Theor Chem Acc (2009) 124:303–317
DOI 10.1007/s00214-009-0623-z

FEATURE ARTICLE

Numerical solution methods for large, difficult kinetic master equations

Terry J. Frankcombe · Sean C. Smith

Received: 5 April 2009 / Accepted: 3 August 2009 / Published online: 19 August 2009
© Springer-Verlag 2009

Abstract The kinetics of gas-phase reactions, including pressure-dependent weak collision and non-equilibrium effects, can be modelled using a master equation. In this paper, we address the practical computational problem of finding solutions to such kinetic master equations. The mathematical structure of the master equation can be utilised to develop a number of specialised numerical techniques that are capable of solving the master equation in the presence of difficult numerics and for large problems. The former is important for modelling low temperature and pressure systems, and the latter is important for modelling the large networks of isomerising species common in combustion chemistry applications. We focus on numerical methods that exhibit particular practical use because of their robust nature or scalability to many isomers, or both. Recent developments in linear-scaling methods are highlighted.

Keywords Master equation · Multi-well · Energy grained · Numerical integration · Collisional energy transfer

1 Introduction

The solution of the unimolecular master equation (ME) [1] and its reversible isomerisation and bimolecular extensions is a challenging numerical problem. Developing robust and general methods for computing the kinetic behaviour of such reactions is driven by the pervasive role they play in combustion, along with other applications. In this paper, our purpose is to explore the numerical issues one must face when solving the ME and thereby simulating the kinetics of the modelled reactions over a wide range of timescales.

The simple energy-resolved ME is an integral–differential equation that statistically describes the time evolution of a population of some reacting species undergoing irreversible unimolecular reaction while dilute within a thermal bath gas. It can be written as

$$\frac{\partial p(t;E)}{\partial t} = \omega \int_0^\infty P(E|E')p(t;E')\mathrm{d}E' - [\omega + k(E)]p(t;E). \tag{1}$$

In this equation, $p(t; E)$ gives the population of the reacting species at energy E and time t. Collisions of the target molecule with the bath gas are defined to be occurring at frequency ω. $P(E|E')$ is a collisional energy transfer (CET) kernel describing the probability that a molecule initially at energy E' is left at energy E after a collision with the bath gas. The unimolecular reaction of the target molecule is described by the microscopic (energy resolved) rate constants $k(E)$. The first term on the right of Eq. 1 accounts for accumulation of population from molecule arriving at energy E after a collision. The second term accounts for the fact that a energy transfer is removing population from energy E, and for reactive loss.

Energy-resolved MEs are not the only types of MEs used in kinetics. Extending Eq. 1 to be resolved in angular momentum as well as energy is straight forward. Going the other way, the relative population of several chemical species without resolving the non-equilibrium distributions

T. J. Frankcombe
Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia
e-mail: tjf@rsc.anu.edu.au

S. C. Smith (✉)
Australian Institute of Bioengineering and Nanotechnology, University of Queensland, St. Lucia, QLD 4072, Australia
e-mail: S.Smith@uq.edu.au

within each species has been described using systems of differential equations for as long as chemical kinetics has been studied. While the general approaches to solving these systems of rate equations can be similar, the computational and numerical effort required is considerably reduced from that required for the MEs discussed here [2–6].

In practice, solutions to Eq. 1 are found after discretization. Usually the energy is simply binned into finite width grains, though other discretizations are possible [7–9]. Typically the number of grains used is hundreds to thousands. A detailed discussion of issues involved with discretization is beyond the scope of this paper. Once the ME is discretized, Eq. 1 can be written in matrix form as

$$\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = A\mathbf{p}(t) \tag{2}$$

where A is a non-symmetric square matrix. It is the solution of Eq. 2, known as the energy-grained master equation (EGME), that this article is primarily concerned with. Note that the solution to Eq. 1 or 2 is only well defined once some initial population $p(0; E)$ or $\mathbf{p}(0)$ (or at some other time t_0) is specified.

The full solution to Eq. 2 is not required to obtain useful information about the reaction. The smallest (least negative) eigenvalue of the matrix A, if well separated from the rest of the spectrum [10], as it usually is, can be equated with the long-time rate constant of the unimolecular reaction. The corresponding eigenvector gives the long-time population distribution, which slowly decays. In 1978, Gaynor et al. [11] adapted Nesbet's method for finding the smallest eigenvalue (and corresponding eigenvector) of a Hamiltonian matrix [12] to find the long-time solution to Eq. 1. We shall discuss this method in more detail in Sect. 4.

The difficulties encountered while solving Eq. 2 are mostly numerical in nature. The ordinary differential equation (ODE) described is *stiff*, describing processes occurring over a wide range of timescales. As the temperature and pressure of the described system go down (say, below 1,000 K and atmospheric pressure or lower), the problem becomes more stiff, leading to greater numerical difficulties. An in-depth discussion of the origin and propagation of numerical errors in finite precision arithmetic—though essential knowledge for any computational scientist—is beyond the scope of this article. The authors particularly recommend the book by Higham [13] for an in-depth coverage of the topic, and the work of Golub and van Loan [14] for more general algorithmic implications.

Many MEs can be solved easily and quickly. A very fast solution to the ME is essential in many applications, such as parameter fitting to experimental data or embedding ME results within a larger fluid dynamics simulation. Our focus here is those MEs that cannot be solved in seconds of computer time, either due to numerical or size issues. Two key considerations are whether standard double precision arithmetic is sufficient to solve the problem, and what is the scaling of the algorithm used to arrive at the solution. Many ME solution methods require computer time scaling with the third power of the system size. To be able to solve very large problems without requiring months or years of computer time, we want better scaling than this. We do not consider the parallelisability of the algorithms.

In this paper, we deal specifically with MEs describing collisional and reactive processes in the gas phase. A similar methodology can be applied to reactions of adsorbents on solid surfaces, as pioneered by Harrison and coworkers [15–17]. Many of the considerations described here apply equally to these recently developed ME applications.

In Sect. 2, we describe the EGME and some common extensions beyond the traditional unimolecular reaction applications. We describe some general approaches to solving the ME in Sect. 3, then discuss the numerics involved in solving the irreversible unimolecular reaction problem in Sect. 4. Sections 5 and 6 discuss how to solve the ME for multiple-well, reversible isomerisation problems and including bimolecular channels, respectively. Important recent developments in scalable solution methods are presented in Sect. 7, before Sect. 8 concludes.

2 Review of the ME

The ME is well known and described in detail elsewhere [1, 18–25], so only some brief details will be given here.

2.1 Simple MEs and basic properties

There is an arbitrariness implicit in writing the ME in the form of Eq. 1, as "collisions" are ill-defined. For some definition of the collision frequency (often taken as the hard sphere value scaled by the Lennard-Jones collision integral $\Omega^*_{2,2}$ [26]) the per-collision probability $P(E|E')$ is related to the energy transfer rate constant $R(E|E')$ by

$$\omega P(E|E') = [\mathrm{M}]R(E|E'). \tag{3}$$

Here, we keep the ω and $P(E|E')$ partitioning, as is common, for its conceptual simplicity. Note that while one can take the collision frequency ω to be dependent on the energy E', it is common to assume a constant collision frequency over the entire energy range, as we do here.

For a simple binning of $p(t; E)$ into a discrete set $\mathbf{p}(t)$, a number of graining schemes are possible. For example, the population element p_i can be interpreted as the value at the centre of the ith energy bin, or at one or the other edge, or as the average over the energy range of the bin. The details

are unimportant, provided the results are converged with respect to decreasing the bin width.

Particularly when employing spectral solution methods, the efficiency and stability of the solution can be improved by symmetrising the matrix before solving the ME (though the symmetric form of the ME is not automatically the best choice [27, 28]). If **f** is the vector describing the Boltzmann population of the system (with the same binning as the vector **p**), then defining the diagonal matrix S through

$$S_{ii} = f_i^{-1/2} \tag{4}$$

means the similarity transform SAS^{-1} yields a symmetric matrix $[SAS^{-1} = (SAS^{-1})^{\mathrm{T}}]$. Transforming the population vector to $S\mathbf{p}$ returns the EGME to the form of Eq. 2, but with a symmetric coefficient matrix SAS^{-1}. Once this transformed problem is solved, the real, observable population distributions are given via the reverse transformation, multiplying by $f_i^{1/2}$.

At this point we briefly mention a property of the EGME that is not often recognised. Applying the similarity transformation that leads to the symmetrised matrix once more yields the transpose of the asymmetric ME matrix A:

$$S(SAS^{-1})S^{-1} = S^2AS^{-2} = A^{\mathrm{T}}. \tag{5}$$

The eigenvectors of A^{T} (the left eigenvectors of A) are the eigenvectors of A divided by the Boltzmann population ($S^2\mathbf{p}$). This has important implications for numerical methods, as shall be discussed in Sect. 4.2.

2.2 Isomerising networks

The reaction schemes that can be described within a ME formulation are not limited to unimolecular reactions. Systems of reversible unimolecular isomerisations, possibly coupled with unimolecular decomposition, are increasingly being modelled with ME methods [22, 24, 29–38]. Because species elements from several different stable isomers are included, these types of problems are often known as *multi-well* MEs.

One possible method to more easily see how large systems of reversible isomerisation reactions (containing many isomers) can be modelled with a ME [and to gain a deeper insight into the process modelled by an equation of the form of Eq. 2 in a ME context] is to slightly modify how one thinks about an EGME for a unimolecular problem. Instead of considering the energy grains as a means to describe a population of a single species, consider each energy grain as a unique species. CET processes are then simply first-order reactions occurring at a rate given by the product of the collision frequency, the energy grain size and the appropriate element of the CET matrix, P. From this point of view, for a gas-phase unimolecular problem there is one *reversible* reaction occurring between each species element and every other species element. Additionally, for a particular set of species elements (those corresponding to energies above the reaction threshold) there are one or more additional first-order reactions occurring: the dissociation or isomerisation channel being modelled. The rate constant for each of these reactions is the appropriate microscopic rate constant.

To model reversible isomerisation, the populations of all of the isomers need to be tracked. Explicitly tracking the population of additional isomers is simply a matter of adding the grained population distribution of each of the additional isomers to the set of species elements being modelled. Thus, the total number of species elements is roughly the several hundred to low thousands for each species well, multiplied by the number of wells. A simplified case (with only a few energy grains) is represented in Fig. 1. Considering the energy-grained population distribution as a set of independent species elements in this way is a more natural extension of the classical two-level

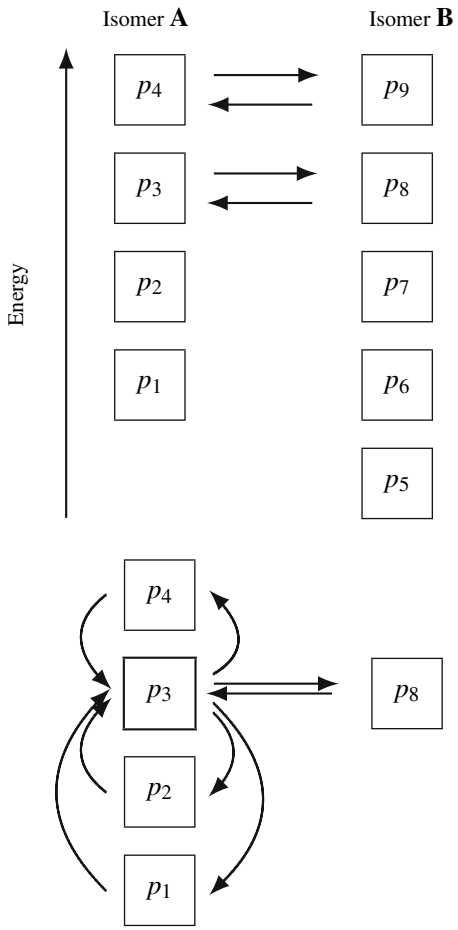

Fig. 1 Very simplified representation of the reversible isomerisation A ⇌ B, showing the explicit generalisation of the Lindemann mechanism. Note that within each isomer each element is reactively coupled with all others. *Lower diagram* highlights first-order processes involving the species element p_3.

Lindemann mechanism to the isomerisation case than is normally given. Clearly, the ME approach can be extended to any number of isomers, with the population elements from each being added to the vector describing the overall population of the system, $\mathbf{p}(t)$.

2.3 Bimolecular reactions

Bimolecular reactions can be modelled with a ME. It is usual to assume that the bimolecular species are sufficiently well mixed not to require energy-resolved modelling, meaning they are incorporated as a single grain in the population vector $\mathbf{p}$. Equation 2 becomes

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = A\mathbf{p} + \sum_r p_{i_r} p_{j_r} \mathbf{k}_r \tag{6}$$

where now the matrix A describes the linear (first order) reactions with zero rows and columns corresponding to the species involved in the bimolecular reaction. The new term on the right describes the bimolecular reactions, indexed by r, to and from the energy-grained unimolecular species and the bimolecular species described by the population elements p_{i_r} and p_{j_r} with the appropriate rate constants given by the vectors $\mathbf{k}_r$.

This EGME is no longer a linear ODE. Linearity can be restored by assuming pseudo-first-order conditions, making one of p_{i_r} or p_{j_r} constant in time. This population grain can be excluded from $\mathbf{p}$ and the additional term on the right of Eq. 6 incorporated into the matrix A, returning the ME to the form of Eq. 2 [39]. Direct time integration (see Sect. 3.2) can easily incorporate such bimolecular reactions and has regularly been used for sophisticated but essentially pressure-independent modelling incorporating such reactions [6, 40, 41]. This remains an open area of investigation. In this work, we consider only pseudo-first-order bimolecular reactions.

2.4 Drift-determined diffusion

The first term on the right of Eq. 1 describes CET within a species population and manifests as a dense block in the ME matrix. One can invoke the diffusion approximation to model CET as a purely local process, replacing the effect of the dense discretized $P(E|E')$ through

$$\sum_j P_{ij} p_j \approx a_i p_{i-1} + b_i p_i + c_i p_{i+1}, \tag{7}$$

where the constants a_i, b_i and c_i depend on the particulars of the diffusion approximation and finite difference scheme used. In terms of the picture presented by the lower part of Fig. 1, the transfer of population between p_1 and p_3 is removed, with a modification of the $p_2 \rightleftharpoons p_3 \rightleftharpoons p_4$ rate constants to compensate. Green et al. [42] conclude that, of the various diffusion approximation formulations available, drift-determined diffusion gave the best results for unimolecular MEs. Applying Eq. 7 transforms the A of Eq. 2, with the dense CET component replaced by a tridiagonal. In a multi-well ME, in which the ME matrix has a dense block for CET within each species, the dense blocks are each replaced by a tridiagonal. We use D for the ME matrix A with the diffusion approximation applied. As discussed in Sect. 5.3, the well-defined sparse structure of the diffusion approximation matrix allows both $D\mathbf{v}$ and $D^{-1}\mathbf{v}$ to be calculated very quickly. Both of these operations scale approximately linearly with the size of the system being modelled, which shall be important in Sect. 7. While solving the diffusion version of the ME is clearly an approximation to the solution of the full ME, this approximation has proved useful [22, 32, 42–46].

2.5 Two-dimensional (angular momentum resolved) MEs

The conservation of angular momentum has important consequences for the calculation of reaction rates and branching ratios. These are most evident in barrierless processes. Angular momentum effects can be included in the ME [1, 24, 25]. One can expect that the effect of angular momentum conservation is reasonably small, and strongest at low temperatures [24].

Formulation of the ME problem in this context requires species to be labelled according to both their total energy and their total angular momentum. The discretized ME can still be expressed as in Eq. 2 and many of the methods discussed in this work are applicable to these "2D" MEs. In particular, the Nesbet method can be applied to 2D problems to give the phenomenological rate constant [47] and the general approaches described in the following section apply equally. One can apply the diffusion approximation to 2D MEs to generate a more structured ME matrix [44, 46], though it is unclear how effective this approximation is for rotational transitions. The 2D ME can be reduced to a 1D ME by assuming some particular model for rotational transitions [1, 48–50]. Another approach is to use basis functions dependent on both energy and angular momentum to discretize the ME [9]. Analytic solutions exist for 2D MEs assuming particular functional forms for the energy transfer kernel [44, 50–52].

In this work we do not specifically treat 2D MEs, discretizing over energy only. For more details we refer the interested reader to the formentioned references.

3 Approaches to the solution

The unimolecular ME can yield far more information that just the classical unimolecular rate constant and long-time

population profile. As for any first-order system of differential equations, Eq. 2 describes the evolution of $\mathbf{p}(t)$ for all times t, once the initial value $\mathbf{p}(0) \equiv \mathbf{p}_0$ is set.

3.1 Spectral solution

The solution of Eq. 2 as a function of time is easily shown to be

$$\mathbf{p}(t) = \exp(At)\mathbf{p}_0. \tag{8}$$

Appearing in this solution is the matrix exponential $\exp(At)$. Calculating the matrix exponential has been explored extensively in the literature, and many approaches exist [53–55]. Quite a few of these approaches have been trialled for the case at hand. It has been found that most of the more sophisticated approaches to calculating either the matrix $\exp(At)$ or the actual desired solution $\exp(At)\mathbf{p}_0$ fail in various ways. One of the more common, well understood and easily calculated solutions can be obtained as [1]:

$$\mathbf{p}(t) = \sum_{j=1}^{n} \alpha_j \exp(\lambda_j t)\mathbf{x}_j, \tag{9}$$

where λ_j and $\mathbf{x}_j$ are the n eigenvalues and eigenvectors of the $n \times n$ matrix A:

$$A\mathbf{x}_j = \lambda_j \mathbf{x}_j. \tag{10}$$

The projection coefficients α_j used in Eq. 9 are determined by projecting the initial population $\mathbf{p}_0$ onto the set of eigenvectors so that

$$\mathbf{p}_0 = \sum_{j=1}^{n} \alpha_j \mathbf{x}_j. \tag{11}$$

Solution via Eqs. 9 and 11 is a *spectral* solution to the EGME, from the term for the set of eigenvalues of a matrix. A spectral solution has the advantage of requiring only a single eigendecomposition of the matrix A, and then a single projection for each initial population of interest.

Coefficients α_i to satisfy Eq. 11 are given by the projection theorem, viz

$$\alpha_i = \langle \mathbf{p}_0, \mathbf{x}_i \rangle. \tag{12}$$

The inner product used in Eq. 12 depends on the orthogonality of the eigenvectors. The eigenvectors of a symmetric matrix form an orthogonal set, meaning that if the matrix is symmetrised before being diagonalised the required inner product is the normal Euclidean inner product,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathrm{T}}\mathbf{v}. \tag{13}$$

If the matrix is not symmetrised before being diagonalised the inner product under which the eigenvectors form an orthogonal set is the Euclidean inner product weighted by the inverse of the Boltzmann distribution,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathrm{T}} S^2 \mathbf{v}. \tag{14}$$

Similarly, for the eigenvectors of A^{T} (the left eigenvectors of the unsymmetrised ME matrix) the appropriate inner product is

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathrm{T}} S^{-2} \mathbf{v}. \tag{15}$$

The interpretation of the smallest eigenvalue of A as the long-time rate constant is obvious from Eq. 9, with certain caveats [10, 34]. Changes in $\mathbf{p}(t)$ at successively shorter timescales are controlled by eigenpairs with successively larger eigenvalues, with contributions from eigenvectors with significantly smaller eigenvalues remaining essentially constant. This means that a truncated version of Eq. 9, with terms involving large eigenvalues discarded, can yield good approximations to $\mathbf{p}(t)$ for all but the shortest times. Generally, discarding from the expansion eigenpairs with eigenvalues of greater magnitude than a cutoff τ can successfully model $\mathbf{p}(t)$ for a range of times t satisfying

$$t \gg 1/\tau. \tag{16}$$

This is an optimistic limit: the smallest t for which a truncated expansion is accurate may be quite a few orders of magnitude larger than $1/\tau$.

The structure of eigenvectors corresponding to larger eigenvalues is less well-known than that of the eigenvector corresponding to the smallest eigenvalue (though descriptions are available in the literature [28, 56]). In the lower part of the spectrum these eigenvectors are also broadly similar to the Boltzmann distribution, decaying exponentially in elements corresponding to high energies. In low energy grains a number of nodes (sign changes) appear corresponding to the eigenpair's position in the spectrum. That is, $\mathbf{x}_2$ has one node, $\mathbf{x}_3$ has two nodes, and so on. The number of eigenvectors of this form is generally about equal to the number of energy grains below the reaction threshold. The eigenvectors from the upper part of the spectrum generally have a "spike" of a single large element or a small region of large elements, surrounded by elements exponentially decreasing in magnitude as one moves away from the spike. In this region of the spectrum the spectral density is considerably lower than that around smaller eigenvalues.

The shape of the eigenvectors is dependent on the temperature and pressure being modelled. Lower temperatures give faster decays for the high energy elements, and lower pressures increase depletion of population from the Boltzmann population at long times, clearly evident in $\mathbf{x}_1$. The spread in the magnitudes of the eigenvector elements is a major source of numerical difficulties in spectral solutions to the ME [23, 28, 57]. Eigenvector elements smaller than the computational unit roundoff times the

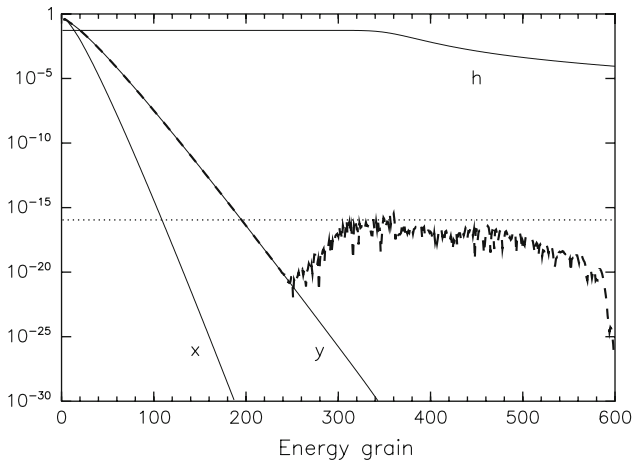

Fig. 2 Eigenvectors corresponding to the smallest eigenvalue for an ethane decomposition problem at 300 K. The three vectors labelled **x**, **y** and **h** represent, respectively, the eigenvectors of the original ME matrix A, the symmetrised version SAS^{-1} and the transpose A^{T}. The "eigenvector" calculated by applying the double precision Lapack routine dsyev [58] to the symmetrised matrix is shown *dashed line*, with the double precision unit roundoff indicated

largest elements cannot be calculated reliably. Yet these tiny elements often completely dominate the projection coefficients of Eq. 12.

In certain applications, one can "scale away" this precision problem by transforming the ME matrix according the similarity transforms described above. The effect of these transforms on the eigenvector corresponding to the long time limit is demonstrated in Fig. 2. However, the more complicated structure of eigenvectors from higher in the spectrum means the wide spread of magnitudes cannot be scaled away globally [27, 28, 57].

One rather heavy-handed remedy is to increase the numerical precision in which the calculation is performed. Quadruple precision arithmetic is supported on a range of computer systems, and this has been used to good effect [23, 28, 59, 60]. Beyond quadruple precision one can resort to *software arithmetic*, such as that implemented in Bailey's Fortran 90 MPFUN package [61]. Such an approach can effectively reduce numerical error to negligible levels. While increasing the precision does not *guarantee* more accurate results [13], in practice pathological cases are rare. Increasing the precision of the calculation comes at the cost of increasing the computer time required, possibly by many orders of magnitude [23, 28, 57, 62, 63]. Thus, this approach is only practical in cases where the ME matrix is not too large.

3.2 Direct integration

An alternative expression for the solution to the first-order linear ODE of Eq. 2 is given by explicit integration in time:

$$\mathbf{p}(t) = \mathbf{p}(0) + \int_0^t \frac{\mathrm{d}\mathbf{p}(\tau)}{\mathrm{d}\tau}\mathrm{d}\tau = \mathbf{p}(0) + \int_0^t A\mathbf{p}(\tau)\mathrm{d}\tau. \qquad (17)$$

This expression fully specifies $\mathbf{p}(t)$ once one sets $\mathbf{p}(0)$, and specifies the *direct integration* solution.

While spectral expansion is the most common approach to solving ME problems, numerical approaches more closely tied to a solution of the form of Eq. 17 are frequently used. For example, integration of the ME can be achieved by Monte Carlo methods [64–67]. This approach appears to be good at simulating very complex dynamics of large systems, but can be very slow to converge. Monte Carlo integration does not require the energy graining of the ME.

Direct numerical integration is also used to solve ME problems [68–71]. A numerical integrator is used to find the time-dependent solution to Eq. 2 via Eq. 17. The stiff nature of the ME matrix (with eigenvalues spanning many orders of magnitude) usually mandates using integration techniques specifically designed for stiff problems [72, 73]. Stiff integrators commonly implement predictor–corrector algorithms, with the correction involving a root finding operation. The root finding is achieved by a form of Newton's method, requiring matrix inversion. The required inverse turns out to be

$$[I - \gamma(t)J]^{-1} \qquad (18)$$

where I is the identity matrix, $\gamma(t)$ is a scalar and J is the *Jacobian* (matrix of derivatives) of the right hand side of the ODE. In the constant coefficient, linear case of Eq. 2, the Jacobian is simply the ME matrix A. Integration of the ME requires the calculation of many matrix-vector products [to give $\mathrm{d}\mathbf{p}(t)/\mathrm{dt}$] and solving many linear systems of equations [to give the inverse given in Eq. 18 for the correction step].

Such integrations with a stiff integrator work very well, even in difficult cases. Direct integration of the ODE (implemented in double precision) has been found to be successful even when the ME is being solved under low temperature and pressure conditions so that spectral methods implemented in double precision fail. Direct integration does not suffer when the populations being modelled span many orders of magnitude (even a greater range than the available precision range) as large relative errors in the very small elements of $\mathbf{p}(t)$ do not contaminate the larger elements, unlike in the projection onto the eigenvector basis of spectral methods. Even the Newton's method used as the corrector is tolerant of errors in the calculation of the required inverse. However, the robustness of the method comes at the price of a substantially larger amount of computational effort necessitated by the repeated matrix inversion in Eq. 18.

Direct integration has a number of other attractive features. Second-order bimolecular reactions (that is, those in which it is not assumed that one of the reactants is in large excess and all populations are modelled explicitly) can be easily included. The EGME can be integrated equally in its symmetrised or asymmetric forms. The latter is important in the case of non-conservative descriptions of isomerisations, in which one cannot easily define a Boltzmann population with which to form the S matrix that symmetrises A. A disadvantage, however, is that for every change in the initial population $\mathbf{p}_0$ the entire calculation must be repeated; a spectral solution allows a simple projection and analytic propagation for any number of initial distributions, once the demanding diagonalisation step has been completed.

3.3 Transient populations

In general, a non-equilibrium initial population $\mathbf{p}_0$ will relax to the long-time steady state of a single exponential decay, $\mathbf{p}(t) \propto \mathrm{e}^{\lambda_1 t}\mathbf{x}_1$, whether reaction takes place ($\lambda_1 < 0$) or not ($\lambda_1 = 0$). The manner of that relaxation, though often short-lived, reveals many important details of the energy transfer within the gas and the microscopic reaction rates. Successfully modelling these transient populations between initiation at $\mathbf{p}_0$ and steady state at $\mathrm{e}^{\lambda_1 t}\mathbf{x}_1$ is essential in many applications, particularly in fitting to high accuracy experimental data.

At high temperatures and pressures, a reasonable range of times can usually be modelled by a very small number of eigenpairs in a spectral solution. There is usually about this number of eigenvalues separated at the small end of the spectrum. As the temperature and, in particular, the pressure of the system is reduced, collisional reactivation of stabilised population becomes more important, and more and more eigenpairs need to be added to the expansion to model chemically interesting times. In the absence of other information, it is usually not clear how many eigenpairs are required to give accurate population profiles for a satisfactory range of times without actually varying the number of eigenpairs used in the expansion of Eq. 9. Equation 16 can be used as a guide only. These considerations do not apply to direct integration approaches, as determining long-time behaviour requires that all shorter time behaviour be modelled accurately first.

It is worth pointing out that the populations calculated by Eq. 8 are the solution to an initial value problem and inherently model *pulsed conditions*. That is, at time $t = 0$ a particular initial population of reactant $\mathbf{p}_0$ is created "as if by magic" (though presumably through some finite process occurring at times $t \leq 0$) and is then left completely alone to evolve according to the rules built into the ME. While this description yields a lot of useful information regarding the processes described by the ME which can be used in many interesting situations, modelling conditions other than pulsed conditions where reactants are created at $t > 0$ may require explicit integration of external reactant fluxes, much like the traditional treatment of *chemical activation* reactions [1, 56]. Experimental measurements often approximate pulsed conditions. However, frequently one cannot know *exactly* what initial population is appropriate to model any particular "pulsed" experiment, unless the process creating or measuring the initial population is significantly faster than all of the processes being modelled by the ME so that the population is effectively created or measured instantaneously. The reactive processes being modelled perturb the nascent population. Thus, the expected population $\mathbf{p}_0$ is never present in the experiment, and the time-dependent population $\mathbf{p}(t)$ calculated from it is slightly perturbed from the population that would be calculated from the population actually present once the initial population creation process has completed. This is represented schematically in Fig. 3. The impact of this effect on the interpretation of a ME calculation is usually negligible compared to other approximations inherent in the ME treatment unless t_c, the timescale of the population creation process, is particularly large. Clearly these interpretation issues do not effect results such as the unimolecular rate constant, where initial populations are not involved.

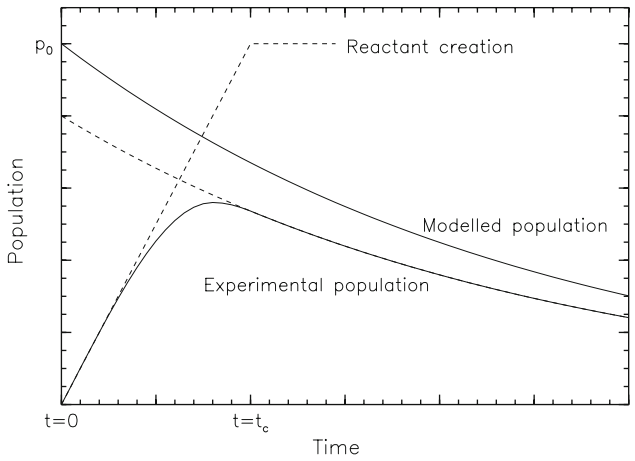

Fig. 3 Schematic representation of the initiation of a "pulsed" experiment with an initial population creation process that occurs over a finite time t_c, not significantly faster than the reactive processes modelled by the ME. The expected population $\mathbf{p}_0$ is never achieved in the experiment due to reaction of the nascent population. ME modelling using $\mathbf{p}_0$ as the initial population therefore solves a "nearby" problem

4 Unimolecular dissociation: the smallest eigenvalue, pseudo-steady-state rate constant

The magnitude of the smallest eigenvalue of the ME has traditionally been the quantity of most interest to

researchers due to its ready interpretation as the classical unimolecular rate constant k_u. Though various methods have been used over the years to calculate this eigenvalue, the Nesbet method has come to dominate the field due to its superior performance.

4.1 Nesbet's method

In 1965, Nesbet [12] proposed a method for calculating the lowest eigenvalue and corresponding eigenvector of a Hamiltonian matrix. In 1978, Gaynor et al. [11] adapted Nesbet's original method to the present case of finding the smallest eigenvalue of the ME matrix. The adapted algorithm in explicitly energy-grained form was given in later papers, such as Gilbert et al. [74].

The basic Nesbet algorithm can be derived in a number of ways, such as a gradient-type search of the *Rayleigh Quotient* surface. While such derivations lead to insights into the convergence of the method to the eigenpair with the desired extremal eigenvalue, the intent of the method is more clearly demonstrated as an iteration to minimise the *residual* vector.

From any trial vector $\tilde{\mathbf{x}}$ one can estimate a corresponding eigenvalue estimate $\tilde{\lambda}$, for example through a Rayleigh Quotient (neglecting, for the time being, any numerical difficulties). The corresponding residual is given by

$$\mathbf{r} = A\tilde{\mathbf{x}} - \tilde{\lambda}\tilde{\mathbf{x}} = (A - \tilde{\lambda}I)\tilde{\mathbf{x}}. \tag{19}$$

If $\tilde{\lambda}$ was the actual eigenvalue of A, and the $\mathbf{r} = \mathbf{0}$ solution of Eq. 19 was found exactly, the corresponding eigenvector of A would be given accurately by adding the solution vector to $\tilde{\mathbf{x}}$. If instead one uses the eigenvalue estimate $\tilde{\lambda}$ and solves for $\mathbf{r} = \mathbf{0}$ approximately using inverse of the trivially invertible diagonal of the diagonally dominant matrix A as an approximation to the true inverse, a better eigenvector estimate is obtained by adding the approximate $\mathbf{r} = \mathbf{0}$ solution to $\tilde{\mathbf{x}}$. If one then accurately updates $\tilde{\lambda}$ from the updated $\tilde{\mathbf{x}}$, $\tilde{\lambda}$ converges monotonically to the desired extremal eigenvalue [11, 12].

In the case of the lowest eigenvalue being small the matrix–vector product $A\mathbf{x}$ cannot be formed accurately, so the Rayleigh Quotient cannot be used for $\tilde{\lambda}$. Fortunately, an alternate way of calculating the eigenvalue estimate is available. Decomposing A into a conservative collisional part and purely reactive part $A = C - K$, with $K = \mathrm{diag}(k_i)$, by summing over the eigenpair relation $A\mathbf{x} = \lambda\mathbf{x}$, we have

$$\sum_j (A\mathbf{x})_j = \sum_j \left(\sum_k C_{jk}x_k - k_j x_j \right) = -\sum_j k_j x_j = \lambda \sum_j x_j \tag{20}$$

as $\sum_j \sum_k C_{jk}x_k = 0$ due to the conservative nature of the collisional transfer process. Hence, provided $\sum_j x_j \neq 0$,

$$\lambda = -\frac{\sum_j k_j x_j}{\sum_j x_j} \tag{21}$$

giving the eigenvalue as a weighted average of the eigenvector. Any representation of the eigenvector such as the symmetrised version $S\mathbf{x}$ or the "regularised" version $\mathbf{h} = S^2\mathbf{x}$ yields equivalent versions of the expression for λ by simple substitution in Eq. 21. These expressions for the eigenvalue are not subject to destructive cancellation errors unless $\sum_j x_j$ (or its equivalent) is small. In practice it is found that $\sum_j x_j$ is not small for the desired eigenvector corresponding to the smallest eigenvalue.[1]

The Nesbet update can be applied sequentially, calculating an update to a single element of $\tilde{\mathbf{x}}$ before updating $\tilde{\lambda}$ and moving on, or simultaneously for all elements with a constant $\tilde{\lambda}$. The latter is more in tune with modern approaches and easier to implement. Any difference in the speed of convergence is immaterial, as the entire calculation converges extremely quickly.

Gaynor et al. [11] made a number of other adaptations in an attempt to minimise the effect of roundoff error. Really, these adaptations are not important. The success of the method in difficult low temperature and pressure cases does not rely on any small increase in accuracy in the calculated residual. The calculated residual will still be completely inaccurate as the true residual gets small, a situation that occurs regularly as the Nesbet method is capable of calculating eigenpairs to high accuracy. The success of the method lies in the fact that the method is a single-vector update iteration: the correction calculated from the residual is added to the existing vector. Any inaccurate update elements will be orders of magnitude smaller than the elements that they are being added to, having no significant detrimental impact on the accuracy of elements that yield a small residual.

For unimolecular reactions with more than one product formed the Nesbet method gives the long-time population and total rate of reactant loss. The unimolecular reaction rate for each channel is simply calculated after the relevant eigenvector has been determined by inserting the channel's microscopic rate constants into Eq. 21.

[1] Equation 20 shows that when $K = 0$ (the pure relaxation case) either $\lambda = 0$ or $\sum_j x_j = 0$. In this $K = 0$ case the smallest eigenvalue is zero so that $\sum_j x_j$ need not be (indeed the normalisation of $\mathbf{x}$ and its strictly positive character imply a large value for the sum). This sum must be zero for the remaining set of eigenvectors with non-zero eigenvalues which control pure relaxation to equilibrium. The reactive $K \neq 0$ case can be seen as a perturbation to the pure relaxation case and changes the sum of the eigenvector elements only a small amount, yielding a small sum calculated from differences between large numbers, a classic candidate for catastrophic cancellation. Subsequently the calculation of all but the smallest eigenvalue from the corresponding eigenvector through these weighted average expressions is unstable.

4.2 Other methods

While the Nesbet method is, for good reason, the dominant method used for finding the smallest eigenpair of a unimolecular reaction problem, other methods are used. In the unimolecular case the most notable of these is the Davidson method [75]. The Davidson method can be viewed as an attempt to speed the convergence of the Nesbet method by not discarding residual information after each iteration. Instead, each Nesbet update vector is added to a basis for a subspace, defining an *orthogonal subspace projection* method [76, 77].

When they work, the Davidson and the closely related Olsen [78] methods converge in a smaller number of iterations than the Nesbet method, particularly when the initial eigenvector estimate is far from the actual eigenvector. However, the method is not a single-vector update iteration. A small eigenproblem must be solved and a larger matrix–vector product is required at each step to form the eigenvector estimate. As such the Davidson and Olsen methods are far more prone to the ravages of roundoff error than the Nesbet method, meaning that difficult low temperature and pressure cases with small eigenvalues and small eigenvector elements cannot be accurately solved with the Davidson or Olsen methods.

Krylov subspace methods (the Lanczos method in the symmetric case, generalised in the Arnoldi method [76, 77, 79, 80]) are orthogonal projection methods in which the subspace grows from a seed vector $\mathbf{v}$ as $A^{m-1}\mathbf{v}$ at the mth iteration. These general methods suffer from similar inaccuracies for small eigenvalues as the Davidson and Olsen methods. However, Frankcombe and Smith [27] have demonstrated that small eigenvalue sensitivity can be returned to the Krylov subspace methods by applying a weighting to the inner product space the method works in. The resulting method is known as a *weighted inner product subspace projection* (WIPSP) method. It can be demonstrated that applying the Lanczos method to the symmetrised matrix SAS^{-1} is applying a WIPSP method, and doing so allows one to solve a wider range of problems than trying to diagonalise the raw A. Diagonalising A^{T} is even better, weighting the population vector space so all elements of the desired eigenvector are of order unity (cf. Fig. 2) [27].

4.3 Transient populations

Transient evolution of the population can be important in modelling unimolecular reactions. Spectral (Eqs. 8–12) and direct integration methods (Eq. 17) can be equally applied to calculating the time-dependent populations that reveal transient behaviour.

While appropriate methods will be mainly described in the following section on isomerisation reactions (where transient behaviours take on a far more central role), we here mention in passing one class of methods that are not applicable to reversible isomerisation reactions. These are generalisations of the Nesbet method for finding eigenvectors corresponding to interior eigenvectors. Two examples are the high-order Nesbet eigenvectors (HONE) and explicitly restricted subspace (ERS) Nesbet methods [28, 57]. These start with low-quality spectral data from a standard eigensolution method and use special properties of the Nesbet iteration to refine the eigenvectors to a level at which they become sufficiently accurate to perform meaningful modelling of transients at medium to long times.

5 Isomerisation reactions

5.1 General considerations

Virtually all of the considerations for solving a unimolecular ME apply to an ME describing a reversible isomerisation, or a series of reversible isomerisations and irreversible decompositions. For some applications a single eigenpair describing the long-time evolution of the system is all that is required. However, generally one wants more detail of the evolution of the system in time, meaning a larger set of eigenpairs are needed. There are usually multiple small eigenvalues describing behaviour occurring on long time scales so that a single eigenpair is often insufficient to model even reasonably long times. There is usually one small eigenvalue for each reactive barrier between wells, and which of these eigenpairs one is interested in is critically dependent on what information one is trying to extract from the ME (i.e. complex interspecies relationships resolved in time or simply a single rate constant for a particular process) [34, 81]. In any case, an efficient algorithm analogous to the Nesbet method has not been developed for multi-well MEs, which do not exhibit the strong diagonal dominance of the unimolecular ME matrix.

The matrices for multi-well MEs are well structured. While the order of the species elements within the total state space $\mathbf{p}(t)$ is arbitrary, certain orderings lead to more obvious structure in the matrix and assist in implementing packed storage schemes. One of the more obvious orderings is to concatenate the state spaces from each individual well, with the grains arranged in ascending energy within each well. This ordering leads to a blocked ME matrix with dense blocks on the diagonal and diagonal matrices for the off-diagonal blocks. An example of the matrix structure for the reaction scheme

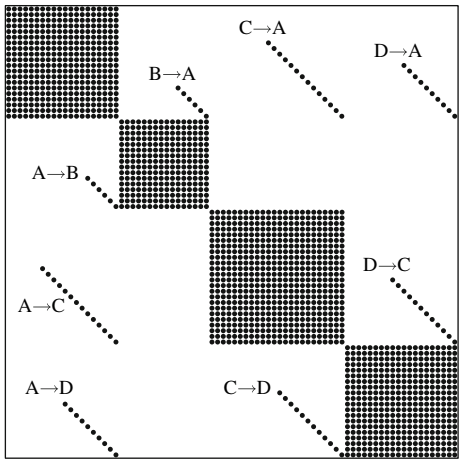

Fig. 4 Structure of a four-well isomerising system. Non-zero elements shown in black. Interconversion rate constants labelled with the reaction they represent

$$\begin{aligned} &A \rightleftharpoons B \\ &A \rightleftharpoons C \\ &A \rightleftharpoons D \\ &C \rightleftharpoons D \end{aligned} \tag{22}$$

with this element ordering is shown in Fig. 4. The block structure of the matrix is clear.

It should be noted that the symmetrisation of the ME, taken for granted in the unimolecular case, can similarly be effected for the multi-well ME. An important detail is that the vector **f** from which to construct the symmetrising transformation $SAS^{-1} = (SAS^{-1})^{\mathrm{T}}$ must be normalised in sections to reproduce the overall Boltzmann distribution, obeying the relevant equilibrium constants $K = [\mathrm{B}]/[\mathrm{A}] = \sum_{i\in\mathrm{B}} f_i / \sum_{i\in\mathrm{A}} f_i$. These K are derived from the microscopic rate constants.

Clearly, the matrix represented in Fig. 4 contains large regions whose elements are zero. Using packed storage (storing only non-zero parts of the matrix and constructing matrix-vector products component-wise) is very attractive when solving large multi-well MEs. The percentage of the matrix that is identically zero increases dramatically with the number of species being modelled.

5.2 Approaches to the solution in the isomerisation case

While the considerations for calculating the solution of the ME are similar for unimolecular and multi-well MEs, the ranges of possible behaviours of $\mathbf{p}(t)$ are not. The modelled population for the unimolecular problem simply relaxes towards equilibrium, with some possible adjustment due to depletion at high energies from reaction. The behaviour of the populations of multi-well systems, on the other hand, is difficult to predict. Multi-well systems can move through a number of steady state-like regimes at different times and energies. Populations at different energies, or of different isomers, can take on a steady-state character at different times. These steady-state behaviours are the basis of direct inversion methods of solving the ME [30, 35, 82–84]. At long times a steady state will generally emerge over the whole of the state space. This final long-time state is similar to the long-time state of the unimolecular ME.

Like unimolecular MEs, how difficult the multi-well ME is to solve numerically is dependent on the conditions being modelled. One of the main guides to the difficulties that will be encountered can be obtained by examining the combined Boltzmann vector, **f**. For low-temperature problems, the elements of this vector corresponding to high energies will be small. The existence of elements within a few orders of magnitude of the prevailing numerical precision indicates that numerical difficulties are likely in a spectral solution.

For small to medium sized, well-behaved problems at not too low temperatures and pressures, standard eigensolution routines in double precision [14, 58] can be used to calculate the spectral expansion, without taking any particular notice of the structure of the spectrum or the eigenvectors. Direct diagonalisation methods such as the QR method can be used in moderately sized cases, though this method scales badly and requires explicit storage of the matrix.

For larger problems, or when packed storage is used, iterative methods relying on the matrix–vector product must be used. While subspace projection methods [27, 76, 77, 79, 80] are popular generally for finding extremal eigenpairs of large matrices in a scalable manner, the nature of the spectrum and desired eigenvectors of multi-well ME matrices means that convergence can be extremely slow, or may stagnate completely. Furthermore, catastrophic loss of precision is common in low temperature and pressure cases.

If the temperature or pressure is too low to allow accurate spectral decomposition, direct integration of the ODE can be used. However, direct integration with off-the-shelf stiff integrators scales with the third power of the system size so that large networks of isomers cannot be modelled with a reasonable computational effort. For a modification to the standard stiff integrator that is scalable for large networks of isomers, see Sect. 7.

5.3 Solving the eigenproblem: the diffusion approximation

Vastly superior convergence of Krylov subspace-based methods can be achieved by implementing a shift and invert transform. The eigenvectors of the matrix

$(A - \sigma I)^{-1}$ are identical to those of A, with the eigenvalues transformed according to $1/(\lambda - \sigma)$. A small or zero σ transforms the desired small eigenvalues to large and well-spaced eigenvalues, a regime in which the Krylov subspace methods perform best. Note that the existence of a zero eigenvalue in the conservative ME means σ must be set non-zero in that case, to avoid attempting to invert a singular matrix.

To implement the shift and invert technique in a scalable manner requires that

$$(A - \sigma I)^{-1}\mathbf{v}, \tag{23}$$

for arbitrary $\mathbf{v}$, be calculated quickly. For a dense matrix finding $(A - \sigma I)^{-1}$ is an operation that scales with n^3, reducing the practicality of the shift and invert technique.

A more tractable approach can be found if the diffusion approximation is applied to the collisional processes. This changes the ME that is being solved, but in many cases the approximation is a small one. With the diffusion approximation the CET part of the ME matrix is reduced to tri-diagonal form. If the elements in $\mathbf{p}(t)$ are ordered so that the elements from all the isomers corresponding to the same energy are together,[2] the matrix becomes banded [32]. The bandwidth (not including the main diagonal) is the same as the number of isomers.

The calculation of $(A - \sigma I)^{-1}\mathbf{v}$ for a banded matrix A can be performed much more quickly and scalably than for a dense matrix. A Cholesky (GG^{T}), LDL^{T} or non-symmetric LU factorisation of a banded matrix of order n and bandwidth p requires computational effort of the order of np^2. The matrix needs only to be factorised once and once factorised, calculating the product of the inverse of the matrix and an arbitrary vector is of the order of np, giving the fast inverse required to allow efficient use of a shift and invert strategy. The speed and accuracy of this method has been demonstrated by Gates et al. [32].

A minor complicating issue is what type of factorisation to use. As the factorisation needs to be performed once only, the difference in the speed of the factorisation is overshadowed by accuracy and stability considerations, which favour the Cholesky factorisation if symmetry is used explicitly. The basic LU factorisation is the most robust if one is willing to abandon symmetry altogether. The LU factorisation does not depend on the definiteness of the matrix and so is stable at low temperatures and allows the spectrum to be shifted in order to focus on interior regions.

[2] A suitable ordering from the example of Fig. 1 is p_5, p_1, p_6, p_2, p_7,…

6 Time-dependent evolution in complex-forming bimolecular reactions

Certain classes of bimolecular reactions can also be modelled using ME methods [25]. A ME approach is relevant when the dynamics of the reaction are significantly influenced by unimolecular processes. This is the case when bimolecular species come together to form a unimolecular intermediate or a long-lived collision complex that behaves like a unimolecular species.

6.1 Linearised bimolecular channels

The most basic way to treat such a reaction is to model the bimolecular reaction as an irreversible source term in a traditional chemical activation ME [1]. The unimolecular or multi-well part of the reaction is treated in isolation to calculate the eigendecomposition. The bimolecular source term is added at the projection and propagation stage of the calculation, either approximately or by explicit integration [1, 56].

This approach to bimolecular reactions with a long-lived unimolecular intermediate is limited to modelling reactions whose bimolecular parts have a predetermined time dependence, such as an unchanging *infinite source*. A more satisfying and flexible approach is applicable under pseudo-first-order conditions [39]. The total time-dependent populations of two species A and B undergoing the reaction

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{C} \tag{24}$$

with reactant B in excess are then described by the equations

$$\begin{aligned}\frac{\mathrm{d}[\mathrm{A}]}{\mathrm{d}t} &= -k^{\mathrm{A}\to\mathrm{C}}[\mathrm{A}][\mathrm{B}] + \sum_i k_i^{\mathrm{C}\to\mathrm{A}}[\mathrm{C}]_i \\ &= -\widehat{k}^{\mathrm{A}\to\mathrm{C}}[\mathrm{A}] + \sum_i k_i^{\mathrm{C}\to\mathrm{A}}[\mathrm{C}]_i\end{aligned} \tag{25}$$

where $\widehat{k}^{\mathrm{A}\to\mathrm{C}} = k^{\mathrm{A}\to\mathrm{C}}[\mathrm{B}], [\mathrm{C}]_i$ is a species element describing the energy-grained population of species C and $k_i^{C\to A}$ is an appropriate microscopic rate constant for the dissociation reaction. The pseudo-first-order conditions give

$$\frac{\mathrm{d}[\mathrm{B}]}{\mathrm{d}t} \approx 0 \tag{26}$$

making $\widehat{k}^{\mathrm{A}\to\mathrm{C}}$ a constant and Eq. 25 a linear differential equation. By assuming that the reactant not in excess remains thermally distributed, the population of the bimolecular state can be incorporated into a linear first-order system of differential equations of the form of Eq. 2 by adding a single species element to $\mathbf{p}$ for the bimolecular state ([A]). One row and column is added to the ME matrix

A. The elements of the additional row of the matrix A corresponding to the bimolecular state are comprised of $k_i^{C\to A}$ and $-\widehat{k}^{\mathrm{A}\to\mathrm{C}}$ terms to reconstruct Eq. 25. The corresponding additional column is constructed of $\widehat{k}_i^{\mathrm{A}\to\mathrm{C}}$ terms (the splitting of $\widehat{k}^{\mathrm{A}\to\mathrm{C}}$ into each energy level of species C, $\sum_i \widehat{k}_i^{\mathrm{A}\to\mathrm{C}} = \widehat{k}^{\mathrm{A}\to\mathrm{C}}$), describing the gain in the population of C. An additional $-k_i^{C\to A}$ term appears on the diagonal of the matrix to account for dissociation from the unimolecular state (retaining the zero column sum property for this conservative ME matrix).

The $k_i^{C\to A}$ terms can be readily calculated from temperature-dependent rate data by inverse Laplace Transform techniques [85]. The detailed balance condition is

$$\widehat{k}_i^{\mathrm{A}\to\mathrm{C}} f^{\mathrm{A}} = k_i^{\mathrm{C}\to\mathrm{A}} f_i^{\mathrm{C}} \tag{27}$$

where f_i^{C} describes the Boltzmann distribution of the unimolecular species normalised against the bimolecular population f^{A} according to the equilibrium constant K for the reaction,

$$K = \sum\nolimits_i f_i^{\mathrm{C}} \Big/ f^{\mathrm{A}}[\mathrm{B}]. \tag{28}$$

6.2 Irreversible product channels

In the reversible isomerisation case of Sect. 5, different portions of the Boltzmann vector for the whole system $\mathbf{f}$ were normalised separately to satisfy the various equilibrium constant ratios. When bimolecular channels are added the same procedure is used, with the element of $\mathbf{f}$ describing the bimolecular state given by a ratio such as Eq. 28. The overall normalisation of $\mathbf{f}$ is once more arbitrary, usually giving $\|\mathbf{f}\| = 1$ for some appropriate norm. Symmetrisation of the matrix is achieved with the usual transform derived from the inverse square root of elements of $\mathbf{f}$. This presents no problems for conservative systems with fully reversible pseudo-first-order bimolecular channels as the elements of $\mathbf{f}$ are well-defined and non-zero.

On the other hand, modelling irreversible bimolecular reactions is not so straight forward [22, 23]. While the linear system of differential equations can be formed in all cases, a ME including irreversible reactions to explicitly included states yields an equilibrium distribution $\mathbf{f}$ with zero elements (as in the long-time limit the reaction proceeds to completion, depleting at least one of the species), so that the inverse square root is undefined. While the eigenproblem could still be solved in asymmetric form, the matrix S—undefined in this $f_i = 0$ case—is still required to apply the projection theorem in order to calculate the projection coefficients for the spectral propagation.

One option for dealing with this is to include the reaction as though it were reversible and setting the fictitious equilibrium constant sufficiently favouring products so that the reverse reaction does not contribute significantly. Careful testing of the effect of the false equilibrium is required in this case.

A second option is not to explicitly include the irreversible states in the modelled state space, as in the traditional unimolecular dissociation case. The various parts of the equilibrium vector $\mathbf{f}$ can then be normalised appropriately, ignoring the irreversible reactions. The non-conservative system can then be solved normally. To calculate the irreversible product populations after propagation, one can apply either conservation of population for a single irreversible product, or explicit integration of the time-dependent flux:

$$P_{\mathrm{irrev}}(t) = \int_0^t \sum_i k_i p_i(\tau)\mathrm{d}\tau \tag{29}$$

where $P_{\mathrm{irrev}}(t)$ is the desired product population, $p_i(\tau)$ is the population modelled by the ME and k_i are appropriate microscopic rate constants for the irreversible reaction.

6.3 Solving the ME

Solving the eigenproblem for MEs including bimolecular channels is very similar to solving the underlying unimolecular or isomerisation problem. In particular, applying the diffusion approximation yields a matrix which can be well packed and quickly inverted. If the grains describing the thermalised bimolecular states are appended to the end of the state space describing the unimolecular system, applying the diffusion approximation and ordering the elements within the unimolecular system appropriately (Sect. 5) yields a banded *arrowhead* matrix. The structure of the main part of the matrix retains the banded form of the unimolecular or isomerising system without the bimolecular channels, while the bimolecular states add non-zero rows and columns on the bottom and right of the matrix. Without pivoting, factorising such a banded arrowhead matrix maintains the arrowhead structure of the factors, again yielding a fast inverse to allow shift and invert iterative methods to be applied [22].

We finish this section by once more mentioning numerical integration of the ODE associated with the discretized ME. Clearly there is no significant theoretical difference between integrating the MEs for unimolecular dissociation or isomerisation reactions, multi-well systems and linearised collision complex-forming bimolecular reactions, as they are all described by the same linear first-order ODE, Eq. 2. Numerical integration indeed works well. One advantage that numerical integration has over spectral methods is that it does not rely on the linear nature of the ODE. MEs including bimolecular channels can be numerically integrated *without* invoking pseudo-first-order conditions. While numerical integration of bimolecular

reactions is common when integrating complex reaction schemes, the authors are unaware of published work integrating non-equilibrium populations in the context of a ME describing a second-order bimolecular system.

When direct integration is used to solve the ME, the advantages of working with the symmetrised ME are considerably reduced. Unlike in the spectral solution case the matrix S is otherwise unneeded. Hence, causing S to be undefined by including irreversible channels in the ME explicitly can be handled gracefully using numerical integration of the ODE.

7 Recent developments in scalable solutions for multi-well MEs

In a recent paper, the current authors directly compared three methods for solving a particular multi-well ME: spectral solution with a direct diagonalisation routine, spectral solution with a shift and invert Lanczos iterative diagonalisation, and direct numerical integration with a stiff integrator [63]. It was found that at high temperatures the Lanczos diagonalisation was the fastest, but the requirement of high precision at low temperatures and pressures made numerical integration favourable in that regime.

The problem with all three methods is scaling; all scale as n^3 as the number of isomers being modelled increases. Thus, one cannot solve truly large problems with these methods as they stand.

7.1 Scalable linear system solves

Both the shift and invert Lanczos and direct integration methods require forming matrix–vector products $A\mathbf{v}$—which scales linearly with the number of isomers if implemented correctly—and finding solutions to linear systems of equations of the form of

$$(aA + bI)\mathbf{x} = \mathbf{b} \tag{30}$$

(from Eqs. 18 and 23). Solving Eq. 30 for $\mathbf{x}$ through a factor and backsolve method is the source of the n^3 scaling in these methods. A scalable way of solving Eq. 30 would yield overall scalable methods.

One such scalable method is the GMRES method [86, 87]. GMRES builds the solution to Eq. 30 in a Krylov subspace that requires only $(aA + bI)\mathbf{v} = aA\mathbf{v} + b\mathbf{v}$ to build. These matrix–vector products can be formed in linear-scaling time, as the most onerous part is forming $A\mathbf{v}$.

However, GMRES applied to Eq. 30 for multi-well ME matrices converges extremely slowly, due to the large spread of the eigenvalues of $(aA + bI)$. A standard technique to apply in that scenario is *preconditioning*. One requires an easily inverted approximation to $(aA + bI)^{-1}$ to use as a preconditioner. An astute reader may realise we have pointed out in Sect. 5.3 that the diffusion approximation matrix D is a good approximation to A, and that it is easily invertible. As pointed out in Sect. 6.3, including bimolecular channels turns the diffusion matrix into a banded arrowhead matrix, which is again easily invertible in linearly-scaling time. The preconditioned version of Eq. 30,

$$(aD + bI)^{-1}(aA + bI)\mathbf{x} = (aD + bI)^{-1}\mathbf{b}, \tag{31}$$

is much easier to converge with the Krylov-based GMRES iteration due to the small spectral width of $(aD + bI)^{-1}(aA + bI) \approx I$. $(aD + bI)^{-1}\mathbf{b}$ is easily formed and the GMRES iteration requires only

$$(aD + bI)^{-1}\mathbf{d} \tag{32}$$

with

$$\mathbf{d} = (aA + bI)\mathbf{x} \tag{33}$$

at each iteration. All of these components can be evaluated in linear-scaling time, yielding an overall linear-scaling GMRES solution method.

7.2 Linear-scaling shift and invert Lanczos and direct integration

Inserting the linear-scaling preconditioned GMRES iterative solution of Eq. 30 in place of the dense solves in the shift and invert Lanczos and direct stiff integration methods has proven very successful [62, 88]. The solutions produced by the GMRES-based methods are practically indistinguishable to those produced by these two methods with direct factor and solve dense system solves. The near-perfect linear scaling of the GMRES-based direct integration has been explicitly demonstrated for up to 12 isomers [88].

The agreement between the dense-solve and GMRES-solve versions of these two methods mean that these linear-scaling methods inherit the advantages and disadvantages of the two methods. Direct integration with a GMRES-based inversion is a very robust method, successfully integrating initial populations under the most extreme conditions tested: 300 K and 130 Pa [88]. The shift and invert diagonalisation procedure in double precision is significantly faster than the direct integration method. However, this is only reliable at relatively high temperatures and pressures. At lower temperatures and pressures, the method quickly loses accuracy due to the effects of finite precision algebra. As with direct diagonalisation to find the spectral solution [23, 28, 59, 63], implementing the entire shift and invert Lanczos algorithm in sufficiently high precision arithmetic allows the ME solution to be

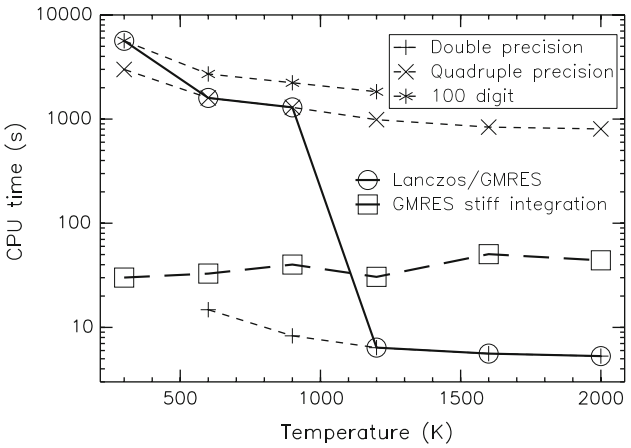

Fig. 5 Effective CPU times to solve a three-well C_3H_4 ME modelling 130 kPa conditions with GMRES-based methods. The *bold Lanczos/GMRES line* traces the solution time at the minimum precision required to reliably find the solution at that temperature. Adapted from Frankcombe and Smith [62]

found accurately at any desired temperature and pressure [62, 63]. As usual, such an approach comes at the cost of substantially increased execution time (but at the same scaling of the algorithm).

These trends are illustrated in Fig. 5. This figure shows the CPU time required to solve a representative $^1CH_2 + C_2H_2 \rightleftharpoons C_3H_4 \rightarrow C_3H_3 + H$ ME, where the C_3H_4 species exists as three inter-converting isomers. When double precision GMRES-based shift and invert Lanczos is sufficient to solve the problem, it does so in the fastest time. However, as the temperature is reduced the Lanczos calculation must be performed in higher and higher precision, at a substantial cost in CPU time. The GMRES-based direct stiff integration algorithm solved the problem in a similar amount of CPU time irrespective of the temperature; indeed a trend towards shorter execution times for lower temperatures is evident. For robustness, reasonable speed and scalability, direct integration is attractive [62, 63, 88].

It should be pointed out that coupling GMRES-based inversions and stiff integration has been applied previously in other contexts (see [89] and references therein). The authors were unaware of this work when developing these GMRES-based ME solution methods [62, 88].

8 Outlook

Methodology for the solution of the ME has developed significantly in recent years, with particular emphasis on extending the formalism to allow for effective modelling of more complex unimolecular and bimolecular reactions involving multiple isomeric species. While the formulation of the multi-isomer problem is relatively straightforward and a number of examples of such studies exist in the literature [22, 24, 29, 30–38], finding the solution to the ME presents considerable numerical challenges. These challenges revolve around the fact that standard eigensolver routines and spectral decompositions can suffer from catastrophic loss of precision due to the properties of the ME matrix. Furthermore, standard methods will not admit extension to large problems due to their poor scaling properties. This causes one to turn to a number of iterative methods typically utilised for large sparse matrix applications.

Many of the standard iterative methods fail for the ME because of the numerical error issues highlighted above. The drive towards general and robust ME simulation programs necessarily involves the development of specially tailored algorithms which are both scalable and do not suffer catastrophic (or insidious) failure in the presence of ill-conditioning in the matrix at lower temperatures.

Recent developments exploiting existing approximations have yielded specialised methods capable of solving ME problems, variously overcoming both loss of precision and poor scaling to large problems [23, 27, 28, 62, 63, 88]. Both rigorous and fast solution methods can now be a part of the kineticist's arsenal as the essential chemistry of elementary reactions is interfaced with the more global modelling of combustion and atmospheric environments.

References

1. Gilbert RG, Smith SC (1990) Theory of unimolecular and recombination reactions. Blackwell Scientific, Oxford
2. Schranz HW, Smith SC, Mebel AM, Lin SH (2002) J Chem Phys 117:7055
3. Macnamara S, Burrage K, Sidje RB (2008) Multiscale Model Simul 6:1146
4. Mebel AM, Kislov VV, Hayashi M (2007) J Chem Phys 126:204310
5. Flower DR, Pineau des Forêts G, Walmsley CM (2005) Astron Astrophys 436:933
6. Susnow RG, Dean AM, Green WH, Peczak P, Broadbelt LJ (1997) J Phys Chem A 101:3731
7. Schranz HW, Nordholm S (1983) Chem Phys 74:365
8. Venkatesh PK, Dean AM, Cohen MH, Carr RW (1997) J Chem Phys 107:8904
9. Venkatesh PK, Dean AM, Cohen MH, Carr RW (1999) J Chem Phys 111:8313
10. Tsang W, Bedanov V, Zachariah MR (1996) J Phys Chem 100:4011
11. Gaynor BJ, Gilbert RG, King KD (1978) Chem Phys Lett 55:40
12. Nesbet RK (1965) J Chem Phys 43:311
13. Higham NJ (1996) Accuracy and stability of numerical algorithms. SIAM, Philadelphia
14. Golub GH, van Loan CF (1996) Matrix computations, 3rd edn. Johns Hopkins University Press, Baltimore
15. Bukoski A, Blumling D, Harrison I (2003) J Chem Phys 118:843
16. Abbott HL, Harrison I (2006) J Chem Phys 125:024704

17. Abbott HL, Harrison I (2008) J Catal 254:27
18. Oref I, Tardy DC (1990) Chem Rev 90:1407
19. Nordholm S, Schranz HW (1995) In: Barker JR (ed) Advances in chemical kinetics and dynamics, vol 2A. JAI, Greenwich
20. Billing GD, Mikkelsen KV (1996) Introduction to molecular dynamics and chemical kinetics. Wiley, New York
21. Holbrook KA, Pilling MJ, Robertson SH (1996) Unimolecular reactions, 2nd edn. Wiley, Chichester
22. Frankcombe TJ, Smith SC, Gates KE, Robertson SH (2000) Phys Chem Chem Phys 2:793
23. Frankcombe TJ, Smith SC (2002) Faraday Discuss 119:159
24. Miller JA, Klippenstein SJ (2006) J Phys Chem A 110:10528
25. Fernández-Ramos A, Miller JA, Klippenstein SJ, Truhlar DG (2006) Chem Rev 106:4518
26. Neufeld PD, Janzen AR, Aziz RA (1972) J Chem Phys 57:1100
27. Frankcombe TJ, Smith SC (2000) J Comput Chem 21:592
28. Frankcombe TJ, Smith SC (2001) Comput Phys Commun 141:159
29. Quack M (1984) Ber Bunsen-Ges Phys Chem 88:94
30. Green NJB, Marchant PJ, Perona MJ, Pilling MJ, Robertson SH (1992) J Chem Phys 96:5896
31. Aguda BD, Pritchard HO (1992) J Chem Phys 96:5908
32. Gates KE, Robertson SH, Smith SC, Pilling MJ, Beasley MS, Maschhoff KJ (1997) J Phys Chem A 101:5765
33. Tsang W, Bedanov V, Zachariah MR (1997) Ber Bunsen-Ges Phys Chem 101:491
34. Klippenstein SJ, Miller JA (2002) J Phys Chem A 106:9267
35. Green NJB, Bhatti ZA (2007) Phys Chem Chem Phys 9:4275
36. Miller JA, Klippenstein SJ, Robertson SH, Pilling MJ, Green NJB (2009) Phys Chem Chem Phys 11:1128
37. Miller JA, Senosiain JP, Klippenstein SJ, Georgievskii Y (2008) J Phys Chem A 112:9429
38. Frankcombe TJ, Smith SC (2007) J Phys Chem A 111:3691
39. Hanning-Lee MA, Green NJB, Pilling MJ, Robertson SH (1993) J Phys Chem 97:860
40. Chevalier C, Warnatz J, Melenk H (1990) Ber Bunsen-Ges Phys Chem 94:1362
41. Sanchez AL, Linan A, Williams FA, Balakrishnan G (1995) Combust Sci Technol 111:277
42. Green NJB, Robertson SH, Pilling MJ (1994) J Chem Phys 100:5259
43. Robertson SH, Pilling MJ, Baulch DA, Green NJB (1995) J Phys Chem 99:13452
44. Robertson SH, Shushin AI, Wardlaw DM (1993) J Chem Phys 98:8673
45. Robertson SH, Pilling MJ, Green NJB (1996) Mol Phys 89:5131
46. Robertson SH, Pilling MJ, Gates KE, Smith SC (1997) J Comput Chem 18:1004
47. Jeffrey SJ, Gates KE, Smith SC (1996) J Phys Chem 100:7090
48. Smith SC, Gilbert RG (1988) Int J Chem Kinet 20:307
49. Miller JA, Klippenstein SJ, Raffy C (2002) J Phys Chem A 106:4904
50. Smith SC, McEwan MJ, Gilbert RG (1989) J Phys Chem 90:4265
51. Troe J (1977) J Chem Phys 66:4745
52. Troe J (1977) J Chem Phys 66:4758
53. Moler C, van Loan C (2003) SIAM Rev 45:3
54. Moler C, van Loan C (1978) SIAM Rev 20:801
55. Sidje RB (1998) ACM Trans Math Soft 24:130
56. Knyazev VD, Tsang W (1999) J Phys Chem A 103:3944
57. Frankcombe TJ (2002) Numerical methods in reaction rate theory. PhD thesis. University of Queensland, St. Lucia
58. Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Croz JD, Greenbaum A, Hammarling S, McKenney A, Sorensen D (1999) LAPACK users guide, 3nd edn. SIAM, Philadelphia
59. Blitz M, Beasley MS, Pilling MJ, Robertson SH (2000) Phys Chem Chem Phys 2:805
60. Gannon KL, Glowacki DR, Blitz MA, Hughs KJ, Pilling MJ, Seakins PW (2007) J Phys Chem A 111:6679
61. Bailey DH (1995) ACM Trans Math Softw 21:379
62. Frankcombe TJ, Smith SC (2003) J Chem Phys 119:12741
63. Frankcombe TJ, Smith SC (2003) J Theor Comput Chem 2:179
64. Barker JR (1983) Chem Phys 77:301
65. Shi J, Barker JR (1990) Int J Chem Kinet 22:187
66. Barker JR, King KD (1995) J Chem Phys 103:4953
67. Vereecken L, Huyberechts G, Peeters J (1997) J Chem Phys 106:6564
68. Miller JA, Chandler DW (1986) J Chem Phys 85:4502
69. Chandler DW, Miller JA (1984) J Chem Phys 81:455
70. Miller JA, Klippenstein SJ (2001) J Phys Chem A 105:7254
71. Hahn DK, Klippenstein SJ, Miller JA (2001) Faraday Discuss 119:79
72. Brown PN, Byrne GD, Hindmarsh AC (1989) SIAM J Sci Stat Comput 10:1038
73. Hindmarsh AC (1983) In Stepleman RS (ed) Scientific computing. North-Holland, Amsterdam
74. Gilbert RG, Luther K, Troe J (1983) Ber Bunsen-Ges Phys Chem 87:169
75. Davidson ER (1975) J Comput Phys 17:87
76. Saad Y (1992) Numerical methods for large eigenvalue problems. Manchester University Press, Manchester
77. Parlett BN (1980) The symmetric eigenvalue problem. Prentice-Hall, Englewood Cliffs
78. Olsen J, Jørgensen P, Simons J (1990) Chem Phys Lett 169:463
79. Cullum JK, Willoughby RA (1985) Lanczos algorithms for large symmetric eigenvalue computations. Birkhäuser, Boston
80. Lehoucq RB, Sorensen DC, Yang C (1998) ARPACK users' guide. SIAM, Philadelphia
81. Miller JA (2002) Faraday Discuss 119:255
82. Setser DW, Rabinovitch BS, Simons JW (1964) J Chem Phys 40:1751
83. Snider N (1984) J Chem Phys 80:1885
84. Schranz HW, Nordholm S (1984) Chem Phys 87:163
85. Davies JW, Green NJ, Pilling MJ (1986) Chem Phys Lett 126:373
86. Saad Y (1996) Iterative methods for sparse linear systems. PWS Publishing Co, Boston
87. Saad Y, Schultz MH (1986) SIAM J Sci Stat Comput 7:856
88. Frankcombe TJ, Smith SC (2003) J Chem Phys 119:12,729
89. Brown PN, Hindmarsh AC, Petzold LR (1994) SIAM J Sci Comput 15:1467