

PAPER • OPEN ACCESS

Fast species ranking for iterative species-oriented skeletal reduction of chemistry sets

To cite this article: Martin Hanicinec *et al* 2020 *Plasma Sources Sci. Technol.* **29** 125024

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Fast species ranking for iterative species-oriented skeletal reduction of chemistry sets

Martin Hanicinec¹ , Sebastian Mohr² and Jonathan Tennyson^{1,*} 

¹ Department of Physics and Astronomy, University College, London, Gower St., London, WC1E 6BT, United Kingdom

² Quantemol Ltd, 320 City Road, London, EC1V 2NZ, United Kingdom

E-mail: j.tennyson@ucl.ac.uk

Received 31 August 2020, revised 10 November 2020

Accepted for publication 24 November 2020

Published 23 December 2020



Abstract

A fast algorithm is developed for ranking the species in a chemistry set according to their importance to the modeled densities of user-specified species of interest. The species ranking can be constructed for any set of user-specified plasma conditions, but here we focus predominantly on low-temperature plasmas, with gas temperatures between 300 and 1500 K covering the typical range of ICP and CCP plasma sources. This ranking scheme can be used to acquire insight into complex chemistry sets for modeling plasma phenomena or for a species-oriented reduction of the given chemistry set. The species-ranking method presented is based on a graph-theoretical representation of the detailed chemistry set and establishing indirect asymmetric coupling coefficients between pairs of species by the means of widely used graph search algorithms. Several alternative species-ranking schemes are proposed, all building on the theory behind different flavors of the directed relation graph method. The best-performing ranking method is identified statistically, by performing and evaluating a species-oriented iterative skeletal reduction on six, previously available, test chemistry sets (including O₂-He and N₂-H₂) with varying plasma conditions. The species-ranking method presented leads to reductions of between 10 and 75% in the number of species compared to the original detailed chemistry set, depending on the specific test chemistry set and plasma conditions.

Keywords: plasma modeling, chemistry reduction, species ranking, chemistry graph

(Some figures may appear in colour only in the online journal)

1. Introduction

Utilizing the unique properties of the low-temperature plasma has become an integral part of almost every industry sector, spanning over a wide range of applications such as medicine, biotechnology, surface modification, microfabrication, har-

vesting energy, thrusters, ozone generation or abatement systems, to name just a few. As an example of the importance of the low-temperature plasma technologies for our every day lives, it has been estimated that as much as one-third of steps involved in the manufacturing of microelectronic technologies are plasma-based [1]. While providing desirable properties, the very complex nature low-temperature plasma systems also poses challenges for describing and understanding plasma phenomena. Understanding the plasma properties is crucial for the optimization of plasma-based processes and technologies and the only way to acquire an insight of any significant depth is through numerical modeling techniques.

* Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

There are many methods available for modeling low-temperature plasma properties and behavior which vary in both accuracy and complexity. No matter what kind of plasma model of whatever spatial dimensionality is considered, each is built around a chemistry set which describes the volumetric interactions between all the species tracked in the model, and additionally, the interactions between the species and surfaces. A volumetric and surface chemistry set is a very important base for every plasma model, accounting for the majority of sources and sinks of species. Many pre-compiled detailed chemistry sets for various feed gases and applications can be found in the literature, see for example references [2–8]. As a consequence of advances in gas kinetics, published chemistry sets are becoming increasingly larger. For plasma physics modeling applications, chemistry sets may routinely include up to a hundred species and many thousands of reactions. For example, Koelman *et al* [9] provide a chemistry set for the splitting of CO₂ in non-equilibrium plasmas which contains 73 unique species involved in 5724 reactions. In the combustion modeling community, where very large chemistry sets have been used for longer than in the plasma modeling community, some applications require sets that may contain several hundred or even thousands of species and ten thousands of reactions [10].

Any chemistry set might contain redundant species or reactions, which could be removed without compromising any plasma model outputs for the space of model parameters linked to the desired application. In fact, almost all published chemistry sets contain redundant species and reactions [11]. There are two main reasons for this. First, chemistry sets tend to cautiously include species and reactions with uncertain importance for the plasma model built around the given set. Second, even if the importance of each species and reaction is somehow tested, it might be done so for a much wider range of input parameters and conditions, than the range of conditions, the chemistry set is finally utilized over. While large chemistry sets do not significantly impact the performance of more approximative plasma models with low computational cost, such as volume-averaged 0D models, this is not true for plasma models resolving several spatial dimensions, where the larger chemistry sets (and particularly larger numbers of species in the set) means that models can rapidly become prohibitively expensive to run.

While one might come across published chemistry sets which have been reduced to smaller sizes to accurately model only a particular application with a particular set of plasma conditions, such as the chemistry set for plasma in air reduced by Bak and Cappelli [12] from the detailed chemistry set by Kossyi *et al* [13], these published reduced sets are, by design, only valid for a narrow domain of plasma conditions and parameters described in the publications and therefore cannot be readily adopted for a more general case. It follows that methods for identifying and eliminating redundant species (and reactions) from very large detailed chemistry sets, or more generally, methods reducing the dimensionality of the underlying systems of partial differential equations are of great importance for more computationally costly plasma models, capable of generating more insight than simple 0D models.

The computational cost of solving spatially resolved plasma models increases greatly with the number of species in the chemistry set up to the point, that solving 3D plasma models with more than a few species might become too time-consuming, costly, and impractical. The number of reactions in the chemistry set N_R on the other hand does not increase the dimensionality of the system of governing equations but might still increase its stiffness. Then there is also a question of the interpretability of a solution, which might become unclear with very complex chemistry sets.

Several review papers have dealt with the problem of chemistry set reduction, mostly in combustion modeling [14–18]. Reduction methods might be classified into three main categories: lumping, time scale analysis, and skeletal reduction [18, 19]. The lumping method [20] may be useful, whenever the species of a chemistry set can be clustered together into groups with very similar chemical properties, and each group can be then treated as one pseudo-species. This method is used very frequently in plasma modeling, for example for lumping of vibrational levels of species A into few groups [2] or a single lumped pseudo-state $A(\nu)$ [21], or for lumping several electronic metastable states into a single compound state A^* [21]. Various reduction methods falling under the time scale analysis category aim to define fast reactions or highly reactive species in the chemistry set and approximate those as partial equilibrium reactions and quasi-steady state species. Their corresponding PDEs can then be solved explicitly as algebraic equations or their solutions tabulated, decreasing the dimensionality of the whole system of PDEs. Various methods based on time scale analysis exist, such as quasi-steady state analysis [22], partial equilibrium analysis and more systematic approaches including the intrinsic low dimensional manifold method [23, 24] and computational singular perturbation (CSP) method by Lam and Goussis [22, 25–27] have been widely adopted in the past in plasma and combustion modeling. Finally, skeletal reduction methods seek to identify and eliminate species and reactions unimportant for a given set of particular simulation conditions and applications. The present work concentrates on skeletal reduction.

The skeletal reduction methods can be further categorized as reactions-oriented and species-oriented, based on whether they aim to reduce the number of reactions or species in the kinetic system. Reactions-oriented methods for elimination of reactions from a chemistry set include among others sensitivity analysis [28, 29] and detailed reduction method [30, 31]. Elimination of reactions from a chemistry set does not reduce the dimensionality of the system of governing PDEs, and therefore generally the decrease in plasma model computation time due to eliminating a number of reactions from a set is only marginal and due to faster computation of the species sources terms (as well as a possible decrease in stiffness of the system). For that reason, species-oriented reduction methods are arguably more important. The simplest method which is frequently used by the plasma modeling community is the elimination of species with very low density. This method was used for reduction by, among others, Van Gaens and Bogaerts [5], Liu *et al* [32], and Turner [21], but this method is rather crude and cannot be generalized. More systematic methods for species

elimination include, for example, sensitivity analysis [11, 33], Jacobian analysis, CSP [34], principal component analysis (PCA) [35], the simulation error minimization connectivity method of Nagy and Turanyi [10] and the directed relation graph method (DRG) of Lu and Law [36–38] as well as its variations, such the DRG with error propagation (DRGEP) method of Pepiot-Desjardins and Pitsch [39] and DRG-aided sensitivity analysis [22, 40]. Again, most of the above-mentioned methods were developed and used specifically for combustion modeling, but some were recently also used in plasma-assisted combustion applications (PCA and DRGEP, both by Bellemans *et al* [35, 41]) or even in plasma physics and astrochemistry modeling, such as DRG by Sun *et al* [42], DRGEP by Venot *et al* [43] and sensitivity analysis by Obrusník *et al* [44]. The majority of the methods mentioned above are even implemented in a single software package by Lebedev *et al* [45].

In this work, we present a fast, automated method of ranking all species in a plasma system according to their importance for modeling a certain subset of selected *species of interest*, together with a simple iterative species-oriented method of skeletal reduction utilizing such a ranking. What we are aiming for is the application of an automated *chemistry set optimizer* which can be attached to a web-based database of plasma kinetic data. Input for such an application would take the form of a detailed chemistry set (which is valid for a wide range of conditions and simulation parameters), set of simulation parameters and set of species of interest, while output would be an optimized chemistry set of a minimal reduced size, reproducing the densities of all the species of interest with the given plasma model with errors lower than pre-defined value compared to the input detailed chemistry set. The detailed chemistry set could be either a pre-compiled feature of the kinetic database, or dynamically assembled from the processes in the database and their kinetic data. While the methods of lumping and time scale analysis each have their domain of application, the skeletal reduction is an obvious choice for the outlined chemistry set optimizer application, where the detailed set of species and reactions is already expected to be present in the database of kinetic processes, and the attached reduction method aims to identify the essential species (and their reactions) and ignore the rest.

This work is divided into several sections. The next section provides the introduction of the *ranking-based species-oriented iterative skeletal reduction method*, which was used to assess different species ranking schemes. In section 3, we present a number of competing ranking schemes based on graph representation of detailed chemistry sets. Different methods of chemistry graph edge weights distribution are considered, all based on a single evaluation of a plasma model with the detailed chemistry set and different flavors of the DRG theory. On this weighted chemistry graph, several methods to evaluate indirect coupling between pairs of species are presented, based on different graph theory search algorithms. Every pair of edge-weighting method and indirect coupling method defines a species ranking scheme, which can be used for a species-oriented skeletal reduction of the detailed chemistry set. Section 4 gives details for a number of test

reduction cases which are used to identify the statistically best-performing ranking scheme in section 5; finally, the last section provides conclusions and outlook.

2. Ranking-based iterative reduction method

The existing species-oriented skeletal reduction methods mentioned in the introduction all use a thresholding parameter in one form or another (such as the direct interaction coefficient threshold ϵ from the DRG theory), with a somewhat arbitrary value. This might pose a problem for some applications, such as the chemistry set optimization tool outlined in the introduction. The arbitrary thresholding coefficients ϵ employed in the existing species-oriented skeletal reduction methods generally have a very uncertain mapping to the errors of the model outputs induced by the reduction with given ϵ . The output errors also are not monotonic with ϵ , forbidding the use of binary search to converge toward the optimal ϵ value. Together with the higher computational cost of some of the existing methods, this fact renders employing these methods problematic for the automated chemistry set optimizer tool we aim for. The development of both the reduction method described in the rest of this section and the species ranking scheme addressed in section 3 was motivated by the outlined web-based kinetic processes database application, and specifically as an add-on module for the plasma chemistry Quantemol database (QDB) [8]. However, the method itself is not exclusive to QDB, but applicable to be used with any existing database of plasma kinetic processes, or indeed in any application requiring a fast species ranking, or fast species-oriented skeletal reduction method. Apart from the QDB, many different online databases of plasma kinetic data exist, such as LXCat database for modeling low-temperature plasmas by Pitchford *et al* [46], Phys4Entry by Celiberto *et al* for modeling re-entry plasmas [47], databases for astrochemical modeling KIDA by Wakelam *et al* [48, 49] and UDfA by McElroy *et al* [50], fusion oriented databases by Murakami *et al* of NIFS, Japan [51] and by Park *et al* of Data Center for Plasma Properties, NFRI, Korea [52], or BASECOL by Dubernet *et al* [53], to name a few.

2.1. Introduction of the method

In this work, we present a simple chemistry set reduction method overcoming some of the difficulties mentioned in the previous paragraph. The method is based on ranking all the species in the model based on their importance for modeling densities of a defined set of species of interest and iteratively eliminating the lowest-ranked species, while periodically checking after each iteration, if the reduction error does not exceed a defined maximum threshold value. This way, the validity of the reduced chemistry set is always ensured with respect to the pre-selected species of interest and the set of reduction conditions. The rest of this section outlines the general framework behind the method and some key concepts, while the species ranking method development is described in section 3.

The following key concepts and definitions are important for the present work:

- *Detailed chemistry set*: the set containing all the species and reactions.
- *Reduced chemistry set*: a set with at least one species eliminated from the detailed set.
- *Species elimination*: removal of all the reactions involving the given species (on either side).
- *Reduction conditions*: a set of application-specific conditions (such as pressure, power, etc), for which the reduced chemistry set should yield a low *reduction error*.
- *Reduction error* δ : the relative difference of *species of interest* densities between the detailed and reduced models.
- *Allowed error* Δ : maximal δ allowed.
- *Species of interest*: set of species, densities of which need to be modeled within Δ with the reduced chemistry set.
- *Species ranking*: species hierarchy reflecting how important each species is for modeling all of the *species of interest*. Rank of each species in the chemistry set is determined by its *ranking score*.

The general structure of our reduction method is summarized by the flow diagram given as figure 1. The method uses a *plasma model* to generate results used as inputs for a species ranking method and also to check the *reduction error* after each species elimination. Although the reduction framework described here is defined very generally, throughout this work we used a 0D global model as a *plasma model* and the *reduction error* took the form of a mixed error function adopted from the work of Nagy and Turanyi [10]

$$\delta = \max_{i,j} 2 \frac{|A_i^{\text{red}}(t_j) - A_i^{\text{full}}(t_j)|}{A_i^{\text{full}}(t_j) + \max_j A_i^{\text{full}}(t_j)}, \quad (1)$$

where index i runs over all of the species of interest with number densities A_i and the index j runs over the set of pre-selected *times of interest* $\{t_j\}$. The concept of the times of interest is further explained in section 3.

The global model used throughout this work was the *PyGMol* (Python global model), which was developed for QDB and specifically for the reduction method presented. Although there are several global models available within the plasma modeling community, such as *GlobalKin* by Kushner [54], *Plasma-R* by Kokkoris [55] or *ZDPlasKin* by Pancheshnyi et al [56], the choice of developing a new code was driven by the need for Python implementation which allowed seamless integration with the rest of the reduction framework.

The *PyGMol* global model calculates the number densities n_i (where the index i runs over all the heavy species in the associated chemistry set), as well as the electron temperature T_e , all as a function of time, by integrating the set of particle density balance and electron energy density balance PDEs. The electron density n_e is not solved for explicitly but rather implicitly by enforcing the charge neutrality. The heavy species temperature T_g is treated as a constant input parameter, rather than being solved for self-consistently.

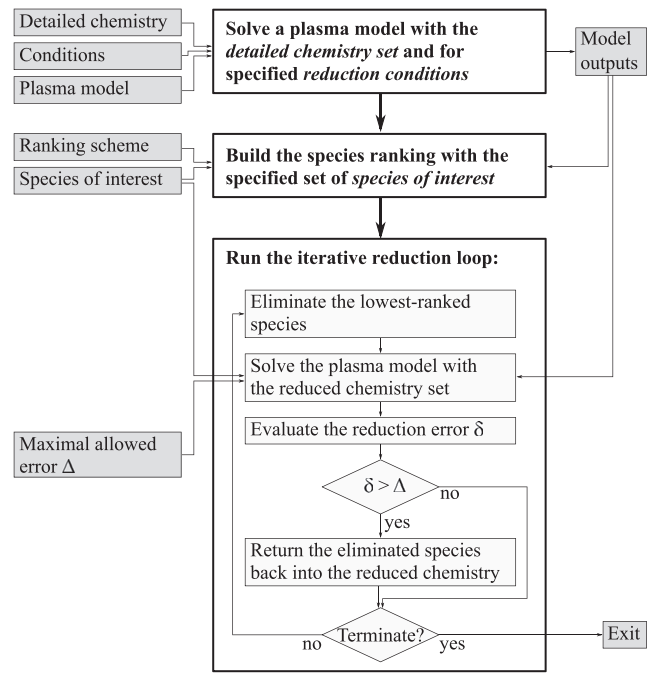


Figure 1. Flow diagram outlining the general structure of our species-oriented reduction method framework. Inputs to the reduction method are on the left-hand side of the diagram.

The particle density balance equation includes contributions from volumetric reactions, flow, and from diffusion (surface sinks) losses as well as surface sources. The volumetric kinetics are defined by the Arrhenius formula for both electron processes (2), and heavy species processes (3),

$$k = A \left(\frac{T_e}{1 \text{ eV}} \right)^n \exp \left(-\frac{E_a}{T_e} \right) \quad (2)$$

$$k = A \left(\frac{T_g}{300 \text{ K}} \right)^n \exp \left(-\frac{E_a}{T_g} \right), \quad (3)$$

where each reaction rate coefficient k is parametrized by factors A , n , and E_a . The surface kinetics are described by a set of species *sticking coefficients* s_i for each species, together with an arbitrary number of *return species* per any stuck species, with *return coefficients* r_{ik} . The surface coefficients are coupled to a simplistic particle diffusion model, where surface particle loss rate S_i and source rate R_i are expressed as

$$S_i = -\frac{D_i}{\Lambda^2} n_i s_i, \quad (4)$$

and

$$R_i = -\sum_k r_{ik} S_i = \sum_k \frac{D_k}{\Lambda^2} n_k s_k r_{ik} \quad (5)$$

respectively, as used (among others) by Kushner in *GlobalKin* [54]. Here, D_i is the diffusivity of the i —the species and Λ is the diffusion length, a function of the plasma geometry.

The electron energy density balance includes contributions of power from any external sources absorbed by the plasma, elastic and inelastic collisions between electrons and heavy species, generation and loss of electrons in volumetric reactions and power lost by electrons and ions surface losses.

A detailed description of the PyGMol global model can be found on the QDB website. While obviously being approximative, the PyGMol model should however suffice here, since the quantitative results of the model are not the aim of the present work; an equally simplistic model was used in a similar study by Turner [21, 57].

The condition for terminating the reduction method from figure 1 was set to be N_δ subsequent iterations resulting in exceeding the maximal reduction error allowed $\delta > \Delta$. An appropriate value of N_δ must be chosen to balance the trade-off between the number of eliminated species and the number of plasma model calls.

Finally, it needs to be noted, that validity of the reduced chemistry set for all the species of interest is strictly ensured within the density error Δ only in the scope of the same plasma model that was used in the reduction method. This plasma model, of course, needs to be sufficiently simple to handle the full detailed chemistry set with very fast run-time, so no reduction is strictly necessary, when using the simple model only. Reduced chemistry sets need to be created for more sophisticated (usually spatially resolved) plasma models where the number of species in the set is of much greater importance. In a case of any *end-application* plasma model (different from the one used in the reduction process), the validity of the reduced chemistry set cannot be readily extrapolated and completely ensured. The use of any reduced chemistry set should therefore always be preceded by careful consideration of the similarity of both the plasma models. As an example, if a reduced set is expected to be used within a plasma model capturing the same *physics* as the plasma model employed in the reduction method, only in multiple dimensions, it might be appropriate to run the reduction method for more than one point of the *reduction conditions* parameter space and/or use lower maximal reduction error Δ within the reduction method, than what is an acceptable accuracy of the end-application model.

By the same token, that validity of the reduced chemistry set is by design only ensured for the single set of plasma conditions that are the reduction conditions. When performing a chemistry set reduction targeting a specific modeling application, it might be appropriate to run the reduction with multiple sets of reduction conditions sampled from the parameter space, and only eliminate species that are flagged redundant in all of those runs. Such a parameter sweep, however, can be performed on top of the reduction method presented in this work and therefore will not form part of this paper.

2.2. Influence of species ranking

It is clear, that the method for ranking the species is the most important part of the reduction method presented. Good species ranking scores should correlate strongly with the error in the species of interest densities induced by removing the species from a chemistry set. Completed reduction runs may be analyzed by plotting the reduction error as a function of the number of species retained in the chemistry set. As an example, figure 2 shows such a reduction plot for an N_2 - H_2 chemistry set (described more closely in section 4.2) and plasma model with arbitrary reduction conditions. The solid line starts with

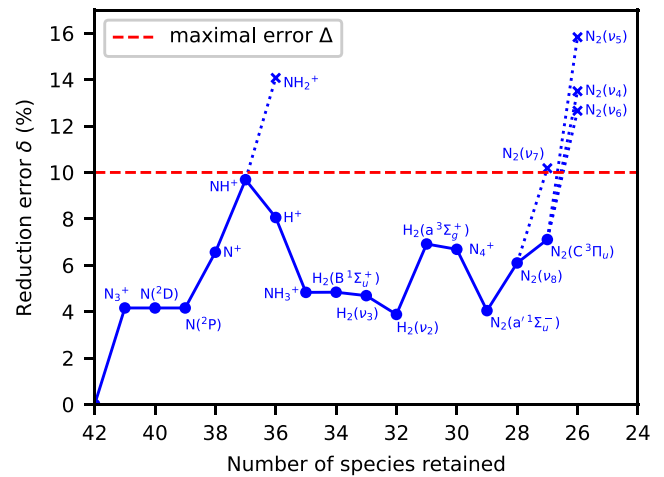


Figure 2. A reduction plot for an N_2 - H_2 chemistry set and plasma model with arbitrary reduction conditions. The error function is plotted against the number of species retained in the chemistry. The solid line follows all the species eliminated from the chemistry set while keeping the reduction error δ below the maximal value allowed (red dashed line). The dotted lines show species that could not be eliminated from the chemistry set with an acceptable reduction error. The species eliminated (or not eliminated) at each iteration are given next to the plotted points.

the detailed chemistry set containing in total 42 species, and follows all the species eliminated from the set while keeping the reduction error δ below the maximal value allowed (in this case, $\Delta = 10\%$). The dotted lines show species, which could only be eliminated with an unacceptable reduction error, and were therefore reinstated back into the chemistry set. The species ranking order coming into the reduction method in this particular case was $[N_3^+, N(^2D), N(^2P), N^+, NH^+, NH_2^+, H^+, NH_3^+, H_2(B^1\Sigma_u^+), H_2(\nu_3), H_2(\nu_2), H_2(a^3\Sigma_g^+), N_4^+, N_2(a^1\Sigma_u^-), N_2(\nu_8), N_2(\nu_7), N_2(C^3\Pi_u), N_2(\nu_6), N_2(\nu_5), N_2(\nu_4), \dots]$, where subsequent unsuccessful elimination of the three species $N_2(\nu_6), N_2(\nu_5)$ and $N_2(\nu_4)$ triggered the termination condition for $N_\delta = 2$. Figure 2 also demonstrates that the reduction error is generally not monotonic with the number of species eliminated.

Finally, some caveats need to be noted regarding the process of species elimination from a chemistry set. The integrity of the chemistry sets needs to be protected at all times. For example, each retained species must have at least one production and consumption channel and at least one positive ion must be retained in the set. This adds extra constraints on species elimination. Similarly, in some cases, elimination of an a species upsets the balance of the chemistry set to the point where the plasma model solution fails. Cases like this are treated as if the iteration resulted in a higher than allowed reduction error. Safeguards against inconsistent chemistry sets after species elimination are coded into the elimination routine of the framework.

3. Species ranking

Several methods for fast species ranking are proposed, each based on a single evaluation of a plasma model with the

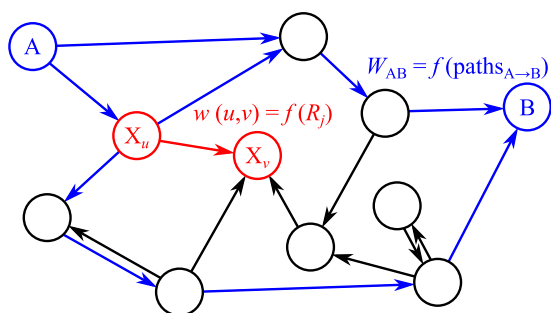


Figure 3. Depiction of a *chemistry graph* with nodes representing species and directional edges representing asymmetric direct interactions between species, weighted by the direct interaction coefficients. A direct interaction coefficient $w(u, v)$ between u th and v th species is depicted in red, while the asymmetric coupling coefficient W_{AB} between species A and B is hinted in blue, depending on all the paths $p(A \rightarrow B)$ in the chemistry graph leading from A to B.

detailed chemistry set, and a graph-theoretical representation of the chemistry. A *chemistry graph* is created, with species as nodes and directed weighted edges between species. The species ranking hierarchy is then built and it reflects the indirect coupling between each species and the species of interest.

3.1. Chemistry graph

Figure 3 shows a schematic depiction of such a chemistry graph, with 11 nodes (species) and 17 directed edges. Each directed edge (u, v) , leading from u th species to the v th species, is weighted by a *direct interaction coefficient* $w(u, v)$, which is generally a function of reaction rates R_j of all the reactions in the detailed chemistry set. By analogy with graph theory nomenclature, when regarding a single edge, or a direct interaction coefficient, u th and v th species will be referred to as *tail species* and *head species* respectively, where the v th head species is directly affected by the presence of all the reactions involving u th tail species. Any direct interaction coefficient $w(u, v)$ is a measure of asymmetric coupling between two species that are directly related through some of the elementary reactions in the detailed chemistry set.

Coupling between two species, however, exists even if they do not share any elementary reactions. The indirect asymmetric *coupling coefficients* W_{AB} between species A and B are therefore defined, reflecting the global (often indirect) effect of the presence of species A (and all its reactions) on the modeled density of the species B. For coupling coefficient W_{AB} , the species A and B will, by convention, be referred to as the *source species* and *target species*, respectively. A coupling coefficient W_{AB} is generally a function of direct interaction coefficients along all the paths leading from the source species A to the target species B in the chemistry graph. This way, in the resulting species ranking, the rank of each species A will reflect its importance (or rather the collective importance of all the reactions involving the species A) for modeling species B belonging to the set of species of interest; even a species with a very low density will rank relatively high, if it acts as an important intermediate species for production or consumption of any of the species of interest.

The following subsection defines several methods for calculating $w(u, v)$ and W_{AB} and how the species ranking is extracted from the set of coupling coefficients.

3.2. Direct interaction coefficients

Several methods for edge weights distribution in chemistry graphs (and for calculating the direct interaction coefficients $w(u, v)$) have been proposed in the literature.

In the original DRG theory [36], Lu and Law propose that $w(u, v)$ takes the form of the sum of absolute values of production and consumption rates for head species by all the reactions involving the tail species, with a normalizing factor of the sum of absolute values of production and consumption rates for head species by all the reactions in the chemistry set:

$$w(u, v) = \frac{\sum_{j=1}^{N_R} |a_{vj}R_j\delta_u^j|}{\sum_{j=1}^{N_R} |a_{vj}R_j|} \tag{6}$$

The index j runs over all N_R reactions in the chemistry and R_j is the reaction rate of the j th reaction

$$R_j = k_j \prod_l n_{lj}^L \tag{7}$$

k_j is a reaction rate coefficient of the j th reaction (2), (3), and n_{lj}^L is the density of the l th reactant of the j th reaction. The term δ_u^j selects only reactions involving the tail species

$$\delta_u^j = \begin{cases} 1 & \text{if } u - \text{th species in } j - \text{th reaction,} \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

and a_{vj} is the net stoichiometric coefficient of head species in j th reaction

$$a_{vj} = a_{vj}^R - a_{vj}^L \tag{9}$$

In their other work [38], Lu and Law propose an alternative definition for the direct interaction coefficient to (6), replacing the denomination factor by the absolute value of the total net production rate of the head species

$$w(u, v) = \frac{\sum_{j=1}^{N_R} |a_{vj}R_j\delta_u^j|}{\left| \sum_{j=1}^{N_R} a_{vj}R_j \right|} \tag{10}$$

This means, that unlike the direct interaction coefficient defined by (6) which is bound between 0 and 1, (10) becomes singular for the head species in equilibrium.

Another definition of the direct interaction coefficient was proposed by Pepiot-Desjardins and Pitch [39] in their DRGEP method. In their definition, the direct interaction coefficients are equal to the absolute value of net production of the head species by all the reactions involving the tail species, normalized by the maximum of total production or consumption of the head species, as

$$w(u, v) = \frac{\left| \sum_{j=1}^{N_R} a_{vj}R_j\delta_u^j \right|}{\max(P_v, C_v)} \tag{11}$$

The range of the direct interaction coefficient (11) also can be shown to be between 0 and 1 [39].

Finally, here we propose another definition of $w(u, v)$ which naturally emerges from (6), (10) and (11) as their combination

$$w(u, v) = \frac{\left| \sum_{j=1}^{N_R} a_{vj} R_j \delta_u^j \right|}{\left| \sum_{j=1}^{N_R} a_{vj} R_j \right|}. \quad (12)$$

Similar to (10) and sharing the same denominator, (12) is not bound between 0 and 1 but rather is singular for the head species in equilibrium.

The direct interaction coefficients defined by (6), (10)–(12) only reflect the importance of volumetric reactions involving tail species toward modeling the head species density. While this is appropriate for combustion modeling, where the DRG method and its variants all originated, it is, in fact, blind to the effect of surface reactions and conversions, which are often of great importance in plasma modeling. Therefore, in the present work, modified definitions are proposed, addressing also the production and consumption rates of head species by diffusion losses and surface conversions, such as

$$w_{\text{DRG}}(u, v) = \frac{\sum_{j=1}^{N_R} |a_{vj} R_j \delta_u^j| + |S_u r_{uv} - S_v (1 - r_{vv})|}{\sum_{j=1}^{N_R} |a_{vj} R_j| + \left| \sum_{i=1}^{N_S} S_i r_{iv} - S_v \right|}, \quad (13a)$$

$$w_{\text{DRG}'}(u, v) = \frac{\sum_{j=1}^{N_R} |a_{vj} R_j \delta_u^j| + |S_u r_{uv} - S_v (1 - r_{vv})|}{\left| \sum_{j=1}^{N_R} a_{vj} R_j + \sum_{i=1}^{N_S} S_i r_{iv} - S_v \right|}, \quad (13b)$$

$$w_{\text{DRGEP}}(u, v) = \frac{\left| \sum_{j=1}^{N_R} a_{vj} R_j \delta_u^j + S_u r_{uv} - S_v (1 - r_{vv}) \right|}{\max(P_v, C_v)}, \quad (13c)$$

$$w_{\text{DRG}''}(u, v) = \frac{\left| \sum_{j=1}^{N_R} a_{vj} R_j \delta_u^j + S_u r_{uv} - S_v (1 - r_{vv}) \right|}{\left| \sum_{j=1}^{N_R} a_{vj} R_j + \sum_{i=1}^{N_S} S_i r_{iv} - S_v \right|}. \quad (13d)$$

Equations (13a)–(13d) correspond to (6), (10)–(12) respectively, with added species consumption and production rates on surfaces. The index i runs over all N_S species in the chemistry set and S_i is the surface sticking rate of i th species, defined by (4). The total production and consumption rates of the head species P_v and C_v are

$$P_v = \sum_{j=1}^{N_R} \max(a_{vj} R_j, 0) + \sum_{i=1}^{N_S} S_i r_{iv} (1 - \delta_v^i), \quad (14a)$$

$$C_v = \sum_{j=1}^{N_R} \max(-a_{vj} R_j, 0) + S_v (1 - r_{vv}), \quad (14b)$$

with

$$\delta_v^i = \begin{cases} 1 & \text{if } i = v, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

These modified direct interaction coefficients also preserve the ranges of the original ones, that means (13a) and (13c) are bound between 0 and 1, while (13b) and (13d) are not.

3.3. Species coupling coefficients

With chemistry graph edges weighted by the direct interaction coefficients according to one of (13a)–(13d), the species coupling coefficients W_{AB} can be defined between any two species A, B, using some well-established graph search algorithms: *shortest path*, *maximum bottleneck path*, *maximal product path*, and *maximal flow* searches.

In the shortest path approach, the coupling coefficient between source species A and target species B is calculated as a reciprocal of the length of the shortest path leading from A to B, with the individual edges being attributed their own *length* equal to reciprocals of weights $w(u, v)$

$$W_{AB}^{\text{sh.path}} = \left[\min_{\text{all paths } p(A \rightarrow B)} \sum_{\text{all edges } (u,v) \in p} w^{-1}(u, v) \right]^{-1}. \quad (16)$$

Since all the edge weights (and therefore lengths) are by definition non-negative, the shortest paths might be calculated very efficiently, using, for example, Dijkstra's algorithm [58] with a very convenient computational complexity $\mathcal{O}(N_S^2)$.

Alternatively, one might define the coupling coefficient W_{AB} as the direct interaction coefficient of a rate-limiting step across all the paths $A \rightarrow B$. For each path $p(A \rightarrow B)$, the path rate-limiting step can be defined as

$$W_{AB,p}^{\min} = \min_{i=1}^{n-1} w(i, i+1), \quad (17)$$

with index i running through all the species in any one path p , and with source and target species A, B corresponding to $i = 0$ and $i = n$ respectively. The coupling coefficient $W_{AB}^{\text{lim.rate}}$ is then defined as the global rate-limiting step

$$W_{AB}^{\text{lim.rate}} = \max_{\text{all paths } p} W_{AB,p}^{\min}. \quad (18)$$

This is equivalent to the maximum bottleneck path problem, widest path problem or a maximum capacity route problem from graph theory. Similarly, as in the case of the shortest path approach, the maximal limiting rate search between any two species can be performed very efficiently, for example by a modified Dijkstra's algorithm with $\mathcal{O}(N_S^2)$ complexity [59].

Another option is to define W_{AB} as a product of all the direct interaction coefficients in a path $p(A \rightarrow B)$ with maximal such a product:

$$W_{AB}^{\text{err.prop.}} = \max_{\text{all paths } p(A \rightarrow B)} \prod_{\text{all edges } (u,v) \in p} w(u, v). \quad (19)$$

This definition of the species coupling coefficients only makes sense for direct interaction coefficients bound between 0 and 1, otherwise, the graph search would favor long, *convoluted* paths involving a maximal number of edges with $w(u, v) > 1$. Therefore, this approach only applies to the w_{DRG} (13a) and w_{DRGEP}

(13c) definitions of direct interaction coefficients, as $w_{\text{DRG}'}$ (13b) and $w_{\text{DRG}''}$ (13d) are not bound between 0 and 1, but rather become singular for head species in total equilibrium. The maximal propagated error search is inspired by the notion of error propagation proposed by Pepiot-Desjardins and Pitch [39] in their DRGEP method. If the direct interaction coefficients as edges are approximating errors induced to the density of a head species by the elimination of the tail species, the errors propagate along the paths from source to target species introducing geometric damping. The idea is that if some error is introduced to the prediction of the source species A, the longer the error needs to propagate along the path $p(A \rightarrow B)$ to reach the target species B, the smaller the effect will typically be. For edge weights bound between 0 and 1, the maximal product path problem is equivalent to the shortest path problem with modified edge lengths, as the coupling coefficient can be redefined as

$$W_{AB}^{\text{err.prop.}} = \exp \left[- \min_{\text{all paths } p(A \rightarrow B)} \sum_{\text{all edges } (u,v) \in p} - \ln w(u,v) \right], \quad (20)$$

which can be solved by any shortest path searching algorithm, which can handle zero-weighted edges, corresponding to $w(u,v) = 1.0$. It needs to be noted that Dijkstra's shortest path algorithm treats edges with zero weight as non-existing edges, which would generally result in an incorrect $W_{AB}^{\text{err.prop.}}$, therefore a different algorithm needs to be used to compute (20), e.g. Bellman–Ford algorithm [60], which would in our case have $\mathcal{O}(N_S^3)$ asymptotic complexity.

The last option considered in this work for expressing the coupling coefficients W_{AB} from a chemistry graph is using a maximum flow search. In analogy to the maximum bottleneck path problem (18), where the weights of the edges can be imagined as flow capacities and the flow capacity of the maximum-bottleneck path $p^{\text{widest}}(A \rightarrow B)$ is searched for, in the maximum flow problem, the maximum flow from A to B is sought, without the restriction of only one single path. The species coupling coefficient $W_{AB}^{\text{max.flow}}$ is then defined as a maximal total flow from A to B, while a partial flow through any single edge $f(u,v)$ does not exceed the edge capacity (or value of the direct interaction coefficient)

$$f(u,v) \leq w(u,v),$$

and flow is conserved in around any node

$$\sum_i f(i,v) = \sum_k f(v,k),$$

except for the source and target species. The maximum flow problem may be solved for example employing the Edmonds–Karp algorithm [61] with $\mathcal{O}(VE^2)$ complexity, where V is the number of vertices in the graph, while E is the number of edges. In the asymptotic case of a fully connected graph, the complexity in terms of the size of the chemistry set might be expressed as $\mathcal{O}(N_S^5)$, for a single pair of (A, B). This makes the maximal flow search relatively slow, in contrast to $W_{AB}^{\text{sh.path}}$, $W_{AB}^{\text{lim.rate}}$ and $W_{AB}^{\text{err.prop}}$ with $\mathcal{O}(N_S^{2-3})$ complexity for evaluating all pairs of $\{(X_i, B)\}$.

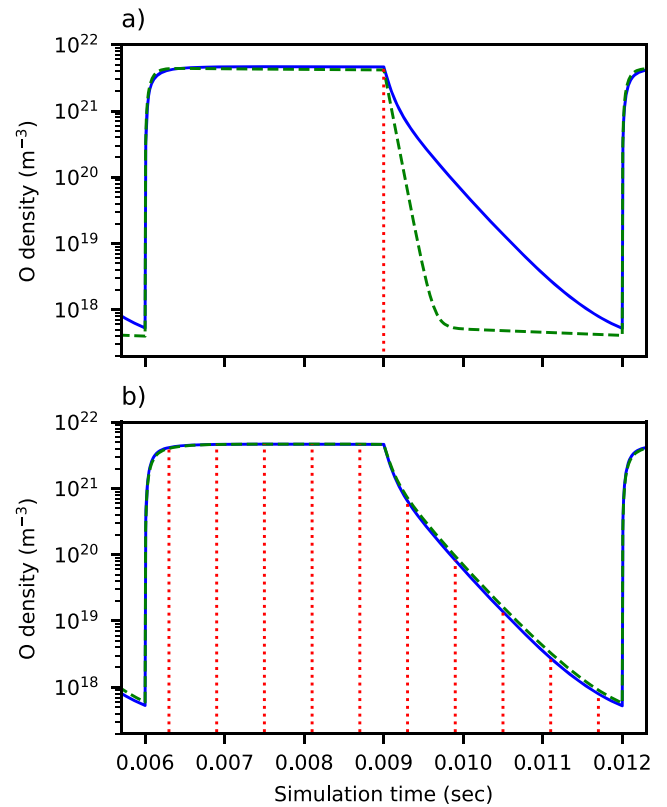


Figure 4. Comparison of the time evolution of atomic oxygen species in pulsed O_2 –He plasma modeled with the detailed and reduced chemistry sets. (a) The reduction was performed for a time of interest, coinciding with the end of the power pulse. (b) The reduction was performed for a set of times of interest, spanning over the whole pulsing period. In both cases, atomic oxygen was set as the only species of interest. Solid blue and green dashed lines show O density modeled using detailed and reduced chemistry set respectively, while the red dotted lines indicate the times of interest used for species ranking and the iterative reduction.

3.4. Species ranking

An instantaneous ranking score C_i^t for i th species X_i can be expressed as a maximum coupling coefficient between X_i and any of the pre-defined *species of interest* at a given simulation time t , such as

$$C_i^t = \max_k W_{X_i X_k}^t, \quad (21)$$

where index k runs over all the species of interest $\{X_k\}$. The instantaneous coupling coefficients C_i^t depend on the reaction rates R_j^t of the detailed chemistry set at any given time t . To ensure the reduction validity over the whole simulation time span, C_i^t are sampled over a sufficiently dense set of *times of interest* $\{t\}$, and the ranking scores are defined as

$$C_i = \max_{k,t} W_{X_i X_k}^t = \max_t C_i^t. \quad (22)$$

This way, the ranking score of any particular species covers the coupling between its presence in the chemistry set and the modeled number densities belonging to the species of interest sampled at the times of interest—which are exactly the outputs the reduced chemistry set is designed to preserve.

Table 1. Summary table of test chemistry sets, and associated reduction pressures p considered for testing of the species ranking methods. The number of species and reactions N_S and N_R are listed for each test chemistry set.

ID	Test set	N_S	N_R	p	Typical appl.	Source
S1	O ₂ –He	25	373	1 atm	Biomedical plasma	Turner [21]
S2	N ₂ –H ₂	42	408	1 atm	NH ₃ synthesis	Hong <i>et al</i> [4]
S3	N ₂ –H ₂	42	408	8 Pa	NH ₃ synthesis	Hong <i>et al</i> [4]
S4	CF ₄ –CHF ₃ –H ₂ –Cl ₂ –O ₂ –HBr	67	563	25 Pa	—	QDB [8]
S5	CH ₄ –N ₂	61	521	1 atm	—	QDB [8]
S6	Ar–NF ₃ –O ₂	39	310	100 Pa	—	QDB [8]

Figure 4 demonstrates the importance of choosing an appropriate set of times of interest $\{t\}$. It shows the time evolution of oxygen radical in an O₂–He atmospheric pressure plasma, comparing the PyGMol outputs with detailed and reduced chemistry for two reduced chemistry sets, each reduced with reduction and species ranking evaluated for a different set of times of interest. In both cases, the reduction was performed with $\Delta = 10\%$ and with only the O atom as a single species of interest. Figure 4(a) shows data for a chemistry set reduced using a single time sample, coinciding with the end of the power pulse. While the atomic oxygen density agrees with the detailed set sufficiently well for the steady-state phase, the reduction error during the afterglow phase might be unacceptable. Figure 4(b) shows that sampling the reaction rates from a larger times of interest set, spanning the pulsing period, results in a (albeit larger) reduced chemistry set which preserves atomic oxygen density even during the afterglow.

The choice of the sampling times of interest will always depend on the specific application and on the computational time budget. A logarithmic distribution of time samples might be more useful for calculations with a constant external power term, while a linear distribution might be preferred for a time-dependent power source, as shown in figure 4. The times of interest might also be sampled based on the detailed chemistry set solution, with the instantaneous sampling frequency proportional to the time derivatives of the densities of species of interest. The computational cost of the ranking algorithm will, however, increase linearly with the size of the samples set, because each time sample will effectively instantiate a separate chemistry graph which needs to be searched.

In section 3.2, we proposed 4 alternative definitions of direct interaction coefficients $w(u, v)$ (13a)–(13d), and in section 3.3, we proposed 4 alternative ways to extract the species coupling coefficients W_{AB} out of a chemistry graph, with directional edges weighted by $w(u, v)$. This defines in total 14 alternative methods of ranking species according to (22), with ranking scores being a function of species of interest, times of interest and time-dependent reaction rates simulated by a plasma model with the detailed chemistry set (the two definitions of the direct interaction coefficient (13b), (13d) are not compatible with the $W_{AB}^{\text{err,prop}}$ coupling coefficient (19).) All the proposed alternative species ranking methods were tested for a chemistry set reduction on an array of test sets, reaction conditions and species of interest (section 4) by the method described in section 2, with the results shown in section 5.

Table 2. Species included in the O₂–He chemistry set as compiled by Turner [21]. Lumped vibrational states for each neutral molecule are denoted by ν .

Neutrals	He, O, O ₂ , O ₃
Excited states	He*, He ₂ *, O(¹ D), O(¹ S), O ₂ (a ¹ Δ _g), O ₂ (a ¹ Δ _g , ν), O ₂ (b ¹ Σ _g ⁺), O ₂ (b ¹ Σ _g ⁺ , ν), O ₂ (ν), O ₃ (ν)
Positive ions	He ⁺ , He ₂ ⁺ , O ⁺ , O ₂ ⁺ , O ₃ ⁺ , O ₄ ⁺
Negative species	e, O ⁻ , O ₂ ⁻ , O ₃ ⁻ , O ₄ ⁻

4. Test reduction cases

Six test chemistry sets are introduced, to provide the basis for a number of test reduction cases generating the data for a statistical assessment of all the species ranking schemes introduced in section 3. Table 1 lists all the chosen test chemistry sets with their sizes (number of species and reactions), typical applications and source publications.

4.1. O₂–He test chemistry set (S1)

The first test chemistry set is one for O₂–He atmospheric pressure plasma taken from the work of Turner [21]. The detailed chemistry set contains 25 species and 373 reactions. Table 2 lists all the species included in the detailed chemistry set, while all the reactions and their kinetics can be found in the source publication. It is evident that the vibrational kinetics is not treated in great detail in the present model, with a single lumped vibrational state (denoted by ν) present for each neutral molecule. This simplification was justified in the source compilation by very dilute mixtures of oxygen in helium and by much lower densities of those vibrational states compared to their ground states. Reduction conditions for this detailed chemistry set should therefore not increase the oxygen ratio beyond the limit set by the source publication.

Perhaps the most topical use of the O₂–He plasma is for a biomedical application at atmospheric pressure. The source publication chose to approximate the so-called micro atmospheric pressure plasma jet (μ APPJ) [62], which is in many ways a typical example. The chemistry set source publication uses a device configuration of $1 \times 1 \text{ mm}^2$ channel with a length of 30 mm for its global model and with a constant power of 1 W supplied to the system for 3 ms (which is a residence time of the gas in the channel at a typical flow rate), followed by a further 3 ms of zero power, modeling the afterglow phase. The source publication does not mention considering any surface losses or species conversions, so a set of *default* surface interaction coefficients was considered for the present work,

Table 3. Default rules for species sticking coefficients s and return species R with return coefficients r considered for detailed chemistry sets and global modeling in present work. The default surface coefficients translate to total quenching of excited species and neutralization of positive ions.

Species	s	R	r
Neutrals	0.0	—	—
Exc. states	1.0	Ground state	1.0
Ions ⁺	1.0	Neutral	1.0
Ions ⁻	0.0	—	—

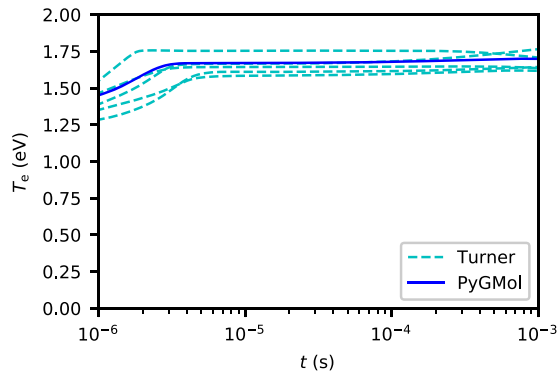


Figure 5. Comparison of the time evolution of electron temperature between the solution of PyGMol global model and the results of Turner [21], for the O₂–He chemistry set. Several lines are plotted from Turner which correspond to different random sets of reaction rate coefficients sampled within their uncertainty distributions.

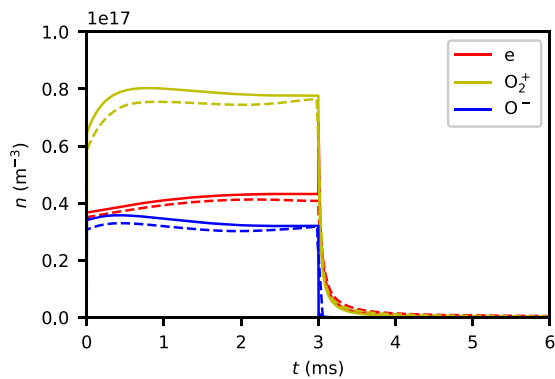


Figure 6. Densities of selected charged species during the power cycle for the O₂–He chemistry set: comparison between the PyGMol global model (solid line, —) and solution by Turner [21] (dashed lines, - - -).

as shown in table 3, for all the species, apart from a few exceptions of ions and excited species, without their neutral ground-state counterparts. The sticking coefficient for both He₂^{*} and He₂⁺ was considered $s = 1.0$, with return species $r_{\text{He}} = 2.0$ and similarly, $s_{\text{O}_4^+} = s_{\text{O}_4^-} = 1.0$, with $r_{\text{O}_2} = 2.0$. Identical surface coefficients were considered for both the on/off phases of the power pulsing.

The source publication [21] used a plasma model with kinetic data in the same functional form as used by PyGMol, providing an opportunity to validate our PyGMol global

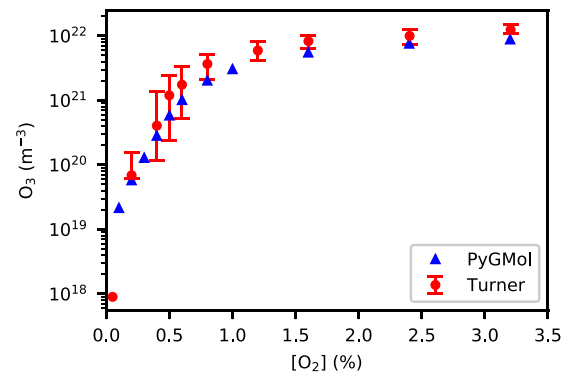


Figure 7. Ozone density at the end of the discharge pulse, as a function of the fraction of O₂ in the feedstock gas for the O₂–He chemistry set. Comparison between the PyGMol solution and that due to Turner [21].

model. Figures 5–7 show very good agreement between the data calculated by Turner [21] and our outputs from PyGMol model, for the same conditions and the same O₂–He chemistry set.

4.2. N₂–H₂ test chemistry sets (S2–S3)

The next two test chemistry sets for N₂–H₂ plasma are based on the source publication of Hong *et al* [4]. The source publication compiles a chemistry set containing 42 distinct species and 408 reactions, building on previous work of Gordiets *et al* [63, 64] and adapting it for atmospheric pressure. In contrast to the O₂–He test chemistry set, this set explicitly tracks the first few vibrationally excited states of the most abundant molecular species N₂ and H₂. All the species included are shown in table 4, while the reactions can be found in the source publication, together with their original sources.

Perhaps the most topical application for both low-pressure and atmospheric pressure is the ammonia synthesis. However, modeling of N₂–H₂ plasmas is also important for a wide range of other applications, such as plasma cutting [65, 66], arcjet thrusters for satellite propulsion [67, 68], AlN, TiN or SiN film deposition [69–71], etching of organic low-permittivity layers [72], nitriding [73], and the interaction of puffed N₂ with the H₂ plasma in fusion reactors [74]. In this work, we base two test chemistry sets on the data from Hong *et al*; one for atmospheric pressure conditions (S2) and one for low-pressure conditions (S3), see table 1.

Kinetic data for the volumetric reactions compiled by Hong *et al* are not all compatible with the PyGMol global model. Most of the electron processes in the source publication cite the cross-sectional data from the ZDPlasKin [56] internal database and many of the heavy species processes follow reaction coefficients described by varying functions of gas temperature, differing from the Arrhenius representation required by PyGMol. The volumetric reactional kinetics were adapted for PyGMol according to the following procedure:

- Electron processes described by cross-sectional data were all fitted to Arrhenius form, assuming the Maxwellian distribution on a grid of electron temperatures. For reproducibility reasons, the cross-sectional data for these

Table 4. Species included in the N₂–H₂ chemistry set as compiled by Hong *et al* [4]

Neutrals	H, H ₂ , N, N ₂ , NH, NH ₂ , NH ₃
Excited states	H ₂ (b ³ Σ _u ⁺), H ₂ (B ¹ Σ _u ⁺), H ₂ (c ³ Π _u), H ₂ (a ³ Σ _g ⁺), H ₂ (ν ₁)–H ₂ (ν ₃), N(² D), N(² P), N ₂ (A ³ Σ _u ⁺), N ₂ (B ³ Π _g), N ₂ (a ¹ Σ _u ⁺), N ₂ (C ³ Π _u), N ₂ (ν ₁)–N ₂ (ν ₈)
Positive ions	H ⁺ , H ₂ ⁺ , H ₃ ⁺ , N ⁺ , N ₂ ⁺ , N ₃ ⁺ , N ₄ ⁺ , NH ⁺ , NH ₂ ⁺ , NH ₃ ⁺ , NH ₄ ⁺ , N ₂ H ⁺
Negative species	e, H [−]

Table 5. Species surface coefficients considered for the N₂–H₂ test chemistry sets: sticking coefficients *s*, return species *R*, and return coefficients *r*. M⁺ denotes a general positive ion, with M being its neutral counterpart. The species not included in the table all have sticking coefficient *s* = 0.

Species	<i>s</i>	<i>R</i>	<i>r</i>	Source
	^a 0.0			
H	^b 4.0 × 10 ^{−2}	H ₂ NH ₃	0.5 0.5	^c ^c
H ₂	2.0 × 10 ^{−4}	H ₂ (ν ₁) NH ₃	0.5 0.5	[76] ^d
H ₂ (ν _{<i>i</i>})	2.0 × 10 ^{−4}	H ₂ (ν _{<i>i</i>−1}) H ₂ (ν _{<i>i</i>+1})	0.5 0.5	[76] [76]
H ₂ (ν ₃)	1.0 × 10 ^{−4}	H ₂ (ν ₂)	1.0	[76]
H ₂ (b), H ₂ (B), H ₂ (c), H ₂ (a)	1.0 × 10 ^{−3}	H ₂	1.0	[4]
N(D), N(P)	1.0	N	1.0	[63]
N ₂	4.5 × 10 ^{−4}	N ₂ (ν ₁) N ₂ (ν _{<i>i</i>−1})	1.0 0.5	[77] [77]
N ₂ (ν _{<i>i</i>})	9.0 × 10 ^{−4}	N ₂ (ν _{<i>i</i>+1})	0.5	[77]
N ₂ (ν ₈)	4.5 × 10 ^{−4}	N ₂ (ν ₇)	1.0	[77]
N ₂ (A)	1.0	N ₂	1.0	[63]
N ₂ (a')	1.0 × 10 ^{−3}	N ₂	1.0	[63]
H ₃ ⁺	1.0	H ₂ H N ₂	0.5 0.5 0.5	
N ₃ ⁺	1.0	N N ₂	0.5 0.5	
N ₄ ⁺	1.0	N ₂ NH ₃	0.5 0.5	
NH ₄ ⁺	1.0	H N ₂	0.5 0.5	
N ₂ H ⁺	1.0	H	0.5	
M ⁺	1.0	M	1.0	

^aOnly used in the atmospheric-pressure case S2.^bOnly used in the low-pressure case S3.^cFitted to match the modeling results reported by Carrasco *et al* [75].^dFitted to match the modeling results reported by Hong *et al* [4].

processes were adopted from the QDB database. The threshold electron energy for each of the processes was taken to be the electron energy of the first non-zero data-point of the corresponding cross-section.

- For electron processes described in the Arrhenius form, the threshold energy was adopted from the corresponding processes cross-sections found in QDB, as described in the previous point.
- All the heavy species processes described in the Arrhenius form were left unchanged.
- The heavy species processes following any other functional dependence were all explicitly expressed for a constant gas temperature T_g .

The surface coefficients (sticking and return coefficients *s* and *r* with return species *R*) are listed in table 5 for both atmospheric-pressure and low-pressure test chemistry sets. The surface coefficients for quenching of the excited states and for *V*–*V* surface transitions are taken from Hong *et al* (with the original sources listed in table 5). Additionally, all the ions are neutralized on surfaces, while ions with their neutral counterparts not present in the system are returned as their most abundant neutral fragments. Finally, a separate category is formed by the species H and H₂. The NH₃ production is reportedly dominated by surface processes both for atmospheric-pressure conditions [4] and for low-pressure conditions [75]. The main channels for NH₃ surface production are reported to be the

Table 6. Species included in the test chemistry set for CF₄–CHF₃–H₂–Cl₂–O₂–HBr plasma, taken from the QDB database (chemistry C27). The * symbol denotes the lowest-energy electronically excited states.

Neutrals	H, H ₂ , C, O, O ₂ , F, F ₂ , Cl, Cl ₂ , Br, CH, CH ₂ , OH, H ₂ O, HF, HCl, HBr, CO, CO ₂ , CF, CF ₂ , CF ₃ , CF ₄ , FO, ClO, CHF, CHF ₂ , CHF ₃ , COF, COF ₂
Excited states	O*, F*, Cl*, Br*
Positive ions	H ⁺ , H ₂ ⁺ , H ₃ ⁺ , C ⁺ , C ₂ ⁺ , O ⁺ , O ₂ ⁺ , F ⁺ , F ₂ ⁺ , Cl ⁺ , Cl ₂ ⁺ , Br ⁺ , CH ⁺ , CH ₂ ⁺ , C ₃ H ₅ ⁺ , C ₃ H ₈ ⁺ , H ₂ O ⁺ , HCl ⁺ , HBr ⁺ , CO ⁺ , CO ₂ ⁺ , CF ⁺ , CF ₂ ⁺ , CF ₃ ⁺ , Cl, O ⁺ , CHF ⁺ , CHF ₂ ⁺
Negative species	e, O ⁻ , F ⁻ , Cl ⁻ , Br ⁻ , CF ₃ ⁻

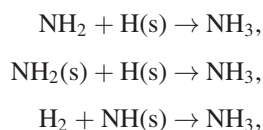
Table 7. Species included in the test chemistry set for CH₄–N₂ plasma, taken from the QDB database (chemistry C28). The * symbol denotes the lowest-energy electronically excited states.

Neutrals	H, H ₂ , C, C ₂ , N, N ₂ , CH, CH ₂ , CH ₃ , CH ₄ , CN, C ₂ H, C ₂ H ₂ , C ₂ H ₃ , C ₂ H ₄ , C ₂ H ₅ , C ₂ H ₆ , C ₃ H ₅ , C ₃ H ₆ , C ₃ H ₇ , C ₃ H ₈ , C ₄ H ₂ , NH, NH ₂ , NH ₃ , N ₂ H, N ₂ H ₂ , N ₂ H ₃ , N ₂ H ₄ , HCN, H ₂ CN, M
Excited states	H*, H ₂ ⁺ , N*, N ₂ ⁺
Positive ions	H ⁺ , H ₂ ⁺ , H ₃ ⁺ , C ⁺ , C ₂ ⁺ , N ⁺ , N ₂ ⁺ , N ₃ ⁺ , N ₄ ⁺ , CH ⁺ , CH ₂ ⁺ , CH ₃ ⁺ , CH ₄ ⁺ , CH ₅ ⁺ , C ₂ H ⁺ , C ₂ H ₂ ⁺ , C ₂ H ₃ ⁺ , C ₂ H ₄ ⁺ , C ₂ H ₅ ⁺ , C ₂ H ₆ ⁺ , NH ⁺ , NH ₂ ⁺ , NH ₃ ⁺ , NH ₄ ⁺
Negative species	e, H ⁻

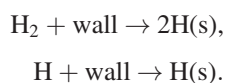
Table 8. Species included in the test chemistry set for Ar–NF₃–O₂ plasma, taken from the QDB database (chemistry C33). The * and ** symbols denote the electronically excited states with the lowest energies.

Neutrals	Ar, O, O ₂ , O ₃ , N, N ₂ , F, F ₂ , NO, N ₂ O, NO ₂ , NF, NF ₂ , NF ₃ , FO, FNO
Excited states	Ar*, Ar**, O*, O ₂ [*] , N*, N ₂ [*]
Positive ions	Ar ⁺ , O ⁺ , O ₂ ⁺ , N ⁺ , N ₂ ⁺ , F ⁺ , F ₂ ⁺ , NO ⁺ , N ₂ O ⁺ , NF ⁺ , NF ₂ ⁺ , NF ₃ ⁺
Negative species	e, O ⁻ , O ₂ ⁻ , O ₃ ⁻ , F ⁻

heterogeneous reactions



with both the surface-adsorbed NH₂(s) and NH(s) species being predominantly controlled by the surface-adsorbed atomic hydrogen H(s), which is in turn produced mainly by adsorption of H and H₂ [4]



In our work, for the sake of brevity, these NH₃ surface production channels were represented by sticking and return coefficients for H and H₂ only, as shown in table 5, with values calibrated to match the modeling results reported by Hong *et al* [4] for atmospheric pressure case and by Carrasco *et al* [75] for the low-pressure conditions, for the plasma conditions and parameters stated in both sources. Surface sticking of the H species was discounted in the case of atmospheric pressure test chemistry set, due to a much lower degree of hydrogen dissociation.

Finally, it needs to be noted that the surface kinetics described by table 5 are highly approximative and should not

replace the full set of heterogeneous reactions (e.g. [4]) in any work aiming for quantitative modeling results. However, in our work, it will suffice to provide important test chemistry sets, with some of the species (mainly NH_x) densities being dominantly controlled by surface diffusion and conversion, rather than volumetric processes.

4.3. Other test chemistry sets (S4–S6)

Three additional test chemistry sets are used in present work to test the reduction technique: sets for plasma in CF₄–CHF₃–H₂–Cl₂–O₂–HBr, CH₄–N₂, and finally Ar–NF₃–O₂. These chemistry sets were chosen with no regard to any specific target application and they were adopted directly from the pre-compiled plasma chemistries given by the QDB database [8]; their consistency and validity was not verified in the present work. They can be regarded as mere *generators* of non-linear outputs needed for analysis of the species ranking methods defined in section 3. Tables 6–8 summarise all the species included in these chemistry sets. The full lists of reactions and their kinetic data are not reproduced here, but can be found in the QDB database. In the instances of reactions described by cross-sectional kinetics, the data from QDB were fitted to the Arrhenius form on a grid on Maxwellian temperatures. Default surface coefficients were used for all the QDB test chemistry sets, as defined in table 3.

Table 9. Summary of all the 36 test reduction cases addressed in the present work. Six sets of species of interest were considered for each test chemistry set.

Case ID	Test set	Species of interest
C1.1	O ₂ -He	e, O ⁻ , O ₃ ⁻
C1.2		O ₂ ⁺ , O ₄ ⁺
C1.3		O ⁺ , He ⁺
C1.4		O, O ₂ (a ¹ Δ _g), O ₃
C1.5		O(¹ D), O(¹ S), O ₂ (b ¹ Σ _g ⁺)
C1.6		O ₃ (ν), O ₂ (ν), He*
C2.1	N ₂ -H ₂ (atm)	e, H ⁻
C2.2		NH ₄ ⁺
C2.3		NH ₃ ⁺ , H ₂ ⁺
C2.4		NH, NH ₂ , NH ₃
C2.5		N ₂ (ν ₁), H ₂ (ν ₁)
C2.6		N ₂ (A ³ Σ _u ⁺), H ₂ (B ¹ Σ _u ⁺)
C3.1	N ₂ -H ₂ (low)	e, H ⁻
C3.2		NH ₄ ⁺
C3.3		NH ₃ ⁺ , H ₂ ⁺
C3.4		NH, NH ₂ , NH ₃
C3.5		N ₂ (ν ₁), H ₂ (ν ₁)
C3.6		N ₂ (A ³ Σ _u ⁺), H ₂ (B ¹ Σ _u ⁺)
C4.1	CF ₄ -...-HBr	e, O ⁻
C4.2		Cl ⁺ , Br ⁺
C4.3		CF ⁺ , CO ⁺
C4.4		OH, CH
C4.5		F, Cl
C4.6		O*, F*
C5.1	CH ₄ -N ₂	e, H ⁻
C5.2		NH ₄ ⁺ , CH ₃ ⁺
C5.3		N ₂ ⁺ , H ₂ ⁺
C5.4		C ₂ H ₃ , C ₂ H ₆
C5.5		CH ₂ , CH ₃
C5.6		N ₂ [*] , H ₂ [*]
C6.1	Ar-NF ₃ -O ₂	e, F ⁻ , O ⁻
C6.2		O ₂ ⁺ , NO ⁺ , O ⁺
C6.3		NF ⁺ , NF ₂ ⁺ , NF ₃ ⁺
C6.4		NF, NF ₂ , NF ₃
C6.5		O, F
C6.6		O*, N*, Ar*

4.4. Reduction cases

Six test reduction cases are considered for each test chemistry set, each with a different set of species of interest. This makes 36 reduction cases, each case specified by the pair

$$\text{Reduction case} = \{\text{chemistry set, important species}\}.$$

All the reduction cases considered in the present work are listed in table 9. For each reduction case, the species of interest consist of a set of selected negative species (always including an electron), two sets of selected positive species (with higher

and lower densities) and three sets of selected neutrals (including one consisting purely of selected excited states.) In some cases, the choice of the species interest is inspired by the chemistry set source publication (such as in the reduction cases C1.4 and C1.5, which feature species identified as important radicals and transient species respectively by Turner [21]), but in most cases, the choice was made purely arbitrarily.

The reduction conditions for each reduction case are summarised in table 10. The reduction conditions were kept the same for each reduction case belonging to any single chemistry set. For the first three chemistry sets, the reduction conditions are taken from Turner *et al* [21], Hong *et al* [4], and Carrasco *et al* [75] respectively. For the chemistry sets sourced from QDB (C4.1–C6.6), the reduction conditions were set arbitrarily, with the exception of the Ar–NF₃–O₂ chemistry set, which adopts the conditions presented in the validation notes in QDB.

Table 11 lists the times of interest used to build the species rankings from the test chemistry sets and to evaluate the reduction error within the iterative reduction algorithm. The same times of interest were used for all reduction cases belonging to any single chemistry set. In the case of the pulsed-power reduction cases with the O₂–He chemistry set (C1.1–C1.6), the times of interest are distributed linearly, covering the whole power cycle. In the remaining cases of constant-power reduction conditions, the times of interest are distributed logarithmically, to capture not only the steady-state conditions but also the evolution from the initial conditions. The simulation times t_{end} (also shown in table 11) were chosen to be sufficiently long for each reduction case to reach a steady state.

5. Results

Altogether, 14 different species ranking methods were introduced in section 3.4, based on path searches in graphs representing chemistry. These ranking methods are built around 4 separate ways to calculate direct interaction coefficients w (13a)–(13d), together with 4 separate methods of devising coupling coefficients W between any two species, based on different graph path search algorithms, introduced in section 3.3; (13b) and (13d) yield direct interaction coefficients which are not bound between 0 and 1, excluding the chemistry graphs from the use of the maximal propagated error search method.

All the species ranking schemes presented were tested with the species-oriented iterative skeletal reduction method introduced in section 2 on an array of 36 test reduction cases listed in table 9 and with further plasma conditions and model parameters listed in tables 10 and 11. The same maximal reduction error $\Delta = 10\%$ was considered in the reduction method with all the test reduction cases. The species ranking scheme yielding the best results with the ranking-based iterative reduction method was identified statistically.

5.1. Species ranking performance metrics

For a direct comparison of the competing species ranking schemes with regards to the ranking-based iterative reduction method, the number of species eliminated from any given test chemistry set provides a good metric for how well the ranking

Table 10. Summary of the reduction conditions for all the chemistry sets considered: pressure p , gas temperature T_g , absorbed power P , the reactor dimensions r and z , the total feedstock gas flow Q and the ratio of the feedstock gases.

Test set	p (Pa)	T_g (K)	P (W)	r (mm)	z (mm)	Q (sccm)	Ratio
O ₂ –He	10 ⁵	305	0.3	0.564	30	300	1:250
N ₂ –H ₂ (atm)	10 ⁵	400	35	25	25	60	1:2
N ₂ –H ₂ (low)	8	300	10	50	340	15	1:9
CF ₄ –...–HBr	25	300	500	100	100	600	1:1:1:1:1:1
CH ₄ –N ₂	10 ⁵	300	5000	20	100	110	1:10
Ar–NF ₃ –O ₂	100	1500	400	40	200	1150	1:2:20

Table 11. Simulation time t_{end} , power pulse duration Δt_{pulse} , and the sets of the times of interest for each one chemistry set considered. Power pulsing was only done for the O₂–He chemistry set, with the pulsing period of 6 ms; the remaining sets were modeled using constant power.

Test set	Δt_{pulse} (ms)	t_{end} (s)	Times of interest
O ₂ –He	3	1.2×10^{-2}	$6.3 \times 10^{-3}, 6.9 \times 10^{-3}, \dots, 11.7 \times 10^{-3}$
N ₂ –H ₂ (atm)	—	100	$10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$
N ₂ –H ₂ (low)	—	10	$10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1$
CF ₄ –...–HBr	—	100	$10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$
CH ₄ –N ₂	—	100	$10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$
Ar–NF ₃ –O ₂	—	0.2	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$

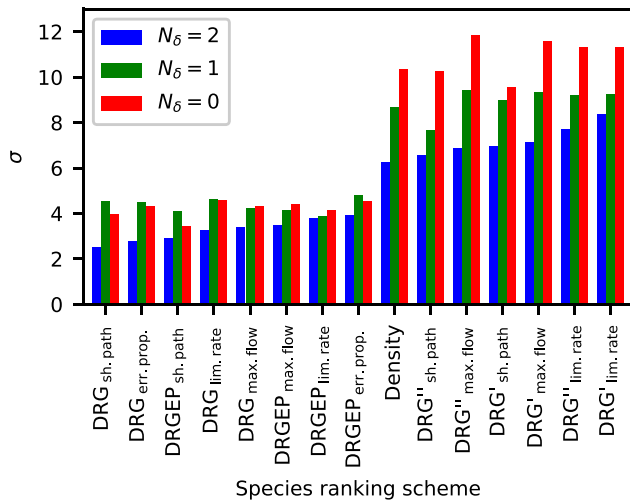


Figure 8. Comparison of the σ values (25) for all of the species ranking schemes considered in the present work and for three choices of $N_\delta \in \{0, 1, 2\}$. The ranking schemes are ordered with increasing σ for $N_\delta = 2$. The bars are color-coded by different values of N_δ used for the ranking-based iterative reduction method. The ranking schemes are labeled with the different methods of calculating the direct interaction coefficients w and the coupling coefficients W ; e.g. the method labeled DRG_{sh.path} uses w_{DRG} (13a) and $W_{\text{AB}}^{\text{sh.path}}$ (16) respectively. The *density* species ranking scheme simply ranks the species according to their density as modeled by the plasma model employed with the detailed chemistry set.

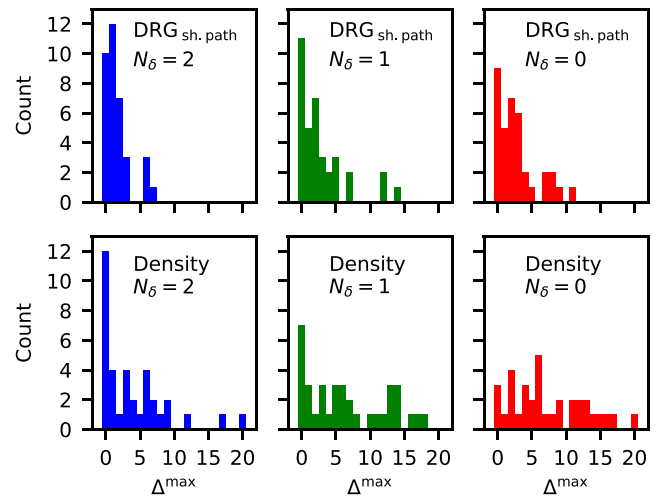


Figure 9. Distribution of Δ^{max} , i.e. the difference in the number of eliminated species between each ranking scheme and the best-performing ranking scheme (24), over all the reduction cases. In the first row, histograms for the DRG_{sh.path} ranking scheme are shown, in comparison to the trivial benchmark of ranking species by their densities, shown in the second row. Three columns of plots correspond to three different choices of $N_\delta = 2, 1, 0$, with the same color-coding as in figure 8.

in the k th reduction case, as

$$N_{\epsilon,k}^{\text{max}} = \max_j N_{\epsilon 3,k}^j. \quad (23)$$

For any j th ranking scheme, the value of

$$\Delta_k^{\text{max},j} = N_{\epsilon,k}^{\text{max}} - N_{\epsilon,k}^j \quad (24)$$

scheme employed works with the reduction method. We define the number of eliminated species by the j th ranking scheme in the k th reduction case as $N_{\epsilon,k}^j$. This is followed with the number of eliminated species by the best-performing ranking scheme

Table 12. Chemistry set sizes (the number of species and reactions) of sets reduced using the DRG_{sh.path} species ranking scheme for all the reduction cases from table 9, compared to the sizes of the detailed chemistry sets. For reference, the last two columns list the *best ranking schemes*, which yielded the most significant reduction in the number of species together with their species reduction magnitude (if different from the DRG_{sh.path} scheme). The reduction cases, where the DRG_{sh.path} ranking scheme was among the best-performing ones, are listed in **bold**.

Case	DRG _{sh.path} species	DRG _{sh.path} reactions	Best ranking schemes	Best sch. species
C1.1	15/25	167/373	DRG' _{max.flow, sh.path} , DRG'' _{all} , DRGEP _{max.flow, lim.rate} , DRG _{max.flow}	14/25
C1.2	15/25	163/373	DRG' _{all} , DRG'' _{all} , DRGEP _{max.flow, lim.rate} , DRG _{max.flow}	14/25
C1.3	18/25	247/373	DRG' _{all} , DRGEP _{all} , DRG _{max.flow}	17/25
C1.4	13/25	118/373	DRGEP _{lim.rate, err.prop., sh.path}	12/25
C1.5	14/25	133/373	DRG' _{all} , DRG'' _{lim.rate, sh.path} , DRGEP _{all} , DRG _{max.flow, sh.path}	
C1.6	13/25	117/373	DRGEP _{lim.rate, err.prop., sh.path} , DRG _{max.flow, sh.path}	
C2.1	26/42	236/408	DRGEP _{max.flow, lim.rate} , DRG _{max.flow, lim.rate, err.prop}	25/42
C2.2	21/42	118/408	DRG _{max.flow}	20/42
C2.3	26/42	193/408	DRGEP _{max.flow}	24/42
C2.4	26/42	236/408	DRGEP _{max.flow} , DRG _{max.flow, err.prop}	24/42
C2.5	10/42	73/408	DRG' _{max.flow, sh.path} , DRG'' _{max.flow, sh.path} , DRG _{err.prop., sh.path}	
C2.6	26/42	210/408	DRG _{err.prop}	25/42
C3.1	23/42	101/408	DRG _{max.flow}	22/42
C3.2	25/42	198/408	DRG _{lim.rate}	19/42
C3.3	36/42	377/408	DRG' _{max.flow, sh.path}	33/42
C3.4	19/42	91/408	DRGEP _{lim.rate, err.prop., sh.path} , DRG _{err.prop., sh.path}	
C3.5	19/42	131/408	DRG' _{sh.path} , DRG _{err.prop.}	18/42
C3.6	28/42	214/408	DRGEP _{all} , DRG _{err.prop}	27/42
C4.1	51/67	300/563	DRG' _{max.flow} , DRG'' _{max.flow}	50/67
C4.2	43/67	165/563	DRGEP _{all} , DRG _{max.flow, err.prop., sh.path}	
C4.3	61/67	475/563	DRG'' _{all}	55/67
C4.4	40/67	147/563	DRGEP _{sh.path}	34/67
C4.5	44/67	187/563	DRG _{err.prop}	37/67
C4.6	49/67	246/563	DRGEP _{err.prop}	47/67
C5.1	36/61	255/521	DRGEP _{all} , DRG _{all}	
C5.2	48/61	380/521	DRGEP _{max.flow, lim.rate, sh.path}	46/61
C5.3	27/61	140/521	DRG'' _{lim.rate}	25/61
C5.4	24/61	168/521	DRG'' _{sh.path} , DRGEP _{max.flow, sh.path} , DRG _{all}	
C5.5	21/61	121/521	DRGEP _{max.flow, sh.path} , DRG _{all}	
C5.6	26/61	165/521	DRG' _{max.flow} , DRGEP _{lim.rate}	24/61
C6.1	20/39	79/310	DRGEP _{lim.rate, sh.path} , DRG _{lim.rate, err.prop., sh.path}	
C6.2	30/39	176/310	DRG _{err.prop}	27/39
C6.3	24/39	116/310	DRG' _{lim.rate, sh.path} , DRG'' _{all}	21/39
C6.4	15/39	43/310	DRG' _{max.flow} , DRG'' _{all} , DRGEP _{all} , DRG _{all}	
C6.5	14/39	32/310	DRG _{max.flow, err.prop}	13/39
C6.6	26/39	138/310	DRG' _{max.flow} , DRG'' _{max.flow}	25/39

will determine how well it performs against the best ranking scheme in the *k*th reduction case. Any well-performing ranking scheme should have consistently low values of $\Delta_k^{\max,j}$ for any reduction case. To assess the performance of the *j*th ranking scheme on the whole array of test reduction cases (table 9), the *deviation from the best* parameter σ^j can be evaluated as

$$\sigma^j = \left[\frac{1}{N} \sum_k (\Delta_k^{\max,j})^2 \right]^{\frac{1}{2}}, \quad (25)$$

where $N = 36$ is the number of reduction cases considered. The σ value is a good measure of a *performance* of a single method in a pool of other species ranking methods, since it quantifies how many more species were eliminated on average by the reduction using *the best* ranking method for every

reduction case. A lower value of σ translates to a better species ranking scheme for species reduction, while the lower bound for σ is 0, for a hypothetical *j*th ranking scheme, which would result in an elimination of the highest number of species (compared to all the other competing ranking schemes), for every single reduction case, or

$$\sigma^j = 0 \Leftrightarrow \forall k : (N_{\epsilon,k}^j = N_{\epsilon,k}^{\max} \text{ and } \Delta_k^{\max,j} = 0).$$

In reality, there is a distribution of $\Delta^{\max,j}$ across the different reduction cases.

Of course both $\Delta_k^{\max,j}$ and σ^j will be influenced not only by how well the *j*th ranking scheme performs in the reduction method, but also by the chosen value of N_δ , or the maximal allowed streak of unsuccessful species eliminations, as the algorithm termination condition. The results were studied for

Table 13. Retained and eliminated species for the reduced O_2 –He chemistry set and the C1.4 reduction case (with species of interest O, $O_2(a^1\Delta_g)$ and O_3). The reduction was performed with the $DRG_{sh,path}$ species ranking scheme, $\Delta = 10\%$, $N_\delta = 2$, and with the reduction conditions from tables 10 and 11.

	Retained species	Eliminated species
Neutrals	He, O, O_2 , O_3 He*, $O(^1D)$	He*, $O(^1S)$
Excited	$O_2(a^1\Delta_g)$, $O_2(b^1\Sigma_g^+)$, $O_3(\nu)$	$O_2(a^1\Delta_g, \nu)$, $O_2(b^1\Sigma_g^+, \nu)$, $O_2(\nu)$
Positive	He ⁺ , O_2^+	He ⁺ , O^+ , O_3^+ , O_4^+
Negative	e, O^-	O_2^- , O_3^- , O_4^-

different values of N_δ , however, it was found that $N_\delta = 2$ offers a good balance between the number of eliminated species and the number of plasma model calls, for the majority of the ranking methods.

5.2. Comparison of ranking schemes

A comparison of the σ value for all of the species ranking schemes considered in this work is shown in figure 8 for three different choices of N_δ . As a benchmark, an additional ranking scheme labeled *density* is included in the results, where the species ranking scores trivially equal to their maximal densities (over all the time samples) solved with the detailed chemistry set, and as such are blind to the choice of the species of interest.

The species ranking scheme using the DRG definition of direct interaction coefficients w_{DRG} (13a) and the shortest path approach for the calculation of the coupling coefficients $W_{AB}^{sh,path}$ (16) can be identified as providing species rankings which perform consistently well with the ranking-based iterative reduction method for $N_\delta = 2$. Regarding constructing the chemistry graph, the DRG (13a) and DRGEP (13c) methods for direct interaction coefficients give significantly better results than DRG' (13b) and DRG'' (13d) methods, while the *shortest path* method for species coupling coefficients $W_{AB}^{sh,path}$ appears to yield marginally better results than the rest.

Finally, the histograms in figure 9 show the distribution of Δ^{max} (24), over all the reduction cases, and for 3 different choices of N_δ . Histograms for the *optimal* $DRG_{sh,path}$ species ranking scheme are plotted, together with the density ranking scheme benchmark. The distribution for the $DRG_{sh,path}$ species ranking scheme is centered reasonably tightly around the optimal value of $\Delta^{max} = 0$, however, there are some instances of reduction cases, where this ranking scheme performs significantly worse than the best-performing ranking scheme. By contrast, the ranking scheme based simply on species density values shows distribution with a significantly wider spread.

5.3. Reduced test chemistry sets

To give an idea of typical sizes of reduced chemistry sets as compared with the detailed ones, table 12 shows the number of species and reactions in sets reduced using the $DRG_{sh,path}$ ranking scheme for each reduction case from table 9. The table also shows for each reduction case all the *best ranking schemes* resulting in the greatest reduction (in a number of species) and the number of species corresponding to those, if the $DRG_{sh,path}$

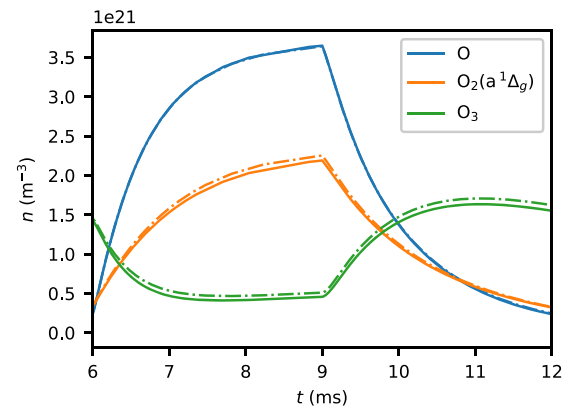


Figure 10. Global model solutions for densities of the species of interest O, $O_2(a^1\Delta_g)$ and O_3 . Comparison between the reduced chemistry set (reduction case C1.4) and the detailed set. The species ranking scheme used for the reduction was $DRG_{sh,path}$. The densities obtained with the detailed chemistry set are plotted with solid lines (—), while the reduced set results are plotted with dash-dotted lines (– · –).

ranking scheme is not among them. The reduction parameters were $\Delta = 10\%$ and $N_\delta = 2$ for each of the reduction cases, and the reduction conditions were as described in tables 10 and 11.

In some cases, such as reduction cases C2.5 or C6.5, the reduced chemistry set was significantly smaller in size than the detailed set. In other cases, such as C4.3, the reduction was far less significant. The difference in size between the detailed chemistry sets and any reduced sets depends on many factors, such as reduction conditions, species of interest, or how well refined the detailed chemistry set is in the first place.

5.4. The C1.4 O_2 –He reduction case

Since it is impractical to analyze every reduction case in this work, a single reduction case, C1.4 O_2 –He chemistry set with O, $O_2(a^1\Delta_g)$ and O_3 as species of interest, was chosen as an example and is considered here in greater detail. Table 13 lists all the species retained in the reduced O_2 –He chemistry set, as well as all the species eliminated, after running the reduction algorithm with the $DRG_{sh,path}$ species ranking scheme for the reduction case C1.4. Figure 10 shows the PyGMol global model solutions for species of interest densities belonging to the same reduced chemistry set. The density evolution is compared to the model with the detailed chemistry

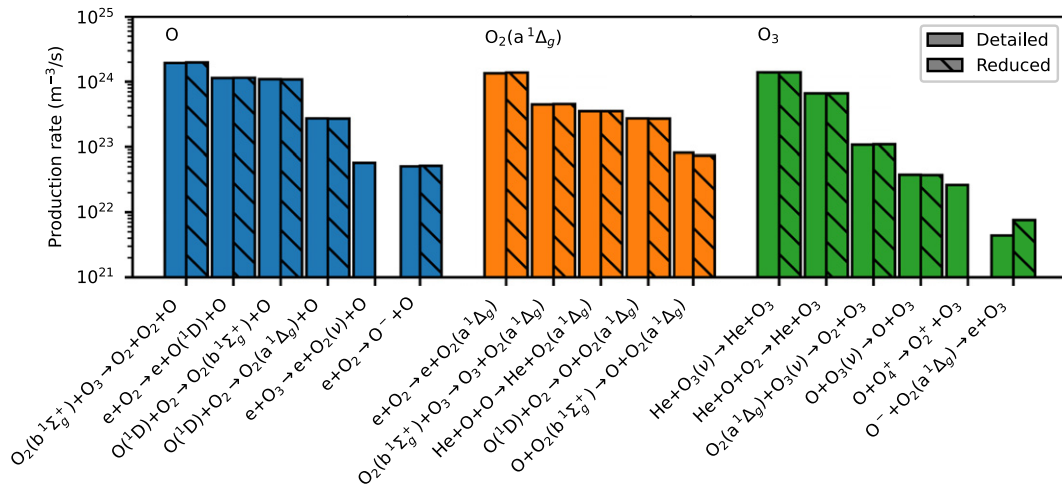


Figure 11. Production rates of the species of interest by the most dominant processes for both detailed and reduced chemistry sets (reduction case C1.4). Values sampled 0.1 ms before the end of the power pulse.

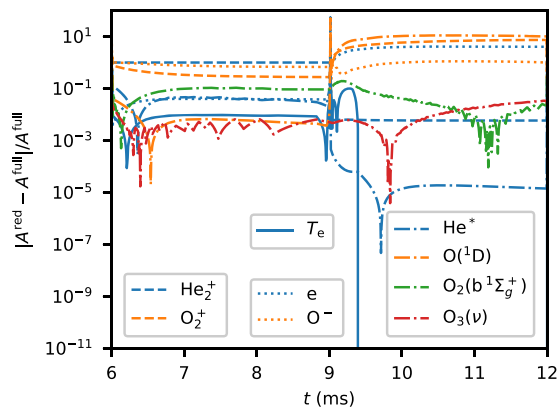
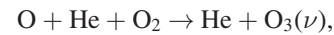


Figure 12. Relative differences of global model outputs between the reduced chemistry set (C1.4) and the detailed set. Outputs plotted are electron temperature and densities of the other species than the selected species of interest. The positive ions are plotted with dashed lines (---), negative species with dotted lines (·····), the excited states with dash-dotted lines (— · —), and the electron temperature is plotted with a solid line (—).

set. It is evident, that the reduced chemistry set indeed preserves the densities of species of interest from the detailed set very well inside the allowed error Δ . Some other facts can be noted about the eliminated species listed in table 13. The He^+ ion has been eliminated by the reduction algorithm despite being the most abundant ion of one of the feed gases. This is due to a relatively large ionization potential of He atoms and consequently the very low He ionization degree ($\approx 10^{-13}$ at the end of a power pulse, as modeled by PyGMol with the detailed chemistry set). The He_2^+ species, on the other hand, has *not* been eliminated by the reduction method, despite being even less abundant than the He^+ species. This is simply a shortcoming of the species ranking scheme employed (DRG_{sh,path}), where He_2^+ is ranked above He^+ , and more importantly, above all of O^- , He^* and O_2^+ , none of which can be eliminated from the chemistry set with an acceptable reduction error δ . He_2^+ can indeed be removed from the reduced set without changing the reduction error significantly and would have

been, if $N_\delta \geq 3$, or if one of the DRGEP species ranking schemes was employed (see table 12). However, for the reduction parameters presented with $N_\delta = 2$ and the DRG_{sh,path} ranking scheme, the reduction algorithm was terminated just before the He_2^+ iteration. Also to be noted, all of the vibrationally excited species were eliminated except the $\text{O}_3(\nu)$ species. This is because $\text{O}_3(\nu)$ is a part of the dominant channel for the consumption of O via the reaction



which could not be removed within an acceptable reduction error.

Figure 11 shows the production rates of each species of interest via the 5 most prominent processes, for both detailed and reduced chemistry sets described in table 13, sampled 0.1 ms before the end of power pulse. It can be seen, that while some of the channels are lost due to the eliminated species, in this particular case $\text{O}_2(\nu)$ and O_4^+ , these are generally much less significant channels, with rates about 2 orders of magnitude or more lower than the most prominent channel. This is also true for consumption channels and for other sample times, including ones sampled for the afterglow phase in the second half of the power cycle.

Finally, figure 12 shows the relative differences of some of the global model outputs between the reduced chemistry set (C1.4) and the detailed set. The outputs plotted are the densities of the species other than pre-selected species of interest, as well as the electron temperature. It can be seen that the density error induced by the reduction for some species, such as $\text{O}(^1\text{D})$ and all the charged species e , O^- , O_2^+ , He_2^+ , is fairly significant. This is unsurprising, since both the species ranking coming into the reduction method, and the reduction error (1) were completely blind to all the global model outputs except the densities of the species of interest. Other outputs, such as densities of He^* , $\text{O}_2(\text{b}^1\Sigma_g^+)$, $\text{O}_3(\nu)$, and the electron temperature T_e are well replicated with the reduced chemistry set. This is, however, not guaranteed by the reduction method presented, but rather can be considered a side effect of preserving the

dominant production and consumption channels of the species of interest (figure 11).

6. Conclusions and outlook

A method of ranking the species in a chemistry set according to their importance for modeling densities of a pre-defined set of species of interest is a crucial component of the presented species-oriented method for skeletal reduction of chemistry sets. In this work, we present several competing fast algorithms for species ranking, based on a graph-theoretical representation of chemistry sets. In all of the species-ranking methods presented, weights of the edges between species in the chemistry graphs (or the direct interaction coefficients) were distributed according to the different flavors of the DRG theory, with modifications to include effects of surface interactions, and several species rankings were proposed, built around different well-established search algorithms from the graph theory. We identify the DRG_{sh,path} species ranking method, which uses the definition of direct interaction coefficients from the original DRG theory [36], with Dijkstra's search for the shortest path [58] in the chemistry graph, as the most suited for species-oriented skeletal reduction of chemistry sets, statistically on an array of diverse test chemistry sets and plasma conditions. The DRG_{sh,path} species ranking method led to reductions of between 10 and 75% in the number of species compared to the original detailed chemistry sets, depending on the specific test set and plasma conditions.

Our primary motivation for the development of a fast species-ranking algorithm is its use in the species-oriented iterative skeletal reduction method. While we demonstrate its successful utilization for reducing the number of species in several test chemistry sets, the resulting reduced sets likely still retain some redundant reactions. It could be desirable in the future to devise a reaction-ranking scheme which would assess the importance of the reactions retained for modeling the set of species of interest, and thus to expand our chemistry set reduction framework with a reaction-oriented skeletal reduction method.

Additionally, regarding the species-oriented iterative skeletal reduction method presented in this work, the validity of each chemistry set reduced using this method is strictly ensured only for the set of plasma model parameters that are the user-specified reduction conditions. It is often the case in a modeling application, that the chemistry set needs to be valid over a *range* of input parameters. Although this might be achieved by running the reduction method on a set of samples from the parameter space, as proposed in section 2.1, this kind of parameter sweep might be unnecessarily inefficient. Therefore, modifying the species-ranking method along with the reduction error function δ , so they reflect a range of reduction conditions rather than a single set, could be an important area of the future research.

Finally, a significant shortcoming of the species-ranking method presented is that the ranking only takes into account the species' importance for modeling *densities* of a set of pre-selected species of interest. The species ranking is blind to any coupling between the presence of the species in the

chemistry set and any other possible plasma model outputs, such as electron density and electron temperature. More sophisticated plasma models can also resolve some additional outputs, such as the gas temperature or particle fluxes at surface, and certain applications might depend on the reduced chemistry set preserving those outputs as well. As an aim for the future research, nodes representing other possible plasma model outputs should be inserted into the chemistry graph with appropriate weights distributed among edges connecting them to the other graph nodes. This will allow for the species ranking to reflect species coupling to the said additional model outputs and the notion of *species of interest* will be expanded to a more general notion of *outputs of interest*.

Acknowledgments

Martin Hanicinec would like to thank EPSRC for a CASE studentship under Grant EP/N509577/1.

ORCID iDs

Martin Hanicinec  <https://orcid.org/0000-0003-3415-2996>

Jonathan Tennyson  <https://orcid.org/0000-0002-4994-5238>

References

- [1] d'Agostino R, Favia P, Oehr C and Wertheimer M R 2005 *Plasma Process. Polym.* **2** 7–15
- [2] Berthelot A and Bogaerts A 2016 *Plasma Sources Sci. Technol.* **25** 045022
- [3] Guerra V, Tejero-del-Caz A, Pintassilgo C D and Alves L L 2019 *Plasma Sources Sci. Technol.* **28** 073001
- [4] Hong J, Pancheshnyi S, Tam E, Lowke J J, Prawer S and Murphy A B 2017 *J. Phys. D: Appl. Phys.* **50** 154005
- [5] Van Gaens W and Bogaerts A 2013 *J. Phys. D: Appl. Phys.* **46** 275201
- [6] Sieck L W, Heron J T and Green D S 2000 *Plasma Chem. Plasma Process.* **20** 235–58
- [7] Herron J T and Green D S 2001 *Plasma Chem. Plasma Process.* **21** 459–81
- [8] Tennyson J *et al* 2017 *Plasma Sources Sci. Technol.* **26** 055014
- [9] Koelman P, Heijkers S, Tadayan Mousavi S, Graef W, Mihailova D, Kozak T, Bogaerts A and van Dijk J 2017 *Plasma Process. Polym.* **14** 1600155
- [10] Nagy T and Turanyi T 2009 *Combust. Flame* **156** 417–28
- [11] Tomlin A S, Pilling M J, Turányi T, Merkin J H and Brindley J 1992 *Combust. Flame* **91** 107–30
- [12] Bak M S and Cappelli M A 2015 *IEEE Trans. Plasma Sci.* **43** 995–1001
- [13] Kossyi I A, Kostinsky A Y, Matveyev A A and Silakov V P 1992 *Plasma Sources Sci. Technol.* **1** 207–20
- [14] Gorban A 2018 *Curr. Opin. Chem. Eng.* **21** 48–59
- [15] Tomlin A S, Turányi T and Pilling M J 1997 Mathematical tools for the construction, investigation and reduction of combustion mechanisms *Comprehensive Chemical Kinetics* vol 35 (Amsterdam: Elsevier) ch 4 pp 293–437
- [16] Okino M S and Mavrouniotis M L 1998 *Chem. Rev.* **98** 391–408
- [17] Law C K 2007 *Proc. Combust. Inst.* **31** 1–29
- [18] Lu T and Law C K 2009 *Prog. Energy Combust. Sci.* **35** 192–215

- [19] Turányi T and Tomlin A S 2014 *Analysis of Kinetic Reaction Mechanisms* (Berlin: Springer)
- [20] Huang H, Fairweather M, Griffiths J F, Tomlin A S and Brad R B 2005 *Proc. Combust. Inst.* **30** 1309–16
- [21] Turner M M 2015 *Plasma Sources Sci. Technol.* **24** 035027
- [22] Zheng X L, Lu T F and Law C K 2007 *Proc. Combust. Inst.* **31** 367–75
- [23] Maas U and Pope S B 1992 *Combust. Flame* **88** 239–64
- [24] Rehman T, Kemaneci E, Graef W and van Dijk J 2016 *J. Phys.: Conf. Ser.* **682** 012035
- [25] Lam S H and Goussis D A 1989 *Symp. Combust.* **22** 931–41
- [26] Lam S H and Goussis D A 1994 *Int. J. Chem. Kinet.* **26** 461–86
- [27] Lam S H 2013 *Combust. Flame* **160** 2707–11
- [28] Rabitz H, Kramer M and Dacol D 1983 *Annu. Rev. Phys. Chem.* **34** 419–61
- [29] Ayilaran A, Hanicinec M, Mohr S and Tennyson J 2019 *Plasma Sci. Technol.* **21** 064006
- [30] Wang H and Frenklach M 1991 *Combust. Flame* **87** 365–70
- [31] Oluwole O, Bhattacharjee B, Tolsma J, Barton P and Green W 2006 *Combust. Flame* **146** 348–65
- [32] Liu D X, Bruggeman P, Iza F, Rong M Z and Kong M G 2010 *Plasma Sources Sci. Technol.* **19** 025018
- [33] Whitehouse L E, Tomlin A S and Pilling M J 2004 *Atmos. Chem. Phys.* **4** 2025–56
- [34] Valorani M, Creta F, Goussis D, Lee J and Najm H 2006 *Combust. Flame* **146** 29–51
- [35] Bellemans A, Deak N and Bisetti F 2020 *Plasma Sources Sci. Technol.* **29** 025020
- [36] Lu T and Law C K 2005 *Proc. Combust. Inst.* **30** 1333–41
- [37] Lu T and Law C K 2006 *Combust. Flame* **144** 24–36
- [38] Lu T and Law C K 2006 *Combust. Flame* **146** 472–83
- [39] Pepiot-Desjardins P and Pitsch H 2008 *Combust. Flame* **154** 67–81
- [40] Sankaran R, Hawkes E R, Chen J H, Lu T and Law C K 2007 *Proc. Combust. Inst.* **31** 1291–8
- [41] Bellemans A, Kincaid N, Deak N, Pepiot P and Bisetti F 2020 *Proc. Combust. Inst.* **38** S1540748920304569
- [42] Sun S R, Wang H X and Bogaerts A 2020 *Plasma Sources Sci. Technol.* **29** 025012
- [43] Venot O, Bounaceur R, Dobrijevic M, Hébrard E, Cavalié T, Tremblin P, Drummond B and Charnay B 2019 *Astron. Astrophys.* **624** A58
- [44] Obrušník A, Bílek P, Hoder T, Šimek M and Bonaventura Z 2018 *Plasma Sources Sci. Technol.* **27** 085013
- [45] Lebedev A V, Okun M V, Chorkov V A, Tokar P M and Strelkova M 2013 *J. Math. Chem.* **51** 73–107
- [46] Pitchford L C et al 2017 *Plasma Process. Polym.* **14** 1600098
- [47] Celiberto R et al 2016 *Plasma Sources Sci. Technol.* **25** 033004
- [48] Wakelam V et al 2012 *Astrophys. J. Suppl. Ser.* **199** 21
- [49] Wakelam V et al 2015 *Astrophys. J. Suppl. Ser.* **217** 20
- [50] McElroy D, Walsh C, Markwick A J, Cordiner M A, Smith K and Millar T J 2013 *Astron. Astrophys.* **550** A36
- [51] Murakami I, Kato D, Kato M, Sakaue H A and Kato T 2007 *Fusion Sci. Technol.* **51** 138–40
- [52] Park J-H, Choi H, Chang W-S, Chung S Y, Kwon D-C, Song M-Y and Yoon J-S 2020 *Appl. Sci. Converg. Technol.* **29** 5–9
- [53] Dubernet M-L et al 2013 *Astron. Astrophys.* **553** A50
- [54] Lietz A M and Kushner M J 2016 *J. Phys. D: Appl. Phys.* **49** 425204
- [55] Kokkoris G, Panagiotopoulos A, Goodyear A, Cooke M and Gogolides E 2009 *J. Phys. D: Appl. Phys.* **42** 055209
- [56] Pancheshnyi S, Eismann B, Hagelaar G and Pitchford L 2008 ZDPlasKin: a new tool for plasmachemical simulations *Conf. on The Eleventh International Symposium on High Pressure, Low Temperature Plasma Chemistry (HAKONE XI)*, (Oleron Island, France)
- [57] Turner M M 2016 *Plasma Sources Sci. Technol.* **25** 015003
- [58] Dijkstra E W 1959 *Numer. Math.* **1** 269–71
- [59] Pollack M 1960 *Oper. Res.* **8** 733–6
- [60] Bellman R 1958 *Q. Appl. Math.* **16** 87–90
- [61] Edmonds J and Karp R M 1972 *J. ACM* **19** 248–64
- [62] Schulz-von der Gathen V, Buck V, Gans T, Knake N, Niemi K, Reuter S, Schaper L and Winter J 2007 *Contrib. Plasma Phys.* **47** 510–9
- [63] Gordiets B, Ferreira C M, Pinheiro M J and Ricard A 1998 *Plasma Sources Sci. Technol.* **7** 363–78
- [64] Gordiets B, Ferreira C M, Pinheiro M J and Ricard A 1998 *Plasma Sources Sci. Technol.* **7** 379–88
- [65] Murphy A B 2012 *Chem. Phys.* **398** 64–72
- [66] Colombo V, Ghedini E and Sanibondi P 2009 *J. Phys. D: Appl. Phys.* **42** 055213
- [67] Sackheim R L 2006 *J. Propul. Power* **22** 1310–32
- [68] Wang H-X, Geng J-Y, Chen X, Pan W X and Murphy A B 2010 *Plasma Chem. Plasma Process.* **30** 707–31
- [69] Ageorges H, Megy S, Chang K, Baronnet J-M, Williams J K and Chapman C 1993 *Plasma Chem. Plasma Process.* **13** 613–32
- [70] Ananthapadmanabhan P V, Taylor P R and Zhu W 1999 *J. Alloys Compd.* **287** 126–9
- [71] Chang Y, Young R M and Pfender E 1987 *Plasma Chem. Plasma Process.* **7** 299–316
- [72] Yamamoto H, Kuroda H, Ito M, Ohta T, Takeda K, Ishikawa K, Kondo H, Sekine M and Hori M 2012 *Japan. J. Appl. Phys.* **51** 016202
- [73] Nayebpashae N, Soltanieh M and Kheirandish S 2016 *Mater. Manuf. Process.* **31** 1192–200
- [74] Oberkofler M et al 2015 *Fusion Eng. Des.* **98–99** 1371–4
- [75] Carrasco E, Jiménez-Redondo M, Tanarro I and Herrero V J 2011 *Phys. Chem. Chem. Phys.* **13** 19561
- [76] Heidner R F and Kasper J V V 1972 *Chem. Phys. Lett.* **15** 179–84
- [77] Black G, Wise H, Schechter S and Sharpless R L 1974 *J. Chem. Phys.* **60** 3526–36