

Cite this: *Phys. Chem. Chem. Phys.*, 2012, **14**, 1131–1155

www.rsc.org/pccp

PAPER

Automatic estimation of pressure-dependent rate coefficients[†]

Joshua W. Allen,^a C. Franklin Goldsmith^{ab} and William H. Green^{*a}

Received 30th August 2011, Accepted 27th October 2011

DOI: 10.1039/c1cp22765c

A general framework is presented for accurately and efficiently estimating the phenomenological pressure-dependent rate coefficients for reaction networks of arbitrary size and complexity using only high-pressure-limit information. Two aspects of this framework are discussed in detail. First, two methods of estimating the density of states of the species in the network are presented, including a new method based on characteristic functional group frequencies. Second, three methods of simplifying the full master equation model of the network to a single set of phenomenological rates are discussed, including a new method based on the reservoir state and pseudo-steady state approximations. Both sets of methods are evaluated in the context of the chemically-activated reaction of acetyl with oxygen. All three simplifications of the master equation are usually accurate, but each fails in certain situations, which are discussed. The new methods usually provide good accuracy at a computational cost appropriate for automated reaction mechanism generation.

1 Introduction

Thermal unimolecular reactions—the set of reactions involving a single reactant molecule—are unusual in that they require an inert third body to proceed. The third body provides or removes the energy needed for reaction *via* 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 activated species (isomer) 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.

Our interest in unimolecular reaction networks is colored by our intended application: automatic reaction mechanism generation. Our automatic mechanism generation code, RMG,¹ utilizes a rate-based approach to model enlargement, and it is therefore important to account for the pressure dependence of chemically-activated reaction networks in order for RMG to generate a valid mechanism. RMG often considers thousands of these reaction networks in the course of constructing a reaction mechanism, so we require a method that is fully automatable, successful over a wide range of conditions, reasonably accurate, and not too computationally intensive. At the time it is making an estimate of $k(T,P)$, RMG has only limited information available about the molecules and reactions, as discussed below. This makes it more challenging to accurately estimate all the $k(T,P)$.

The importance of bimolecular collisions in unimolecular reactions was first proposed by Lindemann in 1922.² It was soon recognized by Hinshelwood and others that a rigorous treatment of these processes required consideration of molecular energy levels.³ The RRKM expression for the micro-canonical rate coefficient $k(E)$ was derived in the early 1950s.^{4–6} In the late 1950s master equation models of chemical systems began appearing,^{7–11} including an early linear integral-differential equation formulation by Widom.¹² Analytical solutions for a variety of simple models soon followed,^{13–15} as did the first numerical approaches.¹⁶ 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 equations¹⁷ and efficient algorithms for calculating the density of states.^{18–20} In the 1990s computing power had increased to the point where it

^a Dept. of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA.
E-mail: whgreen@mit.edu; Fax: 617-324-0066; Tel: 617-253-4580

^b Dept. of Inorganic Chemistry, Fritz Haber Institute, Berlin, 14195, Germany

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1cp22765c

was practical to solve them numerically by discretizing the integrals over energy.

Methods for directly solving the discretized master equations include integration using a stiff ODE solver and stochastic simulation using the Gillespie method. Both of these methods have the benefit of not introducing additional approximations (in the latter only if sufficient sampling of events is taken). However, both methods suffer when wide ranges of time scales are encountered, as they often are in chemical kinetics. Furthermore, with either approach it is not clear how to extract the time-independent rate constants $k(T,P)$ from the full solution. Barker *et al.* have developed the stochastic approach into the software package MultiWell.²¹

Over the last ten to fifteen years a number of numerical techniques have been developed with the goal of simplifying a full master equation formulation into a single set of phenomenological rate coefficients $k(T,P)$. The methods vary in the simplifying assumptions made and the numerical procedures used, which in turn affects the computational expense required and the accuracy and robustness of the result. A sampling of methods and associated software packages is provided in the following paragraphs.

Several approaches are based on computing the eigenmodes of the full master equation matrix. When the conventional phenomenological-kinetics description of the chemical system is applicable, certain “chemically-significant” eigenmodes become separable from the large number of internal energy relaxation eigenmodes. The chemically-significant eigenvalues match the eigenvalues of the matrix of phenomenological rates, and therefore can be used to extract values for each $k(T,P)$, although this may not be straightforward.²² Eigenvalue determination is computationally expensive and numerically challenging for large, stiff systems; *e.g.* at low temperatures and high pressures, solvers often return nonphysical or complex eigenvalues. There are multiple software packages that utilize eigenvalue approaches, including UNIMOL,²³ ChemRate,²⁴ Variflex,²⁵ and the recently-released MESMER.²⁶

On the opposite end of the scale are methods based around the strong collision approximation, wherein all collisions are assumed to irreversibly stabilize the energized species. This decoupling of the energy grains in the system is a huge computational savings, at the cost of neglecting the effects of collisions which cause small energy changes. Despite this gross simplification, the modified strong collision method appears to provide reasonably accurate results, typically within an order of magnitude of the true rate coefficients.²⁷ This may be partially due to the modification of the collision frequency by a temperature-dependent efficiency factor.^{28–30} For multi-well systems, this method was codified by Chang, Bozzelli, and Dean into the software package CHEMDIS.³¹ The modified strong collision method is implemented in the automatic mechanism generation software package XMG, and in early and current versions of RMG.

More recently, Green and Bhatti have proposed a new approximate method for computing $k(T,P)$ based around the pseudo steady-state approximation at high energies coupled with the approximation of a Boltzmann population of low energy levels.³² This method maintains the collision model of the full analysis for the important high-energy states, so a

more accurate result is expected when compared to the modified strong collision method. However, it is not clear yet just how much more expensive and more accurate this method is. Also, the multiple-well formulation of the reservoir state method as described by Green and Bhatti is cumbersome, requiring the method to be repeated using each isomer well and reactant set as a starting point. The present work will develop a formulation that does not require this repetition.

A key input to all $k(T,P)$ calculations is the density of states for each species. In the context of automated reaction mechanism generation it is not yet possible to determine this information by performing quantum chemistry calculations on every species, as these are still too slow for large molecules, and nontrivial to automate. Bozzelli and coworkers developed a method for estimating the density of states from heat capacity data (obtained from group additivity estimates) by fitting three pseudofrequency-degeneracy pairs.³³ The fitted frequencies do not necessarily reflect the real vibrational degrees of freedom of the molecule, and the harmonic-oscillator form is not a good representation for hindered internal rotors.

The primary objective of the current work was to develop a general framework for automatic estimation of pressure-dependent rate coefficients in reaction networks of arbitrary size and complexity. Within this framework, various methods of estimating the density of states and extracting phenomenological rate constants from the master equation matrix could be developed and evaluated. In the former, a new method based on characteristic functional-group frequencies was created to more accurately reflect the real molecular degrees of freedom, including hindered rotors. In the latter, the chemically-significant eigenvalues, modified strong collision, and reservoir state methods were directly compared using identical input parameters.

Section 2 provides a brief review of the formulation of the master equation model. Section 3 describes our framework for estimating the phenomenological rate coefficients for unimolecular reaction networks. Section 4 describes the methods used to estimate the density of states, focusing on a new method based on characteristic functional group frequencies. Section 5 describes the methods used to estimate the phenomenological rate coefficients from the full master equation formulation. Section 6 evaluates the methods of the previous sections using a case study: the reaction of acetyl radical with oxygen, a process important in atmospheric chemistry.

2 Background: the master equation

Descriptions of the full master equation formulation are quite ubiquitous,^{34–38} and so only a brief summary will be given here.

The system of interest is a set of chemically reactive molecular configurations—local minima on a potential energy surface—divided into unimolecular isomers and bimolecular reactants or products. In our vernacular, reactants can associate to form an isomer, while such association is neglected for products. These configurations are connected by chemical reactions to form a network. The system also consists of an excess of inert gas M, representing a thermal bath; this allows

for neglecting all collisions other than those between an isomer and the bath gas.

An isomer molecule at sufficiently high internal energy E can be transformed by a number of possible events:

- The isomer molecule can collide with any other molecule, resulting in a change in energy from E to E'
- The isomer molecule can isomerize to an adjacent isomer at constant E
- The isomer molecule can dissociate into any directly connected bimolecular reactant or product channel

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:

$$\lim_{t \rightarrow \infty} p_i(E, t) = x_{i\infty} \frac{\rho_i(E)e^{-\beta E}}{Q_i(\beta)} = x_{i\infty} b_i(E, T) \quad (1)$$

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 $\beta \equiv (k_B T)^{-1}$. Assuming that inelastic collisions are much more common than reactive collisions, we can treat the bimolecular reactants as thermalized, and represent the concentrations of the n th reactant configuration by $y_{nA}(t)$ and $y_{nB}(t)$, where A and B distinguish between the two reactants.

Neglecting the dependencies of the microcanonical rates and other quantities on the angular momentum quantum number J or any other quantum numbers besides E ,³⁹ and neglecting reactions which do not proceed through one of the isomers, the master equation can be written as

$$\begin{aligned} \frac{d}{dt} p_i(E, t) &= \omega_i(T, P) \int_0^\infty P_i(E, E', T) p_i(E', t) dE' \\ &\quad - \omega_i(T, P) p_i(E, t) \\ &+ \sum_{j \neq i}^{N_{\text{isom}}} k_{ij}(E) p_j(E, t) - \sum_{j \neq i}^{N_{\text{isom}}} k_{ji}(E) p_i(E, t) \\ &+ \sum_{n=1}^{N_{\text{reac}}} y_{nA}(t) y_{nB}(t) f_{in}(E) b_n(E, T) \\ &- \sum_{n=1}^{N_{\text{reac}} + N_{\text{prod}}} g_{ni}(E) p_i(E, t) \end{aligned} \quad (2)$$

$$\begin{aligned} \frac{d}{dt} y_{nA}(t) &= \frac{d}{dt} y_{nB}(t) \\ &= \sum_{i=1}^{N_{\text{isom}}} \int_0^\infty g_{ni}(E) p_i(E, t) dE \\ &\quad - \sum_{i=1}^{N_{\text{isom}}} y_{nA}(t) y_{nB}(t) \int_0^\infty f_{in}(E) b_n(E, T) dE \end{aligned} \quad (3)$$

where $\omega_i(T, P)$ is the collision frequency; $P(E, E', T)$ is the probability of collisional transfer from energy E' to E ; $k_{ij}(E)$, $f_{in}(E)$, and $g_{ni}(E)$ are the microcanonical rate coefficients for isomerization, association, and dissociation, respectively;

$b_n(E, T)$ is the Boltzmann distribution for bimolecular reactant channel n ; and N_{isom} , N_{reac} , and N_{prod} are the numbers of isomers, bimolecular reactant channels, and bimolecular product channels, respectively. In eqn (2), the first pair of terms correspond to collision, the second pair to isomerization, and the final pair to association/dissociation. Eqn (2) applies to isomers, while eqn (3) applies to bimolecular reactants. Note that there are almost always additional reactions creating and destroying all the species, so eqn (2) and (3) are for an idealized situation.

There is some subtlety to the formulation of the master equation, in that some treatments are based on the total energy, while others use the active internal energy. This has been discussed in the literature previously.^{40,41} This work utilizes the one-dimensional total energy model as described by Miller and Klippenstein. In particular, the densities of states contain all of the vibrational and rotational modes of the system, which implies that we are treating all rotational modes as active; this approximation appears to be increasingly accurate as temperature increases.⁴¹

Eqn (2) and (3) are nonlinear, both due to the presence of the bimolecular reactant terms and because both ω_i and $P(E, E')$ depend on the composition, which is changing with time. The rate coefficients can be derived from considering the pseudo-first-order situation where $y_{nA}(t) \ll y_{nB}(t)$, and all $y(t)$ are negligible compared to the bath gas M. From these assumptions the changes in $\omega_i P(E, E')$, and all y_{nB} 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

$$\begin{aligned} \frac{d}{dt} \mathbb{P} &= \mathbb{M} \mathbb{P} \\ \mathbb{M} &= \begin{bmatrix} \mathbf{M}_1 & \mathbf{K}_{12} & \dots & \mathbf{F}_{11}\mathbf{b}_1 y_{1B} & \mathbf{F}_{12}\mathbf{b}_2 y_{2B} & \dots \\ \mathbf{K}_{21} & \mathbf{M}_2 & \dots & \mathbf{F}_{21}\mathbf{b}_1 y_{1B} & \mathbf{F}_{22}\mathbf{b}_2 y_{2B} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ (\mathbf{g}_{11})^T & (\mathbf{g}_{12})^T & \dots & h_1 & 0 & \dots \\ (\mathbf{g}_{21})^T & (\mathbf{g}_{22})^T & \dots & 0 & h_2 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix} \\ \mathbb{P} &= \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ y_{1A} \\ y_{2A} \\ \vdots \end{bmatrix} \end{aligned} \quad (4)$$

where the elements of the vectors \mathbf{p}_i are such that $(\mathbf{p}_i)_r = p_i(E_r)$. The diagonal matrices \mathbf{K}_{ij} and \mathbf{F}_{in} and the vectors \mathbf{g}_{ni} contain the microcanonical rate coefficients for isomerization, association, and dissociation, respectively:

$$(\mathbf{K}_{ij})_{rs} = \begin{cases} \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} k_{ij}(E) dE & r = s \\ 0 & r \neq s \end{cases} \quad (5a)$$

$$(\mathbf{F}_{in})_{rs} = \begin{cases} \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} f_{in}(E) dE & r = s \\ 0 & r \neq s \end{cases} \quad (5b)$$

$$(\mathbf{g}_{ni})_r = \frac{1}{\Delta E_r} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} g_{ni}(E) dE \quad (5c)$$

Above, r is the index for the discretized energies of isomer i , and s is the index for the discretized energies of isomer j or bimolecular channel n . 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:

$$(\mathbf{M}_i)_{rs} = \begin{cases} \omega_i(T, P)[(\mathbf{P}_i)_{rr} - 1] - \sum_{j \neq i}^{N_{\text{isom}}} (\mathbf{K}_{ij})_{rs} & r = s \\ - \sum_{n=1}^{N_{\text{reac}} + N_{\text{prod}}} (\mathbf{g}_{ni})_r & r \neq s \\ \omega_i(T, P)(\mathbf{P}_i)_{rs} & \end{cases} \quad (6)$$

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

$$(\mathbf{P}_i)_{rs} = \frac{1}{\Delta E_r \Delta E_s} \int_{E_r - \Delta E_r/2}^{E_r + \Delta E_r/2} \int_{E_s - \Delta E_s/2}^{E_s + \Delta E_s/2} P_i(E, E', T) dE' dE \quad (7)$$

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

$$h_n = - \sum_{i=1}^{N_{\text{isom}}} \sum_{r=1}^{N_{\text{grains}}} y_{nB} (\mathbf{F}_{in})_{rr} b_n(E_r) \quad (8)$$

3 A general framework for processing reaction networks

The procedure, outlined in Fig. 1, is used to estimate phenomenological rate coefficients $k(T, P)$ for a reaction network of arbitrary connectivity and complexity based exclusively on high-pressure-limit parameters. The parameters needed are:

- **The species in the network.** For each species the thermodynamic parameters (ΔH_f° (298 K), ΔS_f° (298 K), and a model for $C_p(T)$) 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.

- **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 paper 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, 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.) 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. Nonetheless, an automatic procedure for determining E_{\max} is required for our desired application of automatic mechanism generation. Our chosen procedure is as follows:

1. Guess an initial value $E_{\max,0}$ that is a certain number of $k_B T$ above the highest energy on the potential energy surface, usually a transition state or a bimolecular reactant or product channel.

2. Estimate the density of states and the equilibrium distribution for the isomer with the highest ground-state energy. For a calculation performed at multiple temperatures, determine the equilibrium distribution at the maximum temperature, as this gives the widest distribution.

3. Determine the energy at which the tail of the equilibrium distribution is some desired fraction of the maximum of the distribution.

4. If a suitable energy is found, add to it the difference between the highest energy on the potential energy surface and the isomer's ground-state energy. This is the maximum energy E_{\max} .

5. If no energy in the current range qualifies, repeat this procedure starting from a higher value of $E_{\max,0}$ until a suitable E_{\max} is found. In the interest of speed it is preferred to minimize the number of times the density of states is estimated, so our first choice of $E_{\max,0}$ is deliberately chosen to be significantly higher than the maximum energy in most cases.

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. Two methods for estimating the molecular degrees of freedom needed to compute $\rho(E)$, the three-frequency and functional-group frequency methods, are described in Sections 4.1 and 4.2, respectively.

The input includes a set of temperatures and pressures at which to estimate the phenomenological rate coefficients.

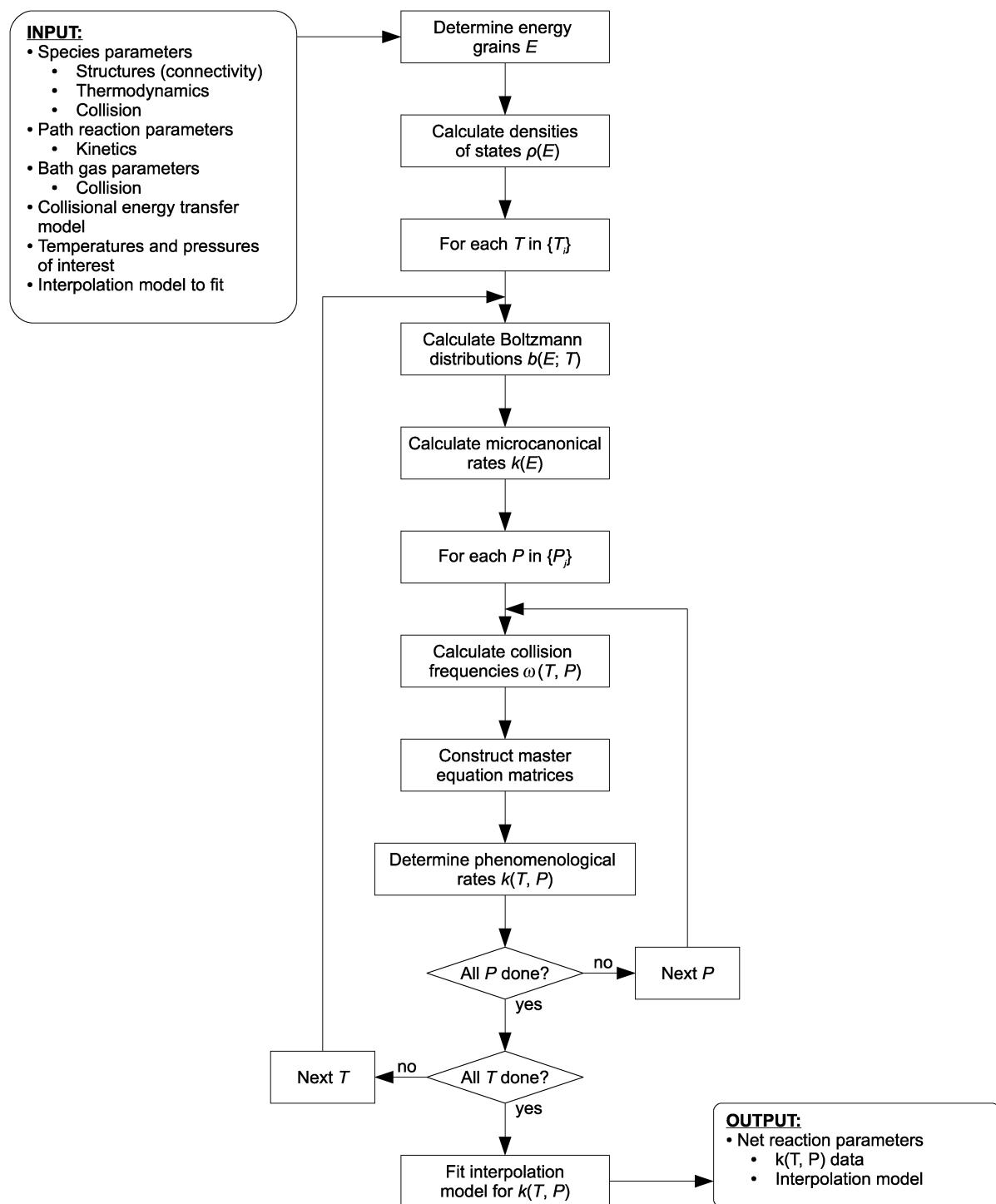


Fig. 1 A general procedure for estimating pressure-dependent rate coefficients for reaction networks of arbitrary size and complexity.

There are certain steps that are only dependent on the current temperature and not on pressure, so we place the iteration over temperature as the outer loop and the iteration over pressure as the inner loop. For example, the equilibrium distributions are functions of temperature only and are calculated using the formula

$$b_i(E_r, T) = \frac{\rho_i(E_r) e^{-\beta E_r}}{\sum_{s=1}^{N_{\text{grains}}} \rho_i(E_s) e^{-\beta E_s}} \quad (9)$$

for isomer i at energy E_r for energy grain r , where $\rho_i(E)$ is the vibrational-rotational density of states for the isomer.

The microcanonical rate coefficients $k(E)$ can be computed using a variety of methods. If reactant and transition state energies, geometries, and vibrational frequencies are available (e.g. from a quantum chemistry calculation), then the Rice-Ramsperger-Kassel-Marcus (RRKM) method 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,⁴² is

$$k(E)\rho(E) = \mathcal{L}^{-1}[Q(\beta)k_\infty(\beta)] \quad (10)$$

Exact formulas exist for simple Arrhenius kinetics³⁷ and modified Arrhenius kinetics $k_\infty(T) = AT^n \exp(-E_a/RT)$ for $n > -1/2$.³¹ 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 (QRKK) 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. The reverse rate coefficient is determined from detailed balance. For an isomerization reaction $A \rightarrow B$ the detailed balance equation is

$$k_f(E)\rho_A(E) = k_b(E)\rho_B(E) \quad (11)$$

and for a dissociation reaction $A \rightarrow B + C$ the equation is

$$k_f(E)\rho_A(E) = k_b(E)\rho_{BC}(E) \quad (12)$$

where $\rho_{BC}(E)$ is the convolved vibrational-rotational density of states for species B and C, and also includes relative translational motion. An alternative formulation incorporates the macroscopic equilibrium coefficient $K_{eq}(T)$ and equilibrium distributions $b_i(E, T)$ at each temperature. For isomerization and dissociation, respectively, the alternative formulation is

$$k_f(E)b_A(E, T) = K_{eq}(T)k_b(E)b_B(E, T) \quad (13)$$

$$k_f(E)b_A(E, T) = K_{eq}(T)k_b(E)b_{BC}(E, T) \quad (14)$$

where $b_{BC}(E, T)$ is the combined equilibrium distribution for species B and C. These two formulations are equivalent; however, there are multiple reasons to use eqn (13) and (14) instead of eqn (11) and (12):

- 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 $k_b(E)b_{BC}(E, T)$, and not the individual values of $k_b(E)$ and $b_{BC}(E, T)$. (In eqn (4) 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 $K_{eq}(T)$. 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 $b(E, T)$. In Section 4 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 eqn (13) and (14).

The collision model is made up of the collision frequencies $\omega_i(T, P)$ and the collisional transfer probabilities model $P_i(E, E', T)$. As is common within the master equation literature, here ω_i is calculated using the Lennard-Jones collision model. This should be a reasonable choice for collisions between polyatomic molecules and weak colliders such as noble gases and diatomics, but less accurate for collisions between two large polyatomics or two very polar molecules.⁴³ The collisional transfer probabilities function $P_i(E, E', T)$ comes in a variety of forms, all of which share the general trend that a collision is more likely to cause a small transfer of energy than a large one. The most common form of $P_i(E, E', T)$ is the single exponential down model

$$P_i(E, E', T) = C_i(E', T) \exp[-\alpha_i(T)(E' - E)] \quad E < E' \quad (15)$$

where the activating equivalent is determined from detailed balance. C is determined by normalization of the probability distribution. The lone parameter $\alpha_i(T)$ is the inverse of $\langle \Delta E_d \rangle$. To be consistent with thermodynamics, detailed balance must be satisfied:

$$P_i(E', E, T) = \frac{\rho_i(E')}{\rho_i(E)} \exp\left(-\frac{E' - E}{k_B T}\right) P_i(E, E', T) \quad E < E' \quad (16)$$

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. Three methods to do this, the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods, are described in detail in Sections 5.1, 5.2, and 5.3, respectively.

4 Methods for estimating the density of states

As described in Sections 1 and 2, formulation of the master equation requires the density of states ρ_{vk} for each isomer. The relevant rovibrational degrees of freedom are those degrees of freedom in which the energy of the excited species can be randomized, *i.e.* the constant J and M modes are not included. In RMG, these degrees of freedom are taken to be the harmonic-oscillator (HO) vibrational frequencies, hindered internal rotors, and a one-dimensional external rotation (*i.e.* the K-rotor for a symmetric top approximation, which is expected to strongly couple with vibrations *via* Coriolis terms). Furthermore, it is assumed that the degrees of freedom are independent. RMG assumes that each molecule may be approximated as a symmetric top and that the density of states for the one-dimensional K-rotor is:

$$\rho_r(E) = \frac{1}{\sigma} \left(\frac{1}{|A - \bar{B}|E} \right)^{1/2} \quad (17)$$

where σ is the rotational symmetry number for this rotor, A is the unique rotational constant, and \bar{B} is the average of the two

most similar rotational constants. As stated in Section 3, since $\sigma|A - \tilde{B}|$ is a proportionality constant in the above expression, we need not be concerned with its actual value, as it will cancel when we evaluate equilibrium distributions. A discussion of the accuracy of the K-rotor model is available from Zhu *et al.*⁴⁴

The density of states for hindered internal rotors is given by the Laplace transform of the hindered-rotor partition function. RMG assumes a semi-classical approximation for the hindered-rotor partition function:

$$\begin{aligned} Q_{\text{HR}_{\text{semi}}} &= Q_{\text{HR}_{\text{classical}}} \frac{Q_{\text{HO}_{\text{quantum}}}}{Q_{\text{HO}_{\text{classical}}}} \\ &= \left(\frac{V_0 \pi}{2\sigma^2 B_r} \right)^{1/2} u^{1/2} e^{-u} I_0(u) \frac{e^{-\alpha u}}{1 - e^{-2\alpha u}} 2\alpha \end{aligned} \quad (18)$$

where B_r is the rotational constant corresponding to the reduced moment of inertia, V_0 is the rotational barrier height assuming a single cosine potential, σ is the symmetry number of the internal rotor, ν is the normal mode frequency corresponding to the torsion, I_0 is the modified Bessel function of the first kind, $u \equiv V_0/2k_B T$, and $\alpha \equiv \nu/V_0$. Like with the K-rotor, the rotational constant B_r appears only as a proportionality constant, so its value will not be important for our master equation calculation. If detailed information about the species or transition state is available, we compute the reduced moment of inertia using the method of Pitzer⁴⁵ which is described as $I^{(2,3)}$ by East and Radom⁴⁶.

The density of states corresponding to the classical hindered rotor partition function was described by Knyazev.⁴⁷ Unfortunately there is no analytical expression for the density of states corresponding to the semi-classical hindered rotor partition function, eqn (18), since there is no analytical inverse Laplace transform for the denominator in the quantum harmonic oscillator partition function. However, the Beyer-Swinehart method gives the exact density of states of the quantum harmonic oscillator. By replacing this denominator with a truncated Taylor expansion, the result from Knyazev can be extended to include the density of states corresponding to a semi-classical hindered rotor partition function:

$$\rho_{\text{HR}_{\text{semi}}}(E) = \begin{cases} \frac{2K(\sqrt{\tilde{E}/V_0})}{\pi\sigma\sqrt{B_r V_0}} & \text{for } E < \max(V_0, h\nu/2) \\ \frac{2K(\sqrt{V_0/\tilde{E}})}{\pi\sigma\sqrt{B_r \tilde{E}}} & \text{otherwise} \end{cases} \quad (19)$$

Above, $\tilde{E} \equiv E - h\nu/2$, and K is the complete elliptical integral of the first kind. The density of states eqn (19) is calculated for each hindered rotor. The combined density states for multiple hindered internal rotors is calculated numerically by repeated application of the convolution integral. Similarly, the density of states for internal rotors and active external rotation is calculated by convoluting the density of states for the active K-rotor with the density of states for all the hindered rotors. Finally, following the method of Astholz,^{20,34} the Beyer-Swinehart algorithm¹⁸ is initialized with the K-rotor hindered-rotor density of states, which convolutes this density of states with the vibrational modes. The result is the complete rovibrational density of states, $\rho(E)$.

Accurate determination of the vibrational frequencies and the potentials for internal rotation requires a three-dimensional geometry and a quantum calculation, which is not routinely available in the RMG framework. Instead, these parameters are estimated from the heat capacity. The heat capacity may be separated into contributions from external translation, external rotation, and the internal rovibrational degrees of freedom:

$$C_V^{\text{total}} = C_V^{\text{translation}} + C_V^{\text{rotation}} + C_V^{\text{rovibration}}$$

In this presentation, the K-rotor is included in the external rotation. The vibrational frequencies and hindered rotor parameters can be calculated by fitting a function to the vibrational contribution to the heat capacity. In most cases, however, the number of vibrational modes exceeds number of known values of the heat capacity as a function of temperature, and therefore these frequencies cannot be determined uniquely. Two methods for approximating the vibrational frequencies are described below.

4.1 The three frequency model

The three frequency model for estimating $\rho(E)$ from heat capacity data was proposed by Bozzelli, Chang, and Dean.³³ This approach assumes that the molecule can be described by the HO model. The heat capacity of a single mode as a function of its HO frequency is given by the following formula from statistical mechanics:

$$\frac{C_V^{\text{HO}}(T; \nu_i)}{R} = e^{\nu_i/k_B T} \left(\frac{\nu_i/k_B T}{e^{\nu_i/k_B T} - 1} \right)^2 \quad (20)$$

where ν_i is the vibrational frequency for the i th mode. Bozzelli *et al.* fit three vibrational frequencies to the heat capacity at seven temperatures (300, 400, 500, 600, 800, 1000, 1500 K), using the following formula:

$$C_V^{\text{vibration}}(T) = \sum_{i=1}^3 g_i C_V^{\text{HO}}(T; \nu_i) \quad (21)$$

$$g_3 = s - g_1 - g_2$$

where s is the number of vibrational modes (*e.g.* $3N - 6$ for a nonlinear molecular with no internal rotors). In the original implementation, the three degeneracies g_i were allowed to be non-integers. Later implementations of this approach forced the degeneracies to be integers, so $\rho(E)$ could be computed efficiently using the Beyer-Swinehart algorithm, and that is what we have done here.²⁷ Thus, five parameters are fit to the heat capacity data: three frequencies and two (integer) degeneracies. This method is simple and computationally efficient. However, there are two problems with this approach. First, the resulting pseudo-frequencies may not be physically realistic. Second, it assumes that the heat capacity necessarily increases monotonically with temperature. For molecules with hindered internal rotors, the true heat capacity may actually have a local maximum with respect to temperature.

4.2 The functional-group frequency model

An alternative to the three-frequency model, inspired by infrared spectroscopy, is to assign frequencies to functional groups.

For each functional group, the number of characteristic frequencies is limited to $3N_{\text{atoms}} - N_{\text{rotors}} - 6$. For example, a heavy atom attached to a methyl group contains 5 atoms and one single bond between heavy atoms, and thus has 9 characteristic frequencies: 3 C–H stretches, 1 R–C stretch, 2 R–C–H bending modes, 2 R–C–H rocking modes, and 1 umbrella mode. Additionally, one of the six subtracted modes is the overall rotation of the methyl group, and this torsional mode is treated as a hindered internal rotor. RMG automatically recognizes that single bonds between heavy atoms that are not within a ring structure may be hindered rotors. Each vibrational mode is associated with a range of frequencies, and RMG automatically selects the appropriate number of frequencies, taking them to be evenly spaced between the two bounds. Thus, for a molecule with two methyl groups, RMG would assume 6 C–H stretching frequencies that are evenly spaced between 2750–2850 cm⁻¹. A list of the functional groups, the corresponding range of frequencies, and the number of frequencies per group are available in the Supporting Materials.[†] By summing over the functional groups, the model predicts the majority of the HO vibrational frequencies. The heat capacity for the remaining degrees of freedom (*e.g.* vibrational frequencies not explicitly accounted for, as well as hindered rotors) is obtained by subtracting the heat capacity of the functional group frequencies from the rovibrational heat capacity:

$$C_V^{\text{remaining degrees of freedom}}(T) = C_V^{\text{rovibration}}(T) - \sum_{i=1}^{N_{\text{functional group frequencies}}} C_V^{\text{HO}}(T; \nu_i) \quad (22)$$

The remaining vibrational frequencies and the parameters for hindered internal rotors are determined from this heat capacity data. Unlike the three frequency model, this method includes a function for hindered internal rotors and therefore can accommodate heat capacities with a local maximum.

$$C_V^{\text{remaining degrees of freedom}} = \sum_{i=1}^{N_{\text{remaining}}} C_V^{\text{HO}}(T; \nu_i) + \sum_{j=1}^{N_{\text{rotors}}} C_V^{\text{HR}}(T; V_0, \nu_j)$$

where $N_{\text{remaining}} \equiv 3N - 6 - N_{\text{rotors}} - N_{\text{functional group frequencies}}$. The heat capacity of a hindered internal rotor is derived from the semi-classical model for the hindered rotor partition function, eqn (18). The heat capacity for this partition function is:

$$\begin{aligned} \frac{C_V^{\text{HR}}(T; V_0, \nu)}{R} &= \frac{d}{dT} \left(T^2 \frac{d \ln Q_{\text{HR semi}}}{dT} \right) \\ &= u^2 \frac{d^2 \ln Q_{\text{HR semi}}}{du^2} \\ &\approx -\frac{1}{2} + u \left[u - \frac{I_1(u)}{I_0(u)} - u \left(\frac{I_1(u)}{I_0(u)} \right)^2 \right] \\ &\quad + \left(\frac{2uxe^{-\alpha u}}{1 - e^{-2\alpha u}} \right)^2 \end{aligned} \quad (23)$$

Although the functional group method will estimate most of the frequencies, it may be the case that not all of the remaining parameters can be determined uniquely from the heat capacity data. Typically, the $C_V(T)$ data from a group additivity calculation has enough information content to determine about six parameters. If there are more than six unknown parameters, then some of the rigid-rotor harmonic-oscillator frequencies or the hindered rotor parameters (or in some cases both) may be lumped into a single frequency or hindered rotor, respectively, so that no more than six parameters are fitted:

$$\begin{aligned} \sum_{i=1}^N C_V^{\text{HO}}(T; \nu_i) &\rightarrow NC_V^{\text{HO}}(T; \nu) \\ \sum_{j=1}^M C_V^{\text{HR}}(T; V_0, \nu_j) &\rightarrow MC_V^{\text{HR}}(T; V_0, \nu) \end{aligned}$$

Alternatively, one can sometimes obtain a better fit and/or fit additional parameters directly by first fitting an interpolation model to the known $C_V(T)$ points, and choosing a higher density of points to use when performing the frequency fitting. We use this approach with the acetyl + oxygen case study discussed in Section 6, using the Wilhoit polynomial model^{48,49} to directly fit up to ten parameters.

In order to make the resulting parameters as realistic as possible, the fitted parameters are given physically meaningful constraints: the harmonic oscillator frequencies are bound between 180 and 4000 cm⁻¹, the hindered rotor torsional frequencies are bound between 40 and 600 cm⁻¹, and the hindered rotor barrier heights are bound between 10 and 10 000 cm⁻¹. The FORTRAN 90 code DQED^{50,51} was used for the bounded, constrained, non-linear least-squares fitting procedure.

5 Methods for estimating phenomenological rate coefficients

The objective of each of the methods described in the following sections is to transform the large master equation matrix from eqn (4) into a small number of phenomenological rate coefficients $k(T, P)$. All of the methods share a common formalism in that they seek to express the population distribution vector \mathbf{p}_i for each unimolecular isomer i as a linear combination of the total populations $x_j(t)$ and $y_{mA}(t)$, y_{mB} of unimolecular isomers C_j and bimolecular reactant channels $A_m + B_m$:

$$\mathbf{p}_i(t) = \sum_{j=1}^{N_{\text{isom}}} x_j(t) \mathbf{u}_{ij} + \sum_{m=1}^{N_{\text{reac}}} y_{mA}(t) y_{mB} \mathbf{v}_{im} \quad (24)$$

The vectors \mathbf{u}_{ij} represent the portion of the population distribution of energy states of unimolecular isomer i that tracks the population of isomer j . In the modified strong collision and reservoir state methods, this is because the energy levels of isomer i are in pseudo-steady-state relationships with isomer j . The interpretation is a bit different in the chemically-significant eigenvalues approach, but the form of the equations is the same. Similarly, the vectors

\mathbf{v}_{im} represent the population distribution of energy states of unimolecular isomer i that tracks the population of bimolecular reactant channel m . Both \mathbf{u}_{ij} and \mathbf{v}_{im} are independent of time.

Notice that, even for the best \mathbf{u}_{ij} and \mathbf{v}_{im} , eqn (24) is an approximation to the true $\mathbf{p}_i(t)$ given by the solution of the full master equation. This approximation is a form of model reduction, and is done so that we can use the conventional phenomenological kinetics equations, and do not have to keep track of the evolving energy distributions of each species.

The phenomenological rate coefficient matrix can be constructed from the full master equation matrix \mathbb{M} shown in eqn (4) and the population distribution vectors \mathbf{u}_{ij} and \mathbf{v}_{im} :

$$k_{ij}(T, P) = \sum_{s=1}^{N_{\text{grains}}} (\mathbf{M}_i \mathbf{u}_{ij})_s + \sum_{\ell \neq i}^{N_{\text{isom}}} \sum_{s=1}^{N_{\text{grains}}} (\mathbf{K}_{i\ell} \mathbf{u}_{\ell j})_s \quad (25a)$$

$$\begin{aligned} k_{im}(T, P) = & \sum_{s=1}^{N_{\text{grains}}} (\mathbf{M}_i \mathbf{v}_{im})_s + \sum_{\ell \neq i}^{N_{\text{isom}}} \sum_{s=1}^{N_{\text{grains}}} (\mathbf{K}_{i\ell} \mathbf{v}_{\ell m})_s \\ & + \sum_{s=1}^{N_{\text{grains}}} (\mathbf{F}_{im} \mathbf{b}_m)_s \end{aligned} \quad (25b)$$

$$k_{nj}(T, P) = \sum_{\ell=1}^{N_{\text{isom}}} \mathbf{g}_{n\ell} \cdot \mathbf{u}_{\ell j} \quad (25c)$$

$$k_{nm}(T, P) = \sum_{\ell=1}^{N_{\text{isom}}} \mathbf{g}_{n\ell} \cdot \mathbf{v}_{\ell m} \quad (25d)$$

Above, the indices i and j represent unimolecular isomers of the initial adduct, m represents bimolecular reactants, n represents bimolecular reactants and products, and s represents an energy grain. Thus, the rate coefficients above are for isomerization, association, dissociation, and bimolecular reactions $A_m + B_m \rightarrow A_n + B_n$, respectively. The various methods provide different values of \mathbf{u}_{ij} and \mathbf{v}_{im} .

5.1 The modified strong collision method

As described by Chang, Bozzelli, and Dean,³¹ the modified strong collision method utilizes a greatly simplified collision model that allows for a decoupling of the energy grains. In the simplified collision 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.^{30,52} By approximating the reactive populations as existing in pseudo-steady state, a matrix equation is formed at each energy through which \mathbf{u}_{ij} and \mathbf{v}_{im} vectors, 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

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} p_1(E, t) \\ p_2(E, t) \\ \vdots \end{bmatrix} = & \begin{bmatrix} m_1(E) & k_{12}(E) & \dots \\ k_{21}(E) & m_2(E) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} p_1(E, t) \\ p_2(E, t) \\ \vdots \end{bmatrix} \\ & + \sum_{i=1}^{N_{\text{isom}}} x_i(t) \begin{bmatrix} \omega_i(T, P) \beta_i(T) b_i(E; T) \delta_{ii} \\ \omega_i(T, P) \beta_i(T) b_i(E; T) \delta_{ii} \\ \vdots \end{bmatrix} \\ & + \sum_{n=1}^{N_{\text{reac}}} y_{nA}(t) y_{nB} \begin{bmatrix} f_{1n}(E) b_n(E; T) \\ f_{2n}(E) b_n(E; T) \\ \vdots \end{bmatrix} \end{aligned} \quad (26a)$$

$$\begin{aligned} \frac{d}{dt} \mathbf{p}(E, t) = & \mathbf{L}(E) \mathbf{p}(E, t) + \sum_{i=1}^{N_{\text{isom}}} x_i(t) \mathbf{z}_i(E) \\ & + \sum_{n=1}^{N_{\text{reac}}} y_{nA}(t) y_{nB} \mathbf{w}_n(E) \end{aligned} \quad (26b)$$

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

$$\begin{aligned} m_i(E) = & -\omega_i(T, P) \beta_i(T) - \sum_{j \neq i}^{N_{\text{isom}}} k_{ji}(E) \\ & - \sum_{n=1}^{N_{\text{reac}}+N_{\text{prod}}} g_{ni}(E) \end{aligned} \quad (27)$$

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

$$\mathbf{L}(E) \mathbf{p}(E) = - \sum_{i=1}^{N_{\text{isom}}} x_i(t) \mathbf{z}_i(E) - \sum_{n=1}^{N_{\text{reac}}} y_{nA}(t) y_{nB} \mathbf{w}_n(E) \quad (28)$$

Solving these gives

$$\begin{aligned} \mathbf{p}^{(i)}(E) &= -[\mathbf{L}(E)]^{-1} \mathbf{z}_i(E) \\ \mathbf{p}^{(n)}(E) &= -[\mathbf{L}(E)]^{-1} \mathbf{w}_n(E) \end{aligned} \quad (29)$$

where $\mathbf{p}^{(i)}(E)$ 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:

$$k_{ij}(T, P) = \omega_i(T, P) \beta_i(T) \sum_{r=r_{0,i}}^{N_{\text{grains}}} p_i^{(j)}(E_r) \quad (30a)$$

$$k_{nj}(T, P) = \sum_{i=1}^{N_{\text{isom}}} \sum_{r=r_{0,i}}^{N_{\text{grains}}} g_{ni}(E) p_i^{(j)}(E_r) \quad (30b)$$

$$k_{im}(T, P) = \omega_i(T, P)\beta_i(T) \sum_{r=r_0,i}^{N_{\text{grains}}} p_i^{(m)}(E_r) \quad (30\text{c})$$

$$k_{nm}(T, P) = \sum_{j=1}^{N_{\text{isom}}} \sum_{r=r_0,j}^{N_{\text{grains}}} g_{nj}(E_r) p_j^{(m)}(E_r) \quad (30\text{d})$$

The indices above have the same meaning as in eqn (25): i and j represent unimolecular isomers, m represents bimolecular reactants, n represents 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, eqn (24):

$$u_{ij}(E_r) = \begin{cases} b_i(E_r)\delta_{ij} & E_r < E_{j,\text{crit}} \\ p_i^{(j)}(E_r) & E_r \geq E_{j,\text{crit}} \end{cases} \quad (31\text{a})$$

$$v_{im}(E_r) = \begin{cases} 0 & E_r < E_{m,\text{crit}} \\ p_i^{(m)}(E_r) & E_r \geq E_{m,\text{crit}} \end{cases} \quad (31\text{b})$$

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 eqn (25) gives identical $k(T,P)$ values as those in eqn (30c), (30d), (30a) and (30b).

5.2 The reservoir state method

An alternative to approximating collisional stabilization as a single-step process, proposed by Green and Bhatti,³² 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 $k_B T$ 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 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. A more detailed treatment, including a demonstration that the resulting $k(T,P)$ values are in fact independent of the initial condition, is provided in the Supporting Materials.[†]

First, a Boltzmann distribution \mathbf{b}_i^r is imposed on the reservoir grains \mathbf{p}_i^r for each isomer i , *i.e.*

$$\mathbf{p}_i^r = x_i(t)\mathbf{b}_i^r \quad (32)$$

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.* $\sum_{s=1}^{N_{\text{res},i}} (\mathbf{b}_i^r)_s < 1$. The quantity \mathbf{b}_i^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^a for each isomer is

$$\frac{d}{dt} \begin{bmatrix} \mathbf{p}_1^a \\ \mathbf{p}_2^a \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1^{\text{aa}} & \mathbf{K}_{12}^a & \dots \\ \mathbf{K}_{21}^a & \mathbf{M}_2^{\text{aa}} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{p}_1^a \\ \mathbf{p}_2^a \\ \vdots \end{bmatrix} + \sum_{j=1}^{N_{\text{isom}}} x_j(t) \begin{bmatrix} \mathbf{M}_1^{\text{ar}} \mathbf{b}_1^r \delta_{1j} \\ \mathbf{M}_2^{\text{ar}} \mathbf{b}_2^r \delta_{2j} \\ \vdots \end{bmatrix} + \sum_{m=1}^{N_{\text{reac}}} y_{mA}(t) y_{mB} \begin{bmatrix} \mathbf{F}_{1m}^a \mathbf{b}_m^a \\ \mathbf{F}_{2m}^a \mathbf{b}_m^a \\ \vdots \end{bmatrix} \quad (33\text{a})$$

$$\frac{d}{dt} \mathbf{p}^a = \mathbf{L} \mathbf{p}^a + \sum_{j=1}^{N_{\text{isom}}} x_j(t) \mathbf{z}_j + \sum_{m=1}^{N_{\text{reac}}} y_{mA}(t) y_{mB} \mathbf{w}_m \quad (33\text{b})$$

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

$$\mathbf{L} \mathbf{p}^a = - \sum_{j=1}^{N_{\text{isom}}} x_j(t) \mathbf{z}_j - \sum_{m=1}^{N_{\text{reac}}} y_{mA}(t) y_{mB} \mathbf{w}_m \quad (34)$$

for which the solution is

$$\begin{aligned} \mathbf{p}^{a(i)} &= -\mathbf{L}^{-1} \mathbf{z}_i \\ \mathbf{p}^{a(m)} &= -\mathbf{L}^{-1} \mathbf{w}_m \end{aligned} \quad (35)$$

where $\mathbf{p}^{a(i)}$ is the pseudo-steady active-state population of each isomer resulting from the thermal activation of isomer i , and $\mathbf{p}^{a(m)}$ 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

$$k_{ij}(T, P) = \sum_r (\mathbf{M}_i^{\text{ra}} \mathbf{p}_i^{a(j)})_r \quad (36\text{a})$$

$$k_{nj}(T, P) = \sum_{i=1}^{N_{\text{isom}}} \mathbf{g}_{ni} \cdot \mathbf{p}_i^{a(j)} \quad (36\text{b})$$

$$k_{im}(T, P) = \sum_r (\mathbf{M}_i^{\text{ra}} \mathbf{p}_i^{a(m)})_r \quad (36\text{c})$$

$$k_{nm}(T, P) = \sum_{i=1}^{N_{\text{isom}}} \mathbf{g}_{ni} \cdot \mathbf{p}_i^{a(m)} \quad (36\text{d})$$

where $\mathbf{p}_i^{a(j)}$ and $\mathbf{p}_i^{a(m)}$ contain the components of $\mathbf{p}^{a(j)}$ and $\mathbf{p}^{a(m)}$ 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.* $y_{mA} \ll y_{mB}$) 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_{mA} and y_{mB} are small enough that $y_{mA}y_{mB}\mathbf{F}_{im}\mathbf{b}_m$ is small.

The terms of the pseudo-steady state vector \mathbf{p}_i^a and the reservoir populations \mathbf{b}_i^r can be used to construct the \mathbf{u}_{ij} and \mathbf{v}_{im} vectors for use in eqn (24):

$$\mathbf{u}_{ij} = \begin{bmatrix} \delta_{ij}\mathbf{b}_j^r \\ \mathbf{p}_i^{a(j)} \end{bmatrix} \quad \mathbf{v}_{im} = \begin{bmatrix} \mathbf{0} \\ \mathbf{p}_i^{a(m)} \end{bmatrix} \quad (37)$$

Again, using the above expressions with eqn (25) gives identical $k(T,P)$ values as those from eqn (36).

Finally, a note on efficient solving of eqn (34). Collisional energy transfer models favor small transfers of energy over large ones. For the single exponential down model, $P(E,E') \rightarrow 0$ exponentially as $|E' - E| \rightarrow \infty$. 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 Fig. 2 for a three-isomer network. The grain-major indexing scheme, Fig. 2b, yields a matrix with a much narrower bandwidth than the isomer-major indexing scheme, Fig. 2a.

5.3 The chemically-significant eigenvalues method

The following description of the chemically-significant eigenvalues method is based on the works of Pilling and Robertson^{53,54} and Miller and Klippenstein.^{22,55,56} First, the eigenvalues and eigenvectors of the master equation matrix in eqn (4), denoted here as \mathbb{M} , are computed:

$$\mathbb{M} = \mathbf{Q}\Lambda\mathbf{Q}^{-1} \quad (38)$$

The eigenvalue decomposition is usually performed after symmetrizing \mathbb{M} , which ensures that the calculated eigenvalues Λ 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. 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:

$$Z_{i\ell} = \sum_{s \in i} Q_{s\ell} \quad i = 1, 2, \dots, N_{\text{isom}}, \quad \ell = 1, 2, \dots, N_{\text{chem}} \quad (39)$$

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

$$Z_{n\ell} = Q_{n\ell} \quad n = 1, 2, \dots, N_{\text{reac}}, \quad \ell = 1, 2, \dots, N_{\text{chem}} \quad (40)$$

Physically the matrix \mathbf{Z} corresponds to the eigenvector matrix obtained from diagonalizing the matrix of phenomenological rate coefficients \mathbf{K} , *i.e.* $\mathbf{K} = \mathbf{Z}\Lambda'\mathbf{Z}^{-1}$. 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:

$$k_{ij}(T, P) = \sum_{\ell=1}^{N_{\text{chem}}} \lambda_\ell Z_{i\ell} Z_{j\ell}^{-1} \quad i, j \in 1, 2, \dots, N_{\text{isom}} + N_{\text{reac}} \quad (41)$$

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

$$Y_{m\ell} = \sum_{i=1}^{N_{\text{isom}}} \mathbf{g}_{mi} \cdot \mathbf{q}_\ell^{(i)} \quad m = 1, 2, \dots, N_{\text{prod}} \quad (42)$$

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

$$k_{mj}(T, P) = \sum_{\ell=1}^{N_{\text{csc}}} Y_{m\ell} Z_{j\ell}^{-1} \quad j = 1, 2, \dots, N_{\text{isom}} + N_{\text{reac}} \quad (43)$$

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*

$$\begin{aligned} u_{ij}(E_r) &= (\mathbf{Q}_{\text{chem}}\mathbf{Z}^{-1})_{(i,r),j} \\ v_{im}(E_r) &= (\mathbf{Q}_{\text{chem}}\mathbf{Z}^{-1})_{(i,r),m} \end{aligned} \quad (44)$$

where \mathbf{Q}_{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

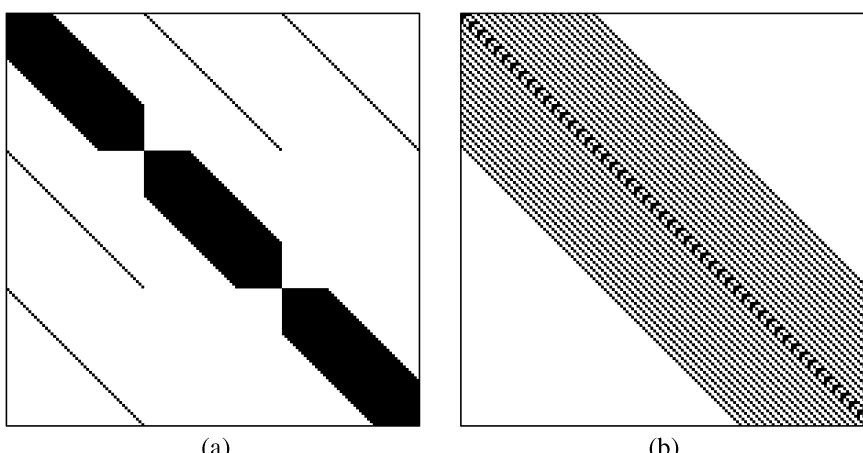


Fig. 2 A typical sparsity pattern for an active-state matrix of a three-isomer network using (a) an isomer-major indexing scheme and (b) an energy grain-major indexing scheme. The latter gives a significantly more banded matrix, which can be used to accelerate the reservoir state linear solve.

matrix \mathbb{M} depends on y_{mB} , in the eigenvalue method the $k(T,P)$, \mathbf{u}_{ij} , and \mathbf{v}_{im} depend on y_{mB} . However, in most cases only small y_{mB} 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 paper.

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. A brief discussion of the initial-rate method is provided in the Supporting Materials.[†]

6 Case study: the acetyl + oxygen system

As an application of the various methods for estimating the density of states and the phenomenological rate coefficients, we turn our attention to the chemically-activated reaction of acetyl radical with oxygen. This system is important to atmospheric chemistry as a step in the conversion of acetaldehyde to peroxyacetyl nitrate (PAN), the latter of which is a

secondary air pollutant.⁵⁷ It is also potentially important in the ignition chemistry of ethanol. The potential energy surface for this reaction was computed originally by Lee, Chen, and Bozzelli.⁵⁸ The values in Fig. 3 are updated results, described below. Our updated values are consistent with those recently published by Villano *et al.*⁵⁹

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.⁶⁰ Later experimental efforts generally agree on the relevant pressure range being about 0.001 to 1 bar at 300 K.^{61–65} Recent experiments have confirmed that the α -lactone is the carbon-containing product associated with OH regeneration.⁶⁶

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.⁵⁸ 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 parameters to vary so as to match the experimental data, and thus were able to match the regeneration of OH.⁶⁷

The calculations of Lee *et al.* 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.⁶³ 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. Additionally, Hou *et al.* found a saddle point geometry for the association reaction and used a rigid-rotor harmonic oscillator model to compute the association rate coefficient. Depending upon which method they chose—MP2, CCSD, G3MP2, CASPT2—the barrier height

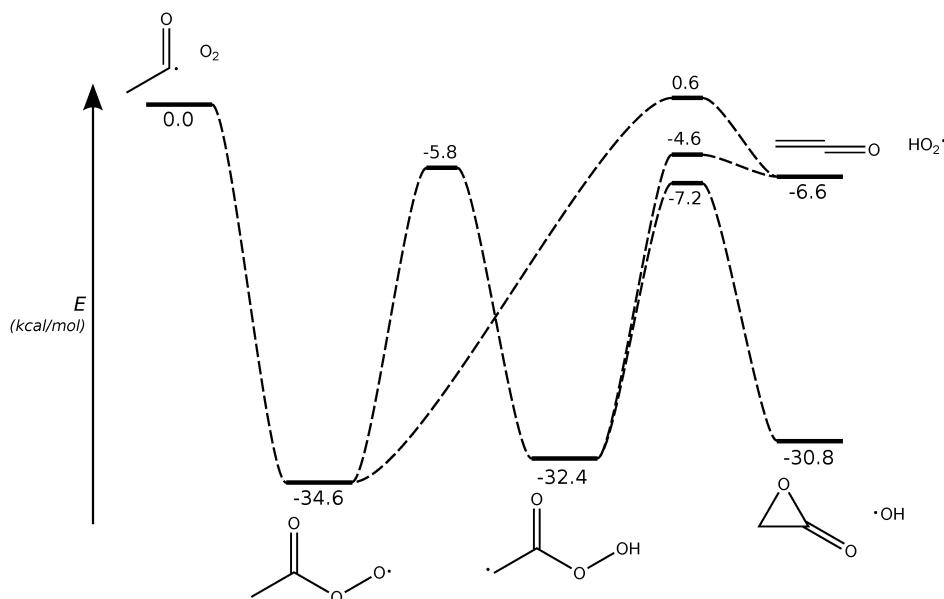


Fig. 3 The RQCISD(T)/CBS//B3LYP/6-311++G(d,p) potential energy surface for the acetyl oxidation network.

varied between $+13 \text{ kcal mol}^{-1}$ and -1 kcal mol^{-1} , relative to the reactants. Given that the C–O bond distance in the saddle point was 2.5 \AA , the CASPT2(7,7) ΔE_0 of $1.7 \text{ kcal mol}^{-1}$ is the most reliable. At these distances the modes between the two reacting moieties will be highly anharmonic, and a rigid-rotor harmonic oscillator model is ill-advised. Using DFT, we were able to locate a saddle point for this reaction as well, but there were four vibrational modes with frequencies less than 20 cm^{-1} . These coupled, anharmonic modes correlate to the relative orientation and separation of the two fragments. A better way to treat this entrance channel would be to use variable reaction coordinate transition state theory.^{68–70} Ideally the interaction potential would be computed at the CASPT2 level, with an active space of at least 7 electrons and five orbitals—six electrons in four orbitals for O_2 , and one electron in one orbital for the unpaired electron in acetyl, with the possibility of adding additional electrons and orbitals to account for the $\text{C}=\text{O}$ π -bond and/or any oxygen lone pair electrons in acetyl. However, these calculations are beyond the scope of this work, and we have adopted the high-pressure limit for acetyl + O_2 recommended by Lee *et al.* for simplicity.

We performed our own RQCISD(T)/CBS//B3LYP/6-311++G(d,p) calculations to generate the PES in Fig. 3

using the methodology of Goldsmith *et al.*⁷¹ All DFT and CBS-QB3 calculations were done using Gaussian03.⁷² All MP2 and QCISD(T) calculations were done using MOLPRO.⁷³ The high-pressure limit for each reaction rate coefficient except acetyl + O_2 was computed using variational transition state theory, as implemented in Variflex.²⁵ These results are provided in Table 1. Although these are high-level calculations, it is doubtful that these numbers are the final word on this important reaction system.

6.1 Density of states comparison

Fig. 4 shows the density of states, partition function, and heat capacity for acetylperoxy and hydroperoxylvinoxy as calculated from vibrational and hindered rotor data estimated using the three-frequency and group frequency approaches and calculated from quantum chemistry, with the goal of judging how the two approximate models compare to the detailed quantum chemistry model. The parameters for the internal modes as determined for each model are given in tables in the Supporting Materials.† The quantum chemistry parameters were obtained from geometry optimization and frequency calculations using density functional theory with the B3LYP functional and 6-311++G(d,p) basis set, as described in the preceding section. The three-frequency and functional group-frequency methods estimate the internal (vibrational and hindered rotor) modes, and a K-rotor for the coupled external rotational mode. For the purposes of this comparison, we do the same with the quantum chemistry calculations.

Both acetylperoxy and hydroperoxylvinoxy have 18 internal degrees of freedom to be represented by the three pseudofrequency-degeneracy pairs. Acetylperoxy contains 16 vibrational modes and two hindered rotor modes, while hydroperoxylvinoxy contains 15 vibrations and three hindered rotors. Nonetheless, the three frequency method will represent all of these as three pseudofrequency-degeneracy pairs. As shown in Fig. 4, we see that this method performs somewhat

Table 1 Calculated high-pressure limit rate coefficients for acetyl + O_2 ^a

Reaction	A	n	E_a
Acetyl + $\text{O}_2 \rightarrow$ acetylperoxy ^b	4.4×10^{-12}	0.0	0.0
Acetylperoxy \rightarrow hydroperoxyl-vinoxy	2.3×10^9	0.75	23.2
Acetylperoxy \rightarrow ketene + HO_2	2.6×10^9	1.2	34.1
Hydroperoxyl-vinoxy \rightarrow ketene + HO_2	5.3×10^{16}	-1.0	29.5
Hydroperoxyl-vinoxy \rightarrow lactone + OH	1.9×10^{17}	-1.1	27.2

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

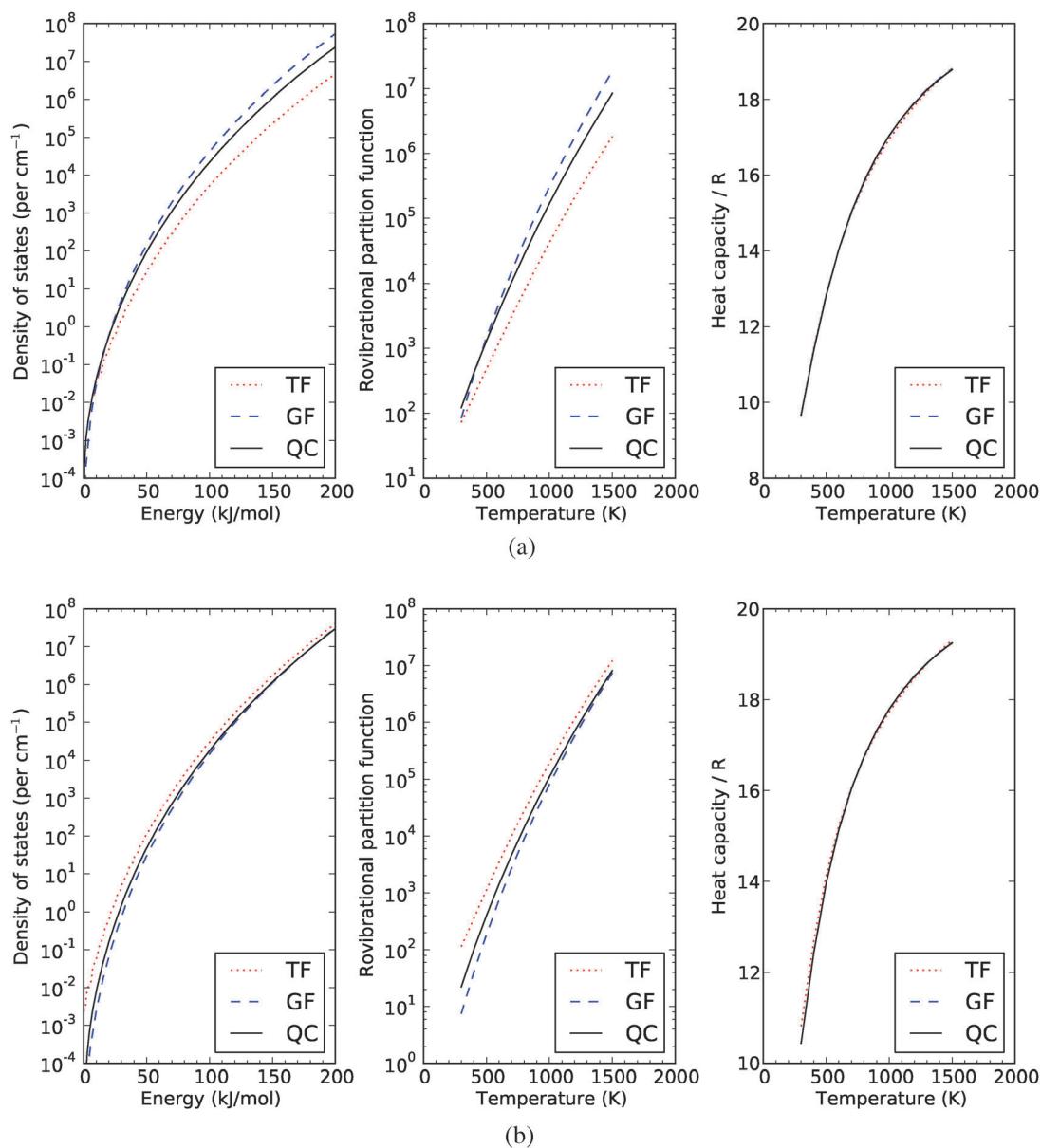


Fig. 4 Comparison of densities of states, partition functions, and heat capacities for internal degrees of freedom for the (a) acetylperoxy and (b) hydroperoxylvinoxy species using a three-frequency model (TF), group frequency model (GF), and quantum chemistry calculations (QC). Both the TF and GF models reproduce the QC heat capacity data, but GF more consistently predicts the density of states than TF.

better for hydroperoxylvinoxy than for acetylperoxy. Our quantum chemistry-calculated barrier heights for the acetylperoxy hindered rotors are 1.22 and 6.12 kcal mol⁻¹, while those for hydroperoxylvinoxy are 4.78, 6.26, and 14.7 kcal mol⁻¹. The high barriers to internal rotation make the harmonic oscillator approximation fairly accurate for hydroperoxylvinoxy.

For acetylperoxy, the group frequency approach provides twelve of the eighteen internal modes as characteristic vibrational frequencies: nine from the CH₃ group and three from the COO[•] group. For hydroperoxylvinoxy, eleven of the eighteen internal modes are set *via* characteristic frequencies: six from the CH₂[•] group and five from the COOH group. In both cases the remaining modes—four vibrations and two hindered rotors for acetylperoxy, four vibrations and three

hindered rotors for hydroperoxylvinoxy—are few enough that their parameters can be fit directly to the remaining heat capacity data. The results in Fig. 4 show an improved estimate of the density of states for both isomers when compared to the three frequency method. In particular, the fitted hindered rotor barrier heights for acetylperoxy using the group frequency method are 0.99 and 4.68 kcal mol⁻¹, correctly predicting a low-barrier and a moderate-barrier hindered rotation.

6.2 Comparison of the phenomenological rate coefficient 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 eqn (24) 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 + oxygen reaction network.

6.2.1 Rate coefficients. 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 same $\rho(E)$ and $k(E)$ from quantum chemical calculations, the same energy discretization (constant 0.5 kcal mol⁻¹ grain spacing), and $\langle \Delta E_{\text{down}} \rangle = 200(T/300)^{0.85} \text{ cm}^{-1}$.⁷⁴ Fig. 5 shows the $k(T,P)$ values for $\text{CH}_3\text{CO} + \text{O}_2 \rightarrow$ products at a variety of temperatures and pressures. (The equivalent values for acetylperoxy → products and hydroperoxylvinoxy → products are available in the Supporting Materials.[†]) 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 + HO₂ channel is dominant at high temperature. All of this is consistent with our physical understanding of the system.

The modified strong collision, reservoir state, and chemically-significant eigenvalue 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 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.

6.2.2 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 $y_{mA}(t)$ 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 + HO₂ 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 chemically-significant eigenvalue calculation.

Fig. 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. For the modified strong collision and reservoir state 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 modified strong collision method does deviate from the others slightly (most notably for hydroperoxylvinoxy), but remains within an order of magnitude of the true profiles.

Also interesting is the effect of high temperature, demonstrated using conditions of 1500 K and 1 bar. This time the primary product is the ketene channel, with relatively little of each isomer produced along the way. All of the reactant and product channels are well-approximated by all of the methods, but show some disagreement for the total concentrations of the isomers. Of the three methods, the chemically-significant eigenvalues method is clearly the most accurate in predicting the isomer profiles, while the modified strong collision and reservoir state methods are only slightly worse.

6.2.3 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 provides the full population distributions $p_i(E,t)$ for each isomer, while the approximate solutions only provide the total populations $x_i(t)$. In comparing concentration profiles, we simply summed the population distributions to obtain $x_i(t)$, but we can also do the reverse: reconstruct the approximate $p_i(E,t)$ from $x_i(t)$ and $y_{mA}(t)$ y_{mB} using eqn (24). Thus, we can also examine how the approximate population distributions evolve in time to see in more detail each method’s strengths and weaknesses.

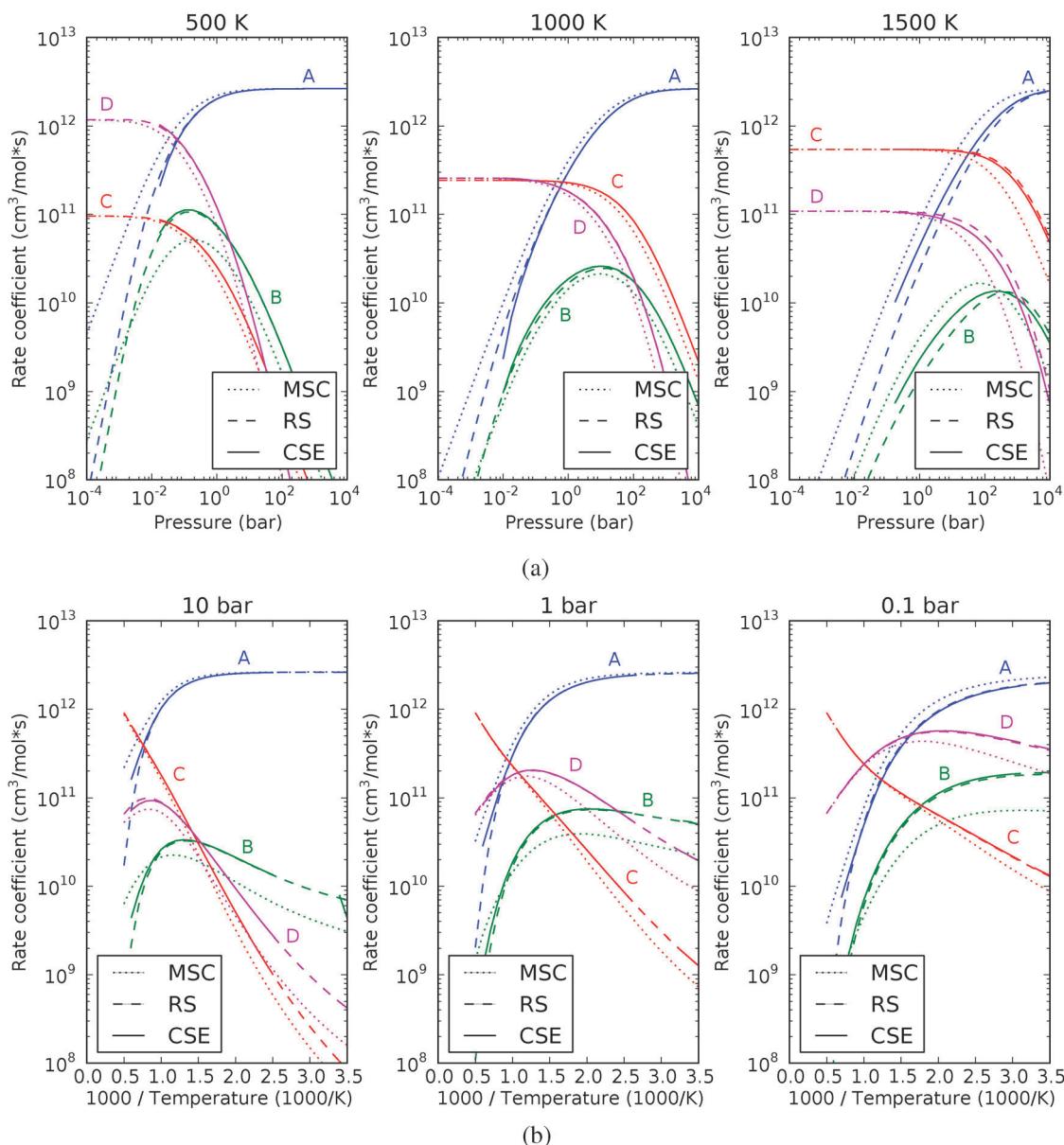


Fig. 5 Comparison of rate coefficients *versus* (a) pressure and (b) temperature for $\text{CH}_3\text{CO} + \text{O}_2 \rightarrow$ products estimated using the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods. In the plots, A = acetylperoxy, B = hydroperoxyl-vinoxy, 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.

Fig. 7 shows the approximate population distributions $p(E,t)$ *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 collision 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 chemically-significant eigenvalue method is the first to accurately match the entirety of the “true” population distribution. As the eigenvalue method makes no assumption about the nature of the energies, it is able to capture the non-Boltzmann, non-steady state behavior of the high reservoir and near threshold grains (about -15 to 0 kcal mol⁻¹). (A side effect of this is that the eigenvalue method predicts that the lowest-energy grains have slightly negative populations at short times; this is why the distribution drops sharply off at certain times.) The reservoir state method distributions have (by assumption) a Boltzmann distribution at low energies, which in reality is only established after many collision times. However, only the most populated low-energy

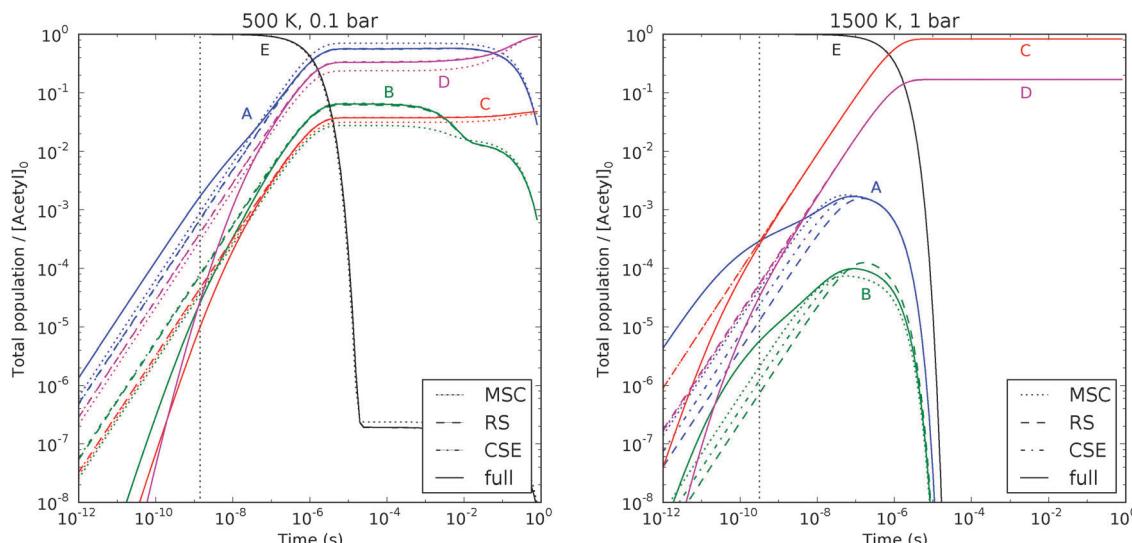


Fig. 6 Concentration profiles at low-temperature and high-temperature conditions, comparing the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods with the full master equation solution, for the acetyl + oxygen system as shown in Fig. 3. In the plots, A = acetylperoxy, B = hydroperoxylvinoxy, C = ketene + HO₂, D = lactone + OH, and E = acetyl + O₂. The reservoir state and chemically-significant eigenvalue methods both perform very well at low temperatures. All three methods are comparably accurate at high temperatures even though they disagree with the full solution at short times. The vertical line indicates the average time between collisions with the bath gas.

grains need to be Boltzmann-distributed before the total population is predicted accurately. The modified strong collision method population distributions are quite similar to the reservoir state method distributions despite the significantly different collision models used. The main discrepancy is in the low reactive energies (around -6 to 0 kcal mol⁻¹ below the reactant energy, but above the lowest isomerization barrier), which the reservoir state method predicts more accurately since it allows for collisional energy transfer between grains.

From the plots in Fig. 6 and 7, we see that the reservoir state method performs most consistently at low temperatures, and is still competitive at moderately high temperatures. The kinetics of the major pathways are consistently predicted across all methods; the disagreement appears only in the minor pathways. The chemically-significant eigenvalue method also performs very well under the conditions shown—often equal to or better than the reservoir state method—but would not be successful at lower pressures due to the blending of the chemical eigenmodes into the internal energy eigenmodes, as shown in Fig. 8.

6.2.4 Eigenvalue analysis. In Fig. 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 Fig. 8a and b 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 Fig. 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.^{75,76} 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 + oxygen.

6.3 Perturbations to the acetyl + oxygen potential energy surface

Although there are a few regions in Fig. 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 what follows. For all perturbations we have provided plots of $k(T,P)$ versus pressure for rate coefficients corresponding to acetyl + oxygen as the reactants. Many other plots of $k(T,P)$ values and some concentration profiles are available in the Supporting Materials.[†]

6.3.1 Acetyl + 20 kcal mol⁻¹. We begin with the case where the acetyl + oxygen ground-state energy is artificially

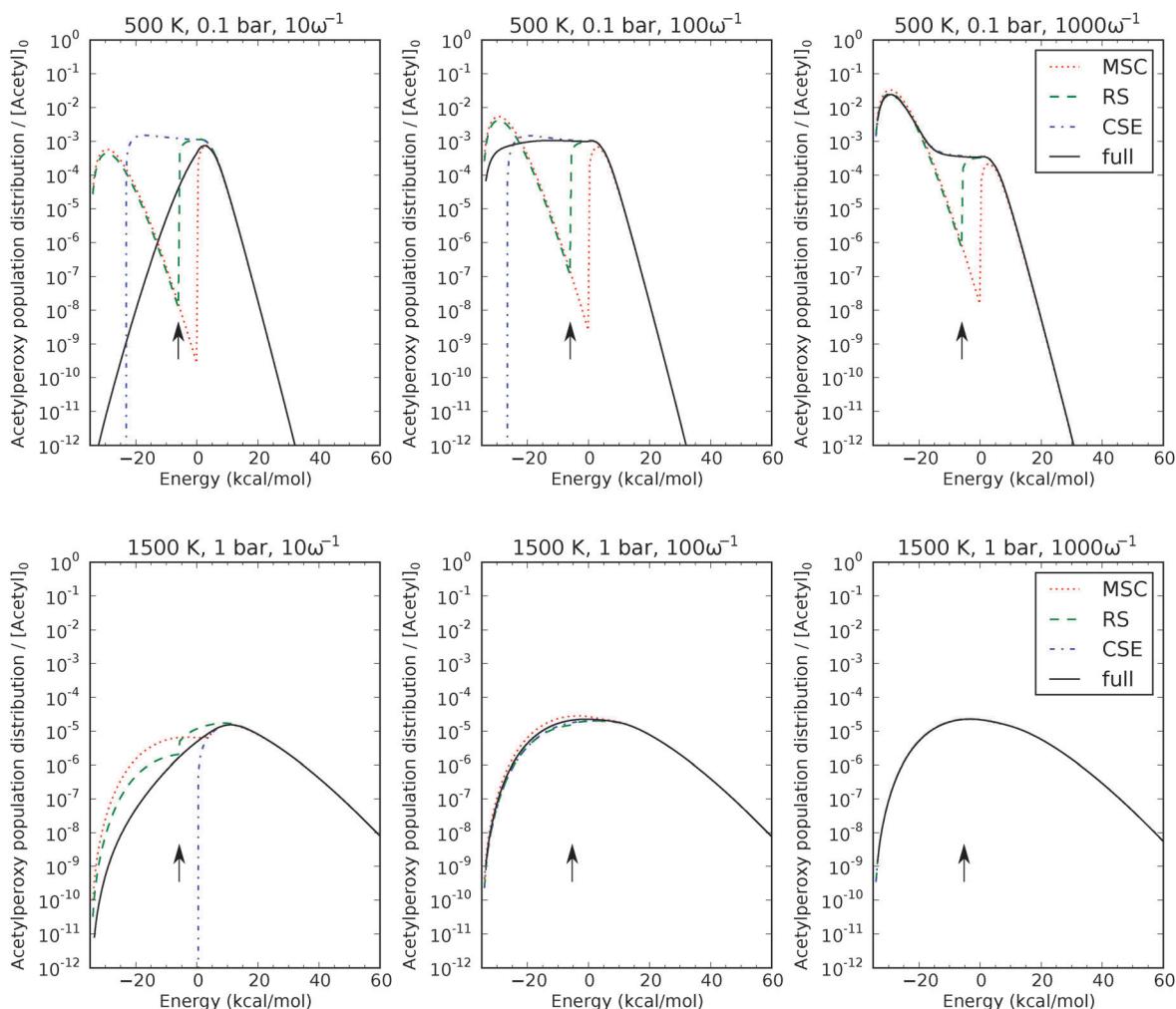


Fig. 7 Population distributions for acetylperoxy *versus* energy after 10, 100, and 1000 collision times, comparing the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods with the full master equation solution. At the conditions used above, ω^{-1} 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 RS and particularly MSC methods underpredict the population of the adduct states with energies just below the reactant energy (-6 to 0 kcal mol $^{-1}$) at lower temperatures. The zero of energy is acetyl + O₂. The arrows indicate the energy of the lowest transition state, which is the top of the “reservoir” in the RS method.

increased by 20 kcal mol $^{-1}$. 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 + oxygen but well above all of the transition barriers. The potential energy surface now reflects a very chemically-activated system, as the entire acetyl + oxygen configuration is 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. Thus we are not surprised to see that very high pressure is required to trap the adducts, as seen in Fig. 9a.

However, only the reservoir state and chemically-significant eigenvalue methods capture this effect, while the modified strong collision method does not. The modified strong collision 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 $^{-1}$ 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 modified strong collision method provides a significantly less accurate prediction for the isomer profiles, while the other methods perform noticeably better, especially at low temperatures.

6.3.2 Acetylperoxy + 20 kcal mol $^{-1}$. Next we perturb the system by artificially raising the ground-state energy of the initial adduct acetylperoxy by 20 kcal mol $^{-1}$, creating 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 into the higher-energy reactive grains, and also encourages

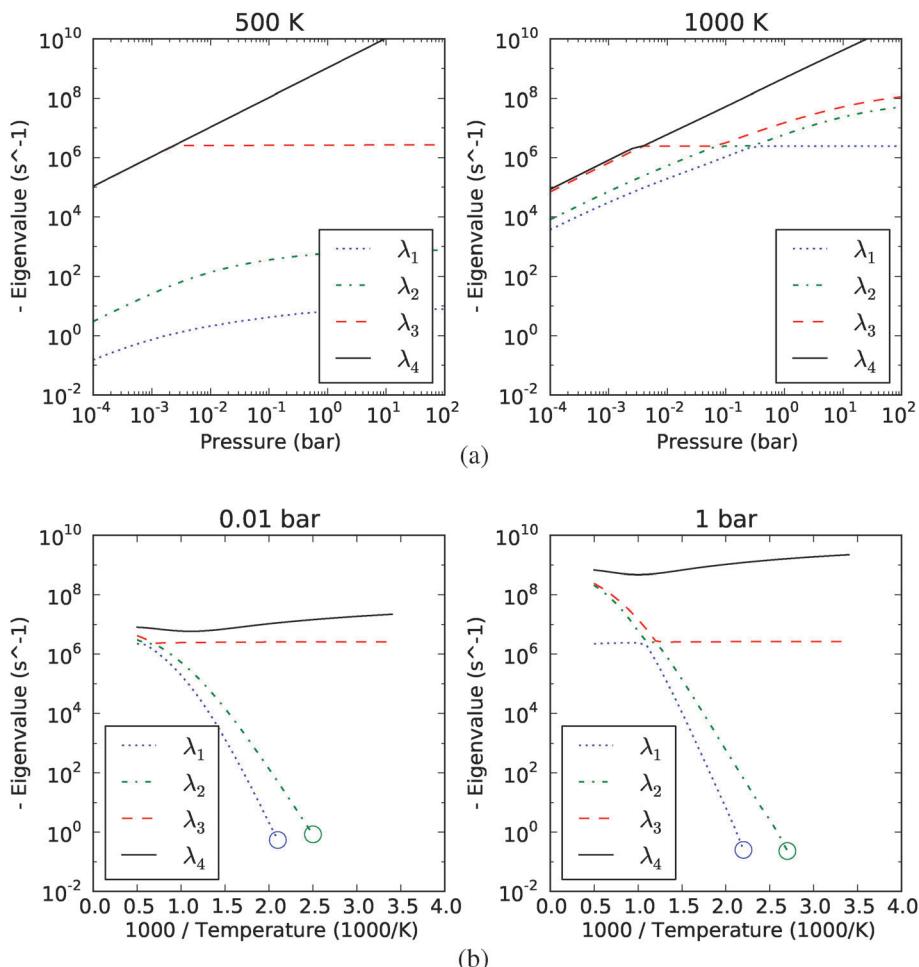


Fig. 8 Eigenvalues of the master equation matrix at (a) several pressures at 500 and 1000 K and (b) several temperatures at 0.01 and 1 bar. As the pressure is decreased or the temperature is increased, one or more of the chemical eigenvalues (λ_1 to λ_3) become indistinguishable from the internal energy eigenvalues (λ_4). At low temperatures the eigensolver gives unphysical positive eigenvalues; the last correctly computed low-magnitude eigenvalues are circled.

chemically-activated well skipping from acetyl + oxygen to hydroperoxyvinoxy. In Fig. 9b we see that the $k(T,P)$ values from acetyl + oxygen to hydroperoxyvinoxy have increased as a result of this perturbation. At 500 K, this increase is so significant that hydroperoxyvinoxy 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 hydroperoxyvinoxy) is often blended into the collisional energy relaxation eigenspectrum, so the chemically-significant eigenvalue method does not succeed with three chemical eigenvalues. For situations like this, Miller and Klippenstein recommend treating the two equilibrated isomers as a single pseudo-species, which would correspond to two chemical eigenvalues.²²

When concentration profiles are compared under low pressure conditions—where the $k(T,P)$ value for acetyl + oxygen \rightarrow acetylperoxy has fallen off—there is some disagreement between the methods on the acetylperoxy profile. This is shown at both low-temperature and high-temperature conditions in Fig. 10.

At low temperature, the modified strong collision method consistently overpredicts the acetylperoxy concentration, while

the reservoir state method consistently underpredicts it. The chemically-significant eigenvalue method is significantly more accurate than both of the other two approximate methods in this instance. However, the modified strong collision and reservoir state methods provide very good profiles for the other species as well, with the reservoir state profiles slightly more accurate. Furthermore, under these conditions acetylperoxy is a minor species, so more error in the observed concentration may be acceptable depending on the application. At high temperature the primary disagreement between the modified strong collision and reservoir state profiles is again for acetylperoxy. It is unsurprising that the reservoir state differs from the full master equation solution, especially at short times, as a significant fraction of the Boltzmann population of acetylperoxy now resides in the reactive grains. What is more surprising is how accurate the modified strong collision method is for the same conditions. The single-step collision model used in the modified strong collision method is perhaps more accurate when there are fewer nonreactive energies, especially when coupled with an accurate collision efficiency $\beta(T)$.

For this perturbed potential energy surface, the high-pressure-limit rate coefficients $k_\infty(T)$ for reactions involving

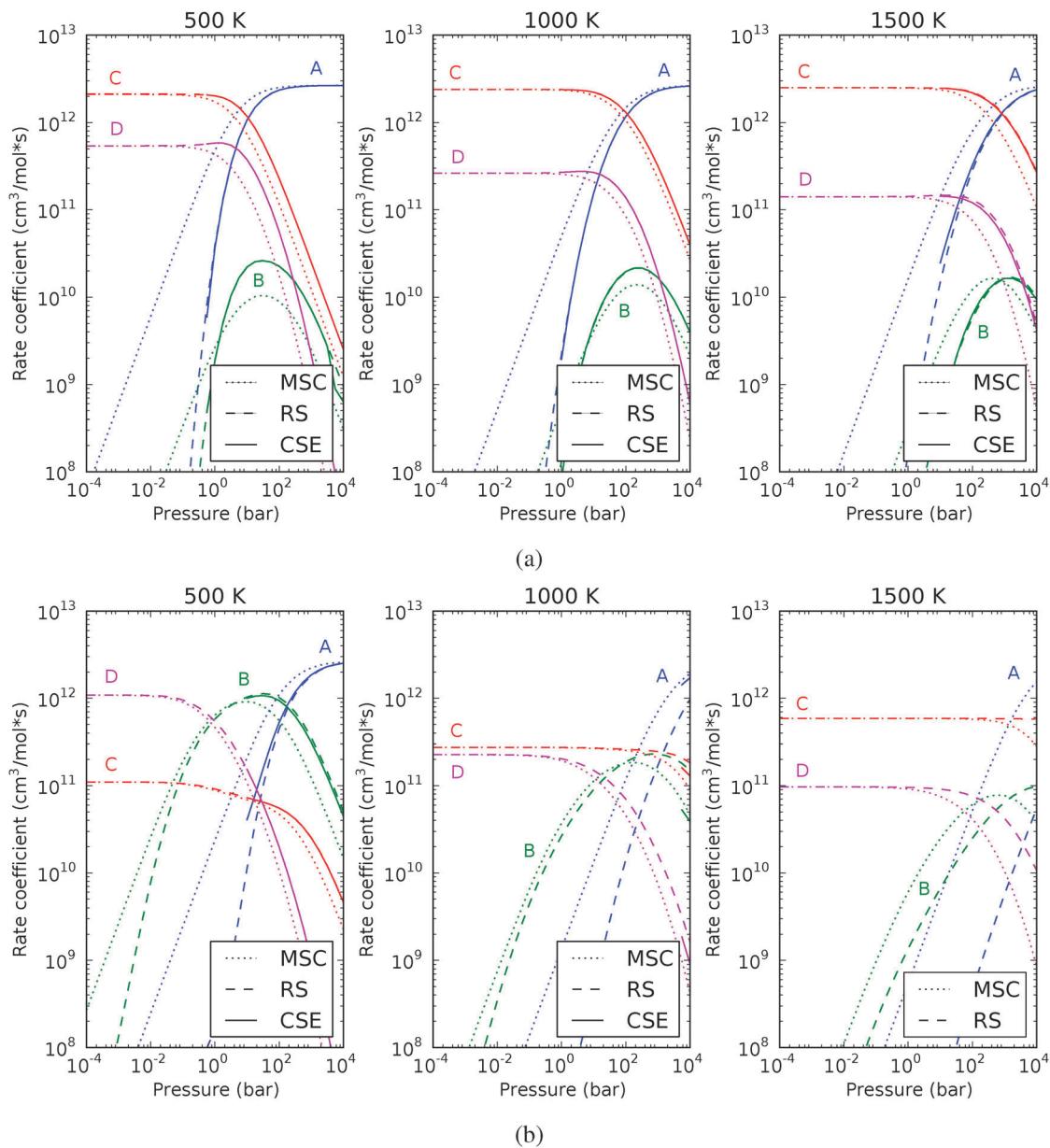


Fig. 9 Comparison of rate coefficients *versus* pressure for $\text{CH}_3\text{CO} + \text{O}_2 \rightarrow$ products estimated using the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods. In the plots, A = acetylperoxy, B = hydroperoxylvinoxy, C = ketene + HO_2 , and D = lactone + OH. (a) Increasing the ground-state energy of acetyl + oxygen by 20 kcal mol^{-1} 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^{-1} 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.

acetylperoxy as the reactant are the largest. At 500 K the $k_\infty(T)$ values for acetylperoxy \rightarrow hydroperoxylvinoxy, acetylperoxy \rightarrow acetyl + O_2 , and acetylperoxy \rightarrow ketene + HO_2 are $9.7 \times 10^9 \text{ s}^{-1}$, $3.4 \times 10^8 \text{ s}^{-1}$ and $2.0 \times 10^6 \text{ s}^{-1}$, respectively. At the pressure of 10 bar used for the profiles in Fig. 10, these are all smaller than the average collision frequency of $\omega = 7.1 \times 10^{10} \text{ s}^{-1}$. This implies that the collision and reaction processes do indeed occur at different timescales, and the chemically-significant eigenvalue method is able to resolve three chemical eigenvalues. However, at 1500 K the $k_\infty(T)$ values for the same reactions are $1.9 \times 10^{11} \text{ s}^{-1}$, $1.5 \times 10^{12} \text{ s}^{-1}$ and $2.0 \times 10^{11} \text{ s}^{-1}$, respectively, which are all well above the

average collision frequency of $\omega = 3.2 \times 10^{10} \text{ s}^{-1}$ at 10 bar. This suggests that the reaction time scale blends into the collision time scale, and thus we are not surprised that the chemically-significant eigenvalues method is unable to resolve three chemical eigenvalues at these conditions.

6.3.3 Hydroperoxylvinoxy + 20 kcal mol⁻¹. A related perturbation is to artificially raise the ground-state energy of the other isomer, hydroperoxylvinoxy, by 20 kcal mol^{-1} , 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

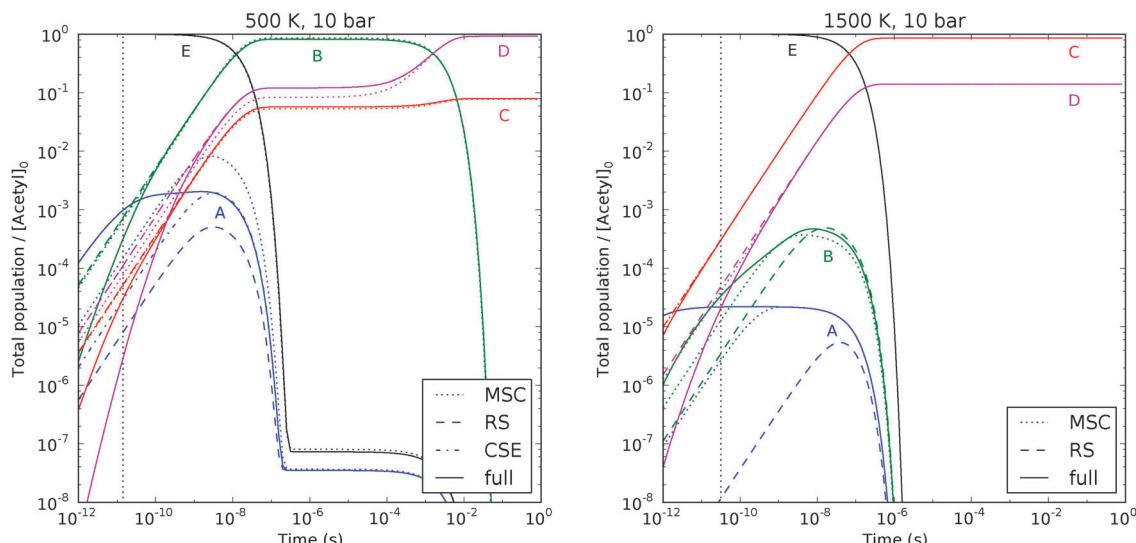


Fig. 10 Concentration profiles at low-temperature and high-temperature conditions, comparing the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods with the full master equation solution for the perturbed acetyl + oxygen system where the ground-state energy of acetylperoxy is artificially increased by 20 kcal mol⁻¹. In the plots, A = acetylperoxy, B = hydroperoxylvinoxyl, C = ketene + HO₂, D = lactone + OH, and E = acetyl + O₂. At low temperature, all three methods agree with each other and the full master equation solution except for acetylperoxy at short times, for which the CSE method provides the best match of the full solution. At high temperature, the MSC profiles are noticeably more consistent with the full master equation solution than the RS profiles, particularly for acetylperoxy. The CSE method is unable to find three chemical eigenvalues at the high temperature conditions, so its profiles are not shown).

population of hydroperoxylvinoxyl to exist in the reactive grains. Overall this encourages well skipping from acetylperoxy to the product channels, and reduces the lifetime of the hydroperoxylvinoxyl. The result is the $k(T,P)$ values for reactions that produce hydroperoxylvinoxyl are depressed, as seen in Fig. 11a. The only major disagreement in $k(T,P)$ values between the modified strong collision and reservoir state methods are for those involving hydroperoxylvinoxyl as a reactant or product. (Again, the chemically-significant eigenvalue 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 reservoir state method performs slightly better under most conditions for all configurations except the very unstable species hydroperoxylvinoxyl, for which the modified strong collision method is slightly more accurate.

6.3.4 Isomerization – 20 kcal mol⁻¹. As a final perturbation, we lower the reaction barrier for the isomerization reaction by 20 kcal mol⁻¹. The results of this perturbation are shown in Fig. 11b. 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 chemically-significant eigenvalues method to have no result with $N_{\text{chem}} = 3$. The modified strong collision and reservoir state methods are competitive, although at high temperatures the former performs noticeably better than the latter in predicting both isomer profiles (figure available in the Supporting Materials†).

6.3.5 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 modified strong collision method performs consistently in all systems, nearly always giving reasonable results, although sometimes noticeably less accurate than the other methods. The modified strong collision method is also numerically inexpensive and robust. For initial explorations of a system, the modified strong collision method is a reasonable choice.

Third, there is a wide range of conditions, especially at low and moderate temperatures, where the reservoir state method provides better estimates for the phenomenological rate coefficients than the modified strong collision 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 reservoir state 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, reservoir state method estimates involving both isomers became less accurate. Thus, we conclude that the reservoir state 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 chemically-significant eigenvalues 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

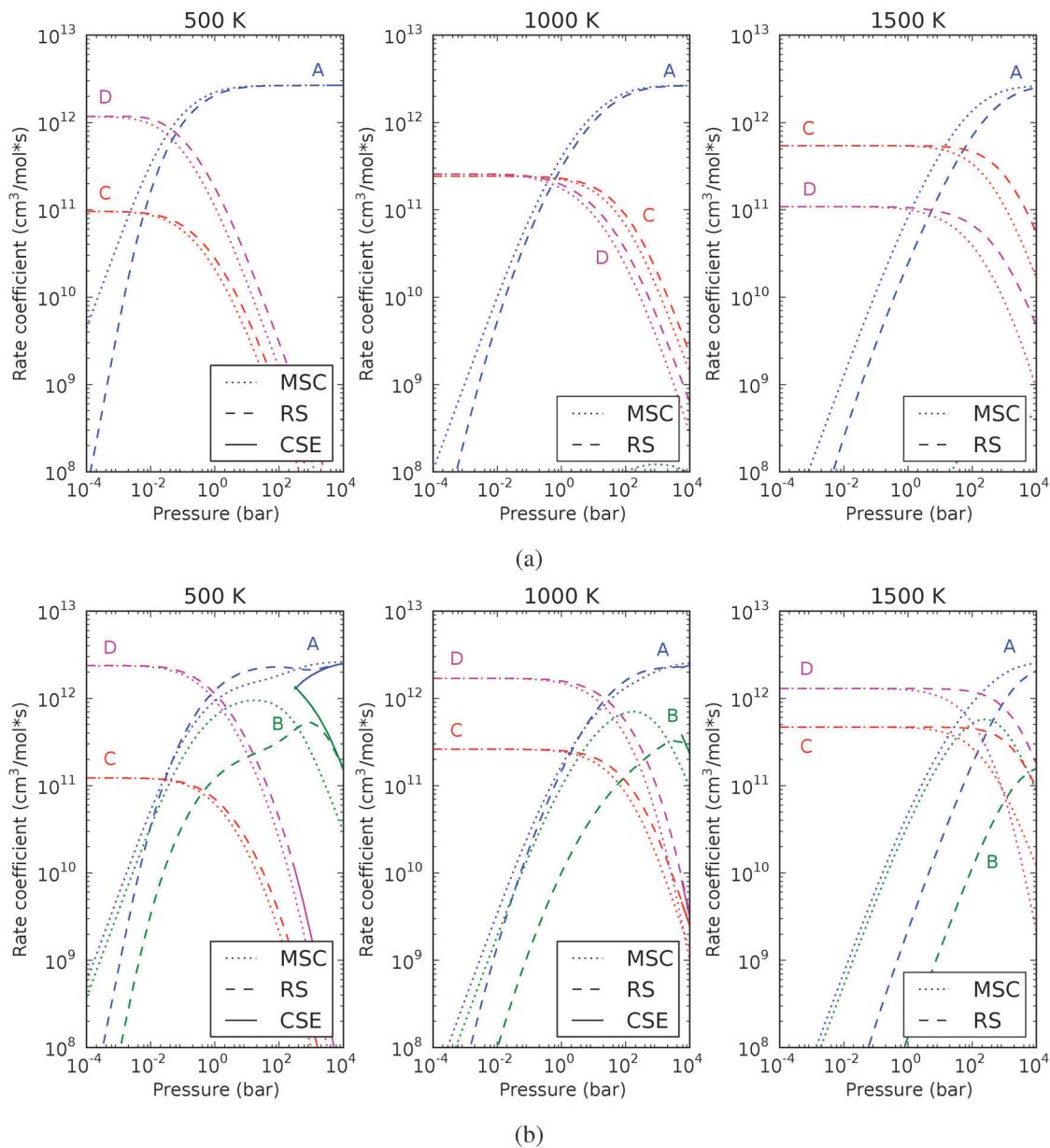


Fig. 11 Comparison of rate coefficients *versus* pressure for $\text{CH}_3\text{CO} + \text{O}_2 \rightarrow$ products estimated using the modified strong collision (MSC), reservoir state (RS), and chemically-significant eigenvalue (CSE) methods. In the plots, A = acetylperoxy, B = hydroperoxyvinoxy, C = ketene + HO_2 , and D = lactone + OH. (a) Increasing the ground-state energy of hydroperoxyvinoxy by 20 kcal mol⁻¹ 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⁻¹ creates a situation of a fast equilibrium between the two isomers except at extreme pressure and low temperature, causing the CSE method to be unable to resolve three chemical eigenvalues. The discrepancies in predicted isomer yields are usually not important, since the thermal equilibrium of the isomers is fast ($\sim 10^{-10} \text{ s}^{-1}$). The discrepancies in predicted yields of the product channels (C and D) are more important; the RS method is more accurate at low temperature, while the MSC method is more accurate at high temperature.

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.²² Unfortunately, both issues arise frequently even for networks of modest size.

6.4 Computational cost considerations

Our results show that the reservoir state and chemically-significant eigenvalues methods are somewhat more accurate

than the modified strong collision method in predicting rate coefficients for minor channels. The chemically-significant eigenvalues method has a stronger theoretical basis, but the reservoir state method is nearly as accurate.

An estimate of the relative execution time of the methods, shown in Table 2, comes from considering the time needed by each method to generate the data in Fig. 5. The reservoir state method has about 2–3x the computational cost of the modified strong collision method, while the cost of chemically-significant

Table 2 CPU time to generate Fig. 5

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

eigenvalues method is nearly 100x the modified strong collision 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 chemically-significant eigenvalues 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 reservoir state method provides enough of an improvement in accuracy compared with the modified strong collision 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 reservoir state method, while generally faster and more robust than the chemically-significant eigenvalues method, is not immune to numerical difficulties. The reservoir state 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 200 active-state grains per isomer). For these reasons, our automatic mechanism generation code, RMG, provides the user the option of either the modified strong collision or reservoir state methods, and automatically falls back to the modified strong collision method if the reservoir state method is unable to complete the Gaussian elimination step.

7 Conclusion

A general framework for estimating the phenomenological rate coefficients for unimolecular reaction networks of arbitrary size and complexity using only macroscopic parameters has been presented. Two important steps in the calculation have been explored in detail: the modeling of the molecular degrees of freedom used to determine the density of states, and the methods for estimating the phenomenological rates from the full master equation model. In the former, the functional group frequency method has been shown to be generally superior to the three frequency method, giving a density of states that is more consistently close to that of the full quantum chemistry-based model. In the latter, the reservoir state method has been shown to be superior to the modified strong collision method except at very high temperatures, giving concentration profiles that more closely follow those of the full master equation solution. These methods are compared with the chemically-significant eigenvalues method

of Klippenstein and Miller and with the full master equation solution in a way that we hope sheds some light on the relationships between these methods.

All three approximations to the master equation perform well at estimating the $k(T,P)$ values of the major pathways of a potential energy surface. However, for strongly chemically-activated systems, significant disagreement between the methods is observed for minor pathways. 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, this method has numerical difficulty at low temperatures due to large separation of chemical and internal energy relaxation timescales, and currently requires human adjustment to handle fewer chemical eigenvalues for systems with a fast equilibration. This could perhaps be automated, but to our knowledge this automation is not currently available. The reservoir state method is much less expensive than the chemically-significant eigenvalues method, and works very well at low and moderate temperatures. However, when the reservoir approximation is not valid—*i.e.* for high temperatures and shallow wells—the reservoir state method gives worse estimates for the $k(T,P)$ values than even the modified strong collision method. This could perhaps be addressed by manually removing the unimolecular wells which no longer correspond to stable species, but such a procedure has not yet been made automated. The modified strong collision method is not as accurate as the other methods in the ranges where those methods work well—particularly for strongly chemically-activated potential energy surfaces with deep unimolecular wells—but is computationally inexpensive and much more robust over the full temperature range.

In practical applications of automatic mechanism generation, most pressure-dependent reaction networks considered are very complex, with a high likelihood of shallow wells being present. For this reason, we recommend the modified strong collision method for general use with automatic mechanism generation. There is room for improvement by using the reservoir state or chemically-significant eigenvalues methods in the ranges where they are reliable, *e.g.* for reactions which the model is sensitive to. In particular, the more rigorous methods could become viable if one could automate removal of the very short-lived wells at high temperature. However, one would need good error control in the software to catch and recover from cases where the more rigorous methods are not reliable.

The result of this work is a pair of modules for estimating molecular degrees of freedom and phenomenological rates that are utilized by and distributed with the open-source automatic mechanism generation code RMG. RMG uses the functional-group frequency method exclusively and offers users a choice between the modified strong collision and reservoir state methods. Preliminary versions of both of these codes, written in Fortran 90, are currently available in the RMG package.¹

Acknowledgements

This work was funded in part by the US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences, under Contract

- 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.
- 73 H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby and M. Schuetz, *et al.*, *MOLPRO, version 2006.1, a package of ab initio programs*, 2006.
- 74 P. Zhang, S. J. Klippenstein, H. Sun and C. K. Law, *Proc. Combust. Inst.*, 2011, **33**, 425–432.
- 75 T. J. Frankcombe and S. C. Smith, *Comput. Phys. Commun.*, 2001, **141**, 39–54.
- 76 T. J. Frankcombe and S. C. Smith, *Faraday Discuss.*, 2001, **119**, 159–171.