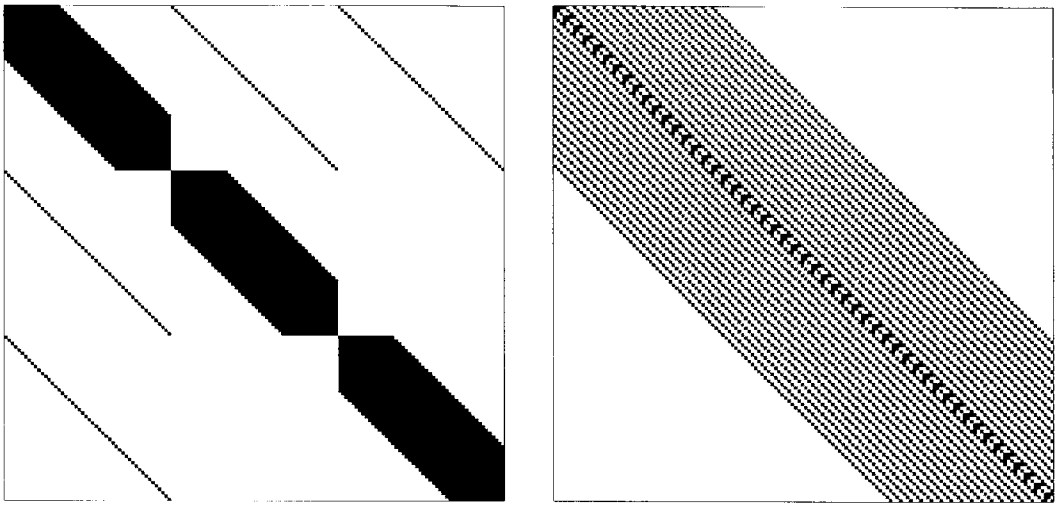
5.2.3 The chemically-significant eigenvalues (CSE) method

The following description of the chemically-significant eigenvalues method is based on the works of Pilling and Robertson [37, 38] and Miller and Klippenstein [39–41]. First, the eigenvalues and eigenvectors of the master equation matrix in Equation (5.21), denoted here as $\mathbb{M}$, are computed:

$$\mathbb{M} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \tag{5.41}$$

The eigenvalue decomposition is usually performed after symmetrizing $\mathbb{M}$, which ensures that the calculated eigenvalues $\mathbf{\Lambda}$ are real. Since this is a physical system, the eigenvalues are nonpositive, with one eigenvalue of exactly zero corresponding to the equilibrium distribution. (If irreversible product channels are included, however, the zero eigenvalue will not be present.)

If the system contains N_{chem} isomers and reactant channels, then we expect the master equation matrix to consist of N_{chem} chemically-significant eigenvalues, including the zero eigenvalue, if present. Under conditions of high pressure and low temperature, these eigenvalues should be distinct and lower in magnitude (*i.e.* less negative) than the other eigenvalues, which correspond to internal energy relaxation.



(a) (b)

Figure 5.2: A typical sparsity pattern for an active-state matrix of a three-isomer network using (a) an isomer-major indexing scheme and (b) an energy grain-major indexing scheme. The latter gives a signficantly more banded matrix, which can be used to accelerate the reservoir state linear solve.

The distinctness criterion is especially important: if the largest magnitude chemical eigenvalue is too close to the smallest magnitude internal energy eigenvalue, the calculated $k(T,P)$ values will be nonsensical. We only consider the case where N_{chem} chemically-significant eigenmodes can be identified.

The separation of the chemical and internal energy eigenvalues means that we can select a time after the initial energy modes have relaxed but before significant chemical transitions have occurred. After this short-time period, the trajectory of the solution will be contained within the manifold defined by the chemically-significant eigenvectors. Therefore we ought to be able to construct the $k(T,P)$ values using only the chemically-significant eigenpairs, independent of the initial condition.

First we construct an $N_{\text{chem}} \times N_{\text{chem}}$ matrix $\mathbf{Z}$ from the chemically-significant eigenvectors by summing all terms in each eigenvector for each isomer:

$$Z_{i\ell} = \sum_{s \in i} Q_{s\ell} \qquad i = 1, 2, \ldots, N_{\text{isom}}, \ell = 1, 2, \ldots, N_{\text{chem}} \tag{5.42}$$

For reactant channels there is only one element in each eigenvector for that channel, so it is transferred to $\mathbf{Z}$ as-is:

$$Z_{n\ell} = Q_{n\ell} \qquad n = 1, 2, \ldots, N_{\text{reac}}, \ell = 1, 2, \ldots, N_{\text{chem}} \tag{5.43}$$

Physically the matrix $\mathbf{Z}$ corresponds to the eigenvector matrix obtained from diagonalizing the matrix of phenomenological rate coefficients $\mathbf{K}$, *i.e.* $\mathbf{K} = \mathbf{Z}\mathbf{\Lambda}'\mathbf{Z}^{-1}$. For systems satisfying the assumed separation of timescales, this allows us to directly express the phenomenological rate coefficients for all reactions between isomers and reactant channels:

$$k_{ij}(T,P) = \sum_{\ell=1}^{N_{\text{chem}}} \lambda_\ell Z_{i\ell} Z_{\ell j}^{-1} \qquad i, j \in 1, 2, \ldots, N_{\text{isom}} + N_{\text{reac}} \tag{5.44}$$

To obtain $k(T,P)$ values for dissociation and bimolecular reactions to a product channel (infinite sink), we will need an additional $N_{\text{prod}} \times N_{\text{chem}}$ matrix $\mathbf{Y}$ with elements

$$Y_{m\ell} = \sum_{i=1}^{N_{\text{isom}}} \mathbf{g}_{mi} \cdot \mathbf{q}_\ell^{(i)} \qquad m = 1, 2, \ldots, N_{\text{prod}} \tag{5.45}$$

Above, $\mathbf{q}_\ell^{(i)}$ contains only the elements of chemically-significant eigenvector $\mathbf{q}_\ell$ involving isomer i. The $k(T,P)$ values for reactions to the product channels are then

$$k_{mj}(T,P) = \sum_{\ell=1}^{N_{\text{chem}}} Y_{m\ell} Z_{\ell j}^{-1} \qquad j = 1, 2, \ldots, N_{\text{isom}} + N_{\text{reac}} \tag{5.46}$$

The time-independent population distribution vector sets $\mathbf{u}_{ij}$ and $\mathbf{v}_{im}$ can be constructed by combining the full and reduced chemically-significant eigenvectors via

$$u_{ij}(E_r) = \left(\mathbf{Q}_{\text{chem}}\mathbf{Z}^{-1}\right)_{(i,r),j} \qquad v_{im}(E_r) = \left(\mathbf{Q}_{\text{chem}}\mathbf{Z}^{-1}\right)_{(i,r),m} \tag{5.47}$$

where $\mathbf{Q}_{\text{chem}}$ is a rectangular matrix containing only the chemically-significant eigenvalues, and the subscripts indicate the row of $\mathbf{Q}_{\text{chem}}\mathbf{Z}^{-1}$ corresponding to isomer i at energy grain r, and the column corresponding to isomer j or reactant channel m. Note that, since the original master equation matrix $\mathbb{M}$ depends on $y_{m\text{B}}$, in the eigenvalue method the $k(T,P)$, $\mathbf{u}_{ij}$, and $\mathbf{v}_{im}$ depend on $y_{m\text{B}}$. However, in most cases only small $y_{m\text{B}}$ are of interest, for which these dependencies are very weak.

The only new assumption utilized by the chemically-significant eigenvalues method is that the chemical and internal-energy relaxation timescales are distinct. When this is the case, as it often is at low and moderate temperatures and high and moderate pressures, the $k(T, P)$ values will accurately reflect the system behavior at chemical timescales. For high temperatures and/or low pressures, one or more chemical timescales become indistinguishable from the internal-energy relaxation timescales. To apply the chemically-significant eigenvalues method in this case, it is necessary to lump the configurations that participate in the fast chemical relaxations together into a single pseudospecies. To our knowledge, no one has yet automated this process; although it certainly would be possible to do so, it is beyond the scope of this work.

Finally, Miller and Klippenstein discuss two ways of extracting the $k(T, P)$ values from the chemically-significant eigenpairs. The above discussion presents the so-called "long-time" method, which has two advantages: it numerically performs better as a chemical eigenvalue approaches the continuum of internal energy eigenvalues, and it makes clear that the $k(T, P)$ values are not dependent on the initial populations. The other method, called the "initial-rate" method, gives identical results at almost all conditions, and may be slightly faster to execute since it avoids the small matrix inversion required by the long-time method.

5.3 A general framework for master equation calculations

In order to evaluate these master equation reduction methods on an equal footing, a general procedure for obtaining phenomenological rate coefficients $k(T, P)$ from a master equation was developed. A flowchart of this general procedure is depicted in Figure 5.3. Within this framework, the master equation reduction methods are entirely interchangeable: they accept the same input parameters and return the same output parameters.

A significant body of input parameters are required in order to assemble the terms of the master equation. Within the context of automatic mechanism generation, where we generally do not have detailed quantum chemistry information available, the following parameters are needed:

- **The species in the network.** For each species the thermodynamic parameters are needed. For those species that are isomers in the network, the chemical structure (connectivity only), molecular weight, and Lennard-Jones parameters are also needed.

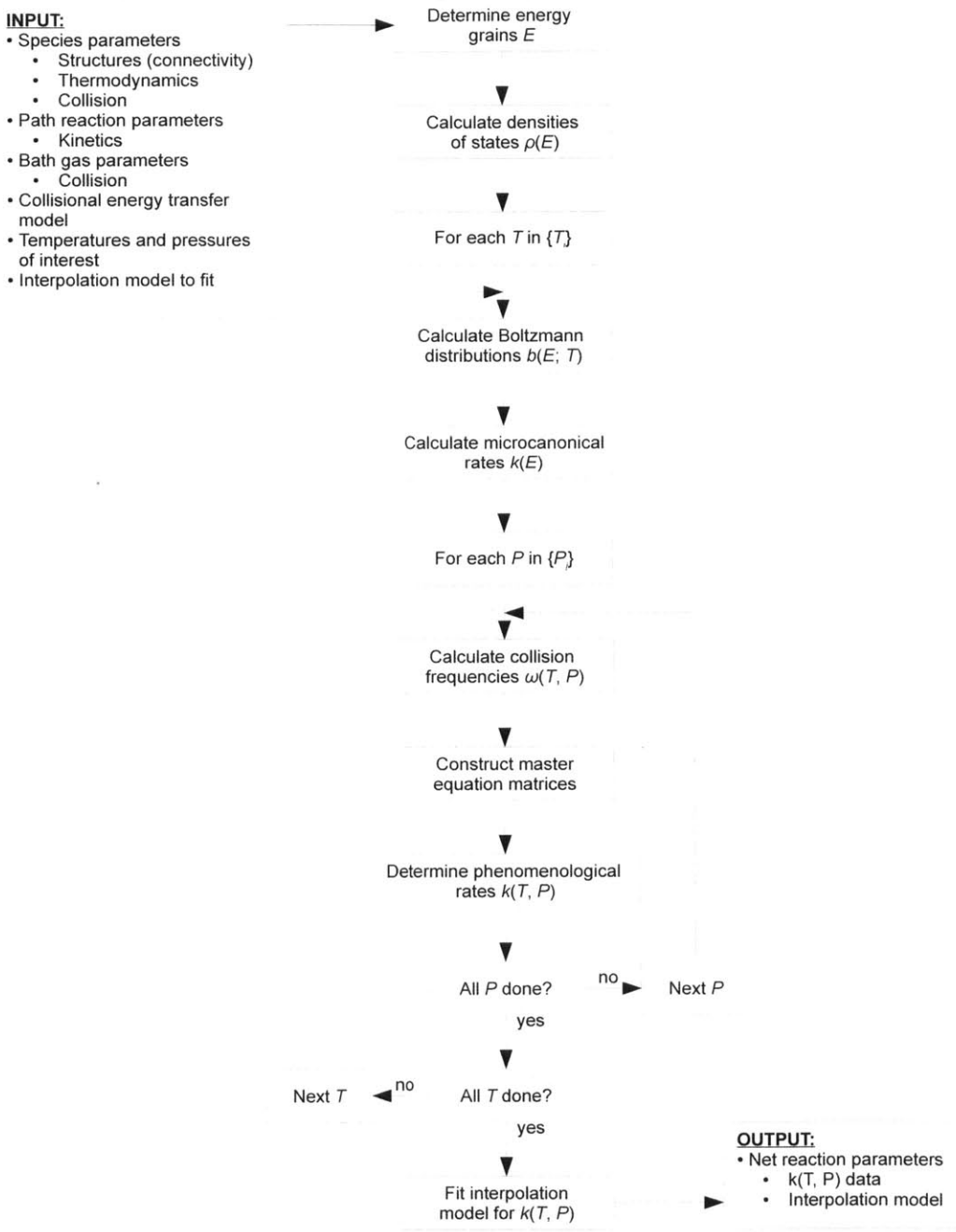


Figure 5.3: A general procedure for estimating pressure-dependent rate coefficients for reaction networks of arbitrary size and complexity. The modular design makes it very easy to swap in and out alternative collision models, reaction models, and master equation reduction methods in order to compare them on an equal basis.

- **The reactions in the network.** For each reaction the high-pressure limit rate coefficient $k_\infty(T)$ is needed.

- **The bath gas.** The molecular weight and the Lennard-Jones parameters are required.

- **The collisional transfer probabilities model.** Throughout this work the single exponential down model is used due to its dependence on a single parameter; more complex models can be used given sufficient information, but often this is not available.

Other inputs include the set of temperatures and pressures at which to estimate the phenomenological rate coefficients, the desired number of energy grains and/or energy grain spacing, and an interpolation model to be fitted to the computed $k(T_n, P_m)$ values to give smooth functions $k(T, P)$.

All of the numerical methods used to estimate the phenomenological rate coefficients require discretization of the energy domain into a vector $\mathbf{E}$. Such a discretization is characterized by a minimum energy $E_{\min}$, maximum energy $E_{\max}$, and either a number of grains N_{grains} or a energy spacing ΔE. The minimum energy $E_{\min}$ is straightforward to determine: it is simply the ground-state energy of the lowest molecular configuration on the potential energy surface. (The ground-state energies on the potential energy surface are estimated as the enthalpy at 0 K using the thermodynamic parameters for each species to extrapolate to 0 K.) The number of grains N_{grains} or energy spacing ΔE depend on the user's desired level of accuracy, and are left as parameters to be input.

However, the choice of $E_{\max}$ is not straightforward. A choice of $E_{\max}$ that is too low may cause the accuracy of the resulting rates to suffer due to neglect of important energies. A choice of $E_{\max}$ that is too high may cause numerical effort to be wasted on energies that do not contribute significantly; worse, the master equation matrix becomes increasingly stiff due the faster reaction rates observed at higher energies. A procedure for determining $E_{\max}$ that ensures that the Boltzmann populations of the energies above $E_{\max}$ are negligible was experimented with, but ultimately we determined that simply placing $E_{\max}$ a set number of $k_{\text{B}}T$ (currently 40) above the highest ground-state energy (usually that of a transition state) was sufficient.

Once a suitable set of energy grains has been selected, the density of states $\rho(E)$ is calculated for all isomers in the network. For multiple temperature and pressure calculations, this step need only be performed once. To obtain the density of states,

we need to estimate the molecular degrees of freedom of each stable species. This work uses the functional-group frequency method developed by Goldsmith to obtain approximate frequencies that both reflect the functional groups of the molecule, account for torsional modes, and are consistent with the estimated thermochemistry [42, 43].

The input includes a set of temperatures and pressures at which to estimate the phenomenological rate coefficients. There are certain steps that are only dependent on the current temperature and not on pressure – such as the calculation of the equilibrium distributions for each isomer – so we place the iteration over temperature as the outer loop and the iteration over pressure as the inner loop.

The collision and reaction models that are required for the master equation were discussed in Sections 5.1.1 and 5.1.2, respectively. This work exclusively uses the Lennard-Jones collision model and the single exponential down model for collisions, while reactions use RRKM theory if detailed transition state is available and the inverse Laplace transform method otherwise.

At this point we have all of the information needed to construct the full master equation matrix, and we are ready to extract the phenomenological rate coefficients. Any of the desired method mentioned in Section 5.2 – modified strong collision (MSC), reservoir state (RS), or chemically-significant eigenvalues (CSE) – can be used interchangeably. Each of these methods returns its computed $k(T, P)$ values, along with the values of $u_{ij}(E)$ and $v_{im}(E)$ that correspond to those values.

The interpolation model is an optional step taken to provide a functional form for the estimated $k(T, P)$ values rather than simply a matrix of values. These are particularly useful when using the results in external applications, e.g. Chemkin. A functional form based on Chebyshev polynomials is common [44], although care must be taken when relatively few temperatures and pressures are used to determine the polynomials in order to avoid unphysical regions. Another option is to fit a modified Arrhenius expression at each pressure and interpolate on a logarithmic pressure scale between these expressions; however, our experience has been that this can give extreme values of the preexponential and temperature exponent, and even then this form does not always fit the computed $k(T_n, P_m)$ accurately in complex reaction networks.

5.4 Acetyl + oxygen: A case study

In order to evaluate the three master equation reduction methods for obtaining phenomenological pressure-dependent rate coefficents $k(T, P)$, we turn to the chemically-

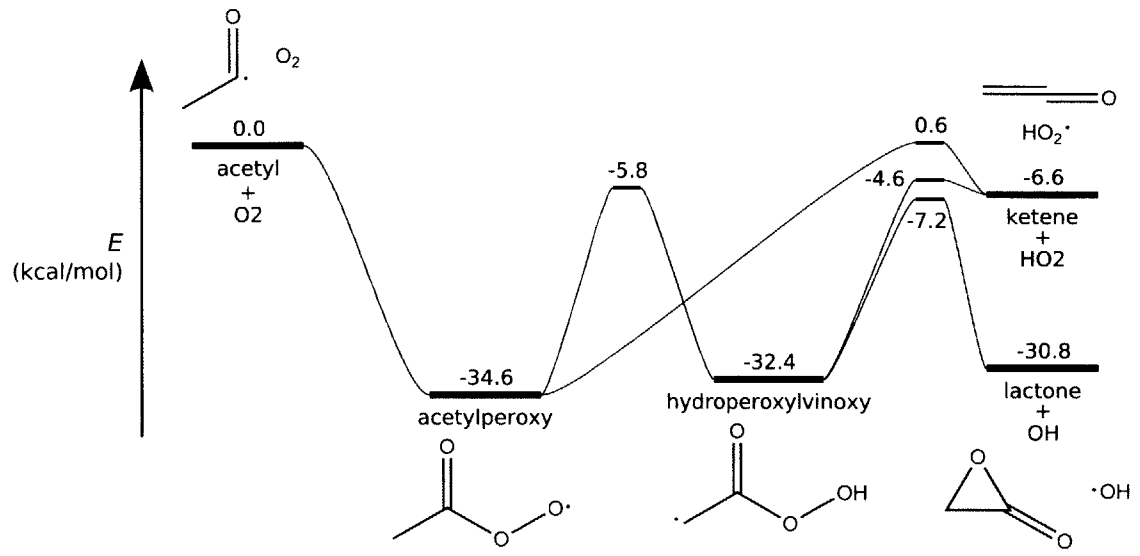


Figure 5.4: The RQCISD(T)/CBS//B3LYP/6-311++G(d,p) potential energy surface for the acetyl oxidation network.

activated oxidation of acetyl radical as a case study. This system is important to atmospheric chemistry as a step in the conversion of acetaldehyde to peroxyacetylnitrate (PAN), the latter of which is a secondary air pollutant [45]. It is also potentially important in the ignition chemistry of ethanol.

Michael, Keil, and Klemm demonstrated that the acetyl radical is a major product of the reaction of acetaldehyde with OH radical, which preferentially abstracts the weakly-bonded aldehydic hydrogen. They also observed pressure-dependent regeneration of OH radical when acetyl radical is reacted with oxygen, ranging from nearly complete regeneration at low pressures to minimal regeneration at high pressures [46]. Later experimental efforts generally agree on the relevant pressure range being about 0.001 to 1 bar at 300 K [47–51]. Recent experiments have confirmed that the α-lactone is the carbon-containing product associated with OH regeneration [52].

On the theoretical side, this system was previously studied by Lee, Chen, and Bozzelli using a three-frequency model to estimate the density of states and the modified strong collision method to estimate the phenomenological rate coefficients. Their calculations significantly underpredicted the regeneration of the OH radical [53]. More recently Maranzana, Barker, and Tonachini studied the system using quantum chemistry calculated modes as input to the density of states and the MultiWell stochastic master equation solver. They allowed the collisional energy transfer param-

Table 5.1: Calculated high-pressure limit rate coefficients for acetyl + O_2.[a]

Reaction			A	n	E_a
acetyl + O_2 [b]	$\longrightarrow$	acetylperoxy	4.4×10^{-12}	0.0	0.0
acetylperoxy	$\longrightarrow$	hydroperoxylvinoxy	3.7×10^{-1}	3.77	21.7
acetylperoxy	$\longrightarrow$	ketene + HO_2	2.7×10^{8}	1.55	33.6
hydroperoxylvinoxy	$\longrightarrow$	ketene + HO_2	1.6×10^{14}	-0.25	28.4
hydroperoxylvinoxy	$\longrightarrow$	lactone + OH	3.0×10^{16}	-0.72	26.7

[a] The units for A are s^{-1} for unimolecular reactions and in cm^3 $molecule^{-1}$ s^{-1} for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$. Computed from RQCISD(T) TST calculations with Eckart tunneling; see text for details.
[b] Taken from reference [53].

eters to vary so as to match the experimental data, and thus were able to match the regeneration of OH [54].

The potential energy surface for this reaction was computed originally by Lee, Chen, and Bozzelli [53]. Unusually, they predicted that the dominant bimolecular product channel was a biradical + OH. Hou *et al.* highlight the unlikeliness of this product after recomputing the potential energy surface using the G3MP2 compound method [49]. Based on their results, we do not consider the biradical as a stable configuration, and do not include it in our potential energy surface or subsequent calculations.

5.4.1 Computational methods

We performed our own RQCISD(T)/CBS//B3LYP/6-311++G(d,p) calculations to generate the PES in Figure 5.4 using the methodology of Goldsmith *et al.* [55] All DFT and CBS-QB3 calculations were done using Gaussian 03 [56]. All MP2 and QCISD(T) calculations were done using MOLPRO [57]. Our updated values are consistent with those recently published by Villano *et al.* [58]

The high-pressure limit for each reaction rate coefficient except acetyl + O_2 was computed using conventional transition state theory with Eckart tunneling and one-dimensional hindered rotors, as implemented in CanTherm. The rate of the acetyl + O_2 association reaction was taken from Hou *et al.*, as the complexity of computing a barrierless association reaction is beyond the scope of this work. These results are provided in Table 5.1.

The phenomenological rate coefficients for the acetyl + oxygen system were calculated by the modified strong collision, reservoir state, and chemically-significant eigenvalue methods using the quantum chemistry parameters from our

RQCISD(T)/CBS//B3LYP/6-311++G(d,p) calculations. RRKM theory was used to compute the microcanonical rate coefficients $k(E)$, with all rotational modes treated as active. The Lennard-Jones and single exponential down collision models were used, with a parameter of $\langle \Delta E_{\text{down}} \rangle = 200(T/300)^{0.85}\,\text{cm}^{-1}$ [59]. Most importantly, identical values were input as parameters to each of the master equation reduction methods, so that the comparison is on an entirely even footing.

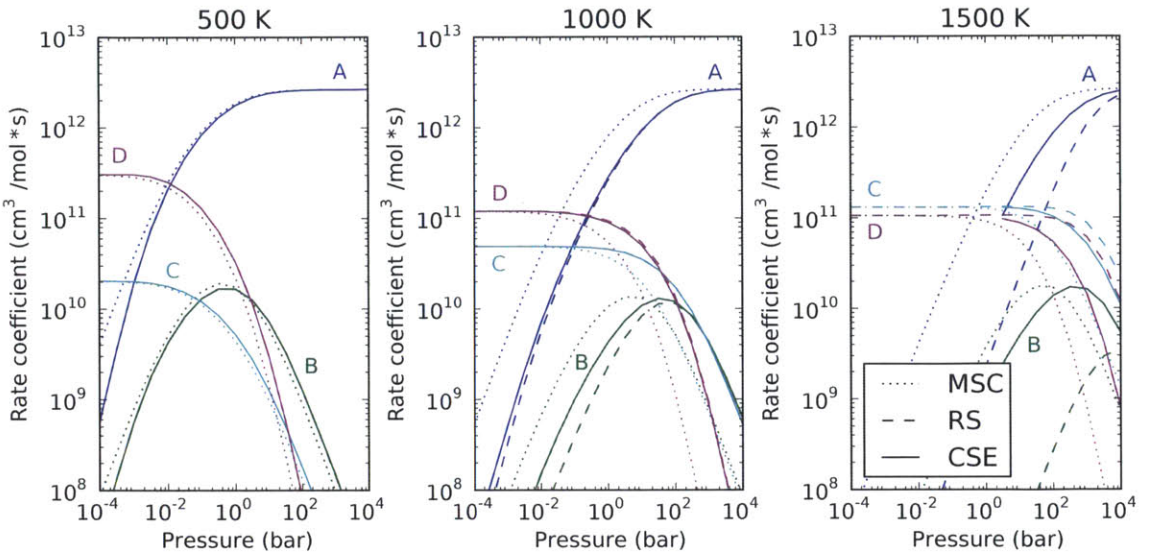
5.4.2 Comparison of the master equation reduction methods

There are several means by which the three methods of extracting phenomenological rate coefficients from the master equation may be evaluated. Most pertinent is direct comparison of the $k(T,P)$ values, as these are what we wish to use in our kinetic models. The downside of this approach is that there is no "true" set of $k(T,P)$ values available to use as a basis for comparison, since the solution of the master equation does not exactly match the conventional reduced-order $k(T,P)$ model. To enable comparison with the true solution, we must turn to comparison of concentration profiles, which we can obtain by direct integration of the master equation. However, we can go one step further by using Equation (5.27) to reconstruct the approximate population distributions $p_i(E,t)$ for each isomer, and compare those with one another and with the full master equation solution. In this section we apply these methods of comparison to the acetyl + O_2 reaction network.

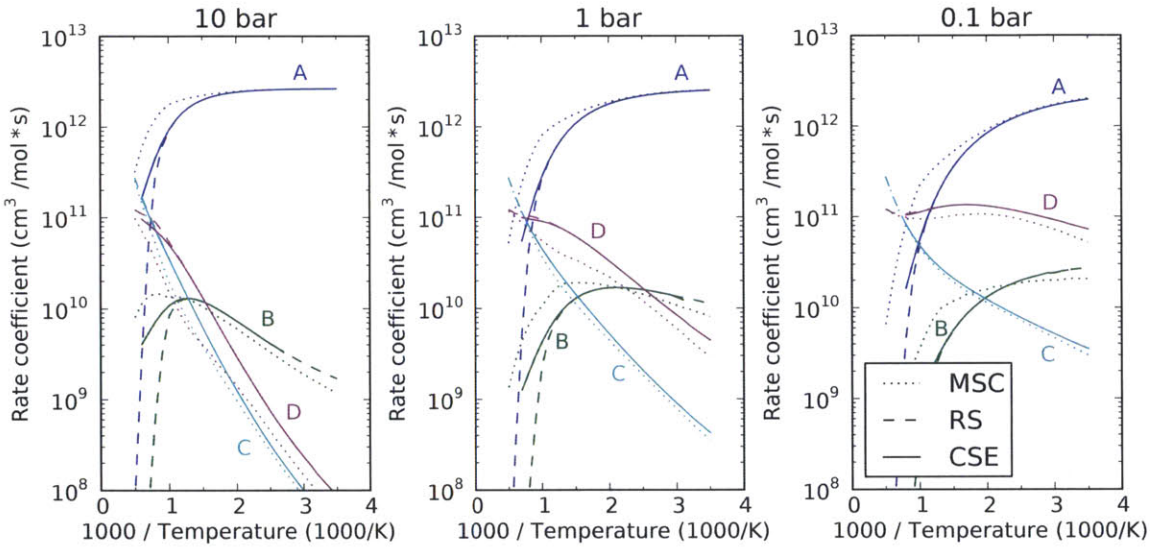
5.4.2.1 Rate coefficients

Figure 5.5 shows the $k(T,P)$ values for $CH_3CO + O_2 \longrightarrow$ products at a variety of temperatures and pressures. As expected, at high pressures and low temperatures, the dominant pathway is collisional stabilization to the acetylperoxy isomer. At low pressures, the lactone + OH product channel is dominant at low and moderate temperatures, while the ketene + HO_2 channel is dominant at high temperature. All of this is consistent with our physical understanding of the system.

The MSC, RS, and CSE methods give very similar values of the rate coefficients at high and moderate pressures and at low and moderate temperatures. However, there are several regions of disagreement. The effects of temperature and pressure on the agreement varies with the type of product. For reaction to the isomers acetylperoxy and hydroperoxylvinoxy, there is good agreement between the methods at low temperature and high pressure; increasing temperature and/or decreasing pressure causes the methods to diverge somewhat. For reactions to the bimolecular product channels, the reverse effect is observed: all methods disagree somewhat at low temperature and high pressure, while increasing temperature and/or decreasing pressure causes the



(a)



(b)

Figure 5.5: Comparison of rate coefficients versus (a) pressure and (b) temperature for $CH_3CO + O_2 \longrightarrow$ products estimated using the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods. In the plots, A = acetylperoxy, B = hydroperoxylvinoxy, C = ketene + HO_2, and D = lactone + OH. The error between the three methods is generally within an order of magnitude. However, over the range where the CSE method can be applied, the RS method gives rate constants that better match the CSE values than the MSC method. The CSE method runs into numerical difficulties at high temperature and low pressure, and at low temperature; see text for details.

methods to converge. Note that the methods always seem to show consensus on the rate coefficients for the major pathways, and only disagree on the minor channels. However, depending on what aspect of the system is of physical interest, having a good value for the minor rate constants can be important.

The CSE method introduces no additional approximations as long as the condition of discrete chemical timescales is met; therefore, we can treat its $k(T,P)$ values as the closest to the true values. At low and moderate temperatures, the RS method $k(T,P)$ values are closer to the CSE values than are the MSC values. However, at high temperature the result is the opposite: the MSC $k(T,P)$ values are closer to the CSE values. Since the MSC and RS methods both utilize the pseudo-steady state approximation, it must be the RS assumption of thermalized low-energy levels that is failing at high temperature. This makes physical sense, as at high temperature a significant fraction of the Boltzmann distribution for the isomer species will reside above the reaction barrier; this population does not need to undergo collisional excitation in order for reaction to proceed, and therefore will not be described by the same physics as the $k(T,P)$ values predicted by the RS method.

5.4.2.2 Concentration profiles

The most direct method of evaluating the performance of the three methods is to use the $k(T,P)$ values to generate predicted concentration profiles for each approximate method and compare them to those produced by solving the full time-dependent master equation with a stiff ODE solver. The objective here is to evaluate how well the $x_i(t)$ and $y_{m\mathrm{A}}(t)$ reflect the total populations of isomer i and reactants/products m. For these calculations, the initial concentrations were set to a small amount of acetyl in an atmospheric ratio of oxygen and nitrogen such that the total was consistent with the temperature and pressure chosen. The lactone + OH and ketene + HO_2 product channels were treated as irreversible sinks. Both low-temperature (500 K) and high-temperature (1500 K) conditions were studied; in each case the pressure was chosen to be as low as possible without obtaining unphysical results with the CSE calculation.

Figure 5.6 compares the concentration profiles from each method with those from the full solution at the same sets of conditions. Under all conditions shown, all three methods exhibit a period at short times before their profiles match those from the full master equation solution. This time period reflects the time scale of collisions, and is generally of the order of $10\omega^{-1}$, where ω is the mean collision frequency at the conditions being studied. Once a suitable number of collisions have occurred, all of the methods predict the total concentrations of reactants, products, and isomers very well.

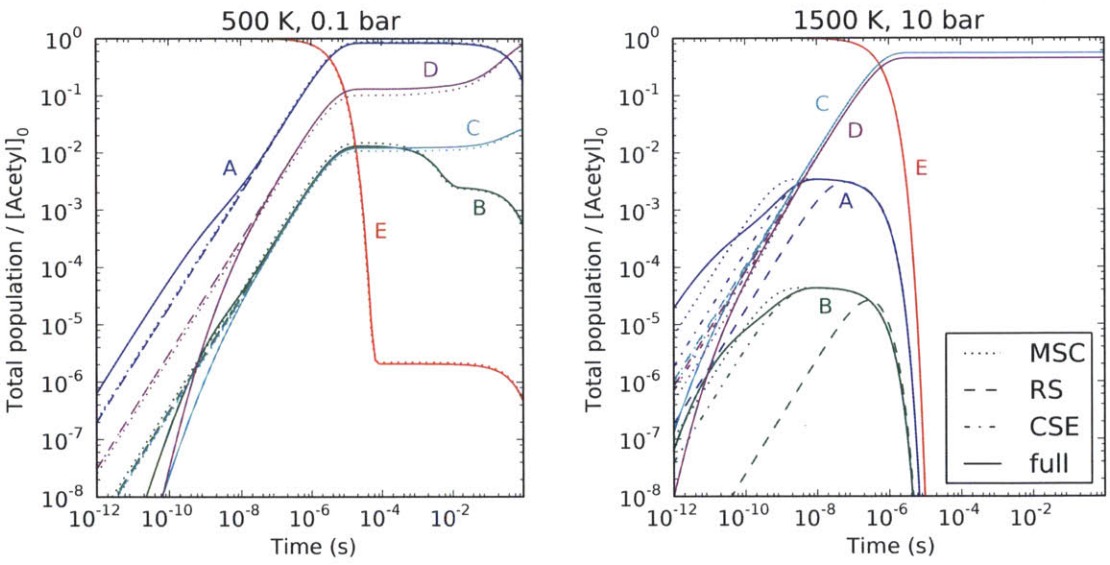


Figure 5.6: Concentration profiles at low-temperature and high-temperature conditions, comparing the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods with the full master equation solution, for the acetyl + oxygen system. In the plots, A = acetylperoxy, B = hydroperoxylvinoxy, C = ketene + HO_2, D = lactone + OH, and E = acetyl + O_2. The RS and CSE methods both perform very well at low temperatures. Although all three methods disagree with the full solution at very short times for the higher temperature case, the RS method takes significantly longer to approach the full solution than the MSC and CSE methods.

For the MSC and RS methods, the poor performance at short times comes from use of the pseudo-steady state approximation, while for the chemically-significant eigenvalues method, this comes from treating the internal energy eigenmodes as completely relaxed.

The first set of conditions, 500 K and 0.1 bar, represent a low temperature, moderately low pressure system. Under these conditions acetylperoxy is the primary product, although the pressure is low enough that a significant amount of lactone is also produced. All of the methods approximate the total concentrations of reactants, products, and isomers very well after the short-time period of about 10^{-8} s. The MSC method does deviate from the others slightly, but remains well within a factor two of the true profiles after the short-time period.

Also interesting is the effect of high temperature, demonstrated using conditions of 1500 K and 10 bar. This time both ketene and lactone are the major products, with relatively little of each isomer produced along the way. All of the reactant and

product channels are well-approximated by all of the approximate methods, but show some disagreement in how quickly they approach the full solution for the isomer concentrations. Of the three methods, the CSE method approaches the full solution the fastest, followed quickly by the MSC method, while the RS method takes significantly longer. Note that the RS method performance is a bit better for acetylperoxy, which is produced in a larger quantity, than for hydroperoxylvinoxy.

5.4.2.3 Population distributions

The concentration profile comparison is valuable because it allows for comparison of the three methods with the "true" solution obtained from the full master equation. However, the "true" solution also provides the full population distributions $p_i(E,t)$ for each isomer. We can also reconstruct the approximate $p_i(E,t)$ from $x_i(t)$ and $y_{m\mathrm{A}}(t)y_{m\mathrm{B}}$ using Equation (5.27). Thus, we can also examine how the approximate population distributions evolve in time to see each method's strengths and weaknesses in more detail.

Figure 5.7 shows the approximate population distributions $p_i(E,t)$ versus energy for the acetylperoxy isomer from each method and the full solution at several times. Under both sets of conditions, the high-energy grains of acetylperoxy are populated before collisional stabilization begins to be important. The collisions tend to cause only small changes in energy, so even after ten collision times there have not been enough collisions to significantly populate the low-energy grains. The result is that the low-energy grains are not yet Boltzmann-distributed, and so none of the approximate methods match the true populations at low energies. As time progresses to 100 and 1000 collsion times, the low-energy grains become increasingly thermalized, and so the approximate methods much more accurately reflect the full solution.

Of the three methods, the CSE method is the first to accurately match the entirety of the "true" population distribution. In the low temperature case, the RS method follows very quickly, which suggests that the RS approximations are valid at these conditions. The MSC method distributions are unable to capture the non-Boltzmann behavior of the energy grains between about -20 to 0 kcal/mol; however, as most of the population is in the lowest-energy grains, which are thermalized, the MSC method is still able to accurately obtain close to the correct total population, as seen earlier.

In the high-temperature case, the MSC method more quickly approaches the true population distribution than the RS method does. However, the RS method continues to perform surprisingly well considering its difficulty at predicting the total concentration shown earlier. This occurs because, in reconstructing $p_i(E,t)$, there is an implicit

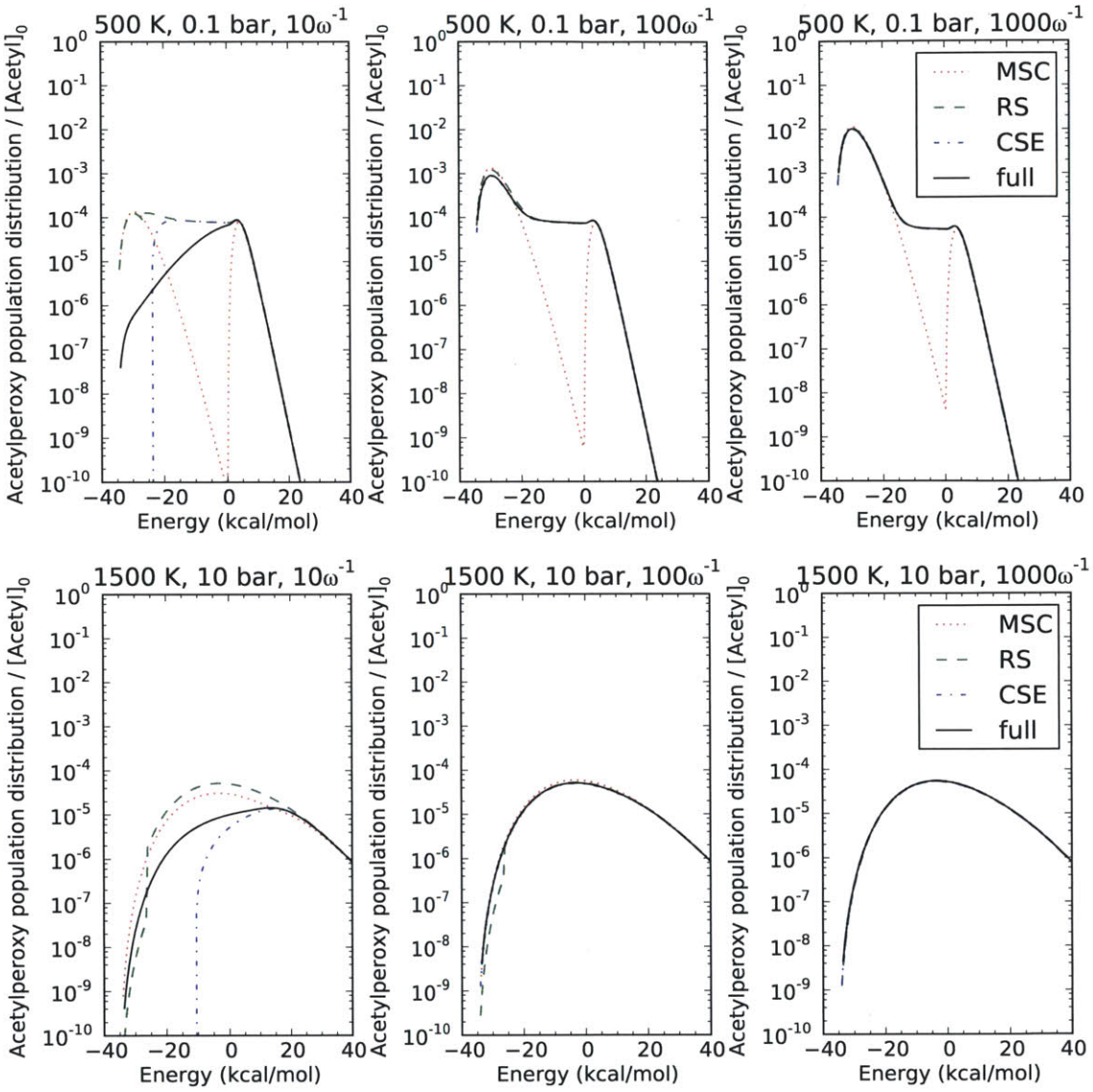


Figure 5.7: Population distributions for acetylperoxy versus energy after 10, 100, and 1000 collision times, comparing the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods with the full master equation solution. At the conditions used above, ω^{-1} is equal to 3.4×10^{-9} and 6.9×10^{-11} s, respectively. The short-lived, high-energy reactive states immediately reach a steady state which is accurately replicated by all three methods. After 100-1000 collision times, each method provides a reasonable approximation of the population distribution of the full solution. The CSE method is the most accurate, while the MSC method underpredicts the population of the adduct states with energies just below the reactant energy (-10 to 0 kcal/mol) at lower temperatures. The zero of energy is the ground-state of acetyl + O_2.

rescaling of the populations that occurs. This implies that we could apply this rescaling directly to obtain better values of the total concentrations. However, the rescaling would require allowing for the possibility of either slightly negative populations or negative $k(T, P)$ values, both of which are nonphysical.

5.4.2.4 Eigenvalue analysis

In Figure 5.5b the $k(T, P)$ curves given by the chemically-significant eigenvalues method appear to stop as temperature decreases at all pressures. To examine why, let us look in more detail at the eigenvalues themselves. Treating the two bimolecular product channels as irreversible sinks, there are three chemically-significant eigenvalues, and no zero eigenvalue. The computed eigenvalues are plotted in Figure 5.8a and 5.8b as a function of temperature and pressure, respectively. The figures show the four lowest-magnitude eigenvalues, with λ_4 representing the lowest internal-energy eigenmode.

In Figure 5.8b the magnitudes of the chemical and internal-energy eigenvalues become increasingly disparate as temperature decreases, indicating that the master equation matrix is becoming increasingly stiff. Once the temperature becomes too low, the large separation in time scales makes the low-magnitude eigenvalues numerically difficult to calculate when using double-precision arithmetic – as was used throughout this work – resulting in one or more unphysical positive eigenvalues [60, 61]. A quadruple-precision eigenvalue algorithm would allow for successful determination of the eigenvalues at low temperature, but we believe this to be unnecessary due to the success of the reservoir state method at low temperatures. The reservoir state method is not immune to numerical issues due to stiffness, but this is generally only problematic for much larger reaction networks than acetyl + O_2.

As the temperature increases or the pressure decreases, we can clearly see one or more of the chemically significant eigenvalues (λ_1 to λ_3) approach the slowest internal energy-relaxation eigenvalue (λ_4). When λ_3 gets too close to λ_4, the discrete separation of chemical and internal energy timescales breaks down, and applying the CSE method often leads to unphysical $k(T, P)$ values. In principle, we can adapt by lumping two or more of the isomers or reactant channels together; however, this lumping technique has not yet been automated, although we believe that it would be possible to do so. Under these conditions the assumptions of the RS method also break down, leaving the MSC method as the best choice.

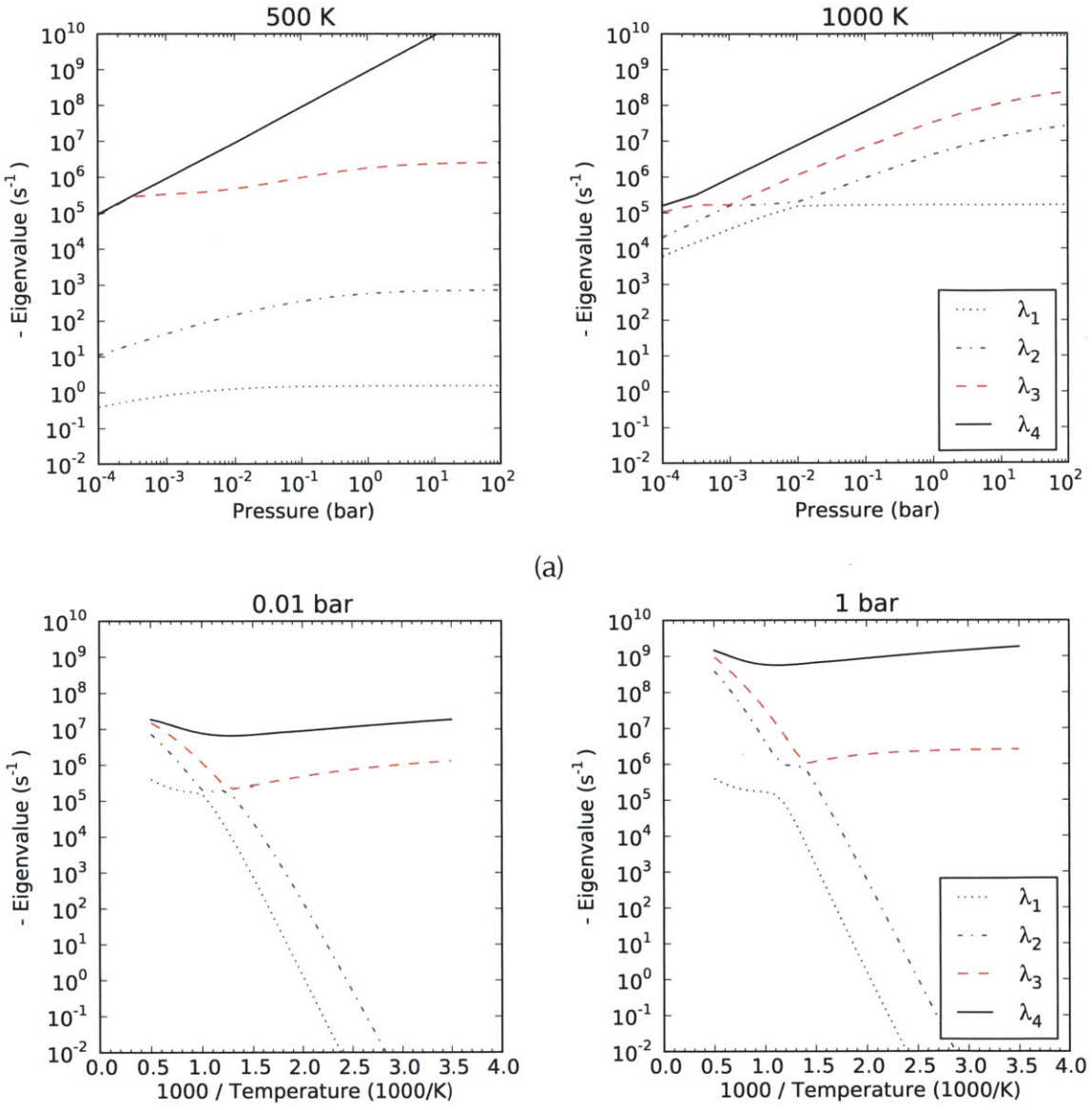


Figure 5.8: Eigenvalues of the master equation matrix at (a) several pressures at 500 and 1000 K and (b) several temperatures at 0.01 and 1 bar. As the pressure is decreased or the temperature is increased, one or more of the chemical eigenvalues (λ_1 to λ_3) become indistinguishable from the internal energy eigenvalues (λ_4).

5.4.3 Perturbations to the acetyl + oxygen potential energy surface

Although there are a few regions in Figure 5.5 where the three approximate methods for estimating $k(T,P)$ values disagree, the magnitude of the disagreement is typically within an order of magnitude. In order to exaggerate the circumstances under which one or more methods differ significantly, we introduce one or more artificial perturbations to the potential energy surface. These perturbations involve raising the ground-state energies or lowering the transition-state barrier heights for one or more configurations on the surface, in order to vary the degree of chemical activation and the relative rates of isomerization. Several such perturbations will be discussed in this section. For all perturbations we have provided plots of $k(T,P)$ versus pressure for rate coefficients corresponding to acetyl + O_2 as the reactants.

5.4.3.1 Acetyl + 20 kcal/mol

We begin with the case where the acetyl + O_2 ground-state energy is artifically increased by 20 kcal/mol. This perturbation results in a system that is much more chemically activated than before by creating a significant range of energies below the ground state of acetyl + O_2 but many $\langle \Delta E_{\text{down}} \rangle$ above all other reaction barriers in the system. This means that an excited acetylperoxy adduct can survive several collisional stabilization events and remain reactive. This means that a very high pressure is required to trap the adducts, as seen in Figure 5.9a. However, only the RS and CSE methods capture this effect, while the MSC method does not.

The MSC method replaces the collision model from the full master equation with a greatly simplified single-step activation/deactivation process. This is not a great model of this perturbed system, as the excited adduct can be stabilized by many kcal/mol without losing reactivity. The collision efficiency $\beta(T)$ is intended as a correction to improve the single-step approximation; this perturbation demonstrates a situation where $\beta(T)$ does not provide an adequate correction. Concentration profile comparisons at both low and high temperature conditions clearly show the MSC method provides a significantly less accurate prediction for the isomer profiles, while the other methods perform noticeably better, especially at low temperatures.

5.4.3.2 Acetylperoxy + 20 kcal/mol

Next we perturb the system by artificially raising the ground-state energy of the initial adduct, acetylperoxy, by 20 kcal/mol. This creates a much shallower well on the potential energy surface with a much smaller density of states, and so faster $k(E)$. This has the effect of pushing more of the Boltzmann distribution of acetylperoxy

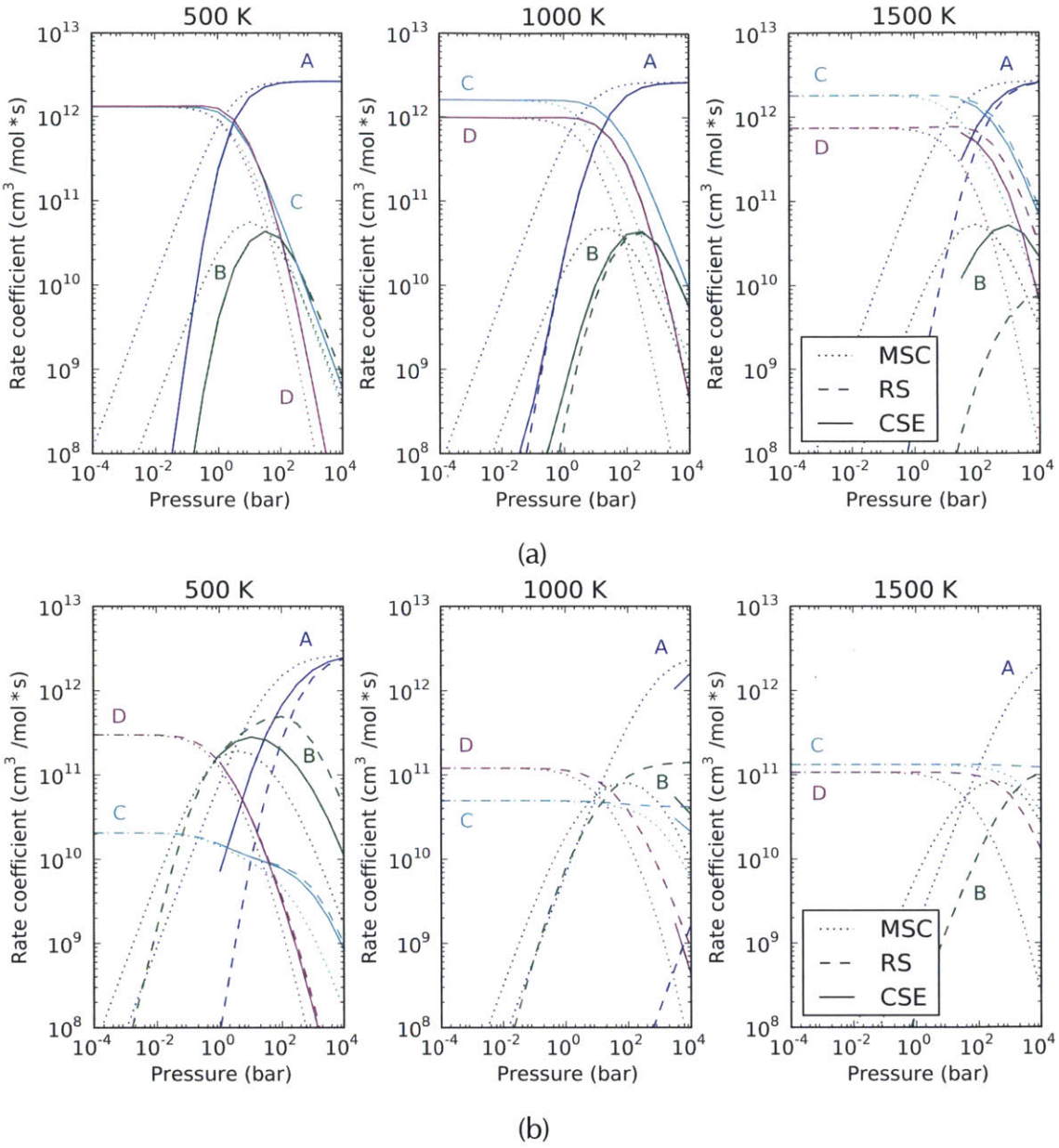


Figure 5.9: Comparison of rate coefficients versus pressure for $CH_3CO + O_2 \longrightarrow$ products estimated using the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods. In the plots, A = acetylperoxy, B = hydroperoxylvinoxy, C = ketene + HO_2, and D = lactone + OH. (a) Increasing the ground-state energy of acetyl + oxygen by 20 kcal/mol causes the MSC method to disagree more with the RS and CSE methods. (b) Increasing the ground-state energy of acetylperoxy by 20 kcal/mol leads to significant discrepancies between the methods, particularly at high temperature and high pressure. The CSE method fails to distinguish three chemical eigenvalues at many of these conditions.

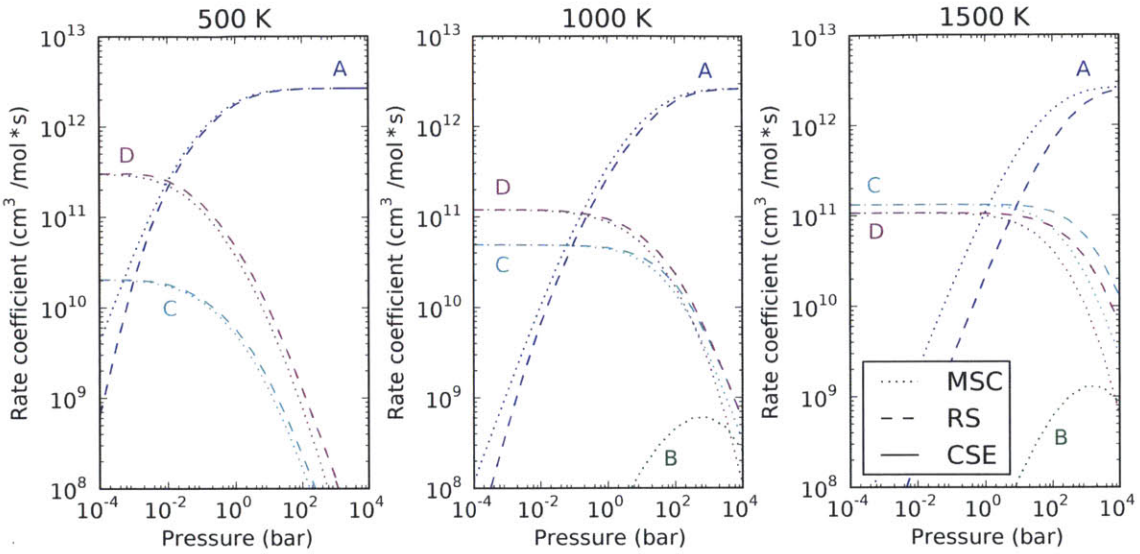
into the higher-energy reactive grains, and also encourages chemically-activated well skipping from acetyl + O_2 to hydroperoxylvinoxy. In Figure 5.9b we see that the $k(T, P)$ values from acetyl + O_2 to hydroperoxylvinoxy have increased as a result of this perturbation. At 500 K, this increase is so significant that hydroperoxylvinoxy is predicted to be the dominant product over the range of 1 to 100 bar. All three methods predict this behavior, although a chemical eigenvalue (equilibration of the isomers acetylperoxy and hydroperoxylvinoxy) is often blended into the collisional energy relaxation eigenspectrum, so the CSE method does not succeed with three chemical eigenvalues, and lumping would be required to obtain $k(T, P)$ values with the CSE method.

5.4.3.3 Hydroperoxylvinoxy + 20 kcal/mol

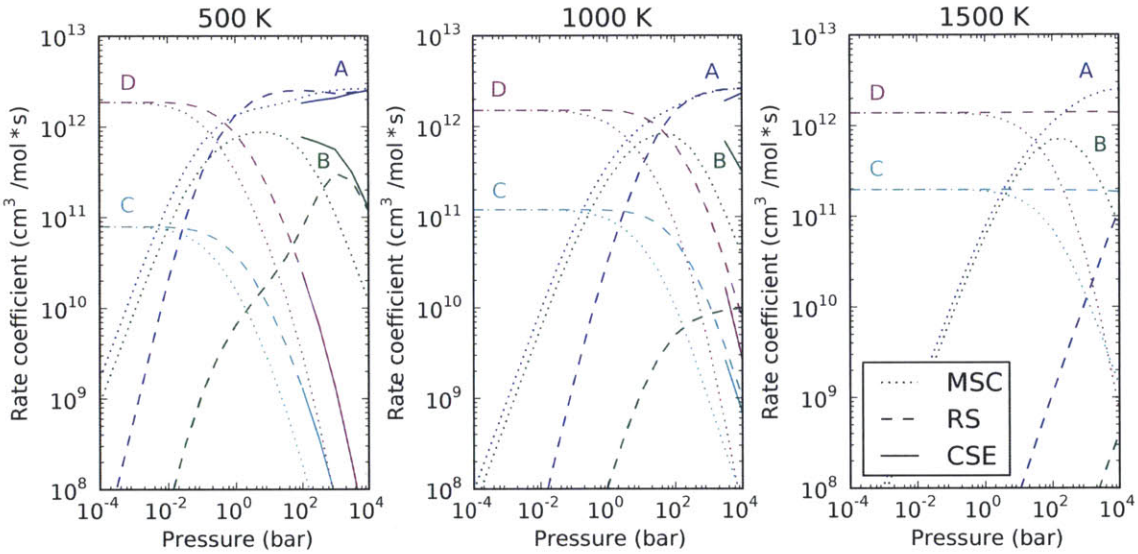
A related perturbation is to artificially raise the ground-state energy of the other isomer, hydroperoxylvinoxy, by 20 kcal/mol, again creating a much shallower well on the potential energy surface. Similar to the last perturbation, this reduces the density of states $\rho(E)$, increases $k(E)$, and causes more Boltzmann population of hydroperoxylvinoxy to exist in the reactive grains. Overall this encourages well skipping from acetylperoxy to the product channels, and reduces the lifetime of the hydroperoxylvinoxy. The result is the $k(T, P)$ values for reactions that produce hydroperoxylvinoxy are depressed, as seen in Figure 5.10a. The only major disagreement in $k(T, P)$ values between the MSC and RS methods are for those involving hydroperoxylvinoxy as a reactant or product. (Again, the CSE method for $N_{\text{chem}} = 3$ has no results because a chemical eigenvalue has merged into the internal energy relaxation eigenspectrum.) Concentration profile comparisons to full solutions of the master equation again show that the RS method performs slightly better under most conditions for all configurations except the very unstable species hydroperoxylvinoxy, for which the MSC method is slightly more accurate.

5.4.3.4 Isomerization - 20 kcal/mol

As a final perturbation, we lower the reaction barrier for the isomerization reaction by 20 kcal/mol. The results of this perturbation are shown in Figure 5.10b. The low barrier causes the isomerization process to be very fast, to the point where a chemical eigenvalue blends into the internal energy relaxation eigenvalues at nearly all conditions shown, causing the CSE method to have no result with $N_{\text{chem}} = 3$. The MSC and RS methods are competitive, although at high temperatures the former performs noticeably better than the latter in predicting both isomer profiles.



(a)



(b)

Figure 5.10: Comparison of rate coefficients versus pressure for $CH_3CO + O_2 \longrightarrow$ products estimated using the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods. In the plots, A = acetylperoxy, B = hydroperoxylvinoxy, C = ketene + HO_2, and D = lactone + OH. (a) Increasing the ground-state energy of hydroperoxylvinoxy by 20 kcal/mol encourages well skipping between acetylperoxy and the two product channels. The CSE method is unable to determine three chemical eigenvalues at all conditions due to the very fast equilibrium between the isomers. (b) Decreasing the isomerization barrier height by 20 kcal/mol creates a situation of a fast equilibrium between the two isomers, causing the CSE method to be unable to resolve three chemical eigenvalues. The RS method is more accurate at low temperature, while the MSC method is more accurate at high temperature.

5.4.3.5 *Discussion*

A few common themes emerge from the above analysis. First, all three methods are fairly accurate for the major channels across a wide range of conditions. Second, the MSC method performs consistently in all systems, nearly always giving reasonable results, although sometimes noticeably less accurate than the other methods. The MSC method is also numerically inexpensive and robust. For initial explorations of a system, the MSC method is a reasonable choice.

Third, there is a wide range of conditions, especially at low and moderate temperatures, where the RS method provides better estimates for the phenomenological rate coefficients than the MSC method. Under these conditions, the equilibrium population distributions for each isomer are narrow enough that only a small fraction of the Boltzmann population exists in reactive grains. When the ground-state energies of either isomer were raised, the RS method provided less accurate estimates for the $k(T,P)$ values involving the unstable isomer, under conditions where the unstable isomer is a minor channel. When the isomerization barrier was lowered, RS method estimates involving both isomers became less accurate. Thus, we conclude that the RS method is most accurate when the isomer wells are "deep", *i.e.* the fraction of the Boltzmann distribution that exists in reactive grains is insignificant.

Fourth, at conditions where the CSE method was able to resolve three chemical eigenvalues, it was consistently the most accurate in its $k(T,P)$ estimates. This is the best method to use for high-accuracy $k(T,P)$ calculations. However, there are a wide range of conditions where this method was unsuccessful due to one or more chemical eigenvalues being merged with the internal-energy relaxation eigenvalues, or because the numerical eigensolver returned an unphysical positive eigenvalue. The eigenvalue-merging issue, *i.e.* the breakdown of the assumed separation of collision and reaction timescales, has been discussed by Miller and Klippenstein; they suggest that under these conditions, the concept of a phenomenological rate coefficient breaks down [41]. Unfortunately, both issues arise frequently even for networks of modest size.

5.4.4 Computational cost considerations

Our results show that the RS and CSE methods are somewhat more accurate than the MSC method in predicting rate coefficients for minor channels, except at high temperatures and/or very low pressures. The CSE method has a stronger theoretical basis, but the RS method is often nearly as accurate.

An estimate of the relative execution time of the methods, shown in Table 5.2,

Table 5.2: CPU time to generate Figure 5.5

Method	Time (s)
Modified strong collision	4.8 s
Reservoir state	9.8 s
Chemically-significant eigenvalues	310 s

comes from considering the time needed by each method to generate the data in Figure 5.5. The RS method has about 2-3x the computational cost of the MSC method, while the cost of the CSE method is nearly 100x that of the MSC method. These factors will only get larger as more isomer wells are considered, or if finer energy discretization is employed. The high computational expense and difficulties inherent in the CSE method at low and high temperatures make it less useful for automatic estimation of large numbers of $k(T, P)$ parameters.

Given the approximate nature of several of the inputs to the master equation – especially the collisional transfer probabilities model and parameters – there is at this point some question as to whether the RS method provides enough of an improvement in accuracy compared with the MSC method to be worth the extra computational effort. This is especially true when we apply these methods using automatically-generated estimated parameters, which themselves can have a significant error. Furthermore, the RS method, while generally faster and more robust than the CSE method, is not immune to numerical difficulties. The RS method involves a linear solve for the pseudo-steady active-state grain populations via Gaussian elimination. For very large networks this elimination can be unsuccessful due to the large matrices involved. We generally observe failures of the LAPACK Gaussian elimination routines when there are more than around thirty unimolecular isomers in the reaction network (typically we allocate about 250 active-state grains per isomer). For these reasons, our automatic mechanism generation code, RMG, provides the user the option of either the MSC or RS methods, and automatically falls back to the MSC method if the reservoir state method is unable to complete the Gaussian elimination step.

5.5 Conclusion

This chapter presented a general framework for estimating the phenomenological rate coefficients $k(T, P)$ for unimolecular reaction networks of arbitrary size and complexity. For detailed investigations of individual reaction networks, this framework

can be used with detailed quantum chemistry information. For automatic mechanism generation, this framework can also be used with parameters estimated form only macroscopic quantities.

Within this framework, methods for estimating the $k(T, P)$ values by reducing the full master equation model have been evaluated for both scenarios. The reservoir state method has been shown to be superior to the modified strong collision method method except at high temperatures, giving rate coefficients, concentration profiles, and population distributions that more closely follow those of the full master equation solution. However, the assumption of thermalized low-energy grains used by the reservoir state method breaks down completely at high temperatures; under these conditions the modified strong collision method is more accurate. The chemically-significant eigenvalues method is noticeably more computationally expensive than the other methods, but also gives the most accurate estimates of the $k(T, P)$ values. However, numerical stiffness issues at low temperature and the blending of collisional and reactive timescales limit the robustness of this method.

The recommended use cases for each method are as follows. The modified strong collison method has the best speed and robustness, and is competitively accurate with the other methods over all conditions; for this reason, it is currently the method of choice for automatic mechanism generation. However, detailed investigations of individual reaction networks – such as the most sensitive reactions from the automatically-generated mechanism – should be refined using either the reservoir state or chemically-significant eigenvalues methods. The reservoir state method is safe to use except in the limits of high temperature, very low pressure, or shallow isomer well depths. Nonetheless, the chemically-significant eigenvalues method remains the most accurate over all conditions which the collision and reaction timescales are distinct, and where the system is not too numerically stiff.

The result of this work is a module for estimating phenomenological rates that are utilized by and distributed with the open-source automatic mechanism generation code RMG [62]. RMG offers users a choice between the modified strong collision and reservoir state methods. A Fortran 90 implementation of this module, named FAME (**F**ast **A**pproximate **M**aster **E**quation) has been included in all releases of RMG-Java since 3.0 in March 2009. A separate Python implementation has been integrated into the CanTherm tool, both of which are part of RMG-Py.

5.6 References

[1] R. G. Gilbert and S. C. Smith. *Theory of Unimolecular and Recombination Reactions*. First edition. Oxford, England: Blackwell Scientific, 1990.

[2] T. Baer and W. L. Hase. *Unimolecular Reaction Dynamics*. Oxford University Press, 1996.

[3] K. A. Holbrook, M. J. Pilling, and S. H. Robertson. *Unimolecular Reactions*. Second edition. John Wiley and Sons, 1996.

[4] J. A. Miller and S. J. Klippenstein. "Solution of Some One- and Two-Dimensional Master Equation Models for Thermal Dissociation: The Dissociation of Methane in the Low-Pressure Limit." *J. Phys. Chem. A* 106, 2002, pp. 4904–4913. DOI: 10.1021/jp0144698.

[5] W. Forst. *Unimolecular Reactions: A Concise Introduction*. Cambridge University Press, 2003.

[6] M. J. Pilling and S. H. Robertson. "Master Equation Models for Chemical Reactions of Importance in Combustion." *Annu. Rev. Phys. Chem.* 54, 2003, pp. 245–275. DOI: 10.1146/annurev.physchem.54.011002.103822.

[7] F. A. Lindemann. "Discussion on "The Radiation Theory of Chemical Action"." *Trans. Faraday Soc.* 17, 1922, pp. 598–606. DOI: 10.1039/TF9221700598.

[8] J. A. Christiansen. PhD thesis. University of Copenhagen, 1921.

[9] C. N. Hinshelwood. "On the Theory of Unimolecular Reactions." *Proc. Royal Soc. A* 17, 1926, pp. 230–233.

[10] A. W. Jasper and J. A. Miller. "Collisional Energy Transfer in Unimolecular Reactions: Direct Classical Trajectories for $CH_4 \rightleftharpoons CH_3 + H$ in Helium." *J. Phys. Chem. A* 113 (19), 2009, pp. 5612–5619. DOI: 10.1021/jp900802f.

[11] J. R. Barker and R. E. Weston. "Collisional Energy Transfer Probability Densities P(E, J; E', J') for Monatomics Colliding with Large Molecules." *J. Phys. Chem. A* 114 (39), 2010, pp. 10619–10633. DOI: 10.1021/jp106443d.

[12] A. W. Jasper and J. A. Miller. "Theoretical Unimolecular Kinetics for $CH_4 + M \rightleftharpoons CH_3 + H + M$ in Eight Baths, M = He, Ne, Ar, Kr, H_2, N_2, CO, and CH_4." *J. Phys. Chem. A* 115 (24), 2011, pp. 6438–6455. DOI: 10.1021/jp200048n.

[13] O. K. Rice and H. C. Ramsperger. "Theories of Unimolecular Gas Reactions at Low Pressures." 49, 1927, pp. 1617–1629. DOI: 10.1021/ja01406a001.

[14] L. S. Kassel. "Studies in Homogeneous Gas Reactions II: Introduction of Quantum Theory." *J. Phys. Chem.* 32, 1928, pp. 1065–1079. DOI: 10.1021/j150289a011.

[15] R. A. Marcus and O. K. Rice. "The Kinetics of the Recombination of Methyl Radical and Iodine Atoms." *J. Phys. Coll. Chem.* 55, 1951, pp. 894–908. DOI: 10.1021/j150489a013.

[16] N. B. Slater. *Proc. Leeds Phil. Soc.* 6, 1955, pp. 268–694.

[17] 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.

[18] S. C. Smith and R. G. Gilbert. "Angular Momentum Conservation in Unimolecular and Recombination Reactions." *Int. J. Chem. Kin.* 20, 1988, pp. 307–329. DOI: 10.1002/kin.550200405.

[19] S. H. Robertson, A. I. Shushin, and D. M. Wardlaw. "Reduction of the Two-Dimensional Master Equation to a Smoluchowsky Type Differential Equation With Application to $CH_4 \longrightarrow CH_3 + H$." *J. Phys. Chem.* 98, 1993, pp. 8673–8679. DOI: 10.1063/1.464474.

[20] A. J. F. Siegert. "On the Approach to Statistical Equilibrium." *Phys. Rev.* 76, 1949, pp. 1708–1714. DOI: 10.1103/PhysRev.76.1708.

[21] A. F. Bartholomay. "Stochastic Models for Chemical Reactions: I. Theory of the Unimolecular Reaction Process." *Bull. Math. Biophys.* 20, 1958, pp. 175–190. DOI: 10.1007/BF02478297.

[22] E. W. Montroll and K. E. Shuler. "The Application of the Theory of Stochastic Processes to Chemical Kinetics." *Adv. Chem. Phys.* 1, 1958, pp. 361–399. DOI: 10.1002/9780470143476.ch11.

[23] I. M. Krieger and P. J. Gans. "First-Order Stochastic Processes." *J. Chem. Phys.* 32, 1960, pp. 247–250. DOI: 10.1063/1.1700909.

[24] P. J. Gans. "Open First-Order Stochastic Processes." *J. Chem. Phys.* 33, 1960, pp. 691–694. DOI: 10.1063/1.1731239.

[25] B. Widom. "Mean-First-Passage Times and the Collision Theory of Bimolecular Reactions." *J. Chem. Phys.* 31, 1959, pp. 1387–1394. DOI: 10.1063/1.1730604.

[26] J. Keck and G. Carrier. "Diffusion Theory of Nonequilibrium Dissociation and Recombination." *J. Chem. Phys.* 43, 1965, pp. 2284–2298. DOI: 10.1063/1.1697125.

[27] J. Troe and H. G. Wagner. "Unimolecular Reactions in Thermal Systems." *Ber. Bunsenges. Phys. Chem.* 71, 1967, pp. 937–979. DOI: 10.1002/bbpc.19670710904.

[28] J. Troe. "Collisional Energy Transfer in Thermal Unimolecular Reactions." *Ber. Bunsenges. Phys. Chem.* 77, 1973, pp. 665–674. DOI: 10.1002/bbpc.19730770903.

[29] D. C. Tardy and B. S. Rabinovitch. "Collisional Energy Transfer: Thermal Unimolecular Systems in the Low-Pressure Region." *J. Phys. Chem.* 45, 1966, pp. 3720–3730. DOI: 10.1063/1.1727392.

[30] C. W. Gear. "The Automatic Integration of Ordinary Differential Equations." *Commun. ACM* 14, 1973, pp. 176–179. DOI: 10.1145/362566.362571.

[31] T. Beyer and D. F. Swinehart. "Number of Multiply-Restricted Partitions." *Commun. ACM* 16, 1973, p. 379. DOI: 10.1145/362248.362275.

[32] S. E. Stein and B. S. Rabinovitch. "Accurate Evaluation of Internal Energy Level Sums and Densities Including Anharmonic Oscillators and Hindered Rotors." *J. Phys. Chem.* 58, 1973, pp. 2438–2444. DOI: 10.1063/1.1679522.

[33] D. C. Astholz, J. Troe, and W. Wieters. "Unimolecular Processes in Vibrationally Highly Excited Cycloheptatrienes. I. Thermal Isomerization in Shock Waves." *J. Phys. Chem.* 70, 1979, pp. 5107–5116. DOI: 10.1063/1.437352.

[34] J. Troe. "Theory of Thermal Unimolecular Reactions at Low Pressures. I. Solutions of the Master Equation." *J. Phys. Chem.* 66, 1977, pp. 4745–4757. DOI: 10.1063/1.433837.

[35] R. G. Gilbert, K. Luther, and J. Troe. "Theory of Thermal Unimolecular Reactions in the Fall-off Range. II. Weak Collision Rate Constants." *Ber. Bunsenges. Phys. Chem.* 87, 1983, pp. 169–177. DOI: 10.1002/bbpc.19830870218.

[36] 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.

[37] M. A. Blitz, K. J. Hughes, M. J. Pilling, and S. H. Robertson. "Combined Experimental and Master Equation Investigation of the Multiwell Reaction H + SO2." *J. Phys. Chem. A* 110, 2006, pp. 2996–3009. DOI: 10.1021/jp054722u.

[38] S. H. Robertson, M. J. Pilling, L. C. Jitariu, and I. H. Hillier. "Master Equation Methods for Multiple Well Systems: Application to the 1-,2-pentyl System." *Phys. Chem. Chem. Phys.* 9, 2007, pp. 4085–4097. DOI: 10.1039/b704736c.

[39] S. J. Klippenstein and J. A. Miller. "From the Time-Dependent, Multiple-Well Master Equation to Phenomenological Rate Coefficients." *J. Phys. Chem. A* 106, 2002, pp. 9267–9277. DOI: 10.1021/jp021175t.

[40] W. H. Miller, Y. Zhao, M. Ceotto, and S. Yang. "Quantum Instanton Approximation for Thermal Rate Constants of Chemical Reactions." *J. Chem. Phys.* 119, 2003, pp. 1329–1342. DOI: 10.1063/1.1580110.

[41] 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.

[42] C. F. Goldsmith. "Predicting Combustion Properties of Hydrocarbon Fuel Mixtures." PhD thesis. Massachusetts Institute of Technology, 2010.

[43] J. W. Allen, C. F. Goldsmith, and W. H. Green. "Automatic Estimation of Pressure-Dependent Rate Coefficients." *Phys. Chem. Chem. Phys.* 14, 2012, pp. 1131–1155. DOI: 10.1039/c1cp22765c.

[44] P. K. Venkatesh, A. Y. Chang, A. M. Dean, M. H. Cohen, and R. W. Carr. "Parameterization of Pressure- and Temperature-Dependent Kinetics in Multiple Well Reactions." *AIChE J.* 43, 1997, pp. 1331–1340. DOI: 10.1002/aic.690430522.

[45] J. H. Seinfeld and S. N. Pandis. *Atmospheric Chemistry and Physics - From Air Pollution to Climate Change*. Second. John Wiley and Sons, 2006.

[46] J. V. Michael, D. G. Keil, and R. B. Klemm. "Rate Constants for the Reaction of Hydroxyl Radicals with Acetaldehyde from 244-528 K." *J. Chem. Phys.* 83, 1985, pp. 1630–1636. DOI: 10.1063/1.449400.

[47] G. S. Tyndall, J. J. Orlando, T. J. Wallington, and M. D. Hurley. "Pressure Dependence of the Rate Coefficients and Product Yields for the Reaction of CH_3CO Radicals with O_2." *Int. J. Chem. Kin.* 29, 1997, pp. 655–663. DOI: 10.1002/(SICI)1097-4601(1997)29:9<655::AID-KIN2>3.0.CO;2-T.

[48] M. A. Blitz, D. E. Heard, and M. J. Pilling. "OH Formation from $CH_3CO + O_2$: A Convenient Experimental Marker for the Acetyl Radical." *Chem. Phys. Lett.* 365, 2002, pp. 374–379. DOI: 10.1016/S0009-2614(02)01484-7.

[49] H. Hou, A. Li, H. Hu, Y. Li, H. Li, and B. Wang. "Mechanistic and Kinetic Study of the $CH_3CO + O_2$ Reaction." *J. Chem. Phys.* 122, 2005, p. 224304. DOI: 10.1063/1.1897375.

[50] R. K. Talukdar, M. E. Davis, L. Zhu, and A. R. Ravishankara. *19th International Symposium on Gas Kinetics*, 2006.

[51] G. Kovács, J. Zádor, E. Farkas, R. Nádasdi, I. Szilágyi, S. Dóbé, T. Bérces, F. Márta, and G. Lendvay. "Kinetics and Mechanism of the Reactions of CH_3CO and $CH_3C(O)CH_2$ Radicals with O_2. Low-Pressure Discharge Flow Experiments and Quantum Chemical Computations." *Phys. Chem. Chem. Phys.* 9, 2007, pp. 4142–4154. DOI: 10.1039/b706216h.

[52] S.-Y. Chen and Y.-P. Lee. "Transient infrared absorption of t-$CH_3C(O)OO$, c-$CH_3C(O)OO$, and α-lactone recorded in gaseous reactions of CH_3CO and O_2." *J. Chem. Phys.* 132, 2010, p. 114303. DOI: 10.1063/1.3352315.

[53] J. Lee, C.-J. Chen, and J. W. Bozzelli. "Thermochemical and Kinetic Analysis of the Acetyl Radical (CH_3CO) + O_2 Reaction System." *J. Phys. Chem. A* 106, 2002, pp. 7155–7170. DOI: 10.1021/jp014443g.

[54] A. Maranzana, J. R. Barker, and G. Tonachini. "Master Equation Simulations of Competing Unimolecular and Bimolecular Reactions: Application to OH Production in the Reaction of Acetyl Radical with O_2." *Phys. Chem. Chem. Phys.* 9, 2007, pp. 4129–4141. DOI: 10.1039/b705116f.

[55] C. F. Goldsmith, S. J. Klippenstein, and W. H. Green. "Theoretical Rate Coefficients for Allyl + HO_2 and Allyloxy Decomposition." *Proc. Comb. Inst.* 33, 2011, pp. 273–282. DOI: 10.1016/j.proci.2010.05.054.

[56] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, G. C. M.; Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko,

P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. *Gaussian 03*. 2004.

[57] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schuetz, et al. *MOLPRO, version 2006.1, A Package of Ab Initio Programs*. 2006.

[58] S. M. Villano, N. Eyet, S. W. Wren, G. B. Ellison, V. M. Bierbaum, and W. C. Lineberger. "Photoelectron Spectroscopy and Thermochemistry of the Peroxyformate Anion." *J. Phys. Chem. A* 114, 2010, pp. 191–200. DOI: 10.1021/jp907569w.

[59] P. Zhang, S. J. Klippenstein, H. Sun, and C. K. Law. "Ab Initial Kinetics for the Decomposition of Monomethylhydrazine (CH_3NHNH_2)." *Proc. Comb. Inst.* 33, 2011, pp. 425–432. DOI: 10.1016/j.proci.2010.05.010.

[60] T. J. Frankcombe and S. C. Smith. "Time Evolution in the Unimolecular Master Equation at Low Temperatures: Full Spectral Solution with Scalable Iterative Methods and High Precision." *Comput. Phys. Commun.* 141, 2001, pp. 39–54. DOI: 10.1016/S0010-4655(01)00298-3.

[61] T. J. Frankcombe and S. C. Smith. "Time-Dependent Master Equation Simulation of Complex Elementary Reactions in Combustion: Application to the Reaction of $^1CH_2$ with C_2H_2 from 300-2000 K." *Faraday Discuss.* 119, 2002, pp. 159–171. DOI: 10.1039/b102562g.

[62] J. W. Allen, R. W. Ashcraft, G. J. Beran, B. A. Buesser, C. A. Class, C. Gao, C. F. Goldsmith, M. R. Harper, A. Jalan, M. Keceli, G. R. Magoon, D. M. Matheu, S. S. Merchant, J. D. Mo, S. Petway, S. Ruman, S. Sharma, K. M. Van Geem, J. Song, Y. Suleymanov, J. Wen, R. H. West, A. Wong, H.-W. Wong, N. W.-W. Yee, P. E. Yelvington, J. Yu, and W. H. Green. *RMG (Reaction Mechanism Generator) version 4.0*. 2013. URL: http://rmg.sourceforge.net/.

6

Detailed modeling of the pressure-dependent reactions of Criegee biradicals with carbonyls

The onset of pressure dependent kinetics for a given reaction network is a function of the temperature of interest and the number of degrees of freedom of the isomers of the network [1]. At the moderate to high temperatures typical of combustion processes, many unimolecular reactions are pressure-dependent, including those that lead to relatively large adducts. However, there are many smaller systems that show pressure-dependent behavior even at atmospheric conditions, and therefore require the master equation-based treatment discussed in Chapter 5.

Carbonyl oxides, such as the Criegee biradical CH_2OO [2], are important intermediates in alkene ozonolysis under atmospheric conditions [3]. They can react with H_2O, NO_x, and SO_x in the atmosphere, which has important implications for atmospheric chemistry [4, 5]. Carbonyl oxide formation is attributed to ring-opening of primary ozonides (POZ) formed directly by 1,3-cycloaddition reactions between ozone and alkenes. Aldehydes and ketones are formed as co-products in these reactions; they can recombine with carbonyl oxides to form secondary ozonides (SOZ), which can subsequently decompose or isomerize into more stable products. Secondary ozonide formation has been previously reported in both gas and solution phase ozonolysis experiments [6–8], but less is known about the reactions between the Criegee intermediate and carbonyl compounds in the gas phase under atmospheric conditions [9, 10].

For decades, carbonyl oxides like the Criegee intermediate and their reactions could not be studied experimentally because of their short lifespan and lack of direct precursors [11]. As a result, experimental rate coefficients for reactions involving Criegee biradicals were often determined indirectly with several assumptions and high uncertainties. Recently, Welz *et al.* discovered that $CH_2I + O_2 \rightleftharpoons CH_2OO + I$ is an efficient route to the formation of the Criegee intermediate with yields high enough

to perform direct kinetic measurements [4, 5]. Taatjes and coworkers have since reported direct kinetic measurements of the Criegee intermediate with acetaldehyde, acetone, and hexafluoroacetone at 293 K and 4 Torr [12]. Secondary ozonide formation was observed in reactions with CF_3COCF_3 and CH_3COCH_3, accompanied by hydroxy alkyl esters as minor products. In contrast, evidence for acetic acid formation was seen in reactions between $CH_2OO + CH_3CHO$, while no significant products were observed at *m/z* values corresponding to SOZ or hydroxy esters. Formic acid formation was also reported in all reactions, but its yield was not quantified.

In this chapter, we apply the master equation methods of Chapter 5, along with high-level *ab initio* quantum chemistry calculations, to theoretically investigate the reactions of the Criegee biradical with small carbonyl compounds: formaldehyde, acetaldehyde, and acetone. We will show that our theoretical analysis is able to reproduce the major experimental observations of Taatjes and coworkers. In order to accomplish this analysis, a new version of the CanTherm software package that integrates master equation analysis of pressure-dependent reaction networks was developed, and will be discussed briefly.

6.1 Integrating pressure dependence functionality with CanTherm

A variety of software packages have been developed for solving master equation models of unimolecular reactions, including UNIMOL [13], ChemRate [14], Variflex [15], CHEMDIS [16], MultiWell [17], SSUMES [18], and MESMER [19]. The original implementation of pressure dependence functionality in RMG-Java by David Matheu utilized CHEMDIS [20]; we developed the FAME module for RMG-Java to replace CHEMDIS, which we were no longer allowed to distribute.

The estimation of pressure-dependent rate coefficients in automatic mechanism generation requires some special functionality, particularly to handle the case where molecular parameters (e.g. vibrational frequencies) are estimated from only macroscopic quantities. FAME was originally designed specifically for this use case. Since the $k(T, P)$ values returned by FAME have a large degree of uncertainty, we soon found the need to refine these estimates using a more rigorous RRKM/master equation treatment, obtaining molecular parameters from quantum chemistry calculations. For this reason, when pressure dependence functionality was developed for RMG-Py, it was designed to be able to handle both situations.

CanTherm was originally developed by Sandeep Sharma and Michael Harper to compute thermodynamic and kinetic properties from the output of quantum chemistry

calculations [21]. This has been extremely useful in refining the values of sensitive parameters while building detailed kinetics models (as will be demonstrated in Chapter 7). Adding pressure dependence functionality to CanTherm seemed like a natural fit, as it would be able to leverage the existing functionality for extracting parameters from quantum chemistry calculations. This also gave the opportunity to clean up the CanTherm codebase a bit, and add a few additional features, e.g. automatic extraction of energies for one-dimensional scans of hindered rotor torsions.

A difficult decision in the development of the new CanTherm was whether to develop it as a standalone package or to integrate it with RMG-Py. The former makes CanTherm easier to distribute, since it relies on fewer dependencies. However, it also requires duplication of a significant amount of code that is shared with RMG-Py, particularly the thermodynamics, kinetics, and statistical mechanics models. Ultimately, we decided to integrate CanTherm with RMG-Py in order to minimize code duplication and to enable the creation of additional functionality that would be difficult with separate codes, such as the drawing of potential energy surfaces with pictures of the molecules instead of just text labels.

6.2 Computational methods

The potential energy surfaces for CH_2OO + HCHO, CH_2OO + CH_3CHO, and CH_2OO + CH_3COCH_3 were computed using geometries optimized with the B3LYP density functional and the MG3S basis set [22] using an ultrafine integration grid, and single-point energies computed at the explicitly correlated RCCSD(T)-F12a level [23–26] using the cc-pVTZ-F12 basis set [27]. Calculated vibrational frequencies and zero-point energies were scaled by 0.99 to partially account for anharmonicities. Relaxed scans along non-methyl torsional degrees of freedom were performed at 10° intervals at the B3LYP/MG3S level to obtain hindered-rotor potentials used in partition functional calculations using the one-dimensional separable rotor model. The Fourier series fit to the scan energies were combined with the reduced moment of inertia $I^{(2,3)}$ [28], and the resulting one-dimensional Schrödinger equation was solved to determine the eigenvalues, using a basis of 200 free rotor functions. The torsional degrees of freedom were projected out from the Hessian at the global minimum conformer and replaced with hindered rotor partition functions, while the remaining normal modes were treated harmonically using scaled B3LYP/MG3S frequencies. Methyl rotors were modeled using a typical three-fold symmetric sinusoidal potential with a 1.2 kcal/mol torsional barrier between successive minima (obtained from rotor scans

performed for CH_3CHO). Gaussian 03 [29] was used for all DFT calculations, while the CCSD(T) energy calculations were done using MOLPRO [30].

Each reaction network was modeled using a one-dimensional RRKM/master equation model as discussed in Chapter 5. Microcanonical rate coefficients for all reactions except the van der Waal's complex-forming entrance reaction were computed using conventional RRKM theory [31–34], including Eckart tunneling [35]. The density of states was computed via inverse Laplace transform of the partition function using the method of steepest descents [36, 37]. Collisions were modeled using the conventional exponential down expression with a single parameter $\langle \Delta E_{\text{down}} \rangle$ for the average energy transferred in a deactivating collision. Two bath gases, nitrogen and helium, were considered. A value of $\langle \Delta E_{\text{down}} \rangle = 200(T/300\ \text{K})^{0.85}\ \text{cm}^{-1}$ was used for nitrogen [38], while a value of $\langle \Delta E_{\text{down}} \rangle = 100(T/298\ \text{K})^{0.8}\ \text{cm}^{-1}$ was used for helium [39]. The collision frequency was computed by assuming a Lennard-Jones potential between the bath gas and the species of interest. Lennard-Jones parameters were estimated by first estimating the critical temperature and pressure using a group additivity method devised by Joback [40] and then using the equations for a Lennard-Jones gas [41] implemented by Harper *et al.* in RMG [42].

Phenomenological rate coefficients $k(T, P)$ were computed from the RRKM/ME model using the reservoir state method, presented in Section 5.2.2. In this work we are primarily concerned with low-temperature conditions, for which the approximation of thermalized low-energy grains used by the reservoir state method is valid. The only exception is for the van der Waals complex between CH_2OO and carbonyl, which is a very shallow well; however, this is also a very minor product, so its $k(T, P)$ values are not important. Species concentration profiles were obtained by direct numerical integration of the full master equation, and agree well with profiles obtained by numerical integration of the phenomenological rate coefficients. All master equation calculations were performed using the updated CanTherm software.

6.3 Results and discussion

6.3.1 CH_2OO + HCHO

The potential energy surface for CH_2OO + HCHO, shown in Figure 6.1, shows several features typical of the CH_2OO + carbonyl systems considered in this work. The secondary ozonide (SOZ) is formed via a van der Waal's (vdW) complex; the activation barrier for complex formation is several kcal/mol below the energy of the infinitely separated CH_2OO+HCHO reactants. This feature, coupled with the depth of the SOZ

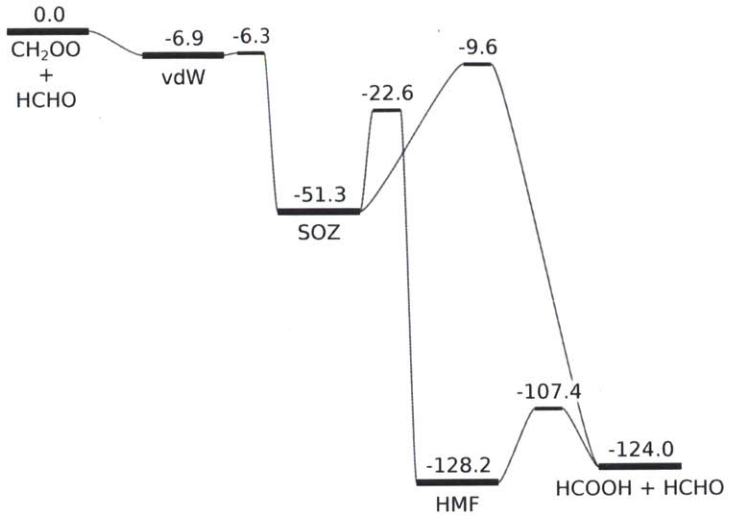


Figure 6.1: The CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S potential energy surface for the CH_2OO + HCHO network. All ground-state energies are given in kcal/mol and include zero-point energies.

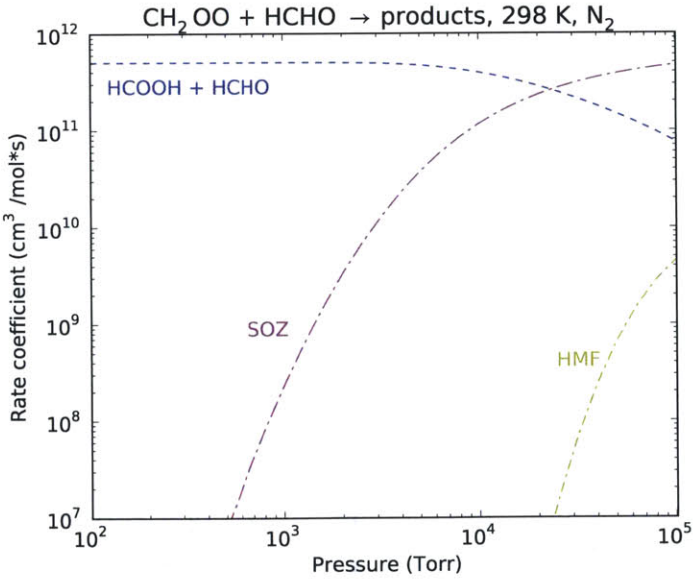


Figure 6.2: Plot of predicted phenomenological rate coefficients $k(T, P)$ versus pressure for $CH_2OO + HCHO \longrightarrow$ products at 298 K in N_2.

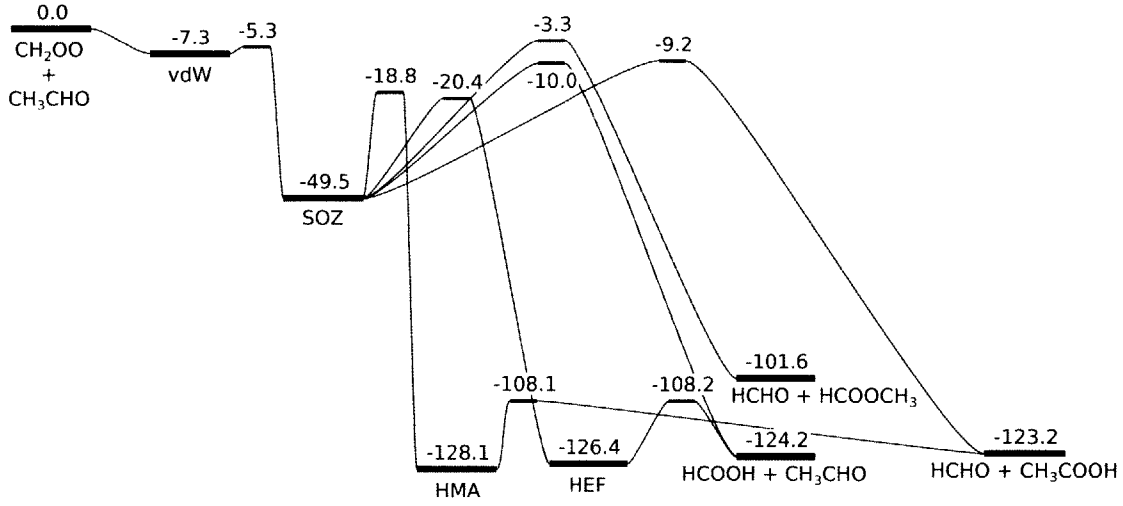


Figure 6.3: The CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S potential energy surface for the CH_2OO + CH_3CHO network. All ground-state energies are given in kcal/mol and include zero-point energies.

isomer well, creates a chemically-activated system. The only major product channel for SOZ is formic acid and HCHO; this can be formed either via a direct concerted elimination and rearrangement or via ring-opening isomerization to hydroxymethyl-formate (HMF). The isomerization pathway has a significantly lower barrier than the direct elimination, so we expect it to be the dominant route to formic acid at the experimental conditions of 293 K and 4 Torr. All of the reaction barriers are well below the ground-state energy of the reactants.

Figure 6.2 shows the phenomenological rate coefficients $k(T, P)$ for the CH_2OO + HCHO $\longrightarrow$ products reactions versus pressure at 298 K in N_2, as predicted using the master equation model and the reservoir state method. At the very high pressures shown (>20000 Torr $\approx$ 25 atm N_2) the major product channel is the secondary ozonide. As the pressure decreases, the branching increasingly favors the formic acid product over SOZ. Below atmospheric pressure (760 Torr) of nitrogen the system is well into the low-pressure limit, and formaldehyde simply serves as a catalyst for the conversion of the Criegee intermediate CH_2OO to its much more stable isomer, formic acid.

6.3.2 $CH_2OO + CH_3CHO$

The potential energy surface for $CH_2OO + CH_3CHO$, shown in Figure 6.3, shows the same chemically-activated nature as the CH_2OO + HCHO surface. This surface

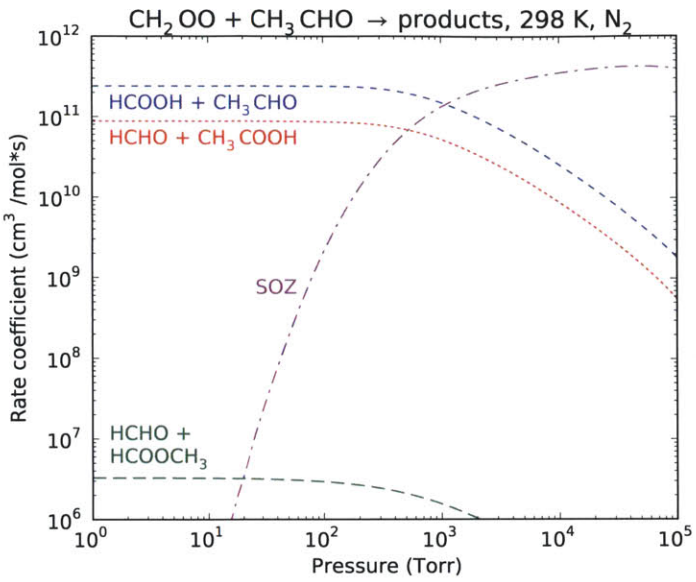


Figure 6.4: Plot of predicted phenomenological rate coefficients $k(T, P)$ versus pressure for $CH_2OO + CH_3CHO \longrightarrow$ products at 298 K in N_2.

appears more complicated due to the presence of three product channels, as the formic acid channel is joined by two pathways to HCHO due to the asymmetry of CH_3CHO: methylformate and acetic acid. Both of the acid products can be formed either by higher-barrier direct elimination reactions or via lower-barrier isomerizations to hydroxyethylformate (HEF) or hydroxymethylacetate (HMA), respectively.

Figure 6.4 shows the phenomenological rate coefficients $k(T, P)$ for the $CH_2OO + CH_3CHO \longrightarrow$ products reactions versus pressure at 298 K in N_2, as predicted using the master equation model and the reservoir state method. This system has larger isomers (with more molecular degrees of freedom) than before, which causes the onset of low-pressure behavior to shift down in pressure. The secondary ozonide is now the major product above 1000 Torr (1.5 atm) of nitrogen. At atmospheric pressure (760 Torr) of nitrogen, all three of SOZ, formic acid, and acetic acid products are formed in comparable yields. The isomerization barrier to formic acid (via HEF) is about 1.7 kcal/mol lower than the isomerization barrier to acetic acid (via HMA), which causes more formic acid to be formed than acetic acid. Methyl formate is predicted to be a very minor byproduct, as it does not have a two-step isomerization and fragmentation pathway, but only a concerted, high-barrier pathway. The significant production of SOZ observed at 760 Torr is consistent with the experiments of Horie *et al.* [43] The SOZ product becomes increasingly minor as pressure is decreased be-

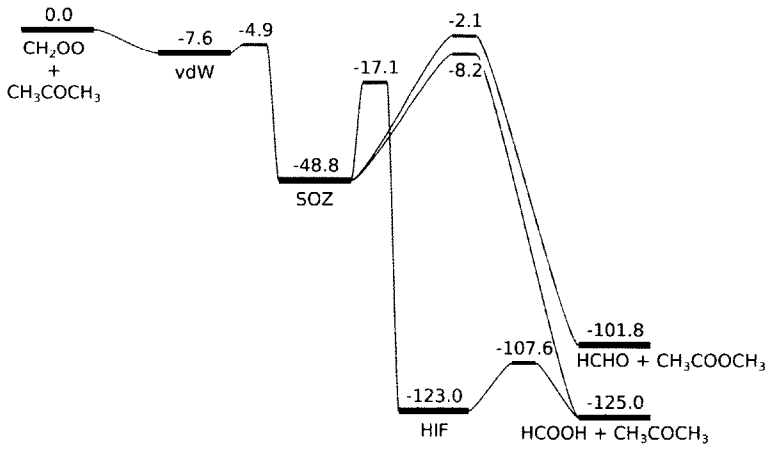


Figure 6.5: The CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S potential energy surface for the CH_2OO + CH_3COCH_3 network. All ground-state energies are given in kcal/mol and include zero-point energies.

low atmospheric conditions, as collisions become so infrequent that negligible SOZ is stabilized before it can react further.

6.3.3 $CH_2OO + CH_3COCH_3$

The potential energy surface for $CH_2OO + CH_3COCH_3$, shown in Figure 6.5, is somewhat simpler than that of CH_3CHO due to the symmetry of acetone. There is now only one product channel that yields HCHO; this channel also produces methyl acetate (CH_3COOCH_3). The only product channel with a lower-barrier isomerization route – which proceeds via hydroxylisopropylformate (HIF) – is the formic acid product.

Figure 6.6 shows the phenomenological rate coefficients $k(T, P)$ for the CH_2OO + CH_3COCH_3 $\longrightarrow$ products reactions versus pressure at 298 K in N_2, as predicted using the master equation model and the reservoir state method. The isomers of this system now have enough degrees of freedom that the secondary ozonide is the major product above 200 Torr (0.25 atm) of nitrogen, and remains a significant minor product down to 4 Torr (0.005 atm). Formic acid is the only major product observed in the low-pressure limit due to there being no lower-barrier route to methyl acetate.

6.3.4 Comparison to experiment

We now apply our master equation model to the experimental conditions of Taatjes and coworkers [12]. Direct numerical simulation of the full master equation at 293

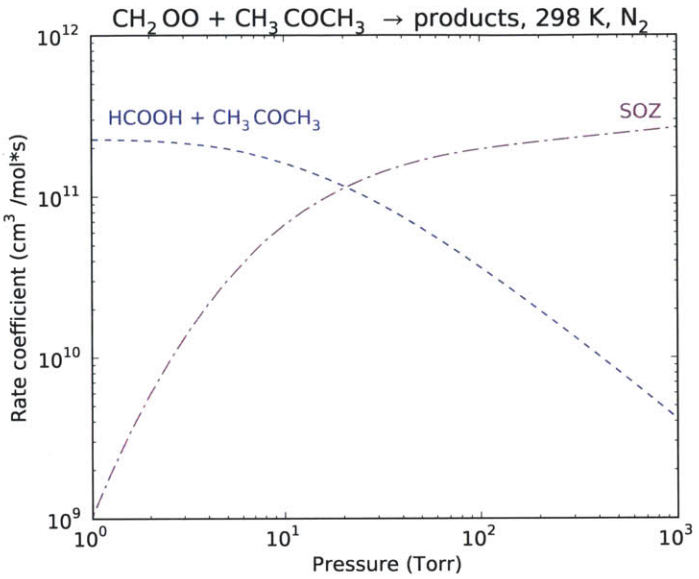


Figure 6.6: Plot of predicted phenomenological rate coefficients $k(T, P)$ versus pressure for $CH_2OO + CH_3COCH_3 \longrightarrow$ products at 298 K in N_2.

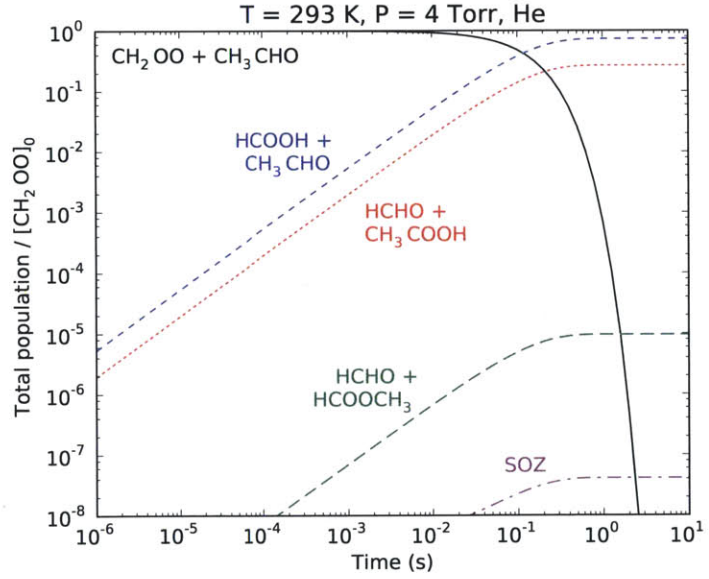


Figure 6.7: Plot of predicted concentration profiles for the $CH_2OO + CH_3CHO$ network at 293 K and 4 Torr He using a CH_3CHO mole fraction of 0.001.

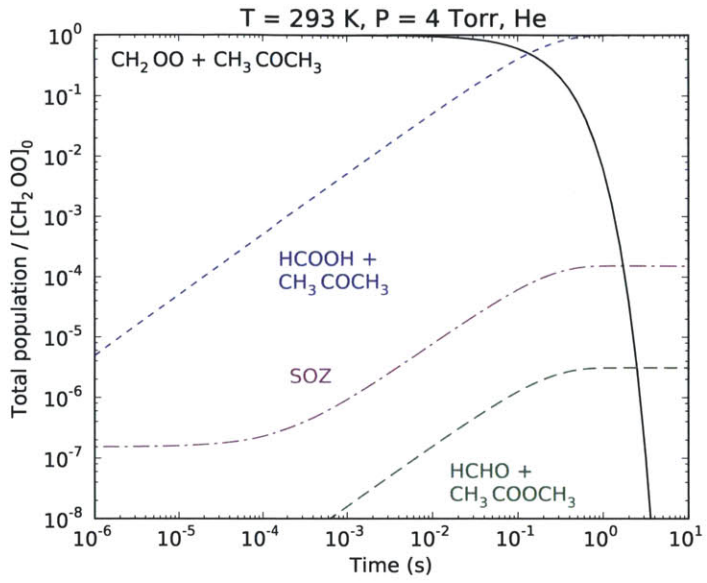


Figure 6.8: Plot of predicted concentration profiles for the $CH_2OO + CH_3COCH_3$ network at 293 K and 4 Torr He using a CH_3COCH_3 mole fraction of 0.01.

K and 4 Torr gives the concentration profiles shown in Figure 6.7. The profiles show that formic acid is the major product, with acetic acid as a significant minor product. SOZ is seen as only a very minor product, behind even that of methyl formate. These results are consistent with the experiments of Taatjes and coworkers, who reported evidence for acetic acid formation in their experiments but did not observe any SOZ. Recent work by Wang *et al.* [44] suggests that the only observable fragment ion from methyl formate below 12 eV occurs at 11.45 eV, while the highest photon energies used in the experiments of Taatjes and coworkers is 11.4 eV [12]. As a result, it is not currently possible to conclusively validate the formation of methyl formate in this system.

Direct numerical simulation of the full master equation at 293 K and 4 Torr gives the concentration profiles shown in Figure 6.8. The yield of SOZ, while still low, is substantially higher in this system than in the CH_3CHO system, which is in qualitative agreement with the results of Taatjes and coworkers. It is unclear if the experiments of Taatjes and coworkers are sensitive enough to detect SOZ at these levels; however, as we will demonstrate in the next section, the yield of SOZ is especially sensitive to the choice of collision model, which is perhaps the largest source of uncertainty in our calculations.

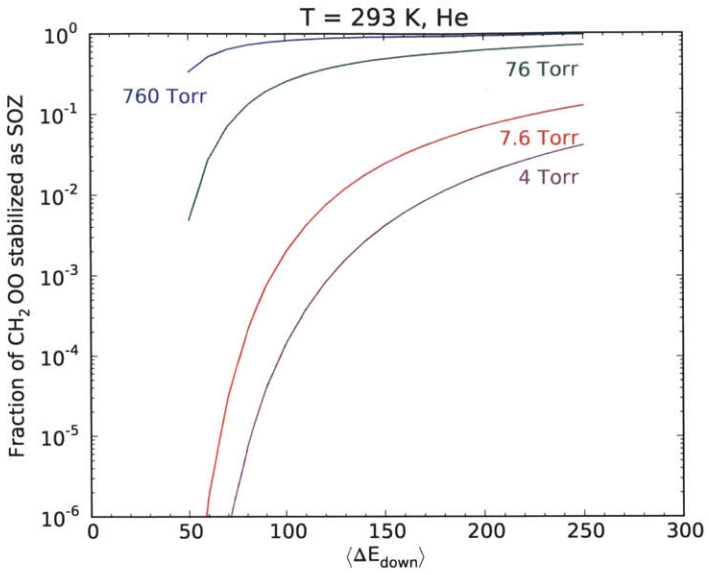


Figure 6.9: Plot of the fraction of CH_2OO stabilized as SOZ to the value of $\langle \Delta E_{\text{down}} \rangle$ parameter in the single exponential down model of collisional energy transfer for the $CH_2OO + CH_3COCH_3$ network at 293 K in He, using a CH_3COCH_3 mole fraction of 0.01. As the pressure decreases, the value of $\langle \Delta E_{\text{down}} \rangle$ has a larger effect on the fraction of SOZ stabilized, especially around the value of 100 cm^{-1} characteristic of helium.

6.3.5 Sensitivity analysis

Goldsmith *et al.* demonstrated that the largest sources of uncertainty for chemically-activated bimolecular reactions that traverse multiple wells were the value of $\langle \Delta E_{\text{down}} \rangle$ used in the single exponential down model of collisional energy transfer, the transition state barrier heights, and the kinetics of the entrance channel [45]. In this section, we examine these sources of uncertainty for the CH_2OO + carbonyl systems.

The predicted yield of SOZ at low pressures is particularly sensitive to the value of $\langle \Delta E_{\text{down}} \rangle$ used in the single exponential down model of collisional energy transfer, as shown in Figure 6.9 for the $CH_2OO + CH_3COCH_3$ system at 293 K in He. A larger value of $\langle \Delta E_{\text{down}} \rangle$ causes collisions to tend to transfer larger amounts of energy, which increases the likelihood that the excited SOZ molecules produced from $CH_2OO + CH_3COCH_3$ will be collisionally stabilized before they can react further. As the figure shows, doubling the value of $\langle \Delta E_{\text{down}} \rangle$ from 100 to 200 cm^{-1} increases the rate of collisional stabilization dramatically; this causes the amount of observed SOZ to increase by two orders of magnitude. By contrast, halving the value of $\langle \Delta E_{\text{down}} \rangle$ from 100 to 50 cm^{-1} causes the amount of observed SOZ to drop to much lower lev-

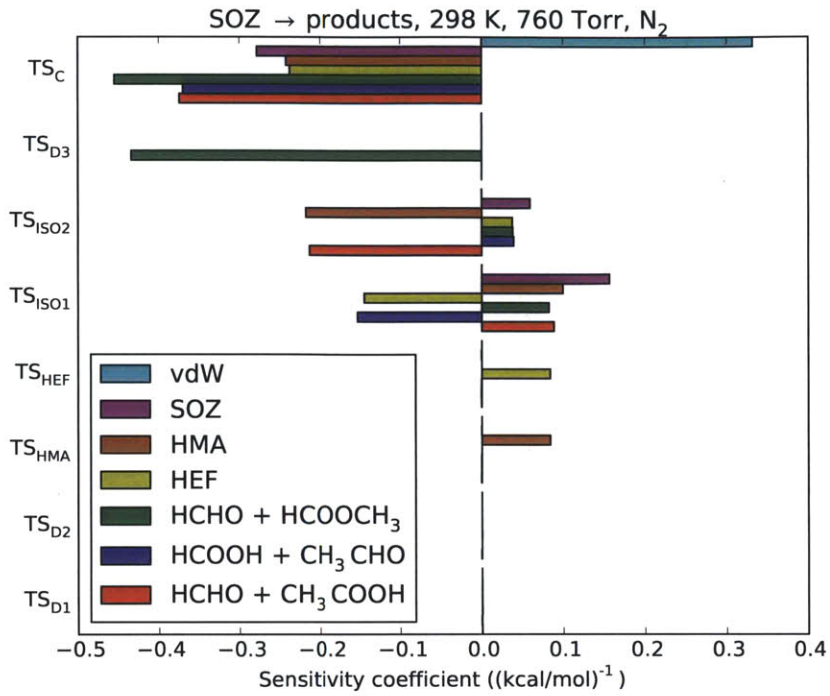


Figure 6.10: Normalized sensitivity coefficients to the transition state barrier heights for the $CH_2OO + CH_3CHO$ network at 298 K and 760 Torr N_2. Each cluster of bars corresponds to the transition state barrier height of one reaction, indicated at left. The bars indicate how the total population of each configuration is affected by adjusting the associated barrier height, and are ordered and colored by configuration as shown in the legend. The sensitivity coefficient is defined as $\partial \ln[\text{total population}]/\partial E_{0,\text{TS}}$, and has units of inverse energy.

els. This strong sensitivity to the collisional energy transfer model at low pressures is consistent with the results of Goldsmith *et al.*

The sensitivity of the products of the $CH_2OO + CH_3CHO$ system to the transition state barrier heights of all reactions except the entrance channel is shown in Figure 6.10. It is not surprising that the SOZ $\longrightarrow$ HCHO + $HCOOCH_3$ (TS_{D3}) barrier height product strongly affects the yield of HCHO + $HCOOCH_3$ despite its high activation barrier because the low-barrier isomerizations do not lead to that product. Since there are lower-energy pathways to $HCOOH+CH_3CHO$ and $HCHO+CH_3COOH$ that pass through HEF and HMA, respectively, the barrier heights of the direct reactions (TS_{D1} and TS_{D1}) are not as important. Instead, the yield of acid products is mainly sensitive to the corresponding isomerization channels of the SOZ leading to the hydroxyalkyl ester intermediates. The barrier height of the SOZ $\longrightarrow$ $CH_2OO + CH_3CHO$ reaction

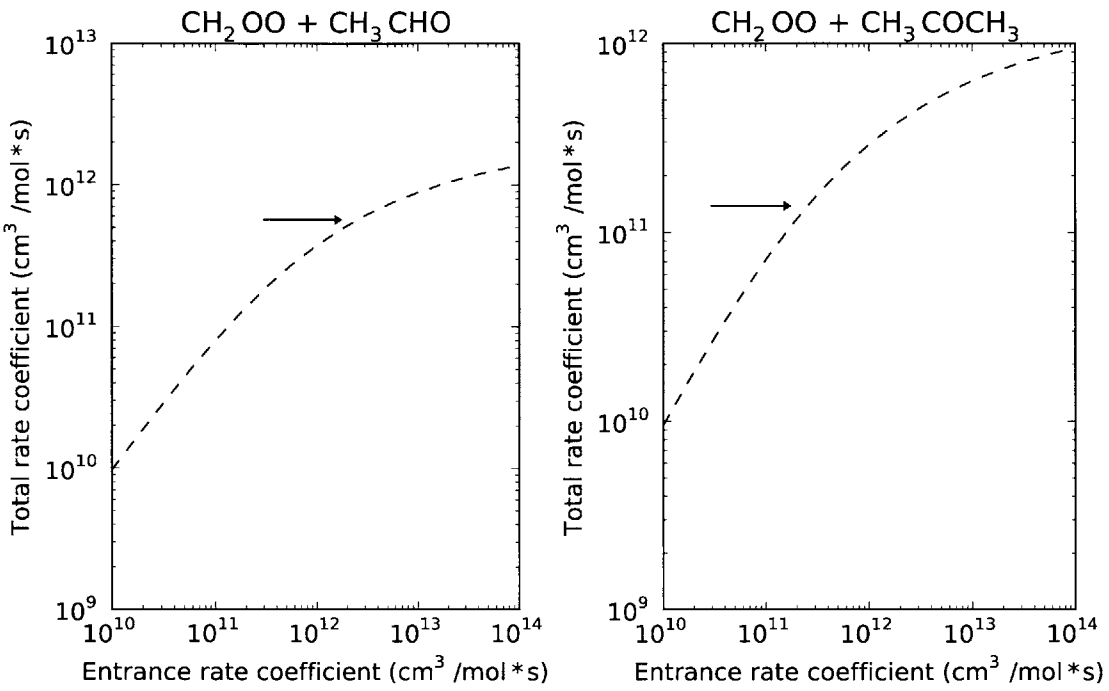


Figure 6.11: Sensitivity of the total rate of $CH_2OO + CH_3CHO \longrightarrow$ products and $CH_2OO + CH_3COCH_3 \longrightarrow$ products to the entrance channel (forming the van der Waal's complex) rate coefficient at 298 K and 760 Torr N_2. The experimentally-observed total rates by Taatjes and coworkers [12] for each reaction are indicated with horizontal arrows. When the entrance rate is low, the corresponding reverse rate is negligible, and the total rate matches the entrance rate. When the entrance rate and its reverse are high, the total rate is less than the forward entrance rate, since a significant amount of the product adduct will simply return to reactants before it can react further.

(TS_C) is also very important; increasing this barrier height causes more of the van der Waal's complex to form at the expense of all other products. The SOZ $\longrightarrow$ HEF pathway has a larger effect on the SOZ formation than SOZ $\longrightarrow$ HMA, and raising the barrier height of either of these will cause flux to divert to the other pathway. Ultimately, the sensitivity of the SOZ yield to the barrier heights is lower than it is to the value of $\langle \Delta E_{\mathrm{down}} \rangle$, which is also consistent with the analysis of Goldsmith *et al.*

Another important source of uncertainty in our calculations is the kinetics of the entrance channel, where CH_2OO and carbonyl come together to form a van der Waal's complex. Increasing the rate of this reaction will certaintly increase the overall consumption rate of CH_2OO + carbonyl, although the correlation is less than one-to-one because of the increased rate of complex falling back to the bimolecular

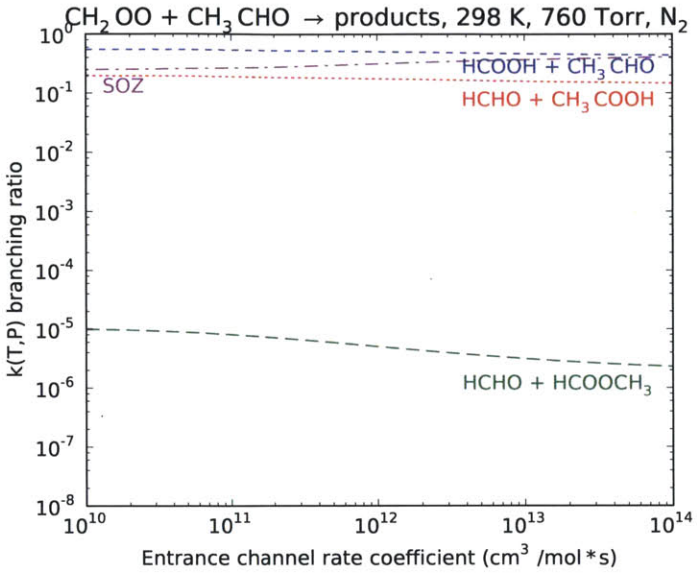


Figure 6.12: Sensitivity of the product branching ratios of CH_2OO + $CH_3CHO \longrightarrow$ products to the entrance channel (forming the van der Waal's complex) rate coefficient at 298 K and 760 Torr N_2. The increased entrance channel rate is not proportionally divided into the various products; the branching to the unimolecular isomer wells increases, while the branching to the bimolecular product channels decreases.

reactants. This overall rate increase is also not necessarily proportionally distributed amongst the different product channels. Figure 6.11 shows the sensitivity of the product branching ratio to the kinetics of the entrance channel for the major products of $CH_2OO + CH_3CHO$. Increasing the entrance channel rate does not affect the branching ratio to the major product channels $HCOOH+CH_3CHO$ and $HCHO+CH_3COOH$; the branching ratio to the minor SOZ and $HCHO + HCOOCH_3$ pathways do change somewhat, with the former increasing and the latter decreasing, but this is not enough to affect the main conclusions of this analysis.

Another important source of uncertainty in our calculations is the kinetics of the entrance channel, where CH_2OO and carbonyl come together to form a van der Waal's complex. Figure 6.11 shows the sensitivity of the total rate of CH_2OO + $CH_3CHO \longrightarrow$ products and $CH_2OO+CH_3COCH_3 \longrightarrow$ products to the entrance channel rate coefficient at 298 K and 760 Torr N_2. As expected, increasing the rate coefficient of the entrance channel will certainly increase the overall consumption rate of CH_2OO + carbonyl, although the correlation is less than one-to-one as the entrance rate increases because of the increased rate of the van der Waal's complex falling back

to the bimolecular reactants. The experimentally observed total rates of Taatjes and coworkers are also shown in Figure 6.11 as horizontal arrows. The entrance channel rate that reproduces these total rates is generally within a factor of three of the nominal value used in our master equation model.

Figure 6.12 shows the effect of increasing the entrance channel rate on the product branching ratio for $CH_2OO + CH_3CHO \longrightarrow$ products at 298 K and 760 Torr N_2. The increased entrance channel rate is not proportionally distributed amongst the different product channels, as the branching to unimolecular isomer wells (SOZ) increases, while the branching to the bimolecular product channels decrease. However, the figure shows that the overall sensitivity of the product branching ratios to the entrance channel rate is small enough that it does not affect the main conclusions of this work.

6.4 Conclusion

We have developed an *ab initio* model of the reaction of the Criegee biradical with formaldehyde, acetaldehyde, and acetone that combines high-level quantum chemistry calculations with master equation analysis. Our model reproduces the major features of recent experimental results. In particular, it shows that the yield of the secondary ozonide in the acetone system is substantially larger than in the acetaldehyde system at 293 K and 4 Torr. We have also demonstrated how the yield of secondary ozonide is strongly sensitive to the collision model used in the master equation analysis.

These results serve as an example of how we can use theoretical methods to generate accurate estimates of chemical reaction kinetics, and how the CanTherm software is a useful tool for doing so. In the next chapter, we will use this idea as part of our process of building detailed kinetics models for the combustion of an alternative fuel.

6.5 Appendix: High pressure-limit rate coefficients

Tables 6.1, 6.2, and 6.3 contain the high pressure-limit kinetics used in this work. All kinetics except the entrance channel were computed using conventional transition state theory with asymmetric Eckart tunneling, and included one-dimensional hindered rotors where appropriate.

Table 6.1: Calculated high-pressure limit rate coefficients for the CH_2OO + HCHO network.[a]

Reaction	A	n	E_a
CH_2OO + HCHO $\longrightarrow$ vdW [b]	8.0×10^{11}	0.0	0.0
SOZ $\longrightarrow$ vdW	2.4×10^{12}	0.851	46.1
SOZ $\longrightarrow$ HMF	6.7×10^{7}	1.97	26.9
SOZ $\longrightarrow$ HCOOH + HCHO	1.1×10^{9}	1.70	40.7
HMF $\longrightarrow$ HCOOH + HCHO	1.7×10^{12}	-0.026	19.1

[a] The units for A are s^{-1} for unimolecular reactions and in cm^3 mol^{-1} s^{-1} for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$. Computed from CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S calculations with Eckart tunneling; see text for details.

[b] Estimated in this work.

Table 6.2: Calculated high-pressure limit rate coefficients for the CH_2OO + CH_3CHO network.[a]

Reaction	A	n	E_a
CH_2OO + CH_3CHO $\longrightarrow$ vdW [b]	8.0×10^{11}	0.0	0.0
SOZ $\longrightarrow$ vdW	4.8×10^{12}	0.731	45.3
SOZ $\longrightarrow$ HEF	1.1×10^{9}	1.64	28.0
SOZ $\longrightarrow$ HMA	4.0×10^{8}	1.75	29.4
SOZ $\longrightarrow$ HCHO + CH_3COOH	1.3×10^{8}	1.90	38.6
SOZ $\longrightarrow$ HCOOH + CH_3CHO	2.7×10^{10}	1.33	39.5
SOZ $\longrightarrow$ HCHO + $HCOOCH_3$	8.8×10^{11}	0.987	47.0
HMA $\longrightarrow$ HCHO + CH_3COOH	1.7×10^{13}	-0.235	18.9
HEF $\longrightarrow$ HCOOH + CH_3CHO	1.9×10^{14}	-0.524	17.6

[a] The units for A are s^{-1} for unimolecular reactions and in cm^3 mol^{-1} s^{-1} for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$. Computed from CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S calculations with Eckart tunneling; see text for details.

[b] Estimated in this work.

Table 6.3: Calculated high-pressure limit rate coefficients for the CH_2OO + CH_3COCH_3 network.[a]

Reaction	A	n	E_a
$CH_2OO + CH_3COCH_3 \longrightarrow$ vdW [b]	8.0×10^{11}	0.0	0.0
SOZ $\longrightarrow$ vdW	9.1×10^{12}	0.617	45.0
SOZ $\longrightarrow$ HIF	3.2×10^{8}	1.81	30.2
SOZ $\longrightarrow HCHO + CH_3COOCH_3$	6.2×10^{11}	0.965	47.3
SOZ $\longrightarrow HCOOH + CH_3COCH_3$	9.4×10^{10}	1.18	40.9
HIF $\longrightarrow HCOOH + CH_3COCH_3$	9.1×10^{14}	-0.733	15.6

[a] The units for A are s^{-1} for unimolecular reactions and in $cm^3\ mol^{-1}\ s^{-1}$ for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$. Computed from CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S calculations with Eckart tunneling; see text for details.
[b] Estimated in this work.

6.6 References

[1] B. M. Wong, D. M. Matheu, and W. H. Green. "Temperature and Molecular Size Dependence of the High-Pressure Limit." *J. Phys. Chem. A* 107 (32), 2003, pp. 6206–6211. DOI: 10.1021/jp034165g.

[2] R. Criegee and G. Wenner. *Justus Liebigs Annalen der Chemie* 564, 1949, p. 9.

[3] L. Vereecken and J. S. Francisco. "Theoretical Studies of Atmospheric Reaction Mechanisms in the Troposphere." *Chem. Soc. Rev.* 41, 2012, pp. 6259–6293. DOI: 10.1039/C2CS35070J.

[4] O. Welz, J. D. Savee, D. L. Osborn, S. S. Vasu, C. J. Percival, D. E. Shallcross, and C. A. Taatjes. "Direct Kinetic Measurements of Criegee Intermediate (CH_2OO) Formed by Reaction of CH_2I with O_2." *Science* 335 (6065), 2012, pp. 204–207. DOI: 10.1126/science.1213229.

[5] L. Vereecken, H. Harder, and A. Novelli. "The Reaction of Criegee Intermediates with NO, RO_2, and SO_2, and their Fate in the Atmosphere." *Phys. Chem. Chem. Phys.* 14, 2012, pp. 14682–14695. DOI: 10.1039/C2CP42300F.

[6] R. Fajgar, J. Vítek, Y. Haas, and J. Pola. "Observation of Secondary 2-butene Ozonide in the Ozonation of trans-2-butene in the Gas Phase." *Tetrahedron Lett.* 37 (19), 1996, pp. 3391–3394. DOI: 10.1016/0040-4039(96)00554-0.

[7] P. Neeb, O. Horie, and G. K. Moortgat. "Formation of Secondary Ozonides in the Gas-Phase Ozonolysis of Simple Alkenes." *Tetrahedron Lett.* 37 (52), 1996, pp. 9297–9300. DOI: 10.1016/S0040-4039(97)82946-2.

[8] R. Fajgar, J. Vitek, Y. Haas, and J. Pola. "Formation of Secondary Ozonides in the Gas Phase Low-Temperature Ozonation of Primary and Secondary Alkenes." *J. Chem. Soc. Perkin Trans. 2*, 1999, pp. 239–248. DOI: 10.1039/A807208F.

[9] P. Aplincourt and M. F. Ruiz-López. "Theoretical Study of Formic Acid Anhydride Formation from Carbonyl Oxide in the Atmosphere." *J. Phys. Chem. A* 104 (2), 2000, pp. 380–388. DOI: 10.1021/jp9928208.

[10] P. Aplincourt and M. F. Ruiz-López. "Theoretical Investigation of Reaction Mechanisms for Carboxylic Acid Formation in the Atmosphere." *J. Am. Chem. Soc.* 122 (37), 2000, pp. 8990–8997. DOI: 10.1021/ja000731z.

[11] C. A. Taatjes, G. Meloni, T. M. Selby, A. J. Trevitt, D. L. Osborn, C. J. Percival, and D. E. Shallcross. "Direct Observation of the Gas-Phase Criegee Intermediate (CH_2OO)." *J. Am. Chem. Soc.* 130 (36), 2008, pp. 11883–11885. DOI: 10.1021/ja804165q.

[12] C. A. Taatjes, O. Welz, A. J. Eskola, J. D. Savee, D. L. Osborn, E. P. F. Lee, J. M. Dyke, D. W. K. Mok, D. E. Shallcross, and C. J.Percival. "Direct Measurement of Criegee Intermediate (CH_2OO) Reactions with Acetone, Acetaldehyde, and Hexafluoroacetone." *Phys. Chem. Chem. Phys.* 14, 2012, pp. 10391–10400. DOI: 10.1039/C2CP40294G.

[13] R. G. Gilbert, S. C. Smith, and M. J. T. Jordan. *UNIMOL Program Suite (Calculation of Fall-Off Curves for Unimolecular and Recombination Reactions)*. Available from: School of Chemistry, Sydney University, New South Wales, Australia. 1993.

[14] V. Mokrushin, V. Bedanov, W. Tsang, M. R. Zachariah, and V. D. Knyazev. *ChemRate Computer Program, Version 1.10*. 1999.

[15] S. J. Klippenstein, A. F. Wagner, R. C. Dunbar, D. M. Wardlaw, and S. H. Robertson. *Variflex Version 1.00*. 1999. URL: http://ftp.tcg.anl.gov/pub/variflex/Summary.vrfx.

[16] 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.

[17] J. R. Barker. "Multiple-Well, Multiple-Path Unimolecular Reaction Systems. I. MULTIWELL Computer Program Suite." *Int. J. Chem. Kinet.* 33, 2001, pp. 232–245.

[18] A. Miyoshi. *SSUMES Software, Rev. 2010.05.23m4*. URL: http://www.frad.t.u-tokyo.ac.jp/~miyoshi/ssumes/.

[19] D. R. Glowacki, C.-H. Liang, C. Morley, M. J. Pilling, and S. H. Robertson. "MESMER: An Open-Source Master Equation Solver for Multi-Energy Well Reactions." *J. Phys. Chem. A* 116(38), 2012, pp. 9545–9560. DOI: 10.1021/jp3051033.

[20] D. M. Matheu. "Integrated Pressure-Dependence in Automated Mechanism Generation: A New Tool for Building Gas-Phase Kinetic Models." PhD thesis. Massachusetts Institute of Technology, 2003.

[21] S. Sharma, M. R. Harper, and W. H. Green. *CanTherm*. 2010. URL: https://github.com/GreenGroup/CanTherm.

[22] B. J. Lynch, Y. Zhao, and D. G. Truhlar. "Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory." *J. Phys. Chem. A* 107(9), 2003, pp. 1384–1388. DOI: 10.1021/jp0215901.

[23] T. B. Adler, G. Knizia, and H.-J. Werner. "A Simple and Efficient CCSD(T)-F12 Approximation." *J. Chem. Phys.* 127(22), 2007, p. 221106. DOI: 10.1063/1.2817618.

[24] G. Knizia, T. B. Adler, and H.-J. Werner. "Simplified CCSD(T)-F12 Methods: Theory and Benchmarks." *J. Chem. Phys.* 130(5), 2009, p. 054104. DOI: 10.1063/1.3054300.

[25] T. B. Adler, H.-J. Werner, and F. R. Manby. "Local Explicitly Correlated Second-Order Perturbation Theory for the Accurate Treatment of Large Molecules." *J. Chem. Phys.* 130(5), 2009, p. 054106. DOI: 10.1063/1.3040174.

[26] T. B. Adler and H.-J. Werner. "Local Explicitly Correlated Coupled-Cluster Methods: Efficient Removal of the Basis Set Incompleteness and Domain Errors." *J. Chem. Phys.* 130(24), 2009, p. 241101. DOI: 10.1063/1.3160675.

[27] K. A. Peterson, T. B. Adler, and H.-J. Werner. "Systematically Convergent Basis Sets for Explicitly Correlated Wavefunctions: The Atoms H, He, B–Ne, and Al–Ar." *J. Chem. Phys.* 128(8), 2008, p. 084102. DOI: 10.1063/1.2831537.

[28] A. L. L. East and L. Radom. "Ab initio statistical thermodynamical models for the computation of third-law entropies." *J. Chem. Phys.* 106(16), 1997, pp. 6655–6674. DOI: 10.1063/1.473958.

[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, G. C. M.; Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. *Gaussian 03*. 2004.

[30] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schuetz, et al. *MOLPRO, version 2006.1, A Package of Ab Initio Programs*. 2006.

[31] O. K. Rice and H. C. Ramsperger. "Theories of Unimolecular Gas Reactions at Low Pressures." 49, 1927, pp. 1617–1629. DOI: 10.1021/ja01406a001.

[32] L. S. Kassel. "Studies in Homogeneous Gas Reactions II: Introduction of Quantum Theory." *J. Phys. Chem.* 32, 1928, pp. 1065–1079. DOI: 10.1021/j150289a011.

[33] R. A. Marcus and O. K. Rice. "The Kinetics of the Recombination of Methyl Radical and Iodine Atoms." *J. Phys. Coll. Chem.* 55, 1951, pp. 894–908. DOI: 10.1021/j150489a013.

[34] R. A. Marcus. "Unimolecular Dissociations and Free Radical Recombination Reactions." *J. Chem. Phys.* 20 (3), 1952, pp. 359–364. DOI: 10.1063/1.1700424.

[35] C. Eckart. "The Penetration of a Potential Barrier by Electrons." *Phys. Rev.* 35, 1930, pp. 1303–1309. DOI: 10.1103/PhysRev.35.1303.

[36] W. Forst and Z. Prasil. "Comparative Test of Approximations for Calculation of Energy-Level Densities." *J. Chem. Phys.* 51 (7), 1969, pp. 3006–3012. DOI: 10.1063/1.1672449.

[37] M. R. Hoare and T. W. Ruijgrok. "Inversion of the Partition Function: The First-Order Steepest-Descent Method." *J. Chem. Phys.* 52 (1), 1970, pp. 113–120. DOI: 10.1063/1.1672655.

[38] P. Zhang, S. J. Klippenstein, H. Sun, and C. K. Law. "Ab Initial Kinetics for the Decomposition of Monomethylhydrazine (CH_3NHNH_2)." *Proc. Comb. Inst.* 33, 2011, pp. 425–432. DOI: 10.1016/j.proci.2010.05.010.

[39] A. W. Jasper and J. A. Miller. "Collisional Energy Transfer in Unimolecular Reactions: Direct Classical Trajectories for $CH_4 \rightleftharpoons CH_3 + H$ in Helium." *J. Phys. Chem. A* 113 (19), 2009, pp. 5612–5619. DOI: 10.1021/jp900802f.

[40] K. G. Joback. "A Unified Approach to Physical Property Estimation Using Multivariate Statistical Techniques." MA thesis. Massachusetts Institute of Technology, 1984.

[41] J. R. Welty, C. E. Wicks, R. E. Wilson, and G. L. Rorrer. *Fundamentals of Momentum, Heat, and Mass Transfer*. John Wiley and Sons, 2001.

[42] J. W. Allen, R. W. Ashcraft, G. J. Beran, B. A. Buesser, C. A. Class, C. Gao, C. F. Goldsmith, M. R. Harper, A. Jalan, M. Keceli, G. R. Magoon, D. M. Matheu, S. S. Merchant, J. D. Mo, S. Petway, S. Ruman, S. Sharma, K. M. Van Geem, J. Song, Y. Suleymanov, J. Wen, R. H. West, A. Wong, H.-W. Wong, N. W.-W. Yee, P. E. Yelvington, J. Yu, and W. H. Green. *RMG (Reaction Mechanism Generator) version 4.0*. 2013. URL: http://rmg.sourceforge.net/.

[43] O. Horie, C. Schäfer, and G. K. Moortgat. "High Reactivity of Hexafluoro Acetone Toward Criegee Intermediates in the Gas-Phase Ozonolysis of Simple Alkenes." *Int. J. Chem. Kinet.* 31 (4), 1999, pp. 261–269. DOI: 10.1002/(SICI)1097-4601(1999)31:4<261::AID-KIN3>3.0.CO;2-Z.

[44] J. Wang, B. Yang, T. A. Cool, and N. Hansen. "Absolute Cross-Sections for Dissociative Photoionization of Some Small Esters." *Int. J. Mass. Spectrom.* 292 (1-3), 2010, pp. 14–22. DOI: 10.1016/j.ijms.2010.02.010.

[45] C. F. Goldsmith, A. S. Tomlin, and S. J. Klippenstein. "Uncertainty Propagation in the Derivation of Phenomenological Rate Coefficients from Theory: A Case Study of n-propyl Radical Oxidation." *Proc. Combust. Inst.* 34 (1), 2013, pp. 177–185. DOI: 10.1016/j.proci.2012.05.091.

7

AUTOMATIC GENERATION OF A DETAILED KINETICS MODEL OF DIISOPROPYL KETONE PYROLYSIS AND OXIDATION

Biofuels offer the possibility of a sustainable, carbon-neutral replacement for petroleum-based liquid fuels that are also compatible with current infrastructure and engine technology [1]. The first generation of biofuels – chiefly ethanol and biodiesel – are less than ideal due to lower energy density, additional transport complications, and a linking of food and fuel prices [2]. Significant effort is therefore underway to investigate lignocellulosic biofuels as an alternative to the first generation grain-based biofuels [3].

There are two important questions to answer when evaluating a cellulosic biofuel candidate. The first question is one of production: can the candidate biofuel be produced sustainably at large scale? Due to the recalcitrance of lignocellulosic materials, new means for breaking down biomass are a central area of biofuels research, particularly via the metabolic engineering of various microorganisms [2]. The second question is one of consumption: does the candidate biofuel burn cleanly, controllably, and efficiently in current and future engine technology? Advanced clean, efficient combustion strategies, particularly those that rely on compression ignition, are often very sensitive to fuel chemistry [4, 5]. A positive answer to both questions is required to make the candidate biofuel practically viable as an alternative to petroleum-based fuels. To address these issues efficiently and comprehensively, coordinated efforts towards biofuel-engine co-development are required.

Figure 7.1 depicts a collaborative biofuel development process currently being spearheaded by Sandia National Laboratories. 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 fuels and mixtures and yield recommendations for the bioengineering scale-up of specific metabolic pathways. This coupling of fundamental and applied combus-

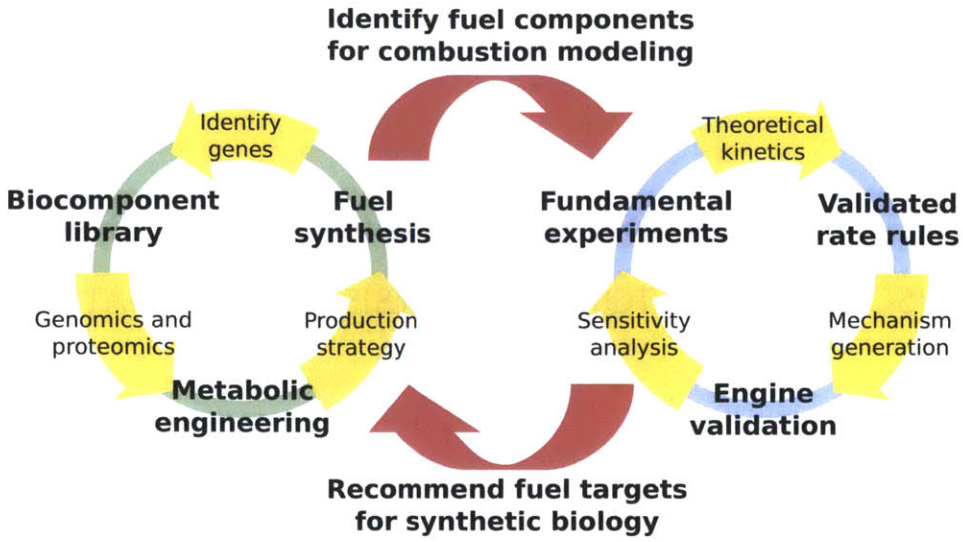


Figure 7.1: A framework for collaborative effort toward lignocellulosic biofuel research. The synthetic biology and combustion modeling development cycles inform and adapt to each other's results.

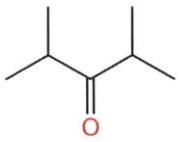
tion chemistry and synthetic biology is a strategy to identify and investigate the most promising fuel compounds through mutual feedback. Moreover, the development of combustion models will provide the predictive capability needed for eventual efficient utilization of the new biofuel stream. A similar framework involving chemical biomass processing is the heart of the Cluster of Excellence "Tailor-Made Fuels from Biomass" founded in 2007 at RWTH Aachen University [6].

Many endophytic fungi directly convert lignocellulosic material into a variety of volatile organic compounds, and harnessing their metabolic processes is a promising platform for cellulosic biofuel generation [7–10]. These fungal pathways are particularly suited for synthetic bioengineering due to their relatively minimal genomes. The natural products of fungal conversion of biomass include a variety of ketones, cyclic ethers and other complex oxygenates. Both fundamental measurements of representative compounds and detailed theoretical kinetics efforts are needed to enable models that can provide feedback on their desirability as fuels.

This chapter presents a detailed kinetics model for the pyrolysis and oxidation of diisopropyl ketone (2,4-dimethylpentan-3-one, denoted DIPK), a representative molecule for branched ketones:

Table 7.1: A summary of experimental conditions used to evaluate the DIPK model.

Experiment	Temperatures	Pressures	Eq. ratios
Flow reactor pyrolysis	800 - 1400 K	0.04 - 1 bar	n/a
Cl-initiated oxidation	550 - 700 K	0.01 bar	0.02
Rapid compression machine	600 - 700 K	10 bar	1



diisopropyl ketone (DIPK)

The central goal of this detailed investigation of DIPK was to derive rules that will enable automatic creation of general, accurate global kinetic models for branched ketone ignition and combustion. Development of the DIPK model also brought together many of the techniques and methodologies discussed in previous chapters, including automatic mechanism generation, parameter estimation, and pressure-dependent kinetics estimation. The DIPK model was evaulated against several experimental measurements obtained by our collaborators. Improvements to the model were made using sensitivity and rate-of-production analysis of areas where the model and experiment disagree to systematically identify parameters in need of refinement. In keeping with the predictive chemical kinetics paradigm, these important parameters were not "fitted" to the data; rather the individual reactions were subjected to an increased level of theoretical analysis or direct experimental measurement. The resulting more accurate parameters were then added to the database of chemistry rate rules to improve future model predictions for both the current fuel molecule and similar molecules.

7.1 Experimental

Several sets of experimental measurements were provided by our collaborators for use in evaluating the automatically-generated DIPK model. A summary of the different experiments is shown in Table 7.1. The three primary experiments span a wide range of oxygen partial pressures (including with no oxygen at all), and so probe very different chemistries. A detailed description of each experimental design is given in the following sections.

7.1.1 Pyrolysis

The pyrolysis of DIPK was investigated at the U10 beamline of the National Synchrotron Radiation Laboratory in Hefei, China [11]. The experimental setup consists of an alumina flow tube with an inner diameter of 7.0 mm mounted in a furnace such that the heated region of the flow tube is 150 mm and the distance between the outlet of the tube and the sampling nozzle is about 10 mm. An inlet stream of 2 vol% DIPK in argon is supplied to the pyrolysis tube at a rate of 1.0 SLM. The product species are sampled using a molecular beam produced by a quartz cone with a $\sim$500 μm (30 Torr) or $\sim$150 μm (760 Torr) orifice at the tip. The molecular beam was passed through a nickel skimmer to a photoionization chamber, where the products were crossed with the synchrotron beam and analyzed using a reflectron time-of-flight mass spectrometer (RTOF-MS). Seven energies of 16.65, 11.70, 11.00, 10.00, 9.45, 9.10, and 8.80 eV were chosen to obtain near threshold ionization. Pyrolysis experiments were run at several temperatures over the ranges of 960 to 1360 K at 30 Torr and 800 to 1200 K at 760 Torr.

DIPK was purchased from Aladdin Reagent Co., Ltd. with a purity of 98%. Argon was purchased from Nanjing Special Gases Factory, Ltd. with a purity of 99.99%.

7.1.2 Cl-initiated oxidation

The products of pulsed-laser chlorine-initiated oxidation of DIPK have been monitored as a function of reaction time, mass, and photoionization energy by using Multiplexed Photoionization Mass Spectrometry (MPIMS) with tunable ionizing radiation provided by the Chemical Dynamics Beamline at the Advanced Light Source (ALS) [12, 13]. The apparatus consists of a heated 1.05 cm inner diameter quartz flow tube with a $\sim$650 μm aperture through which gas escapes. A collimated molecular beam is formed by passing the effluent through a skimmer. Continuous, tunable ionizing radiation provided by the synchrotron intersects the molecular beam orthogonally and ions are accelerated into a linear time-of-flight mass spectrometer. Products are detected as a function of mass, time, and energy, yielding a three-dimentional dataset that can be sliced and integrated accordingly to obtain photoionization spectra and temporal profiles for individual species. Photolysis of Cl_2 by a pulsed excimer laser operating at 351 nm and 10 Hz yields an initial Cl atom concentration of approximately 7×10^{12} molecules/cm^3. Average pre-photolysis background signal is subtracted, resulting in a difference mass spectrum that shows negative signal for the parent compound and its daughter ions. Mass flow controllers (MKS) are used to introduce reactants and carrier gas into the flight tube. A bubbler held at 20 °C provides a steady source of

DIPK. The DIPK has a vapor pressure of 10.4 Torr at 20 °C and an initial concentration of 5.7×10^{13} molecules/cm^3 in the flight tube is estimated at 550 K. Approximately 13% depletion of the parent is observed within 5 ms of photolysis laser firing, followed by a more gradual depletion for the extent of reaction. Pressure in the tube is maintained at 8.0 Torr by feedback control of a butterfly valve in the exhaust line. Initial mole fractions are estimated to be 0.0004 for DIPK, 0.002 for Cl_2, 0.2 for O_2 and the balance He.

7.1.3 Rapid compression machine

The autoignition experiments were conducted for stoichiometric DIPK/O_2/N_2 mixtures in a rapid compression machine (RCM) [14, 15]. The RCM is pneumatically driven and hydraulically controlled with compression times of 25-30 ms. Creviced reactor pistons are used for these experiments to suppress the rollup vortex and provide an adiabatic core regime. Compression ratios were varied to cover a range of compressed gas temperatures between 591-720 K. For all experiments, the compression stroke was fixed to 241 mm, and the compression ratios could be varied between 13 to 31 by varying the clearance through a specially constructed movable end-wall. The initial fill pressures ranged from 0.30 to 0.52 bar, leading to final pressures at the end of compression (EOC) around 10 bar. The reaction chamber was heated to 343 K by electrothermal jackets. Sufficient time was allowed for mixture homogenization and temperature stabilization in the reaction chamber for every new experimental measurement and N_2 was the only diluent used for all measurements. Pressure-time profiles were measured using a dynamic pressure transducer (Kistler 603B) from the start of compression until about 200 ms after the end of compression. The ignition delay time was defined as the time from the end of compression to the maximum rate of pressure rise during ignition.

DIPK was obtained from Sigma Aldrich at 98% purity, while all high purity gases were supplied by Westfalen AG (99.95% N_2) and (99.995% O_2) Praxair which were used without further purification. Test mixtures were prepared manometrically in the combustion chamber. Care was taken to ensure homogeneous mixing of fuel and gases, which was monitored by measuring the pressure with STS 105112 and STS 105107 pressure transmitters with total error 0.8% of full scale.

From the experimental pressure trace and initial experimental conditions, e.g. initial temperature and initial pressure, the compressed gas temperature, T_c, was calculated using the relation [16]

$$\int_{T_0}^{T_c} \frac{\gamma}{\gamma - 1} \frac{dT}{T} = \ln\left(\frac{P_c}{P_0}\right) \tag{7.1}$$

where T_c and P_c are temperature and pressure at top dead center (TDC), T_0 and P_0 are initial temperature and pressure before compression begins, and γ is the ratio of specific heat at constant pressure to specific heat at constant volume.

7.2 Computational methods

7.2.1 Mechanism development

The DIPK model was generated automatically using the Reaction Mechanism Generator (RMG) software [17]. The model was built using the `DFT_QCI_thermo` [18] and `CBS_QB3_1dHR` [19–21] thermodynamics libraries distributed with RMG. The `KlippensteinH202` [22] and `Marinov` [23] libraries were used as seed mechanisms, along with newly-created seed mechanisms for the propargyl [24], C_3H_4 (this work), C_3H_5 [25], and C_3H_7 (this work) pressure-dependent reaction networks.

For networks other than the ones given above, the group frequency method was used to approximate the vibrational frequencies of each species, and the modified strong collision method was used for subsequent pressure dependence calculations [26]. Only reaction networks with isomers comprised of 16 or fewer atoms were treated as potentially pressure-dependent; all larger systems were assumed to be in the high-pressure limit. The final RMG-generated model comprised 172 species and 4396 reactions.

7.2.2 Quantum chemistry calculations

Thermodynamics and kinetics of sensitive species and reactions were calculated using quantum chemistry calculations at the CBS-QB3 level [27, 28]. One-dimensional hindered rotors were included where appropriate; the potentials for these rotors as a function of torsion angle were determined via relaxed scans at the B3LYP/CBSB7 level. All quantum chemistry calculations were performed using Gaussian 03 [29]. Rate coefficients were computed in CanTherm using conventional transition state theory with one-dimensional Eckart tunneling correction.

7.2.3 Reactor models

Simulations of each experiment using the RMG-generated model were performed in CHEMKIN-PRO [30]. The pyrolysis flow reactor was modeled as a plug flow reactor with the same dimensions and flow rates as given in the experimental section.

The simulation also used axial temperature profiles provided by the experimental collaborators. The simulated species concentrations at the reactor outlet were used to compare with the experimental measurements.

The Cl-initiated oxidation experiments were simulated using a homogeneous, isothermal, isobaric batch reactor. Estimated rates for hydrogen abstraction from DIPK by Cl were used for the initiation; with these rates, the simulation matched the experimental observation of approximately 13% depletion of DIPK within 5 ms. The species concentrations after 20 ms were used to determine the product distributions.

The rapid compression machine was simulated as a homogeneous batch reactor at constant volume. Only the portion of the experiment immediately after the compression stroke was simulated; this allowed us to initialize the simulation at the estimated temperature T_c and pressure P_c at top dead center. Some variable volume simulations including the compression stroke were also run to confirm that the estimated T_c and P_c were reasonable.

7.3 Results and discussion

7.3.1 Pyrolysis

Ignition in real systems is a combination of pyrolytic breakdown of fuels and radical chain oxidation. Predictions of the RMG-generated model for DIPK were compared to pyrolysis experiments at 30 Torr and 760 Torr conducted at the National Synchrotron Radiation Laboratory in Hefei, China. Figures 7.2 and 7.3 show plots of the experimentally observed and simulated mole fractions for the major species observed at both pressures. The predicted DIPK consumption matches the experiment closely except at the highest temperatures, where the model DIPK consumption becomes too fast. At all temperatures considered the dominant removal of DIPK is via hydrogen abstraction by the H radical to form either the tertiary or primary radical; the branching ratio (see Figure 7.12) is strongly in favor of the tertiary radical at lower temperature, but becomes more balanced at higher temperature. Fission of the C–C bond adjacent to the carbonyl group becomes a minor but significant pathway at the temperatures where the predicted DIPK consumption is too fast.

The model predicts the final mole fractions of the major pyrolysis products to within a factor of two except at the highest temperatures at 30 Torr. Above 1200 K at 30 Torr, the model predicts significantly less propene and significantly more methane and acetylene than observed experimentally. A sensitivity analysis of these products indicates the pressure-dependent chemically activated reaction of H + propene is

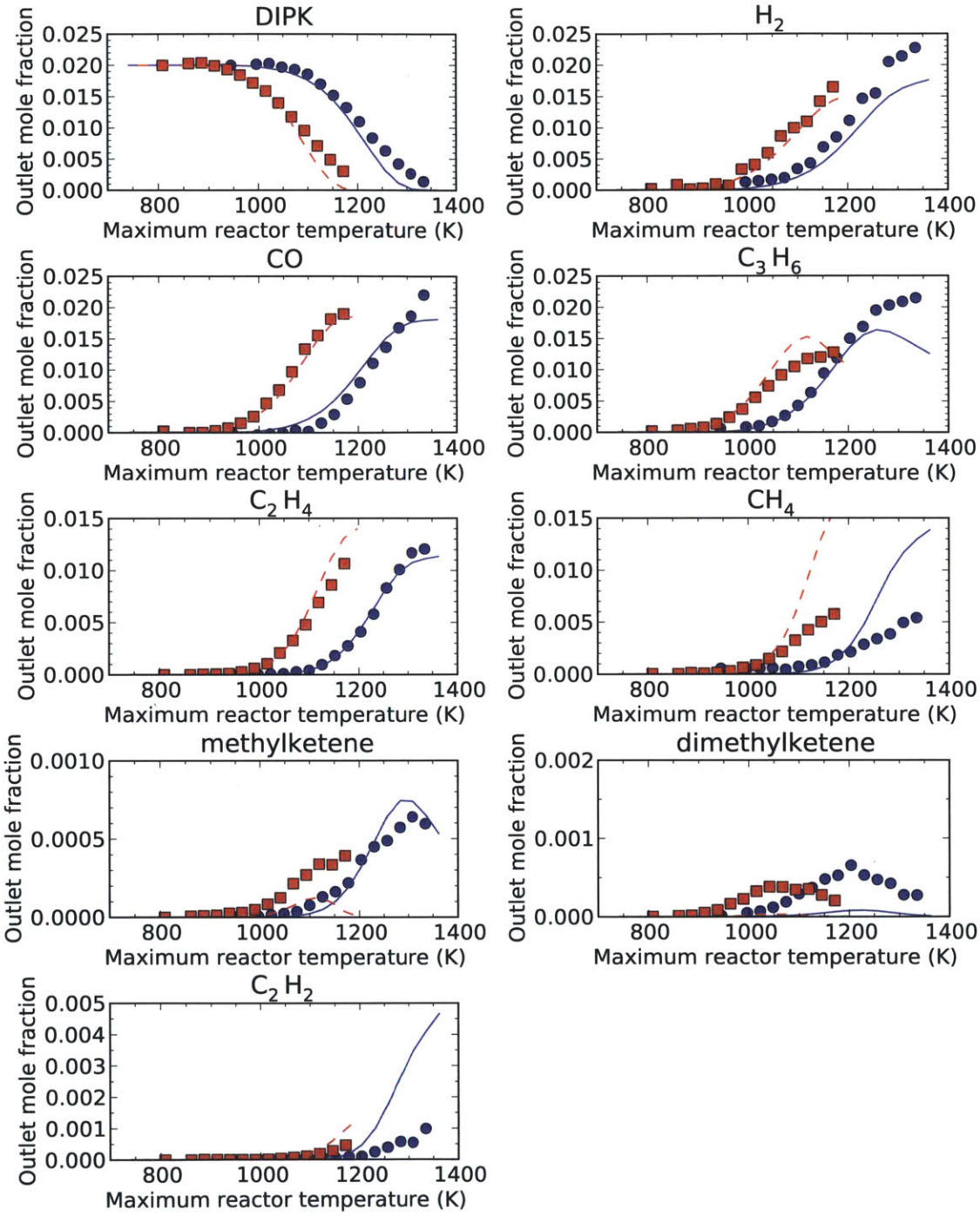


Figure 7.2: Plots of experimentally observed (markers) and modeled (lines) mole fractions for selected major products of DIPK pyrolysis at 30 Torr (blue) and 760 Torr (red). The model matches the experiment well except at the highest temperatures, which show too little C_3H_6 and too much CH_4. This suggests that the model is predicting subsequent pyrolysis of the C_3H_6 that the experiment does not show.

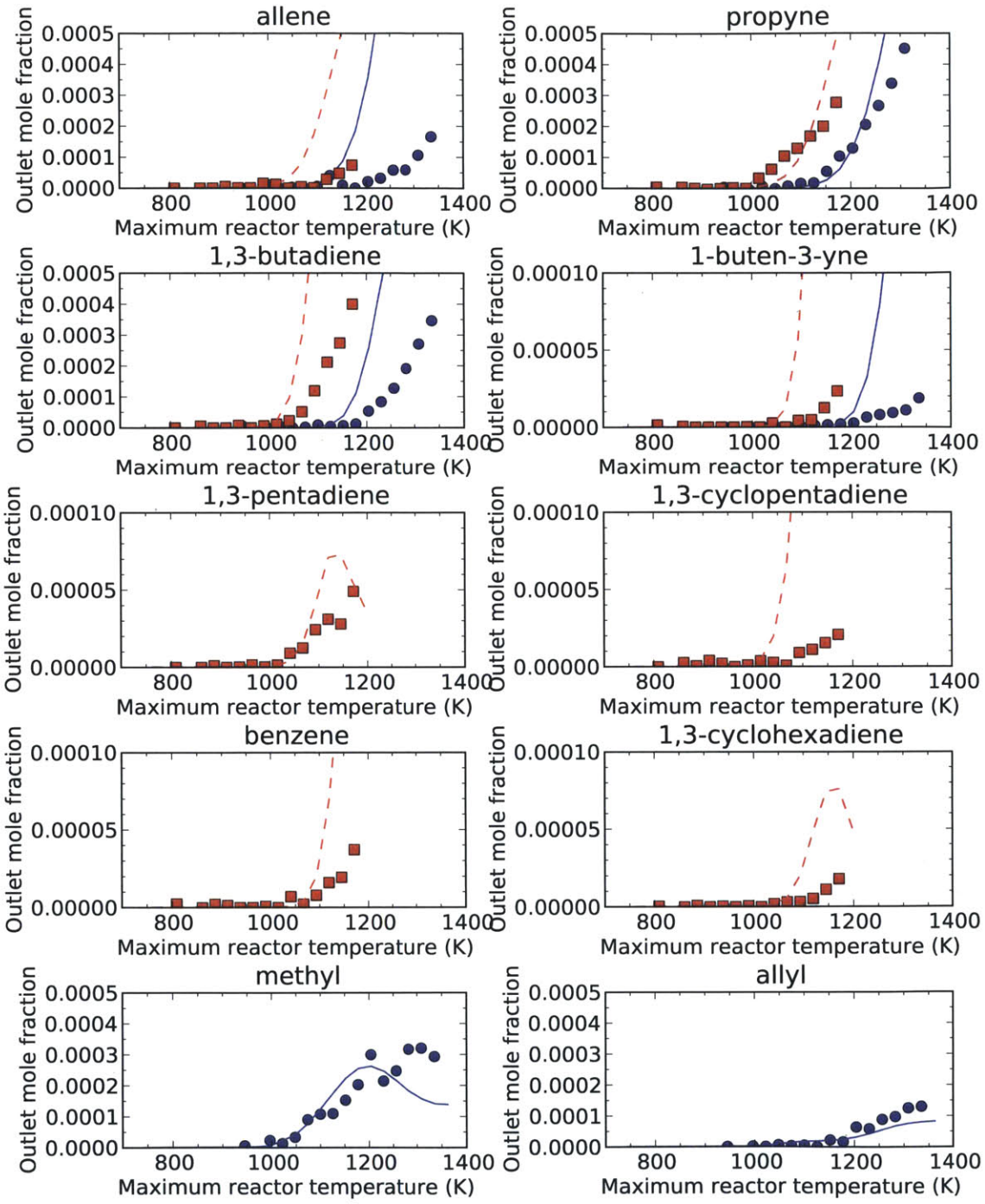


Figure 7.3: Plots of experimentally observed (markers) and modeled (lines) mole fractions for selected minor products of DIPK pyrolysis at 30 Torr (blue) and 760 Torr (red). Many of the minor products are significantly overpredicted at high temperature, likely from subsequent pyrolysis of C_3H_6.

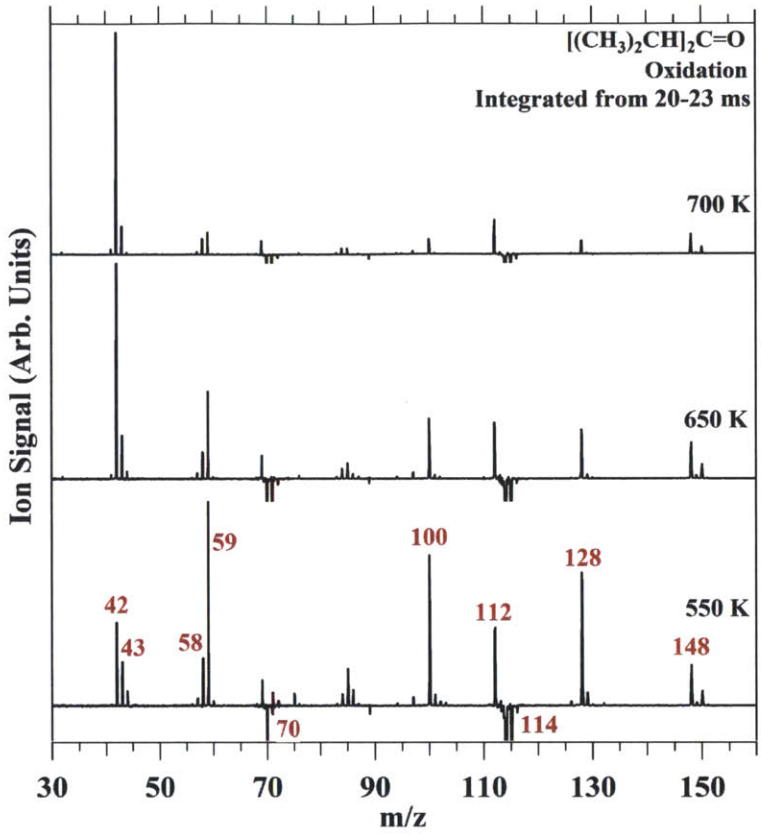


Figure 7.4: Experimental product mass spectra obtained from Cl-initiated oxidation of DIPK (m/z 114) at 550, 650, and 700 K and 8 Torr. The major peaks at 550 K are m/z 42 (C_3H_6), 58 (acetone and propylene oxide), 112 ($C_7H_{12}O$), and 128 (cyclic ether). The peaks at m/z 59 and 100 are daughter ions of the cyclic ether, while m/z 148 likely represents chlorination of a DIPK radical.

particularly important. This suggests that the model is predicting secondary pyrolysis of propene that is not observed experimentally. This may also explain many of the discrepancies in the minor products – particularly for allene, propyne, 1,3-butadiene, 1-buten-3-yne, 1,3-cyclopentadiene, and benzene – which are consistently overpredicted at the same conditions at which propene is underpredicted. The current model includes updated pressure-dependent rates for the H + propene reaction network (discussed in Section 7.3.6) from quantum chemistry calculations at the CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S level, which show some improvement but leave a significant discrepancy. The high level of sensitivity of these reactions may warrant an even more thorough quantum chemistry treatment.

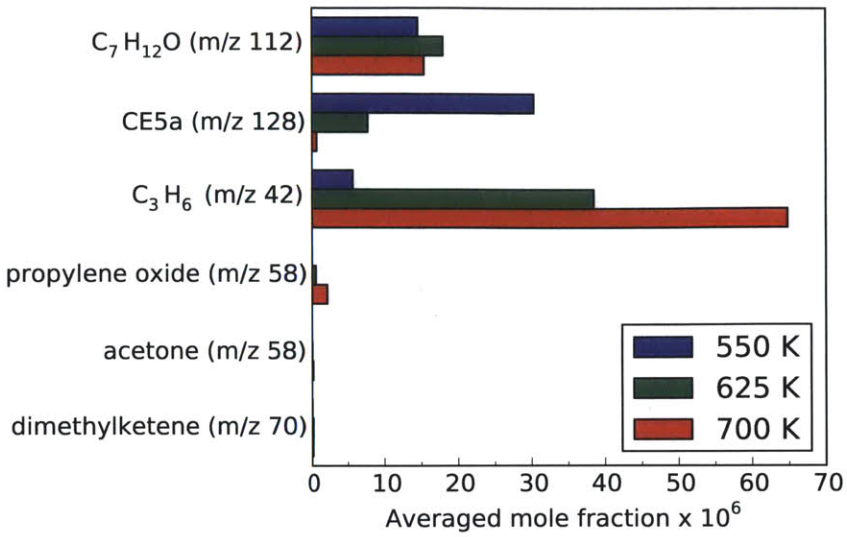


Figure 7.5: RMG-predicted product distributions for the simulated Cl-initiated oxidation of DIPK at 8 Torr. The numbers of the cyclic ethers correspond to those in Table 1. As temperature increases, the cyclic ether formation pathways are turned off in favor of the β-scission pathways to C_3H_6, as experimentally observed. The model predicts very low formation of m/z 58 (acetone and propylene oxide) or 70 (dimethylketene).

7.3.2 Cl-initiated oxidation

The initial steps in low-temperature DIPK oxidation were probed using pulsed-laser chlorine-initiated oxidation experiments employing time-resolved tunable synchrotron Multiplexed Photoionization Mass Spectrometry (MPIMS) at the Advanced Light Source. Figure 7.4 shows a series of mass spectra obtained from experiments at 550, 650, and 700 K and 8 Torr. Individual isomeric compounds can be identified from their photoionization spectra [31, 32].

The spectrum at 550 K shows a variety of products, which have been identified as propene (m/z 42), acetaldehyde (m/z 44), acetone and propylene oxide (m/z 58), 2,4-dimethyl-1-penten-3-one (hereafter $C_7H_{12}O$) (m/z 112), DIPK (m/z 114), and one or more cyclic ethers (m/z 128). The photoionization curves of the peaks at m/z 59 and 100 show the same characteristics as the curve at m/z 128, indicating that these peaks represent daughter ions of the cyclic ethers. The peaks at m/z 148 and 150 have a 3:1 ratio characertistic of the ^{35}Cl:^{37}Cl isotopic ratio, suggesting that these represent chlorination products. As the temperature increases to 650 and 700 K, the propene peak at m/z 42 becomes the dominant peak, while the cyclic ether peak at m/z 128 disappears, along with the daughter ion peaks at m/z 59 and 100.

A diagram of many of the pathways in the DIPK mechanism relevant under the

conditions of the ALS experiments is shown in Figure 7.6. Oxidation is initiated by abstraction of a hydrogen atom from DIPK, yielding either a tertiary or primary fuel radical (labeled R_{tert} and R_{pri}, respectively). On a per-site basis, formation of the tertiary radical is favored due to its relative stability compared to the primary radical [33]. However, there are six times as many primary hydrogens as tertiary, and both initial radicals contribute to the subsequent oxidation chemistry. The radicals R_{tert} and R_{pri} can decompose, rearrange, or react with O_2 via a peroxy radical ($R_{tert}OO$ and $R_{pri}OO$). The fate of R_{tert} and R_{pri} is determined by the system temperature. At low temperature the $R + O_2 \rightleftharpoons ROO$ equilibrium will lie toward the peroxide radicals ROO, and the peroxy chemistry pathways discussed in the next paragraph dominate. At high temperatures the equilibrium shifts to the alkyl radicals R, and the alkyl radical decomposition pathways become dominant. In the DIPK system, the most important decomposition pathway is the β-scission of R_{pri} to produce propene.

The fate of peroxy radicals is critical for determining autoignition properties. One particularly important ROO reaction channel is isomerization via internal H-atom transfer to hydroperoxyalkyl radical (QOOH) isomers, which can eliminate OH and form cyclic ethers. The OH radical is a highly reactive, chain-propagating species, and its QOOH precursor is responsible for low-temperature chain branching [34]. The cyclic ethers observed in the ALS experiments are markers of QOOH species and therefore indicate likelihood of low- and intermediate-temperature heat release, which is advantageous for high-load operation in advanced engine strategies [35, 36]. Another important reaction path for either ROO intermediate is elimination of an unreactive HO_2 radical to form an unsaturated, stable coproduct ($C_7H_{12}O$, m/z = 112), an effectively chain-terminating pathway at low temperature [34].

Figure 7.5 show the predicted product distributions at 550, 625, and 700 K and 8 Torr using the RMG-generated model to simulate the ALS experiment. As observed in the experiment, the model predicts the cyclic ether formation pathways to turn off in favor of the β-scission pathway to propene as temperature increases. The HO_2 elimination pathways to $C_7H_{12}O$ remain mostly unchanged over this temperature range. Interestingly, the model does not predict a significant production of either acetone or propylene oxide; it also predicts that significantly more propylene oxide is formed, while the experimental photoionization curves suggest that the two m/z 58 species are formed in close to a 1:1 ratio. The model also predicts a minor but significant amount of isopropyl hydroperoxide at m/z 76 which is not seen in the experiment. Overall, however, the model and experiment are in very good agreement with the major features.

Excluding unimolecular rearrangement of the initial DIPK radicals, seven possible QOOH species can lead to five distinct cyclic ether species, as shown in Figure 7.6. The experimental photoionization curve at m/z 128 has an onset threshold of 8.9 eV. Ab initio calculations at the CBS-QB3 level suggest that the four-membered-ring and five-membered-ring cyclic ethers (CE4 and CE5, respectively) are closest to this ionization threshold, at 8.9 and 8.8 eV, respectively.

The RMG-generated model predicts that the dominant cyclic ether observed under the ALS experimental conditions is the five-membered-ring cyclic ether CE5a, which is formed via the mechanism shown in Figure 7.7. This mechanism involves a 1,2-acyl migration that proceeds through an unstable three-membered-ring intermediate. At the lower O_2 levels of the ALS experiments, this pathway is much faster than the O_2 addition pathway, and so the peroxy chemistry proceeds through the secondary radical R_{sec}. O_2 addition to R_{sec} leads to cyclic ether CE5a via intra-H migration to form $Q_{st}OOH$, followed by ring-closing OH elimination.

7.3.3 Rapid compression machine

Finally, the RMG model developed based on the combined pyrolysis and Cl-initiated oxidation experiments was used to study the autoignition of DIPK. Experimental measurements of the ignition delay were provided from rapid compression machine (RCM) experiments conducted at Aachen University. A plot of the experimental data, along with the corresponding predictions from the RMG-generated model, is shown in Figure 7.8. The observed ignition delay is substantially larger than those for structurally similar hydrocarbons (*e.g.* 2,4-dimethylpentane), reflecting the lower reactivity of the resonance-stabilized tertiary DIPK radical R_{tert}. Moreover, the experiments show an exceptionally prominent region of negative temperature coefficient (NTC) behavior – where the reactivity decreases with increasing temperature – from 650 to 750 K.

Both the long ignition delay and the prominent NTC region are predicted by the RMG model, although the NTC region is shifted by about 100 K in temperature. Mechanistically, NTC occurs because the shift of the $R + O_2 \rightleftharpoons RO_2$ equilibrium towards reactants causes an increase in the chain-terminating decomposition pathways relative to the chain-branching and chain-propagating pathways from peroxy radical chemistry. For DIPK, the depth of the NTC feature is very sensitive to the branching ratio of DIPK + OH to primary vs. tertiary radicals. As temperature increases, the branching ratio shifts toward making more primary radical, which increasingly decomposes via β-scission rather than following the peroxy pathways. The β-scission pathway enhances the NTC region.

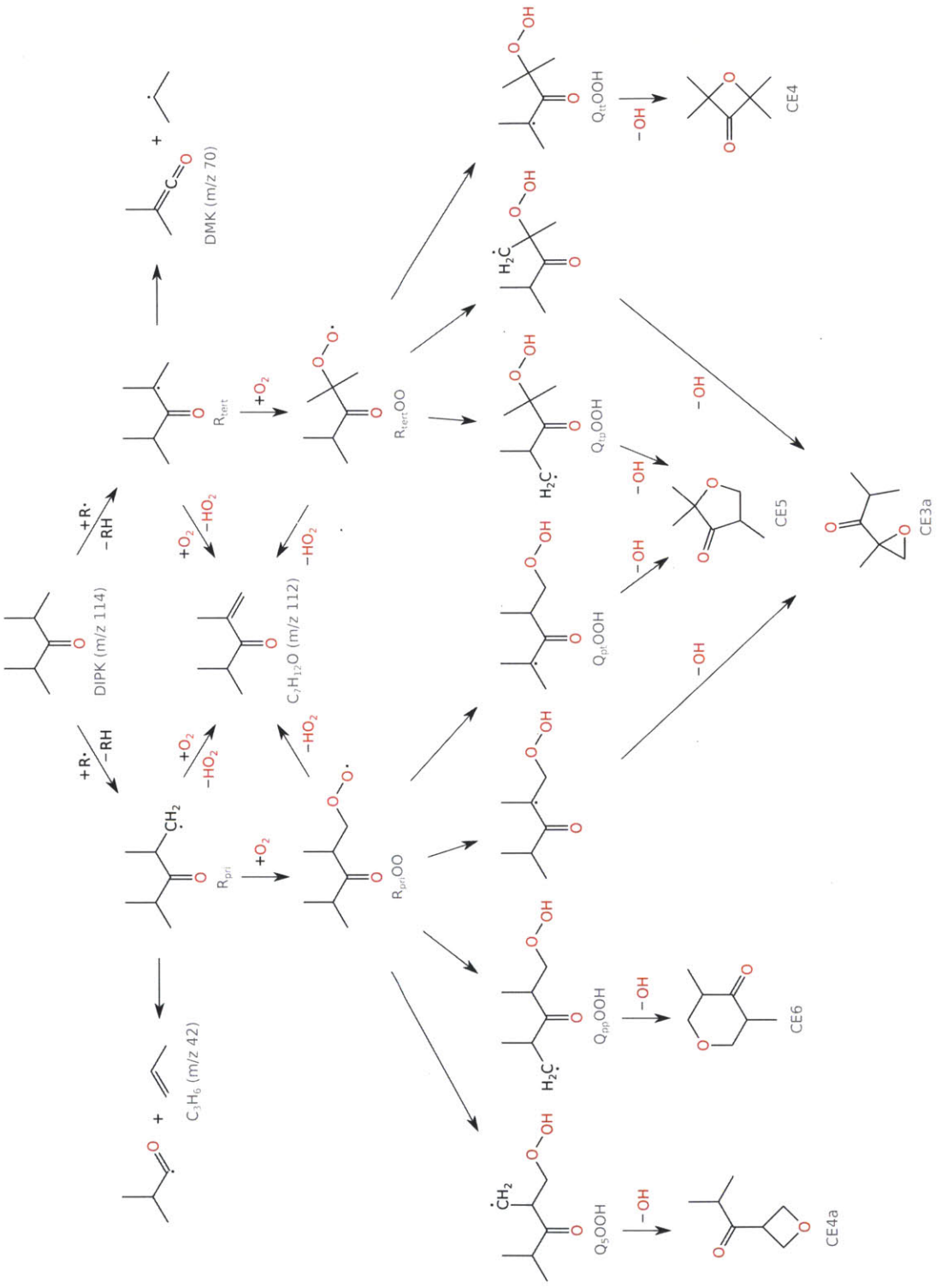


Figure 7.6: Low-temperature peroxy chemistry from the tertiary and primary radicals formed via H-abstraction from DIPK. The two RO_2 radicals lead to seven QOOH radicals, which in turn lead to five cyclic ethers. The competing β-scission and alkene formation pathways are also shown. The names (and m/z values if appropriate) of relevant species discussed in the text are given.

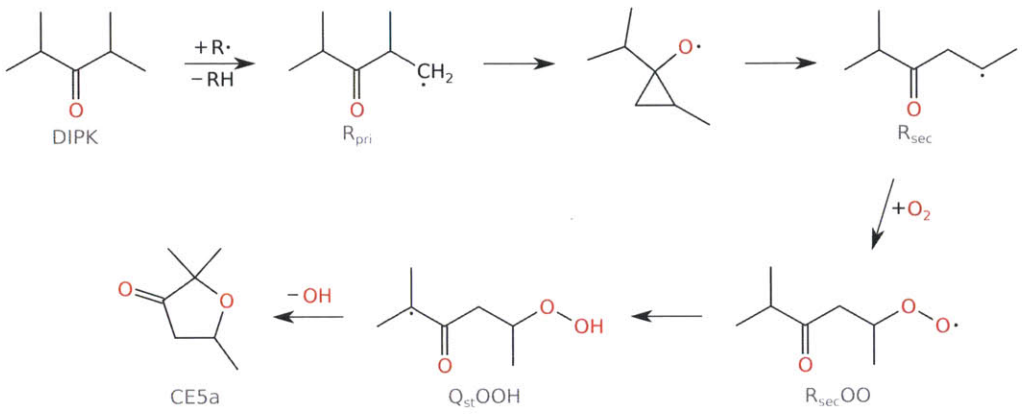


Figure 7.7: Cyclic ether formation pathway via 1,2-acyl migration from primary radical formed via H-abstraction from DIPK. This is the major cyclic ether formation pathway predicted by the RMG-generated model under the conditions of the ALS experiments.

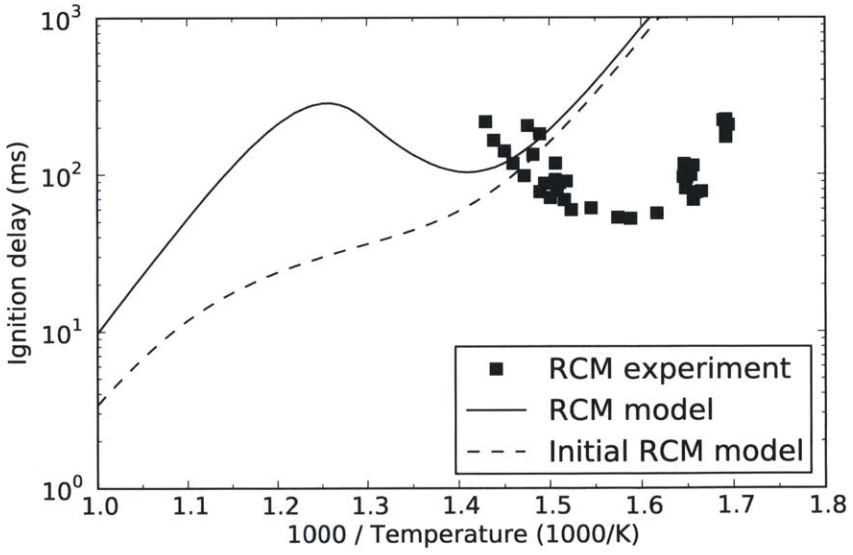


Figure 7.8: Plot of the experimentally measured and simulated ignition delay times from DIPK oxidation in a rapid compression machine (RCM). Both the experiment and the model show a significant region of negative temperature coefficient behavior, although the model is shifted by about 80 K in temperature.

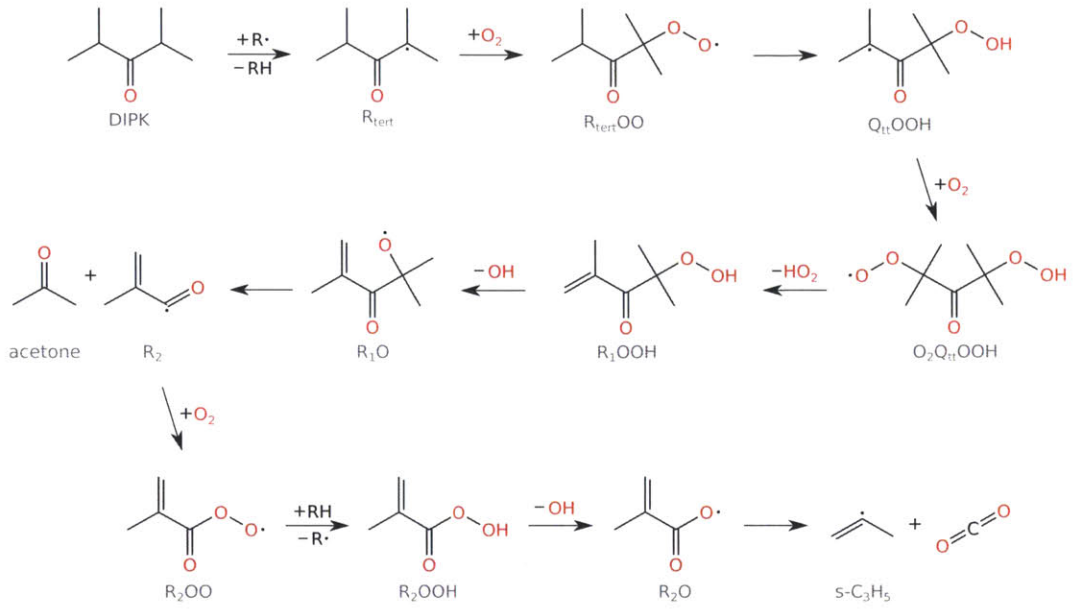


Figure 7.9: An important low-temperature peroxy pathway in DIPK oxidation at high pressure. At the RCM conditions, there is such a high concentration of oxygen that, at low temperature, it readily adds a second time to the QOOH radical instead of eliminating to a cyclic ether. This quickly leads to a peroxide, labeled R_1OOH, that serves as a bottleneck to OH radical production.

One mechanism predicted by the RMG model to be especially important at the lower temperatures (below 700 K) of the RCM conditions is shown in Figure 7.9. At these temperatures, the DIPK hydrogen abstraction branching ratio strongly favors formation of the resonance-stabilized tertiary alkyl radical R_{tert}. The $R + O_2 \rightleftharpoons RO_2$ equilibrium lies well to the products, so the peroxy chemistry is important. In fact, the O_2 concentration is so high in the RCM system that, after $R_{tert}OO$ isomerizes to $Q_{tt}OOH$ – itself a resonance-stabilized alkyl radical – a second O_2 readily adds to form $O_2Q_{tt}OOH$. The hydrogen atoms on $O_2Q_{tt}OOH$ are not as easily accessible for isomerization, so instead the second peroxy eliminates as HO_2 to form the peroxide R_1OOH, which acts as a bottleneck for the subsequent chemistry. Nonetheless, the O−O bond in R_1OOH is relatively weak, so the peroxide falls apart to R_1O and OH. R_1O then has a fast β-scission pathway to form acetone and a C_4H_5O radical, hereafter called R_2. Another sequence of peroxy chemistry leads to a decomposition to form $s-C_3H_5$ and CO_2. Many of the reactions in this sequence appear as sensitive reactions in the RMG-generated model at 650 K.

A sensitivity analysis of the DIPK RCM model predicted ignition delay time τ_{ign}

to the kinetic parameters was conducted to better understand which reactions were controlling the behavior in the different regimes. The most sensitive reactions at temperatures of 650 K (below the NTC region), 750 K (within the NTC region), and 850 (above the NTC region) are shown in Figure 7.10. The sensitivity coefficient plotted is the direct sensitivity of the ignition delay to the rate coefficient, in the form

$$S_i(T) \equiv \frac{\partial \ln \tau_{\text{ign}}(T)}{\partial \ln k_i(T)} \tag{7.2}$$

At 650 K we are in a low-temperature regime, and the major sensitive reactions correspond to reactions in the sequence shown in Figure 7.9. At these temperatures, a reaction of DIPK with OH is much more likely to produce a tertiary radical R_{tert} instead of a primary radical R_{pri}. The $R_{tert} + O_2 \longrightarrow R_{tert}OO$ equilibrium favors the formation of the peroxy radical, and so the reactions of the $R_{tert}OO$ peroxy pathways appear as sensitive. Interestingly, the most sensitive reaction at this temperature is the O−O bond cleavage of the species R_1OOH; this reaction is acting as a bottleneck to the subsequent chemistry, so its rate has a very strong effect on the observed ignition delay.

At 750 K we have entered the NTC region (as predicted by the model). Not surprisingly, there are many more sensitive reactions than at the other conditions. The increase in temperature has caused both a shifting of the $R_{tert} + O_2 \longrightarrow R_{tert}OO$ equilibrium toward R_{tert} and a shifting of the DIPK + OH branching ratio toward R_{pri}. The $R_{pri} + O_2 \longrightarrow R_{pri}OO$ continues to favor the products, and so many of the reactions in the $R_{pri}OO$ peroxy pathways appear as important. Some of the high-temperature pathways involving H_2O_2 are already appearing as important, further complicating the picture.

At 850 K we have entered a high-temperature regime. Both the $R_{tert}OO$ and $R_{pri}OO$ peroxy pathways have been turned off by the equilibria of the $R + O_2 \longrightarrow ROO$ reactions. Instead, the sensitive reactions are now those that involve production of H_2O_2, which leads to OH formation via $H_2O_2 \longrightarrow 2\,OH$, the most sensitive reaction at this temperature.

The prediction of strong NTC behavior is a significant success for the automatically-generated RMG model, and a validation of the predictive kinetics paradigm. Initial RMG models, before the explicit determination of rate rules for the initial oxidation steps of ketones, showed only a weak inflection in the NTC region. The dramatic improvement in the NTC prediction was brought about by calculation of important reaction kinetics and equilibria that were identified by sensitivity analysis and comparison

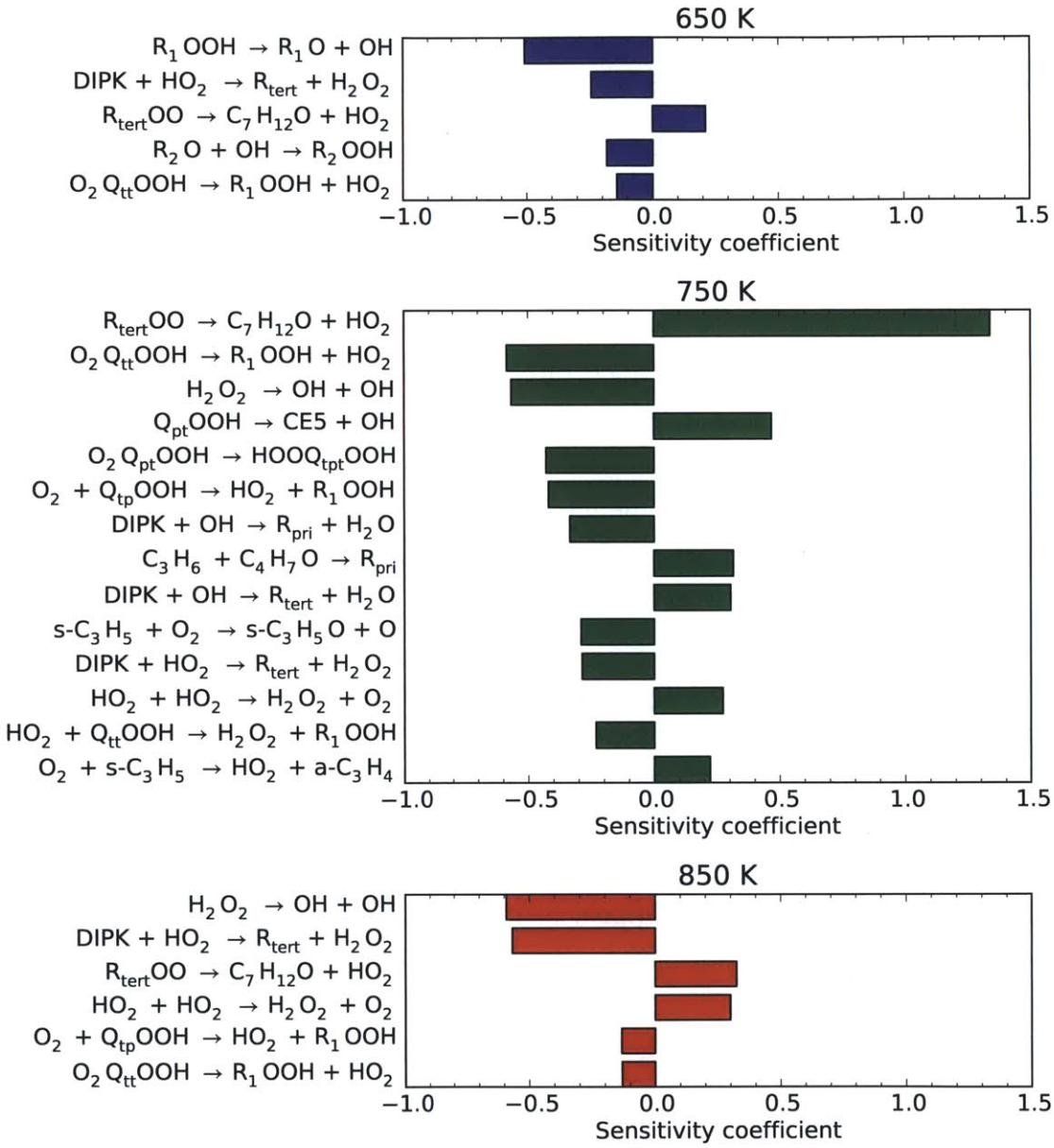


Figure 7.10: Sensitivity analysis of the ignition delay to kinetics for the DIPK RCM model at several temperatures. The value plotted is the sensitivity of the ignition delay to the rate coefficient, in the form $S \equiv \partial \ln \tau_{\text{ign}} / \partial \ln k$. Based on the model predictions, the 650 K temperature represents a point before the NTC, the 750 K temperature occurs within the NTC, and the 850 K temperature exists above the NTC region.

to experiment. These included branching ratios for the initial abstraction from DIPK, equilibrium constants of both DIPK radical reactions with O_2, and the β-scission reactions of both DIPK radicals. These improvements will be discussed in the subsequent sections.

7.3.4 Updated thermochemistry

Thermodynamic parameters for several species were updated during the iterative construction of the DIPK model. In particular, the thermochemistry of the tertiary and primary alkyl and peroxy radicals were refined due to the importance of correctly predicting the equilibria of the $R + O_2 \rightleftharpoons ROO$ reactions that initiate the peroxy chemistry. Quantum chemistry calculations at the CBS-QB3 level, with one-dimensional hindered rotor treatment at the B3LYP/CBSB7 level, were used to refine the thermodynamic parameters for these species, as well as for the parent molecule DIPK. Note that the two internal C–C bonds adjacent to the carbonyl are in fact highly coupled torsions due to the steric interactions of the isopropyl groups. A very accurate calculation of the thermochemistry of these molecules, which was beyond the scope of the present work, may require treating these more rigorously than the one-dimensional approximation allows.

The equilibrium of a $R + O_2 \rightleftharpoons ROO$ reaction can be expressed as the ceiling temperature T_{ceil}, which is the temperature at which the ratio of the alkyl and peroxy radicals at equilibrium are equal:

$$K_c(T_{ceil})[O_2] = 1 = \frac{[ROO]}{[R]} \tag{7.3}$$

Below this temperature, the equilibrium lies toward the products, and peroxy radical ROO will be abundant. Above this temperature, the equilibrium lies toward the reactants, and alkyl radical R will be abundant. A plot of the ceiling temperature as a function of oxygen concentration $[O_2]$ using the CBS-QB3 computed values is shown in Figure 7.11 for both the tertiary and primary DIPK radicals. The tertiary $R + O_2 \rightleftharpoons ROO$ reaction reaches the ceiling temperature at a lower temperature than the primary radical at all oxygen concentrations. The RCM experiments were all run well below the ceiling temperatures of both $R + O_2 \rightleftharpoons ROO$ reactions. Note that the onset of the RMG-predicted NTC region occurs at around 700 K, which is very near the ceiling temperature of $R_{tert} + O_2 \rightleftharpoons R_{tert}OO$. The ALS experimental conditions reflect a range from just below the ceiling temperatures of both reactions to just above; accordingly, the experiments and the model show a turning off of the peroxy chemistry over this range.

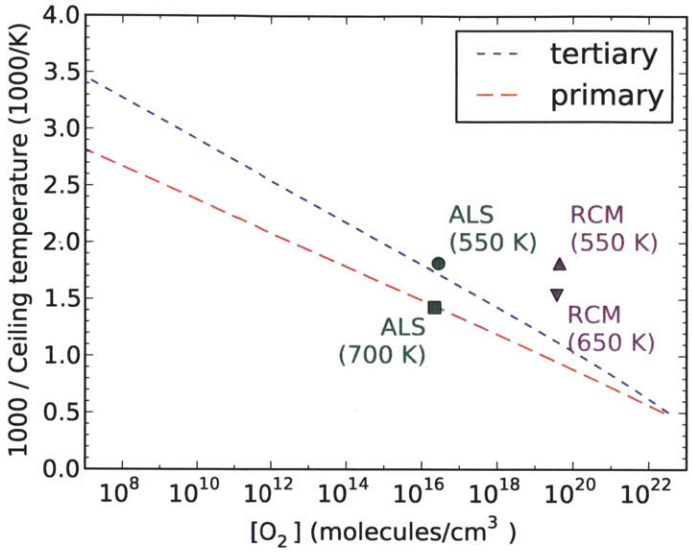


Figure 7.11: Plot of the ceiling temperatures as a function of oxygen concentration for the peroxy formation reactions $R + O_2 \rightleftharpoons RO_2$ involving the tertiary and primary DIPK radicals. The ceiling temperature denotes the point where the ratio of peroxy and alkyl radicals is unity. Below this temperature, the equilibrium lies toward peroxy radicals; above this temperature, the equilibrium lies toward alkyl radicals. The coordinates of the ALS and RCM experiments discussed in this work are also shown.

A new thermodynamics group `C2CJCO` was also added to RMG to represent a tertiary radical adjacent to a ketone carbonyl. The group additivity parameters for this group were set to match those for a tertiary radical adjacent to an aldehyde carbonyl, which was already present as the `C2CJCHO` group. The definition of `C2CJCO` actually includes both ketones and aldehydes, so `C2CJCHO` was placed as a child of the new `C2CJCO` group in the tree. Before this change, RMG would use the generic tertiary radical correction `Cs_T` instead of the `C2CJCHO` values. The new `C2CJCHO` group caused the enthalpy of formation at 298 K of all affected molecules, including R_{tert}, $Q_{tt}OOH$, and $Q_{pt}OOH$, to decrease by 6.7 kcal/mol, a much better reflection of the stability imparted by resonance with the adjacent carbonyl.

7.3.5 Updated rate coefficients

Many reaction rate coefficients were also refined using quantum chemistry calculations or updated literature values during the course of constructing the DIPK model. Particular attention was paid to the DIPK hydrogen abstraction branching ratios, as the model was often very sensitive to the amount of tertiary and primary radicals produced. In the pyrolysis experiments, H (major) and CH_3 (minor) are the important

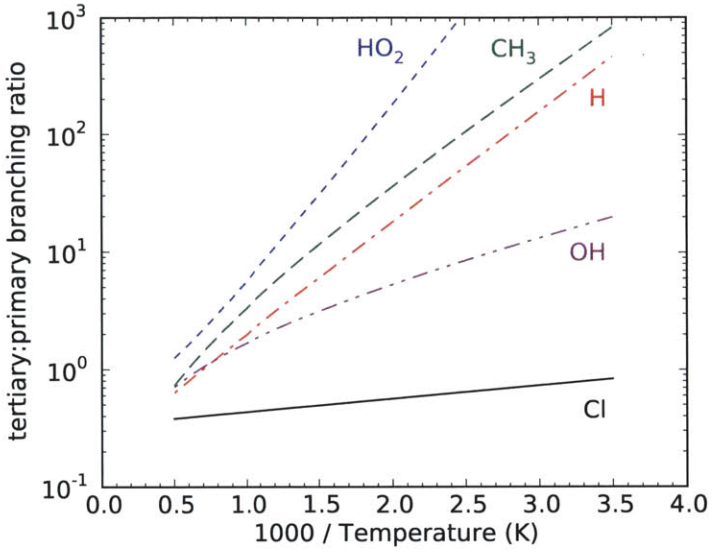


Figure 7.12: Plot of the DIPK hydrogen abstraction branching ratios for several radicals, based on the DIPK model parameters. Cl is significantly less selective than OH, which in turn is significantly less selective than H, CH_3, and HO_2. These branching ratios include reaction-path degeneracy, which means that a completely unselective radical would have a branching ratio of 1:6.

radicals. In the oxidation experiments, OH and HO_2 are the important radicals (as is Cl in the ALS experiments). A plot of the computed branching ratios as a function of temperature for each of these radicals is shown in Figure 7.12. (The figure includes the reaction-path degeneracy, so a completely nonselective radical would have a branching ratio of 1/6.) The plot clearly shows the sharp decrease in selectivity from HO_2 to CH_3 to H to OH to Cl. Cl radical is so nonselective that, due to the high degeneracy of primary hydrogen atoms on DIPK, DIPK + Cl hydrogen abstraction overall forms more R_{pri} than R_{tert} over the entirety of the range 300-2000 K.

Tables 7.2 and 7.3 provide the kinetics for many of the important reactions updated while building the DIPK model. Table 7.2 focuses on the non-peroxy chemistry (the DIPK hydrogen abstraction and subsequent β-scission reactions), while Table 7.3 reflects the peroxy reactions. The DIPK + OH hydrogen abstraction reaction kinetics were taken from Zhou *et al.* [33] while all of the peroxy rates were adapted from the rate rules of Miyoshi [37]. The other hydrogen abstraction and β-scission reactions were calculated by us at the CBS-QB3 level, with several reactions also getting one-dimensional hindered rotor treatments. Both tables compare the updated kinetics to the original RMG-estimated kinetics, giving values of the (high-pressure limit) rate coefficient at several temperatures for easy comparison.

Although we generally expect RMG-estimated rate coefficients to be around order-of-magnitude accurate, the tables show several examples for which this is not the case. The largest disagreements, some of which span several orders of magnitude, correspond to cases where RMG does not have an exact rate rule that matches the reaction, and instead is constructing an approximate rate rule by averaging together rules for the most similar reactions it can find. If the reaction is sensitive, this can cause a dramatic change in the quality of the model predictions. It can also cause RMG to both miss important chemistry or include additional chemistry that ought to be not important. An example of the latter is the intra-H migration of R_{tert} to form a $C_7H_{13}O_3$ radical with an enol group, which RMG originally predicted to be very significant before an improved rate estimate caused this pathway to become negligible. Thus, for automatic mechanism generation software such as RMG, the database of chemical parameters is just as important as the mechanism generation algorithms to obtaining a reasonable detailed kinetics model.

7.3.6 The C_3H_7 potential energy surface

Given that DIPK has two isopropyl groups, it is perhaps not surprising that its high-temperature pyrolysis and oxidation yields a significant amount of propene. Several sensitive reactions under pyrolysis conditions involve the reaction of propene with hydrogen radicals, which can either abstract a hydrogen or add across the double bond to form chemically-activated *n*-propyl or isopropyl radicals. The activated *n*-propyl can then undergo β-scission to form ethylene and methyl radical. We investigated this system in detail to attempt to address the deviation in propene concentration seen between the RMG model and the pyrolysis experiments.

Quantum chemistry calculations were performed at the CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S level to characterize the local minima and transition states for the hydrogen abstraction and chemically activated reactions of H + propene. One- dimensional hindered rotors were included, with scans performed at the B3LYP/MG3S level. The resulting potential energy surface for the chemically- activated reaction network is shown in Figure 7.13. High-pressure limit rate coefficients were calculated using conventional transition state theory with Eckart tunneling; the computed reaction rates are shown in Table 7.4. Plotting the high-pressure limit kinetics for the H + propene reactions in Figure 7.14 shows that the addition reactions are favored over the abstraction reactions over the entire temperature range considered.

The pressure-dependent rate coefficients were calculated from the quantum chemistry results using the RRKM/ME analysis presented in the previous chapter. The

Table 7.2: Calculated reaction rates for DIPK non-peroxy chemistry updated in this work.

Reaction / Source	A[a]	n	E_a	$k(300\ K)$[a]	$k(600\ K)$[a]	$k(1000\ K)$[a]
DIPK + OH $\longrightarrow$ R_{tert} + H_2O						
RMG rate rule	5.3×10^{9}	0.95	0.57	4.6×10^{11}	1.4×10^{12}	2.8×10^{12}
Zhou *et al.*[b]	9.8×10^{3}	2.70	-4.04	4.2×10^{13}	9.2×10^{12}	9.5×10^{12}
DIPK + OH $\longrightarrow$ R_{pri} + H_2O						
RMG rate rule	1.4×10^{7}	2.0	0.86	3.1×10^{11}	2.5×10^{12}	9.3×10^{12}
Zhou *et al.*[b]	3.1×10^{1}	3.54	-2.91	2.4×10^{12}	2.4×10^{12}	5.6×10^{12}
DIPK + HO_2 $\longrightarrow$ R_{tert} + H_2O_2						
RMG rate rule	1.4×10^{-4}	4.30	3.68	1.3×10^{4}	5.8×10^{6}	1.8×10^{8}
CBS-QB3[c]	5.3×10^{-6}	5.24	6.31	1.3×10^{3}	9.7×10^{6}	1.2×10^{9}
DIPK + HO_2 $\longrightarrow$ R_{pri} + H_2O_2						
RMG rate rule	2.0×10^{0}	3.29	1.07	4.8×10^{7}	1.1×10^{9}	8.8×10^{9}
CBS-QB3[c]	4.9×10^{-3}	4.57	14.1	5.3×10^{-2}	1.8×10^{5}	2.1×10^{8}
DIPK + H $\longrightarrow$ R_{tert} + H_2						
RMG rate rule	6.0×10^{6}	2.16	-0.45	2.9×10^{12}	8.8×10^{12}	2.3×10^{13}
CBS-QB3[c]	1.3×10^{6}	2.32	3.38	2.6×10^{9}	2.2×10^{11}	2.2×10^{12}
DIPK + H $\longrightarrow$ R_{pri} + H_2						
RMG rate rule	7.5×10^{8}	1.75	7.51	5.5×10^{7}	1.0×10^{11}	3.1×10^{12}
CBS-QB3[c]	2.3×10^{6}	2.45	7.58	8.0×10^{6}	2.5×10^{10}	1.1×10^{12}
DIPK + CH_3 $\longrightarrow$ R_{tert} + CH_4						
RMG rate rule	5.5×10^{13}	0.0	8.90	1.8×10^{7}	3.1×10^{10}	6.2×10^{11}
CBS-QB3[d]	1.8×10^{-1}	3.98	3.84	2.1×10^{6}	8.3×10^{8}	2.3×10^{10}
DIPK + CH_3 $\longrightarrow$ R_{pri} + CH_4						
RMG rate rule	3.3×10^{6}	1.90	11.1	1.5×10^{3}	6.0×10^{7}	6.4×10^{9}
CBS-QB3[d]	4.7×10^{-4}	4.92	7.28	3.6×10^{3}	4.9×10^{7}	6.9×10^{9}
C_3H_6 + C_4H_7O $\longrightarrow$ R_{pri}						
RMG rate rule	5.2×10^{11}	0.0	6.56	8.7×10^{6}	2.1×10^{9}	1.9×10^{10}
CBS-QB3[c]	1.5×10^{0}	3.38	3.64	7.9×10^{5}	1.7×10^{8}	3.3×10^{9}
DMK + i-C_3H_7 $\longrightarrow$ R_{tert}						
RMG rate rule	1.1×10^{9}	0.83	6.83	1.3×10^{6}	7.0×10^{8}	1.1×10^{10}
CBS-QB3[d]	1.5×10^{0}	2.98	3.80	6.2×10^{4}	1.2×10^{7}	1.9×10^{8}
R_{tert} $\longrightarrow$ $C_7H_{13}O_3$-3						
RMG rate rule	1.9×10^{10}	0.70	27.3	1.4×10^{-8}	1.9×10^{2}	2.6×10^{6}
CBS-QB3[d]	8.3×10^{-10}	6.04	29.3	3.6×10^{-16}	1.1×10^{-3}	4.4×10^{2}

[a] The units for A and $k(T)$ are s^{-1} for unimolecular reactions and in $cm^3\ mol^{-1}\ s^{-1}$ for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$.

[b] Taken from reference [33].

[c] Computed from CBS-QB3 TST calculations with Eckart tunneling and 1D hindered rotors; see text for details.

[d] Computed from CBS-QB3 TST calculations with Eckart tunneling; see text for details.

Table 7.3: Calculated reaction rates for DIPK peroxy chemistry updated in this work.

Reaction / Source	A[a]	n	E_a	$k(300\ K)$[a]	$k(600\ K)$[a]	$k(1000\ K)$[a]
$R_{tert} + O_2 \longrightarrow R_{tert}OO$						
RMG rate rule	1.4×10^{13}	0.0	0.0	4.5×10^{12}	4.5×10^{12}	4.5×10^{12}
Miyoshi[b]	9.8×10^{11}	0.33	-0.42	1.3×10^{13}	1.1×10^{13}	1.1×10^{13}
$R_{pri} + O_2 \longrightarrow R_{pri}OO$						
RMG rate rule	4.5×10^{12}	0.0	0.0	4.5×10^{12}	4.5×10^{12}	4.5×10^{12}
Miyoshi[b]	6.9×10^{16}	-1.63	0.20	4.6×10^{12}	1.8×10^{12}	8.2×10^{11}
$R_{tert} + O_2 \longrightarrow C_7H_{12}O + HO_2$						
RMG rate rule	1.8×10^{14}	-0.20	0.0	5.7×10^{13}	5.0×10^{13}	4.5×10^{13}
Miyoshi[b]	8.7×10^{13}	0.0	16.0	1.9×10^{2}	1.3×10^{8}	2.8×10^{10}
$R_{pri} + O_2 \longrightarrow C_7H_{12}O + HO_2$						
RMG rate rule	2.7×10^{12}	-0.06	0.0	1.9×10^{12}	1.9×10^{12}	1.8×10^{12}
Miyoshi[b]	1.4×10^{12}	0.0	14.4	4.8×10^{1}	8.2×10^{6}	1.0×10^{9}
$R_{tert}OO \longrightarrow C_7H_{12}O + HO_2$						
RMG rate rule	5.9×10^{8}	1.44	28.5	4.0×10^{-9}	2.5×10^{2}	7.4×10^{6}
Miyoshi[b]	7.6×10^{13}	0.0	31.2	1.5×10^{-9}	3.4×10^{2}	1.2×10^{7}
$R_{pri}OO \longrightarrow C_7H_{12}O + HO_2$						
RMG rate rule	1.3×10^{11}	0.50	25.0	1.3×10^{-6}	2.3×10^{3}	1.3×10^{7}
Miyoshi[b]	3.4×10^{12}	0.0	31.9	2.0×10^{-11}	8.3×10^{0}	3.7×10^{5}
$R_{tert}OO \longrightarrow Q_{tt}OOH$						
RMG rate rule	1.1×10^{6}	1.48	18.9	8.7×10^{-5}	1.9×10^{3}	2.2×10^{6}
Miyoshi[b]	3.6×10^{11}	0.0	16.9	1.8×10^{-1}	2.5×10^{5}	7.3×10^{7}
$R_{tert}OO \longrightarrow Q_{tp}OOH$						
RMG rate rule	9.4×10^{9}	0.0	22.4	4.9×10^{-7}	6.8×10^{1}	1.2×10^{5}
Miyoshi[b]	3.0×10^{11}	0.0	22.7	8.9×10^{-6}	1.6×10^{3}	3.3×10^{6}
$R_{pri}OO \longrightarrow Q_{pt}OOH$						
RMG rate rule	1.4×10^{5}	1.38	16.9	1.8×10^{-4}	6.6×10^{2}	3.9×10^{5}
Miyoshi[b]	1.5×10^{10}	0.0	15.3	1.0×10^{-1}	4.0×10^{4}	6.8×10^{6}
$R_{pri}OO \longrightarrow Q_{pp}OOH$						
RMG rate rule	5.4×10^{5}	1.51	20.0	8.8×10^{-6}	4.6×10^{2}	8.0×10^{5}
Miyoshi[b]	1.2×10^{11}	0.0	23.1	1.8×10^{-6}	4.6×10^{2}	1.1×10^{6}
$R_{pri}OO \longrightarrow Q_5OOH$						
RMG rate rule	5.1×10^{6}	1.55	21.0	1.7×10^{-5}	2.3×10^{3}	5.8×10^{6}
Miyoshi[b]	9.2×10^{11}	0.0	23.0	1.7×10^{-5}	4.0×10^{3}	8.8×10^{6}

[a] The units for A and $k(T)$ are s^{-1} for unimolecular reactions and in $cm^3\ mol^{-1}\ s^{-1}$ for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$.

[b] Taken from reference [37].

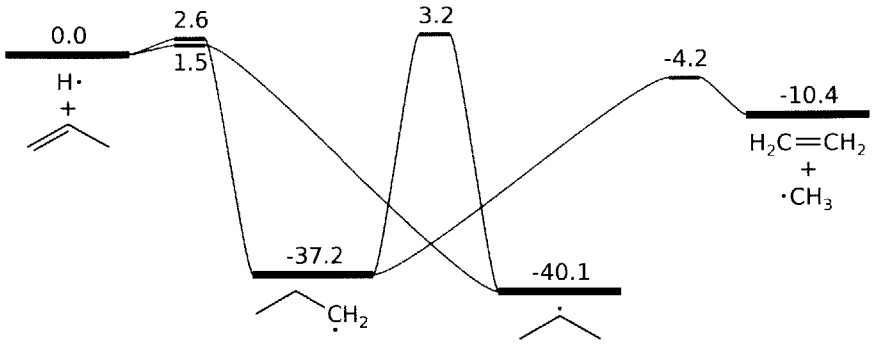


Figure 7.13: The CCSD(T)-F12/cc-pVTZ-F12//B3LYP/MG3S potential energy surface for H + propene. All relative energies are given in kcal/mol.

Table 7.4: Calculated high-pressure limit rate coefficients for H + propene.[a]

Reaction			A	n	E_a
H + C_3H_6	$\longrightarrow$	*i*-C_3H_7	4.93×10^{8}	1.55	0.334
H + C_3H_6	$\longrightarrow$	*n*-C_3H_7	6.34×10^{8}	1.51	1.53
H + C_3H_6	$\longrightarrow$	a-C_3H_5 + H_2	3.71×10^{4}	2.88	5.58
H + C_3H_6	$\longrightarrow$	s-C_3H_5 + H_2	2.07×10^{6}	2.27	12.1
H + C_3H_6	$\longrightarrow$	p-C_3H_5 + H_2	1.30×10^{7}	2.17	15.4
n-C_3H_7	$\longrightarrow$	*i*-C_3H_7	2.07×10^{-10}	6.54	28.1
C_2H_4 + CH_3	$\longrightarrow$	*n*-C_3H_7	6.24×10^{4}	2.39	4.50

[a] The units for A are s^{-1} for unimolecular reactions and in cm^3 mol^{-1} s^{-1} for bimolecular reactions, with E_a in kcal/mol. The rate coefficient is $k = A(T/1[K])^n \exp(-E_a/RT)$. Computed from CCSD(T)-F12 TST calculations with Eckart tunneling; see text for details.

modified strong collision method was used to approximate the $k(T,P)$ values, as the chemically-significant eigenvalue method was not able to resolve enough chemically-distinct timescales at the higher temperatures of interest. A plot of the computed $k(T,P)$ values versus temperature at pressures of 30 and 760 Torr – the two pressures used in the pyrolysis experiments – for the net reactions H + propene $\longrightarrow$ products is shown in Figure 7.15. At both pressures, the well-skipping reaction $H + C_3H_6 \longrightarrow C_2H_4 + CH_3$ becomes increasingly important as temperature increases, eventually becoming the dominant pathway. This reaction is the primary reason for the model prediction of significant secondary pyrolysis of propene.

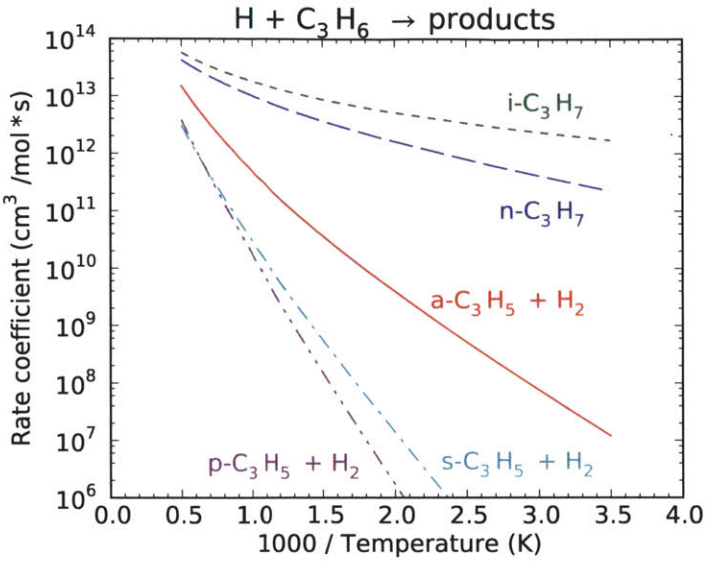


Figure 7.14: Plot of the high-pressure-limit kinetics versus temperature for H + propene reactions. The addition reactions are favored over the abstraction reactions over all temperatures considered.

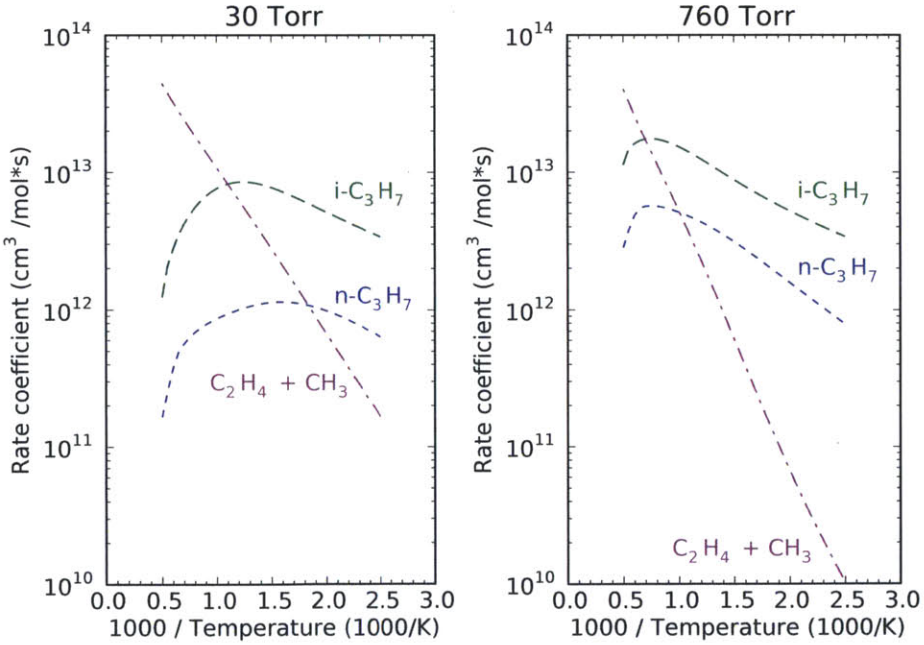


Figure 7.15: Plot of pressure-dependent kinetics versus temperature at 30 Torr and 760 Torr for H + propene reactions. The rate of the well-skipping pathway to C_2H_4 + CH_3 is significantly enhanced at the lower pressure. This reaction drives the model prediction of significant secondary pyrolysis of propene.

7.4 Conclusion

In summary, as part of an investigation of possible biofuel production based on metabolic pathways of endophytic fungi, we have developed a detailed, predictive kinetics model for diisopropyl ketone, a representative compound of the natural products of such fungi. This model has been refined and validated against a variety of experimental data, spanning a range of temperatures, pressures, and oxygen concentrations. Many thermochemistry and kinetics parameters were updated during the development of the DIPK model; these insights have been incorporated within RMG's databases and allow more accurate general prediction of combustion of DIPK and will be extended to other ketones. In the next steps, RMG will be used to predict and model the combustion performance of DIPK in homogeneous-charge compression-ignition (HCCI) engines, one example of emerging clean and efficient combustion strategies. This collaboration between combustion and biofuel production research efforts provides a structured means to screen and recommend future biofuel candidates.

This chapter also demonstrated how the methods discussed in the previous chapters come together to enable the automatic generation and evaluation of detailed kinetics models. The original RMG database of chemical parameters enabled us to quickly generate of an initial model. By comparing this model to experimental measurements, and using techniques such as sensitivity and rate-of-production analyses, we were able to systematically identify model parameters in need of refinement. We gave these parameters further attention using more rigorous experimental measurements or theoretical calculations instead of simply fitting to the experiment; this way, we can more readily use the new information for analogous species and reactions. We were also able to both estimate pressure-dependent kinetics for unimolecular reactions during model generation and to calculate more accurate pressure-dependent kinetics for individual systems of particular interest using master equation methods. Our understanding of the strengths and weaknesses of the master equation reduction methods informs our choice for evaluating $k(T, P)$ values in both situations. The end result was a detailed kinetics model that gives us significant mechanistic insight into the oxidation of diisopropyl ketone and similar ketones.

7.5 References

[1] J. L. Fortman, S. Chhabra, A. Mukhopadhyay, H. Chou, T. S. Lee, E. Steen, and J. D. Keasling. "Biofuel Alternatives to Ethanol: Pumping the Microbial Well."

Trends Biotech. 26 (7), 2008, pp. 375–381. DOI: 10.1016/j.tibtech.2008.03.008.

[2] P. P. Peralta-Yahya, F. Zhang, S. B. del Cardayre, and J. D. Keasling. "Microbial Engineering for the Production of Advanced Biofuels." *Nature* 488 (7), 2012, pp. 320–328. DOI: 10.1038/nature11478.

[3] G. P. Robertson, V. H. Dale, O. C. Doering, S. P. Hamburg, J. M. Melillo, M. M. Wander, W. J. Parton, P. R. Adler, J. N. Barney, R. M. Cruse, C. S. Duke, P. M. Fearnside, R. F. Follett, H. K. Gibbs, J. Goldemberg, D. J. Mladenoff, D. Ojima, M. W. Palmer, A. Sharpley, L. Wallace, K. C. Weathers, J. A. Wiens, and W. W. Wilhelm. "Sustainable Biofuels Redux." *Science* 322 (5898), 2008, pp. 49–50. DOI: 10.1126/science.1161525.

[4] 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.

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

[6] A. J. Janssen, F. W. Kremer, J. H. Baron, M. Muether, S. Pischinger, and J. Klankermayer. "Tailor-Made Fuels from Biomass for Homogeneous Low-Temperature Diesel Combustion." *Energy Fuels* 25 (10), 2011, pp. 4734–4744. DOI: 10.1021/ef2010139.

[7] 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* (NRRL 50072)." *Microbiology* 154 (11), 2008, pp. 3319–3328. DOI: 10.1099/mic.0.2008/022186-0.

[8] 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.

[9] S. K. Singh, G. A. Strobel, B. Knighton, B. Geary, J. Sears, and D. Ezra. "An Rndophytic *Phomopsis* sp. Possessing Bioactivity and Fuel Potential with its

Volatile Organic Compounds." *Microbial Ecol.* 61, 2011, pp. 729–739. DOI: 10.1007/s00248-011-9818-7.

[10] 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.

[11] W. Wang and Y. Zhao. "Quantum Instanton Calculation of Rate Constant for $CH_4 + OH \longrightarrow CH_3 + H_2O$ Reaction: Torsional Anharmonicity and Kinetic Isotope Effect." *J. Chem. Phys.* 137 (21), 2012, p. 214306. DOI: 10.1063/1.4768874.

[12] 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.

[13] C. A. Taatjes, G. Meloni, T. M. Selby, A. J. Trevitt, D. L. Osborn, C. J. Percival, and D. E. Shallcross. "Direct Observation of the Gas-Phase Criegee Intermediate (CH_2OO)." *J. Am. Chem. Soc.* 130 (36), 2008, pp. 11883–11885. DOI: 10.1021/ja804165q.

[14] C. Lee, S. Vranckx, K. A. Heufer, S. Khomik, Y. Uygun, H. Olivier, and R. X. Fernandes. "On the Chemical Kinetics of Ethanol Oxidation: Shock tube, Rapid Compression Machine and Detailed Modeling Study." *Z. Phys. Chem.* 226, 2012, pp. 1–27. DOI: 10.1524/zpch.2012.0185.

[15] S. Vranckx, C. Lee, H. K. Chakravarty, and R. X. Fernandes. "A Rapid Compression Machine Study of the Low Temperature Combustion of Cyclohexane at Elevated Pressures." *Proc. Combust. Inst.* 34 (1), 2013, pp. 377–384. DOI: 10.1016/j.proci.2012.06.071.

[16] G. Mittal and C.-J. Sung. "Aerodynamics Inside a Rapid Compression Machine." *Combust. Flame* 145 (1-2), 2006, pp. 160–180. DOI: 10.1016/j.combustflame.2005.10.019.

[17] J. W. Allen, R. W. Ashcraft, G. J. Beran, B. A. Buesser, C. A. Class, C. Gao, C. F. Goldsmith, M. R. Harper, A. Jalan, M. Keceli, G. R. Magoon, D. M. Matheu, S. S. Merchant, J. D. Mo, S. Petway, S. Ruman, S. Sharma, K. M. Van Geem, J. Song, Y. Suleymanov, J. Wen, R. H. West, A. Wong, H.-W. Wong, N. W.-W. Yee,

P. E. Yelvington, J. Yu, and W. H. Green. *RMG (Reaction Mechanism Generator) version 4.0.* 2013. URL: http://rmg.sourceforge.net/.

[18] C. F. Goldsmith, G. R. Magoon, and W. H. Green. "Database of Small Molecule Thermochemistry for Combustion." *J. Phys. Chem. A* 116 (36), 2012, pp. 9033–9057. DOI: 10.1021/jp303819e.

[19] K. M. V. 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.

[20] N. Hansen, M. R. Harper, and W. H. Green. "High-Temperature Oxidation Chemistry of n-Butanol – Experiments in Low-Pressure Premixed Flames and Detailed Kinetic Modeling." *Phys. Chem. Chem. Phys.* 13, 2011, pp. 20262–20274. DOI: 10.1039/C1CP21663E.

[21] M. R. Harper, K. M. V. 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.

[22] M. P. Burke, M. Chaos, Y. Ju, F. L. Dryer, and S. J. Klippenstein. "Comprehensive H_2/O_2 Kinetic Model for High-Pressure Combustion." *Int. J. Chem. Kin.* 44 (7), 2012, pp. 444–474. DOI: 10.1002/kin.20603.

[23] N. M. Marinov. "A Detailed Chemical Kinetic Model for High Temperature Ethanol Oxidation." *Int. J. Chem. Kin.* 31 (3), 1999, pp. 183–220. DOI: 10.1002/(SICI)1097-4601(1999)31:3<183::AID-KIN3>3.0.CO;2-X.

[24] J. A. Miller and S. J. Klippenstein. "The Recombination of Propargyl Radicals and Other Reactions on a C_6H_6 Potential." *J. Phys. Chem. A* 107 (39), 2003, pp. 7783–7799. DOI: 10.1021/jp030375h.

[25] J. A. Miller, J. P. Senosiain, S. J. Klippenstein, and Y. Georgievskii. "Reactions over Multiple, Interconnected Potential Wells: Unimolecular and Bimolecular Reactions on a C_3H_5 Potential." *J. Phys. Chem. A* 112 (39), 2008, pp. 9429–9438. DOI: 10.1021/jp804510k.

[26] J. W. Allen, C. F. Goldsmith, and W. H. Green. "Automatic Estimation of Pressure-Dependent Rate Coefficients." *Phys. Chem. Chem. Phys.* 14, 2012, pp. 1131–1155. DOI: 10.1039/c1cp22765c.

[27] J. J. A. Montgomery, 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.

[28] J. J. A. Montgomery, M. J. Frisch, J. W. Ochterski, and G. A. Petersson. "A Complete Basis Set Model Chemistry. VII. Use of the Minimum Population Localization Method." *J. Chem. Phys.* 112 (15), 2000, pp. 6532–6542. DOI: 10.1063/1.481224.

[29] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, G. C. M.; Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. *Gaussian 03*. 2004.

[30] Reaction Design. *CHEMKIN-PRO*. 2012. URL: http://www.reactiondesign.com/products/open/chemkin-pro.html.

[31] O. Welz, J. D. Savee, D. L. Osborn, S. S. Vasu, C. J. Percival, D. E. Shallcross, and C. A. Taatjes. "Direct Kinetic Measurements of Criegee Intermediate (CH_2OO) Formed by Reaction of CH_2I with O_2." *Science* 335 (6065), 2012, pp. 204–207. DOI: 10.1126/science.1213229.

[32] O. Welz, J. Zádor, J. D. Savee, M. Y. Ng, G. Meloni, R. X. Fernandes, L. Sheps, B. A. Simmons, T. S. Lee, D. L. Osborn, and C. A. Taatjes. "Low-temperature Combustion Chemistry of Biofuels: Pathways in the Initial Low-Temperature (550 K-750 K) Oxidation Chemistry of Isopentanol." *Phys. Chem. Chem. Phys.* 14, 2012, pp. 3112–3127. DOI: 10.1039/C2CP23248K.

[33] C.-W. Zhou, J. M. Simmie, and H. J. Curran. "Ab Initio and Kinetic Study of the Reaction of Ketones with OH for T = 500-2000 K. Part I: Hydrogen-Abstraction from $H_3CC(O)CH_{3-x}(CH_3)_x$, x = 0 $\mapsto$ 2." *Phys. Chem. Chem. Phys.* 13, 2011, pp. 11175–11192. DOI: 10.1039/C0CP02754E.

[34] J. Zádor, 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.

[35] J. E. Dec. "Advanced Compression-Ignition Engines – Understanding the In-Cylinder Processes." *Proc. Combust. Inst.* 32 (2), 2009, pp. 2727–2742. DOI: 10.1016/j.proci.2008.08.008.

[36] Y. Yang, J. Dec, N. Dronniou, and B. Simmons. "Characteristics of Isopentanol as a Fuel for HCCI Engines." *SAE Int. J. Fuels Lubr.* 3, 2010, pp. 725–741. DOI: 10.4271/2010-01-2164.

[37] A. Miyoshi. "Systematic Computational Study on the Unimolecular Reactions of Alkylperoxy (RO_2), Hydroperoxyalkyl (QOOH), and Hydroperoxyalkylperoxy (O_2QOOH) Radicals." *J. Phys. Chem. A* 115 (15), 2011, pp. 3301–3325. DOI: 10.1021/jp112152n.

8

RECOMMENDATIONS FOR FUTURE WORK

This thesis has made several significant contributions toward facilitating the automatic generation of predictive detailed kinetics models. Particular focus has been given to the rapid and accurate estimation of reaction rate coefficients, including pressure-dependent rate coefficients. The success of these methods, and the value of predictive kinetic modeling, was demonstrated by developing a model of the pyrolysis and combustion of diisopropyl ketone, a candidate biofuel. The progress made towards predictive chemical kinetics has uncovered several new challenges that require attention in order to further achieve the vision laid out at the beginning of this thesis. In this chapter we will discuss several such avenues for future work.

8.1 Add uncertainty estimation to RMG

An important but often overlooked output of any model of a physical process is the uncertainty in the model predictions. Knowledge of the uncertainty gives the model user an idea of the reliability of the model. When the model gives predictions that differ from experiment, the uncertainty range helps tell us whether the model can be improved by refining the parameter values, or if we have omitted some important physical behavior (such as missing reactions). For instance, effort to refine the estimate of a physical parameter is only worthwhile if (1) the model is sensitive to that parameter value and (2) the uncertainty in that parameter is significant. Despite the utility of uncertainty information, it is not currently available from RMG-generated models.

Adding uncertainty functionality to RMG will require two phases. In the first, estimates must be added for the uncertainty of each parameter in the RMG database. Since most entries have no uncertainty information, this will be a significant undertaking. Some entries may be able to recover their uncertainty values by consulting the original reference. However, there will likely be many entries for which this is not possible, and some general procedure will need to be created to decide on un-

certainty values for these cases. If nothing else, the overall uncertainties observed in Chapter 3 could be used, although they are probably an overestimate of the uncertainty of individual values. The representation of the uncertainty data will also need to be chosen, though in most cases, the amount of knowledge of the uncertainty is small enough that the most appropriate representation of the uncertainties is simple interval form.

The second phase is to enable RMG to propagate these parameter uncertainties toward the overall model predictions. Some inspiration can be taken from the work of Goldsmith *et al.* at propagating uncertainty through the master equation model of pressure-dependent kinetics [1]. Uncertainty propagation during numerical simulation of reaction systems within RMG will likely require use of a more advanced differential algebraic equation solver than the DASSL solver [2] currently used in RMG. The IDA solver in the SUNDIALS package [3] may be a reasonable choice.

8.2 Add more kinetics data to the RMG database

In constructing the detailed kinetics model for diisopropyl ketone, the vast majority of the time was spent in the refining of the thermodynamic and kinetic parameters for the sensitive species and reactions in the model. As the first effort to model ketone combustion and pyrolysis, there were many rules that needed to be updated in order to obtain reasonable parameter estimates. Indeed, the uncertainties in estimated kinetic parameters observed in Chapter 3 are so large that they are expected to cause a very large uncertainty in the overall model predictions. Reducing these uncertainties and accelerating the development of models for similar molecules requires that we take full advantage of this new knowledge by adding it to our database of kinetics rate rules.

The new format for the RMG database discussed in Chapter 3 makes it easier to add new kinetics entries. First, users can add entries in either the forward or reverse direction, as RMG is able to determine the direction automatically. For instance, users no longer need to compute the reverse rate coefficient of a β-scission reaction in order to add it to the radical addition to multiple bond family, which eliminates one possible source of human error. Second, users have much more flexibility in the choice of units for all parameters. This enables users to provide entries in the exact form given in the reference, which minimizes the possibility of transcription errors and makes such errors much easier to track down. Finally, users no longer need to manually determine the functional groups for the reactions they are adding, although they do

need to label the atoms of the reacting sites so that RMG can automatically assign the functional groups. We have made some effort to automate this tagging as well on the RMG website. These steps eliminate much of the tedium and opportunities for error from the process of adding entries to the database, which signficantly lowers the barrier for growing the database.

Once a substantial training set has been assembled, some effort can be made to optimize the kinetics estimation procedure. One possible means of doing so is to reorganize the hierarchical tree of functional groups in order to more closely associate functionalities with similar effects on the kinetics, especially for multifunctional reaction sites. This is now possible because we have decoupled the kinetics training set from the functional group tree for each reaction family, as discussed in Chapter 3. An example of a potential tree rearrangement can be seen in the current hydrogen abstraction tree for the radical group, Y·. One of the higher levels of the tree splits the monoradical functional groups that are adjacent to another radical from those that are not. However, the group additivity correction at 1000 K for C · ≡C· ($10^{3.52}$) is much closer to that of C≡C· ($10^{2.62}$) than that of O · −O· ($10^{-7.82}$). This suggests that the presence of an adjacent radical is a smaller effect on the rate than the triple bond to the reactive radical, which implies that we should rearrange the tree to bring C · ≡C· closer to C≡C·. This also demonstrates one value of the kinetics group additivity method: by isolating the effect of each functional group independently, it helps us learn how to organize our functional group tree to improve the kinetics estimates.

8.3 Develop method for automatic transition state structure searches

Even with the lower barrier to adding kinetics entries to the new RMG database, the determination of the parameters to add remains a significant source of manual labor. One way to significantly reduce this would be to enable the ability to automatically search for and optimize transition state structures using quantum chemistry calculations. The ability to automatically determine transition state structures would move us much closer to truly predictive chemical kinetics, as it would enable the computer to automatically identify and update the values of parameters that have both a high sensitivity and high uncertainty without any manual intervention from the human user.

Perhaps the most important step in finding a transition state structure is in assembling a reasonable initial guess for the geometry, as the optimization is generally highly nonlinear. One possible route to automate this would utilize the same hierar-

chy of functional groups currently used throughout the RMG database, only here we would associate these functional groups with the three-dimensional coordiates that describe the reactive sites. This would allow us to fall up the hierarchy to a more general geometry when a more specific geometry is not available, and we could easily update the tree with new geometries as they are determined. Nonreactive functional groups could then be attached to the reactive sites with standard 2D-to-3D algorithms.

Conducting even very simple quantum chemistry calculations for the transition state of all reactions is today far too computationally expensive without access to a large supercomputer. Fortunately, since only a small subset of the reactions are sensitive, the number of such calculations required should be far more tractable. Nonetheless, part of the challenge will be in identifying the appropriate level of theory at which to conduct the quantum chemistry calculation. It may be that the appropriate model chemistry is dependent on the sensitivity of the parameter, as more sensitive parameters would demand more rigorous quantum chemistry treatments.

8.4 Improve models of collisional energy transfer

In developing the model of the reaction of the Criegee biradical with small carbonyls in Chapter 6, we showed that the amount of secondary ozonide produced is very sensitive to the model of collisional energy transfer used in the master equation. Most master equation models in the literature to date have used the single exponential down model simply because it only has one parameter to determine; this parameter is typically set to some empirical value or fitted to reproduce the experimental data. Parameter fitting is anathema to the ideal of predictive chemical kinetics; given the importance of this parameter to certain model predictions, it would be beneficial to have some theoretical justification for the values used. Some effort towards this has been made by Barker and Weston [4] and by Jasper and Miller [5] using classical molecular dynamics trajectory simulations of collision events to gain insight into the dependence of collisional energy transfer on both energy and angular momentum.

8.5 Improve the automatic estimation of the potential energy surface for pressure-dependent reaction networks

The predictions of the Criegee biradical + carbonyl models were also quite sensitive to many of the transition state barrier heights on the potential energy surface. In these models we have some confidence in the chosen barrier heights due to their values

coming from rigorous quantum chemistry calculations. However, we also need to estimate transition state barrier heights when estimating pressure-dependent kinetics within RMG, where detailed transition state information is generally not available. At the moment we are simply using the high-pressure limit Arrhenius activation energy E_a to determine the corresponding barrier E_0, as required by our use of the inverse Laplace transform method to compute the microcanonical rate coefficient $k(E)$ from the high pressure-limit rate $k(T)$. However, there is no guarantee that these quantities are closely related, as evidenced by the large amount of scatter on Evans-Polanyi plots like those in Chapter 3. The sensitivities in the Criegee biradical models show that an error of only a few kcal/mol can cause significant changes in the computed $k(T, P)$ values. It would be worth revisiting the automatic construction of the potential energy surface and estimation of the microcanonical rate coefficients $k(E)$ in RMG jobs in the future to see if the accuracy can be improved.

8.6 Use RPMD rate theory to calculate microcanonical rate coefficients

In Chapter 4 we demonstrated the general applicability of ring polymer molecular dynamics (RPMD) rate theory for the accurate calculation of bimolecular rate coefficients $k(T)$ which incorporate the quantum mechanical effects of zero-point energy and tunneling through an activation barrier. However, there are many important reaction families that are unimolecular, and it would be valuable to apply RPMD rate theory to incorporating quantum effects in their kinetics as well. Given that these reactions are also potentially pressure-dependent, we would generally prefer to compute the microcanonical rate coefficient $k(E)$ rather than the high pressure-limit rate coefficient $k(T)$ so that we can input it into the master equation models of Chapter 5.

A microcanonical version of RPMD rate theory should be straightforward. For instance, the true quantum mechanical microcanonical rate coefficient

$$k(E) = \frac{1}{\rho_{\text{reac}}(E)} \lim_{t \to \infty} \tilde{c}_{fs}(t; s) \tag{8.1}$$

and corresponding microcanonical flux-side correlation function

$$\tilde{c}_{fs}(t; s) = \text{tr}\left[\delta(E - \hat{H})\hat{F}e^{+i\hat{H}t/\hbar}\hat{h}e^{-i\hat{H}t/\hbar}\right] \tag{8.2}$$

are very similar to their canonical counterparts, except that they substitute the reactant density of states for the reactant partition function and the delta function $\delta(E - \hat{H})$ for the canonical Boltzmann factor. The classical and RPMD microcanonical rate coefficients can be developed using the same methodology as for the canonical cases. Probably most of the actual implementation can be achieved by simply turning off the thermostats used to sample from a canonical distribution. Some effort will likely need to be made to refine our definition of the transition state dividing surface, which currently expects there to be pairs of breaking and forming bonds, as is typical of bimolecular reactions but not of unimolecular reactions.

8.7 Include species pruning in RMG model enlargement algorithm

As we continue to push RMG towards developing detailed kinetics models of more complex fuel molecules, we are often running into a memory limitation, even on computer clusters. For instance, we have yet to fully converge a model for JP-10 (exo-tetrahydrodicyclopentadiene), a tricyclic C_{10} hydrocarbon used as jet fuel, even with 32 GB of memory. Although we could simply continue to wait for Moore's law to catch up with us, it is very likely that within all those gigabytes of memory are a large body of very unstable species which are almost certain to never be part of the final model. Eliminating these species from further consideration could conceivably recover significant memory, and also accelerate execution time.

Species pruning is complicated somewhat by the presence of pressure-dependent reaction networks. The process of expanding pressure-dependent networks, called the activated species algorithm [6], effectively adds an intermediate layer between the core and edge for unimolecular isomers whose reactions have been explored within a network, but that are not necessarily candidates for adding to the core because of well-skipping at low pressure. The pruning algorithm must be careful to not prune these explored isomers, or else we will not be able to correctly predict the $k(T, P)$ values of the well-skipping net reactions.

We have done some preliminary work in both RMG-Java and RMG-Py on implementing a pruning algorithm based on species fluxes. The algorithm only prunes species with maximum fluxes below a certain tolerance when compared to the characteristic flux of species in the model core. Pruning also only occurs when necessary to keep the total number of species on the edge below a certain threshold. Although this implementation is in a basic working state, further effort is needed to probe the effect of pruning on the generated model. A better implementation of species pruning

would also save a list of all pruned species somewhere on disk, and then return to these species at the very end of the model generation to check that they are in fact still insignificant.

8.8 Parallelize RMG

The high memory demands of RMG jobs for complex fuel molecules often necessitate running jobs on a computer cluster. Most computer clusters offer many processors per node for running jobs; for instance, the cluster regularly used for RMG jobs by our group has many eight-processor, 32 GB nodes. However, RMG is currently not designed to take advantage of these additional processors, meaning they usually sit idle so that we can occupy the entire memory of the node without competition from other jobs. As we continue to add CPU-intensive features to RMG, the ability to scale jobs to multiple processors will become increasingly necessary.

Efficient parallelization requires tasks that are independent from one another and require little to no communication between processes. Fortunately, many of the tasks that RMG currently performs should be straightforward. For example, estimating the thermodynamics of a set of species can be easily spread across a pool of worker processes. Each process can be given a copy of the database once at initialization; after that the process simply needs to be input the species of interest and will return the estimated thermodynamic parameters. Other obvious possibilities for parallelization include kinetic parameter estimation for a set of reactions, calculation of pressure-dependent rate coefficients for a set of unimolecular reaction networks, and simulating of multiple reaction systems.

An important decision to be made when selecting a parallelization strategy is the choice of a shared-memory or separate-memory model. The former has the advantage of requiring much less communication between processes, but also limits the amount of parallelization to single nodes. The latter requires more careful consideration of the amount of communication between processes, but also can easily scale across many nodes.

8.9 Final thoughts

In my thesis work I have had the opportunity to contribute to and eventually oversee the development of a large-scale scientific software project. The RMG software is positioned to become an increasingly important component of predictive kinetic modeling. One of the reasons for the success of RMG has been its adoption of an

open and collaborative philosophy. The distribution of RMG as a free, open source software with no registration wall is one part of this, as it minimizes the barrier for new users. However, the real benefits have come from also doing our development in public. Everyone has access to the entire version history of the RMG source code, which means they can easily reproduce a result from any version of the software. Each developer can make changes to the code independently of the others; the merging together of these changes provides an inherent opportunity for democratic peer review. Thus, our public development process provides an accountability and reproducability consistent with the principles of scientific inquiry. It is my hope that RMG will continue to serve as an example of how computational science ought to be performed.