A novel atom tracking algorithm for the analysis of complex chemical kinetic networks

C. Laurent, C.J. Frewin, P. Pepiot*

Sibley School of Mechanical and Aerospace Engineering, Cornell University, NY 14853, USA

A B S T R A C T

A novel numerical algorithm is presented that enables the tracking of specific atoms during the simulation of simple combustion reactors as those atoms are transferred from initial reactants to final products. Tracking is performed by labeling individual atoms or groups of atoms of interest, and solving, in addition to the coupled ordinary differential equations describing the time evolution of species concentrations, appropriate transfer equations providing the concentration of tracked atoms at each possible location on all chemical species involved. The transfer equations for a given chemical kinetic mechanism are automatically generated using simple structural and energy-based arguments to characterize and quantify how, in each elementary reaction, atoms are re-organized as reactants are converted into products. The capabilities of the tracking algorithm are illustrated by analyzing soot precursors formation in a constant-volume homogeneous reactor, providing new, quantitative evidence of well-known behaviors, such as the link between the molecular structure of a fuel and its sooting propensity.

© 2016 The Combustion Institute. Published by Elsevier Inc. All rights reserved.

1. Introduction

Today, worldwide energy needs for transportation are overwhelmingly met with petroleum-based fuels [1]. One aspect of growing concern with this fact is the emission of pollutants, such as soot particulates, generated by the massive use of fossil fuels. Bio-derived fuels are expected to progressively displace conventional fuels over time in an effort to reduce our dependence on crude oil and to mitigate negative environmental impact [2]. Due to the wide variety of sources and production routes, bio-fuels are inherently diverse, with chemical functionalities potentially distinct from their fossil-derived counterparts [3], resulting in an explosion in the number of potential fuels and fuel blends available. This, in turn, provides new opportunities to optimize fuel blends for higher-efficiency, lower-emissions engines [4,5].

Transportation fuels, from fossil or renewable sources, often consist of a large number of fuel constituents [6,7], with each one potentially playing a very specific role during the combustion process. The observables, typically some global measure of engine and combustion performance, result from complex, multi-level physical and chemical interactions between those constituents. Numerous studies, both experimental and numerical, provide strong evidence that fuel molecular structures and fuel component interactions both impact combustion in non-trivial ways. For example, the sooting propensity of hydrocarbons fuels, which can be measured through Threshold [8] or Yield [9] Sooting Indices, has been shown to tightly correlate with the structural groups present in the fuel initially, such as rings, primary, secondary or tertiary carbons, and type of carbon–carbon bonds [10–14]. Adding oxygenated compounds to hydrocarbons, even in small amounts, has been shown to lead to a reduction in particulate emissions, but with varying extents depending on the type and location of the oxygen atoms in the additives [11,15,16].

Some common techniques used to investigate the link between fuel structure and combustion characteristics often rely on group contribution analysis [13,17] that relates output observations, such as sooting propensity, to input parameters, such as fuel structural groups, using statistical regression techniques. While convenient in establishing correlations, such methods cannot be used to characterize the role of the individual chemical processes contributing to the observations. The relatively recent advances in chemical kinetic modeling now allow for a more fundamental analysis of the combustion dynamics. In particular, sensitivity analysis [18,19] and reaction flux analysis [20,21] have been used extensively to identify the major reaction pathways and sensitive steps in chemical kinetic models. While these are powerful analysis tools, such techniques often only provide quantitative information focused on individual reactions and species, such as how sensitive a global combustion quantity is to a given reaction parameter, or how much a reaction contributes to the overall production or consumption of

---

* Corresponding author.
   E-mail address: pp427@cornell.edu (P. Pepiot).

a species. Therefore, some additional manipulation is required to get a broader picture, for instance, identifying the main pathways leading from the fuel all the way to the products themselves, and this is generally accompanied by a switch from quantitative measures to a more qualitative analysis.

A very attractive way to elucidate the mechanisms behind complex chemical reaction processes is to use atom tracking. Commonly used in experiments, these tracing techniques consist by selectively replacing some atoms in the initial reactants with isotopes, e.g. deuterium (^{2}H), carbon-13 (^{13}C), or carbon-14 (^{14}C) while searching for and quantifying the amount of these isotopes in products, thus allowing one to deduce what kinetic pathways are activated during reaction. Homan et al. [22], for instance, labeled carbon atoms in molecules typical of those found in modern fuels, and observed that while all carbon atoms contributed to soot regardless of the fuel, some contributed distinctively more than others. For example, primary carbons showed close to no preferential distribution in the products, while phenyl or naphthyl carbons had a much higher probability to end up in soot. Buchholz et al. [23] performed combustion experiments with ^{14}C-labeled diesel oxygenate dibutyl-maleate (DBM). DBM is a symmetric molecule that combines carbon–carbon single bonds, and carbon–oxygen single and double-bonds. By selectively labeling carbons in the molecules, they were able to determine how oxygenated functional groups within the fuel molecule impacted CO and CO_2 production, showing for example that in contrast to the C–O single bonds, which are not retained in the products, the C=O double-bonds in DBM do not break during combustion. A related study was conducted by Eveleigh et al. [24], in which seven oxygenated molecules with size ranging from 2 to 7 carbons were labeled using ^{13}C, and subjected to pyrolysis conditions at temperatures ranging from 1200 to 1450 °C. A similar conclusion was drawn, stating that carbon atoms contribute to particulate matter to various extents, depending on the nature of the neighboring moiety. Oxygen-containing functional groups have a significant influence on the formation of particulate matter through a reduction in the contribution of adjacent carbon atoms. For example, the methyl carbon in ethanol contributed two-third by weight to the resulting soot, compared to barely a fourth for the carbons linked to the hydroxyl group. As a comparison, all carbon atoms in toluene were found to contribute nearly equally to soot regardless of their nature: methyl carbon, tertiary carbon, or phenyl (ring) carbon.

As atom tracking techniques can provide valuable experimental information, the capability to label atoms in simulations also appears as a highly desirable feature for kinetic mechanism development, validation, and analysis. Such capability was previously developed by Bunev et al. [25,26] to investigate specific features of hydrogen–methane chemistry and hydrogen/nitric oxides interaction during ignition. Their approach, which they defined as a numerical tracers approach, consists of adding any molecule containing a labeled atom at a given location to the chemical kinetic mechanism of interest, and treating them as individual species. This requires duplicating all reactions involving species that contain at least one labeled atom, and adjusting the rates of these reactions accordingly to recover the appropriate overall rate of the original reaction. These rate adjustments were done using symmetry numbers or statistical factors commonly found in the transition state literature. By treating molecules with labeled atoms as individual species, evolving during combustion based on their own set of reactions, the labeled atoms can effectively be tracked during combustion, and their interactions with the remaining species identified. Bunev et al. used their tracking technique to determine, for instance, whether species such as H_2, O_2, CO_2, and CH_4, added as a diluent in the initial system, were actively participating in combustion, or acting as an inert diluent, or behaving as a combination of both. While functional when considering small fuel molecules and mechanisms (e.g., methane), this approach requires extensive and non-systematic changes to the chemical model used in the simulation, thus rendering any extension to molecules and mechanisms relevant for transportation fuels very difficult, if not impossible.

The objective of this paper is to present a novel automatic algorithm to numerically track labeled atoms or group of atoms during combustion, thereby creating new capabilities and perspectives to analyze and validate combustion kinetic models for complex fuels. The framework and guiding principles behind the proposed method are simple. We restrict the scope of this paper to time-evolving, homogeneous (0D) reactors, and focus on the kinetic network itself in the absence of transport. The user input simply consists in selecting which atom to label in the initial system. The output contains the time evolution of the concentration of labeled atoms at any location in any given species involved in the chemical mechanism used for the simulation. The emphasis is placed on ensuring that the algorithm can be directly applied to any chemical kinetic mechanism in conventional format (e.g. Chemkin), making the only additional information required being the actual structure of each molecular species appearing in the mechanism. Tracking equations are formulated automatically and systematically from the set of reactions contained in the chemical mechanism, using simple structural and energy-based arguments to characterize and quantify how, in each reaction, atoms are re-organized as the reactants are converted into products. These tracking equations are solved in parallel to the coupled ordinary differential equations (ODE) that describe the time evolution of species concentrations, and require minimum modification to the stiff ODE solver typically used for combustion kinetic simulations.

The focus of this paper is placed on the description of the method, followed by a number of examples aiming at illustrating some of the key analysis capabilities it enables. While these examples do provide some qualitative numerical evidence of the experimental observations mentioned above, more in-depth analysis of specific systems is left to future work.

The remainder of this paper is organized in the following sequence. First, some important definitions and notations are provided (Section 2), followed by the derivation of the governing equations (Section 3). Section 4 describes the core assumption of the method, which relies on minimizing a so-called reconfiguration enthalpy, and its numerical implementation through the Constrained Hierarchical Pattern Matching Algorithm (CHPMA), which automatically determines, for any given reaction, how atoms are transferred from reactants to products. Finally, Section 5 provides examples of how to post-process the raw data generated by the tracking algorithm in the context of soot precursors formation in rich hydrocarbon oxidation configurations. It is important to note that the automation does require the introduction of a number of assumptions, which are sometimes significant: those assumptions and the associated potential limitations and future directions of the technique, especially in terms of comparison with experimental data, are discussed in detail in Section 5.4.

2. Definitions and notations

We consider a reacting mixture of ideal gases, consisting of n^S chemical species. For simplicity of exposition, the mixture is assumed to evolve at a fixed volume, or constant density ρ, so that, given an initial pressure p, the full thermochemical state, or composition, of the mixture is completely characterized by the n^S-vector of species concentrations $\mathbf{C}$ (with C_i referring to the concentration of species S_i) and the mixture temperature T. The mixture evolves in time according to a chemical kinetic mechanism containing n^R reactions, with each one being uniquely defined by a list
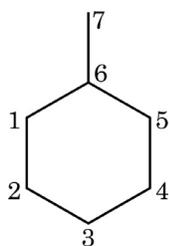
Fig. 1. Indexing of a methyl-cyclohexane molecule: only skeleton atoms, in this case all carbon atoms, are assigned an index from 1 to 7. Order is arbitrary, and symmetry properties are ignored.

of reactants, a list of products, and an appropriate set of Arrhenius rate coefficient parameters.

2.1. Atom labeling and atom concentrations

For simplicity of exposition, and without loss of generality, we restrict ourself to hydrocarbon combustion. In this case, any species S_i consists of carbon, C, oxygen, O, and hydrogen, H, atoms only, and its formula can be written:

$$S_i : C_{n_i^C} O_{n_i^O} H_{n_i^H} \tag{1}$$

We use the term *skeleton atom* for any atom able to form more than one covalent bond, in our case C and O, and denote by n_i^A the total number of skeleton atoms in species S_i:

$$n_i^A = n_i^C + n_i^O \tag{2}$$

Skeleton atoms in species S_i are indexed from 1 to n_i^A, as illustrated in Fig. 1 for methyl-cyclohexane. Note that the order in which the atoms in a molecule are indexed is irrelevant, and that for simplicity later on, all skeleton atoms are indexed, regardless of the symmetry properties of the molecule.

We denote by $\kappa_{i;j}$ the concentration of a skeleton atom, C or O, at a given location j in species S_i. This concentration is trivially equal to that of the species itself for any i and j:

$$\kappa_{i;j} = C_i . \tag{3}$$

We define κ_i to be the n_i^A-vector containing skeleton atom concentrations at all possible locations in species S_i. The total concentration of skeleton atoms in species S_i is easily obtained as the L_1-norm of κ_i, and denoted by K_i:

$$K_i = |\kappa_i|_1 = n_i^A C_i , \tag{4}$$

where the Einstein summing convention is not implied.

The above variables are extended to account for labels affixed to atoms in order to distinguish them from one another and to track how they are transferred from species to species as combustion proceeds. Atoms can be labeled at the very beginning of a simulation, or at any time thereafter. When labeling, all atoms at user-specified locations are assigned a specific label, with the total number of labels being n^L. For clarity, labels are chosen to be contiguous integers, from 1 to n_L. Labels are identified with a $\star$ superscript, e.g. $\star k$ designates the kth label. Skeleton atoms at non-labeled locations are assigned a $\star 0$ default label.

Labeling examples are shown in Figs. 2 and 3. Figure 2(a) illustrates how to label atoms in a methyl-cyclohexane (MCH)/air mixture to track a single atom from the fuel, Fig. 2(b) illustrates how to label atoms in order to track each atom in the initial MCH molecules independently from one another, and Fig. 3 demonstrates how to label atoms in a multi-component mixture in order to track the fate of a specific fuel component, in this case, MCH in a MCH/n-heptane (NC7) mixture.

At this point, label-specific atom concentrations can now be defined. Following the notation introduced above, we denote by $\kappa_{i;j}^{\star k}$

(a) $n_L = 1$ (b) $n_L = 7$
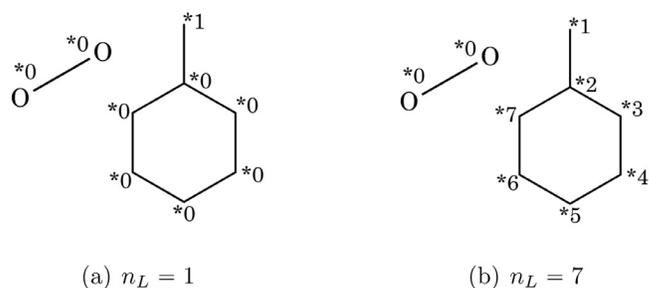
Fig. 2. Single and multiple atoms labeling for MCH oxidation. Note that in the multiple atoms labeling case, results will be identical for labels $\star 3$ and $\star 7$ on the one hand, and $\star 6$ and $\star 4$ on the other hand, due to the symmetry of the methyl-cyclohexane molecule.
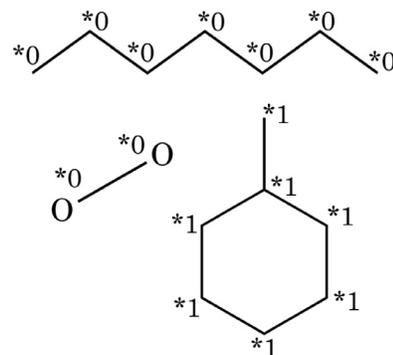
Fig. 3. Molecule labeling in a multi-component (NC7 and MCH) mixture: all carbons from MCH will be tracked ($n_L = 1$).

the concentration of $\star k$-labeled atom at location j in species S_i, and $\kappa_i^{\star k}$ the vector of size n_i^A containing those concentrations at all locations in species S_i. For example, the initial concentrations of $\star 2$-labeled atoms in MCH corresponding to Fig. 2(b) are:

$$\kappa_{MCH}^{\star 2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ C_{MCH} \\ 0 \end{bmatrix} , \tag{5}$$

where the skeleton atom indexing of Fig. 1 is used. Note that for clarity, MCH is used in place of the species index i. The total concentration of $\star k$-labeled atoms in species S_i is defined by:

$$K_i^{\star k} = |\kappa_i^{\star k}|_1 = \sum_{j=1}^{n_i^A} \kappa_{i;j}^{\star k} . \tag{6}$$

The objective of this tracking methodology can now be redefined more clearly as the following: given an initial mixture of hydrocarbon species and oxygen, in which all or part of the skeleton atoms have been labeled using n^L distinct labels, compute $\kappa_{i;j}^{\star k}(t)$, which is the time evolution of the concentration of $\star k$-labeled atom at each location on each species involved in the chemical system.

2.2. Atom transfer probabilities

The next step in establishing our numerical tracking framework is to describe the exchange of atoms between reactants and products at the individual reaction level. For this purpose, we consider a simple example, namely, the reaction of toluene (TOL) with

Fig. 4. Realizable atom mappings for Reaction 8. Mapping $\mathcal{M}_1$, or $\mathcal{M}_{\mathrm{TOL,A_1}}$ describes how atoms in toluene are transferred to benzene, while mapping $\mathcal{M}_2$, or $\mathcal{M}_{\mathrm{TOL,CH_3}}$ describes how they are transferred to the methyl radical. $\mathcal{M}_1$ and $\mathcal{M}_2$ combined form a permutation matrix.

hydrogen (H), yielding benzene (A_1) and a methyl radical (CH_3):

$$TOL + H \longrightarrow A_1 + CH_3 . \tag{7}$$

This reaction can also be written using the molecular structures directly:

$$(8)$$

where indices have been assigned to all skeletal atoms involved. Basic atom conservation states that each atom in a toluene molecule undergoing reaction is transferred to one and only one location in the products benzene and methyl. This transfer can be represented mathematically as a permutation matrix $\mathcal{P}$ of rank equal to the total number of skeleton atoms on either reactant or product side (this rank is equal to 7 for Reaction 8). Each row corresponds to a skeleton atom location in the reactant side, while each column corresponds to a skeleton atom location in the products. Elements of the permutation matrix can only take two values: 1 if an atom is effectively transferred between the two locations during reaction, 0 otherwise, with only one non-zero entry allowed per row and per column. It is convenient to treat this permutation matrix as a block matrix, each block $\mathcal{M}_{A,B}$ corresponding to the re-arrangement, or *mapping* of atoms from reactant A to product B. Figure 4 provides an example of such mappings for Reaction 8. The full permutation matrix is decomposed into two mapping matrices, $\mathcal{M}_{\mathrm{TOL,A_1}}$ and $\mathcal{M}_{\mathrm{TOL,CH_3}}$.

In this example, the methyl carbon in toluene is transferred to the methyl radical, as indicated by the first row of $\mathcal{M}_{\mathrm{TOL,A_1}}$ containing only zeros, and the first element of $\mathcal{M}_{\mathrm{TOL,CH_3}}$ being unity,

while all carbons in the aromatic cycle are transferred to the benzene molecule. This mapping is consistent with the expected reaction pathway, in which the ring is preserved and $^1C - ^2C$ bond is the only bond broken, and is therefore classified as *realizable*. Due to the symmetry properties of the aromatic cycle, any mapping that corresponds to a rotated cycle are also realizable (Fig. 5(a)). However, any set of mappings that would require breaking more than the $^1C - ^2C$ bond in toluene, for example, through a re-organization among the ring carbons, are not expected to occur and are deemed "unrealizable" (Fig. 5(b)).

Generalizing notations, $\mathcal{P}^{n,(m)}$ will be used to designate the mth realizable atom permutation for reaction R_n. $\mathcal{P}^{n,(m)}$ is associated with a corresponding set of mapping matrices for all reactant i-product j pairs in R_n. Those mapping matrices will be denoted by $\mathcal{M}_{i,j}^{n,(m)}$. The total number of realizable permutations (or set of mappings) for reaction R_n is denoted by $n_{\mathcal{P}}^n$. We assume that *i*) realizable permutations can reliably be identified, and *ii*) those realizable permutations occur with equal probability. Accordingly, we finally define for each reaction R_n a set of *transfer probability matrices* $\{\mathcal{T}^n\}$, with each entry $\mathcal{T}_{i,j}^n$ defined as the matrix whose elements (k, p) contain the probability of transferring a skeleton atom at location k in the ith reactant to location p in jth product in reaction R_n, computed by averaging all $n_{\mathcal{P}}^n$ realizable mappings between those two species:

$$\mathcal{T}_{i,j}^n = \frac{1}{n_{\mathcal{P}}^n} \sum_{m=1}^{n_{\mathcal{P}}^n} \mathcal{M}_{i,j}^{n,(m)} \tag{9}$$

Coming back to Eq. (8), it is easily shown that there are only twelve realizable mappings, differing from one another through simple rotation or symmetry of the aromatic cycle. The resulting set of transfer matrices for that reaction are shown in Fig. 6.

This simple example also illustrates the fact that transfer matrices should fully reflect the symmetry properties of the various molecules participating in the reaction, with atoms structurally indistinguishable from one another in the reactants being assigned identical rows in the transfer matrix.

3. Formulation of transfer equations

The conservation equation governing the time evolution of a species concentration C_i can be written:

$$\frac{dC_i}{dt} = \sum_{n=1}^{n^R} \nu_{i;n} \omega_n , \tag{10}$$

where $\nu_{i;\,n}$ is the stoichiometric coefficient of species S_i in reaction R_n (negative for a reactant, and positive for a product), and ω_n is the rate of reaction R_n, expressed as

$$\omega_n = k_n \prod_{j=1}^{n^S} C_j^{\gamma_{j;n}^R} . \tag{11}$$

In the previous equation, k_n is the rate constant of reaction R_n in Arrhenius form, and

$$\gamma_{j;n}^R = \max\left(0, -\nu_{j;n}\right) \tag{12}$$

is non-zero and equal to $|\nu_{j,\,n}|$ only if S_j is a reactant in R_n. For later convenience, we also define the similar quantity for products:

$$\gamma_{j;n}^P = \max\left(0, \nu_{j;n}\right), \tag{13}$$

non-zero and equal to $|\nu_{j,\,n}|$ only if S_j is a product in R_n.

The conservation equations describing the time evolution of the labeled atom concentrations are similar to those for the species concentrations themselves, but modified to account for the transfer probabilities $\mathcal{T}$ defined above. Accordingly, the time derivative of $\kappa_i^{\star k}$, the vector containing the concentration of $^\star k$-labeled atoms

$$\mathcal{M}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathcal{M}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathcal{M}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathcal{M}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

(a) Realizable mappings.　　　　　　　　　(b) Unrealizable mappings.

Fig. 5. Additional mapping matrices for reaction 8. $\mathcal{M}_1$ is used for $\mathcal{M}_{\mathrm{TOL},A_1}$, and $\mathcal{M}_2$ is used for $\mathcal{M}_{\mathrm{TOL},CH_3}$.

$$\mathcal{T}_{\mathrm{TOL},A_1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

$$\mathcal{T}_{\mathrm{TOL},CH_3} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Fig. 6. Set of transfer matrices for Reaction 8.

at each location in species S_i, can be written as the difference between production ($\mathcal{P}$) and consumption ($\mathcal{C}$) terms:

$$\frac{d\kappa_i^{\star k}}{dt} = \left.\frac{d\kappa_i^{\star k}}{dt}\right|_{\mathcal{P}} - \left.\frac{d\kappa_i^{\star k}}{dt}\right|_{\mathcal{C}}. \tag{14}$$

3.1. Consumption term

The rate of consumption of $\star k$-labeled atoms at location j in species S_i, itself appearing as reactant in reaction n, is equal to the full reaction rate ω_n, multiplied by the fraction of $\star k$-labeled atoms at this location, $\alpha_{i;j}^{\star k}$:

$$\alpha_{i;j}^{\star k} = \frac{\kappa_{i;j}^{\star k}}{\kappa_{i;j}} = \frac{\kappa_{i;j}^{\star k}}{C_i}, \tag{15}$$

or $\boldsymbol{\alpha}_i^{\star k}$ in vector form (collecting values for all locations in species S_i). The consumption term for $\boldsymbol{\kappa}_i^{\star k}$ can therefore be written as:

$$\left.\frac{d\boldsymbol{\kappa}_i^{\star k}}{dt}\right|_{\mathcal{C}} = \sum_{n=1}^{n_R} \gamma_{i;n}^R \omega_n \boldsymbol{\alpha}_i^{\star k}, \tag{16}$$

where γ^R has been defined in Eq. (12). Eq. (16) is inconvenient to use directly, since $\alpha_{i;j}^{\star k}$ becomes ill-defined when C_i, the concentration of species S_i, is small or zero. Instead, we note that all non-zero contributions in Eq. (16) come from reactions involving S_i as a reactant, and are therefore proportional to $C_i^{\gamma_{i;n}^R}$, with $\gamma_{i;n}^R \geq 1$ if R_n is elementary. Hence, the division by C_i can readily be absorbed into the reaction rate expression to yield a well-defined quantity:

$$\omega_n \alpha_{i;j}^{\star k} = \frac{\omega_n}{C_i} \kappa_{i;j}^{\star k}$$

$$= k_n \left(\prod_{\substack{\text{reactants } S_j \\ S_j \neq S_i}} C_j^{\gamma_{j;n}^R} \right) C_i^{\gamma_{i;n}^R - 1} \kappa_{i;j}^{\star k} \tag{17}$$

3.2. Production term.

The rate of production of labeled atoms is somewhat more complicated to express, since it involves the atom transfer probability matrices $\mathcal{T}$ specific to each reaction. In vector form, following the same notations as before, this term can be written as:

$$\left.\frac{d\boldsymbol{\kappa}_i^{\star k}}{dt}\right|_{\mathcal{P}} = \sum_{n=1}^{n^R} \gamma_{i;n}^P \omega_n \left(\sum_{m=1}^{n^S} \delta_{m;n}^R \left(\mathcal{T}_{m,i}^n\right)^{\mathbf{T}} \boldsymbol{\alpha}_m^{\star k} \right), \tag{18}$$

where the transpose of the transfer probabilities matrix is used and

$$\delta_{m;n}^R = \begin{cases} 1, & \text{if } S_m \text{ is a } \textbf{reactant} \text{ in reaction } R_n \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

The jth term of the vector Eq. (18) corresponds to the production rate of $\star k$-labeled atoms at location j in species S_i, $\kappa_{i;j}^{\star k}$. This is obtained by summing over each reaction n in which S_i is a *product* (as indicated by the use of γ^P), the contributions from all possible sources of $\star k$-labeled atoms in that reaction. These contributions come from all locations in the *reactants*, and are proportional to the fraction of $\star k$-labeled atoms available and to the transfer probability between these locations and the jth location in S_i. Again, this is a well-conditioned equation even for zero or low concentration species because S_m is necessarily a reactant in reaction n, so that the division by C_m in $\boldsymbol{\alpha}_m$ can be absorbed conveniently in the reaction rate, ω_n, as shown in Eq. (17).

Some tedious, but straightforward algebra can show that Eq. (14), in conjunction with Eqs. (16) and (18), satisfies the following two conservation properties:

- At any given time, the sum of all labeled and $\star$0-labeled atom concentrations at any location j in any species i is equal to the concentration of species S_i:

$$\sum_{\star k=0}^{n_L} \kappa_{i;j}^{\star k} = C_i \text{ for all } i, j \tag{20}$$

- For constant volume systems, for any label $^\star k$, the total concentration of $^\star k$-labeled atoms in the system, that is, summed over all possible skeleton atom locations, is constant at all times and equal to its initial value $K^{\star k,0}$:

$$\sum_{i=1}^{n^S}\sum_{j=1}^{n_i^A} \kappa_{i;j}^{\star k} = \sum_{i=1}^{n^S} K_i^{\star k} = K^{\star k,0} \text{ for all } {}^\star k \tag{21}$$

The number of atom tracking equations that have to be solved in addition to the conservation equations for species concentrations is directly proportional to the number of labels being used and to the number of skeleton atom locations in the system. The total number of variables can be written:

$$n_{tot} = n^S + n^L \times \sum_{i=1}^{n^S} n_i^A. \tag{22}$$

Because the tracking atom equations are linear, a simulation tracking n_L labels can be replaced by n_L simulations tracking one label, for which the total number of variables is reduced to $n^S + \sum_{i=1}^{n^S} n_i^A$. While significant, this is still within the capabilities of recent stiff ODE solvers for most kinetic mechanisms.

4. Transfer probabilities calculations

As demonstrated above, the key ingredient to the tracking methodology is the definition of the transfer probability matrices $\mathcal{T}$ for each reaction in the kinetic mechanism under consideration. With typical mechanisms containing between 10^3 and 10^4 reactions, manual inspection of the structural changes resulting from each and every reaction is impossible, and a fully automatic approach is essential. For this purpose, we designed a robust methodology that relies on minimizing a so-called reconfiguration energy across the reaction. The numerical implementation is accelerated by a careful analysis of the structure of both reactants and products to efficiently generate the necessary set of transfer matrices for each reaction without the need for external expertise. The methodology and underlying assumptions are detailed below.

4.1. Main assumptions

During a chemical reaction, reactant molecules collide and transform into products. The rate at which collision and conversion occur is quantified by the law of mass action and the Arrhenius parameters associated with the reaction. As the reaction proceeds, the reactant molecules are affected by structural changes: covalent bonds are created or broken, hydrogen atoms are transferred, carbon atoms are reorganized, and the geometry of the reactant molecules change to that of the products. Determining the fate of a given atom in the reactants and its final location in the products usually requires precisely knowing the structure of the transition states involved in the reaction. These structures may become quite complex, and can be determined for example from *ab initio* quantum chemistry calculations [27]. While it is essential to determine reaction constants, no information about the transition states is retained in the chemical kinetic mechanism formats most commonly used by the combustion community, which typically consist of a list of reactions specifying reactants and products, and for each reaction, a set of Arrhenius parameters.

In the context of this work, we make two important simplifying assumptions:

- The chemical kinetic mechanisms on which the method is applied are formed for the most part by *elementary* reactions, that is, reactions describing actual molecular collisions, and for which the transition states are assumed to be relatively simple and close in structure to both reactants and products.

- Under these conditions, the actual transition state occurring during the reaction is assumed to be the one that minimizes the extent of molecular re-arrangements, measured by the number and type of covalent bonds broken or created between skeleton atoms, and the number of hydrogen atoms transferred during the reaction.

While significant, those assumptions allow us to automatize the calculation of the transfer probability matrices. Exceptions, however, are expected to occur, and appropriate treatment is provided for non-elementary reactions or identified reactions with complex transition states. Note that it is always possible to specify the transfer probabilities manually for those reactions for which the automatic procedure is deemed not valid by an expert user.

4.2. Species structural representation

A graph-based approach is used to record the structural characteristics of chemical species involved in a given kinetic mechanism and to track changes during the reaction. For this, a two-level representation is adopted. The first level, called *skeleton representation*, only tracks skeleton atoms and how they are linked to one another. This is done using simple non-oriented graphs where vertices are created for each skeleton atom in the molecule, distinguishing between the type of atoms (C or O). An edge is placed between two vertices if a bond exists between the two corresponding atoms. The second representation, or *enhanced representation*, adds attributes to each vertex to account for radicals (through a deficiency in the number of linked hydrogen atoms), and to each edge to account for the nature of the covalent bond (simple, double, or triple bonds). Examples for each representation level is provided in Fig. 7.

Whenever available, the graph representations are directly obtained from the Chemical Abstracts Service (CAS) Registry Number [28] and corresponding *molfile*, a widely used chemical structure format file originally developed at Molecular Design Limited [29]. In the absence of a CAS number, structures are postulated from context and any supplementary material provided with the mechanism, and a proper molfile is built manually. The molfiles are then converted to a suitable graph representation containing a list of vertices (skeleton atoms), adjacency list (representing the bonds between those atoms), and attributes, and are stored for use in the subsequent steps.

4.3. Reconfiguration enthalpies

Transfer probability matrices are computed from the realizable permutations and corresponding reactants-products mappings in any given reaction, and therefore a criterion must be defined to distinguish realizable permutations from unrealizable ones. In line with our previously stated assumptions, we assign to each permutation $\mathcal{P}$ a *reconfiguration enthalpy*, $\mathcal{H}_\mathcal{P}$, that quantifies the extent of the structural changes necessary to convert the set of reactants into the set of products according to the mapping matrices $\mathcal{M}$ contained in $\mathcal{P}$. More specifically, the enhanced graph representations of each reactant and product involved in the reaction, combined with the atom transfer information contained in $\mathcal{P}$ is used to determine the fate of all covalent bonds in the reaction, and to classify them into the following categories: broken, created, transformed into another bond type, or transferred unchanged. Assigning a bond dissociation enthalpy value to each bond type, we can write $\mathcal{H}_\mathcal{P}$ as an effective change in bond enthalpy from reactants to products:

$$\mathcal{H}_\mathcal{P} = \sum_{b\in\mathcal{B}_c} H_{b,p} + \sum_{b\in\mathcal{B}_b} H_{b,r} + \sum_{b\in\mathcal{B}_t} |H_{b,p} - H_{b,r}| \tag{23}$$

(a) Actual molecule   (b) Skeleton representation   (c) Enhanced representation
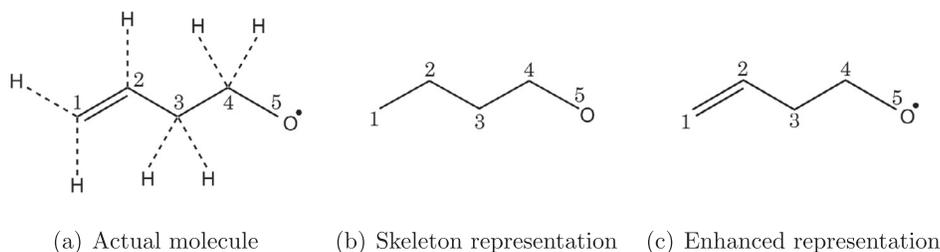
Fig. 7. Molecule representations used in the tracking methodology. Skeleton representations of the molecule keep track of skeleton atoms and covalent bonds only, while enhanced representations contain additional attributes such as bond type or radical location.

Table 1
Bond dissociation enthalpies used to evaluate the reconfiguration enthalpies of permutations, adapted from [30].

Bond type	Generic bond dissociation Enthalpy (KJ/mol)
O–O	146
C–C	348
C–O	358
C–H	413
O–H	463
O=O	495
C=C	614
C=O	799
C≡C	839
C≡O	1072

where

- $\mathcal{B}_c$ is the set of bonds in the products being *created* by the permutation $\mathcal{P}$.
- $\mathcal{B}_b$ is the set of bonds in the reactants being *broken* by the permutation $\mathcal{P}$.
- $\mathcal{B}_t$ is the set of bonds changing type across reaction by the permutation $\mathcal{P}$.
- $H_{b,p}$ and $H_{b,r}$ are the assigned dissociation enthalpies of a bond b in products and reactants, respectively. Note that if a bond is transferred unchanged from a reactant to a product, its net contribution $|H_{b,p} - H_{b,r}|$ is zero, and the bond is effectively ignored in the calculation of $\mathcal{H}_{\mathcal{P}}$.

A permutation $\mathcal{P}^{n,(m)}$ and corresponding set of mapping matrices $\mathcal{M}_{i,j}^{n,(m)}$ are deemed *realizable* if its reconfiguration enthalpy is minimum:

$$\mathcal{H}_{\mathcal{P}}^{n,(m)} = \mathcal{H}_{\mathcal{P}}^{n}\big|_{\min}, \tag{24}$$

where the minimum is evaluated across all n possible permutations for reaction R_n, where n is equal to the factorial of the number of skeleton atoms on either side of the reaction. Note that Eq. (23) is similar to the calculation of a heat of reaction using bond dissociation enthalpies, but only account for the bonds affected by a given permutation $\mathcal{P}$. Bond dissociation enthalpies typically depend on steric factors linked to the the molecules geometries and thermochemical state. However, in the simplified context of this work, constant values have been found sufficient to characterize the realizability of a permutation, and the values provided in Table 1 are used.

As an example, we calculate the reconfiguration enthalpy for several possible atom permutations for the decomposition reaction of C_7H_{11} into C_4H_5 and C_3H_6. The enhanced representations of each of those molecules is provided in Fig. 8.

A schematic of four representative sets of mapping matrices are shown in Fig. 9. For each of them, the structure of the reactant C_7H_{11} is shown vertically, across rows, while the structure of the

Fig. 8. Enhanced representations of each molecule involved in the reaction $C_7H_{11} \rightarrow C_4H_5 + C_3H_6$.

two products C_4H_5 and C_3H_6 is shown horizontally. Each row corresponds to one atom location in the reactant, while each column corresponds to one atom location in the products. Shaded cells indicate where, according to each specific mapping, reactant atoms are transferred in the products. The necessary bond reconfigurations to match those transfers are indicated on the molecules.

The numerical values for all four mappings are provided in Fig. 10. Taking Mapping $\mathcal{P}_1$ as an example, the detail of the reconfiguration enthalpy calculation gives:

$$\mathcal{H}_{\mathcal{P}_1} = \underbrace{H_{1_p-H} + H_{5_p-6_p} + H_{7_p-H}}_{\text{bonds created}}$$
$$+ \underbrace{H_{1_r-H} + H_{1_r-2_r} + H_{3_r-H} + H_{5_r-6_r}}_{\text{bonds broken}}$$
$$+ \underbrace{|H_{1_p-2_p} - H_{2_r-3_r}| + |H_{6_p-7_p} - H_{6_r-7_r}|}_{\text{bonds transformed}} \tag{25}$$

The total reconfiguration enthalpies are shown to vary widely across mappings, with a clear minimum obtained for $\mathcal{P}_4$. This minimum reconfiguration enthalpy is associated with the expected β-scission of the $3_r - 4_r$ covalent bond to form a propylene molecule.

Lumped species treatment

The procedure outlined above to identify realizable mappings requires accurate enhanced representations that identify without ambiguity the type of bonds between skeleton atoms, and the location of hydrogen atoms. However, especially in the case of hydrocarbon chemistry, it is customary to lump individual isomer molecules into a single representative species in order to reduce the complexity of the mechanisms. Since bond type, or radical location, become undetermined for those lumped species, enhanced representations cannot be used anymore in the evaluation of $\mathcal{H}_{\mathcal{P}}$. The same holds true for resonance-stabilized species with mesomeric forms: the location of radical and/or special bonds in these species is also ambiguous, and they are considered as lumped from CHPMA's perspective. Reconfiguration enthalpies for reactions involving *at least one lumped species* will be computed from the skeleton representation of the molecules, by considering the *number of bonds broken* only. This can be achieved in a straight forward manner by using Eq. (23) with all bond dissociation enthalpies replaced by a constant, unity value.

4.4. Hierarchical pattern matching algorithm

The minimum reconfiguration enthalpy criterion is in theory sufficient to determine the transfer probabilities matrices for all

(a) Mapping $\mathcal{P}_1$

(b) Mapping $\mathcal{P}_2$

(c) Mapping $\mathcal{P}_3$

(d) Mapping $\mathcal{P}_4$
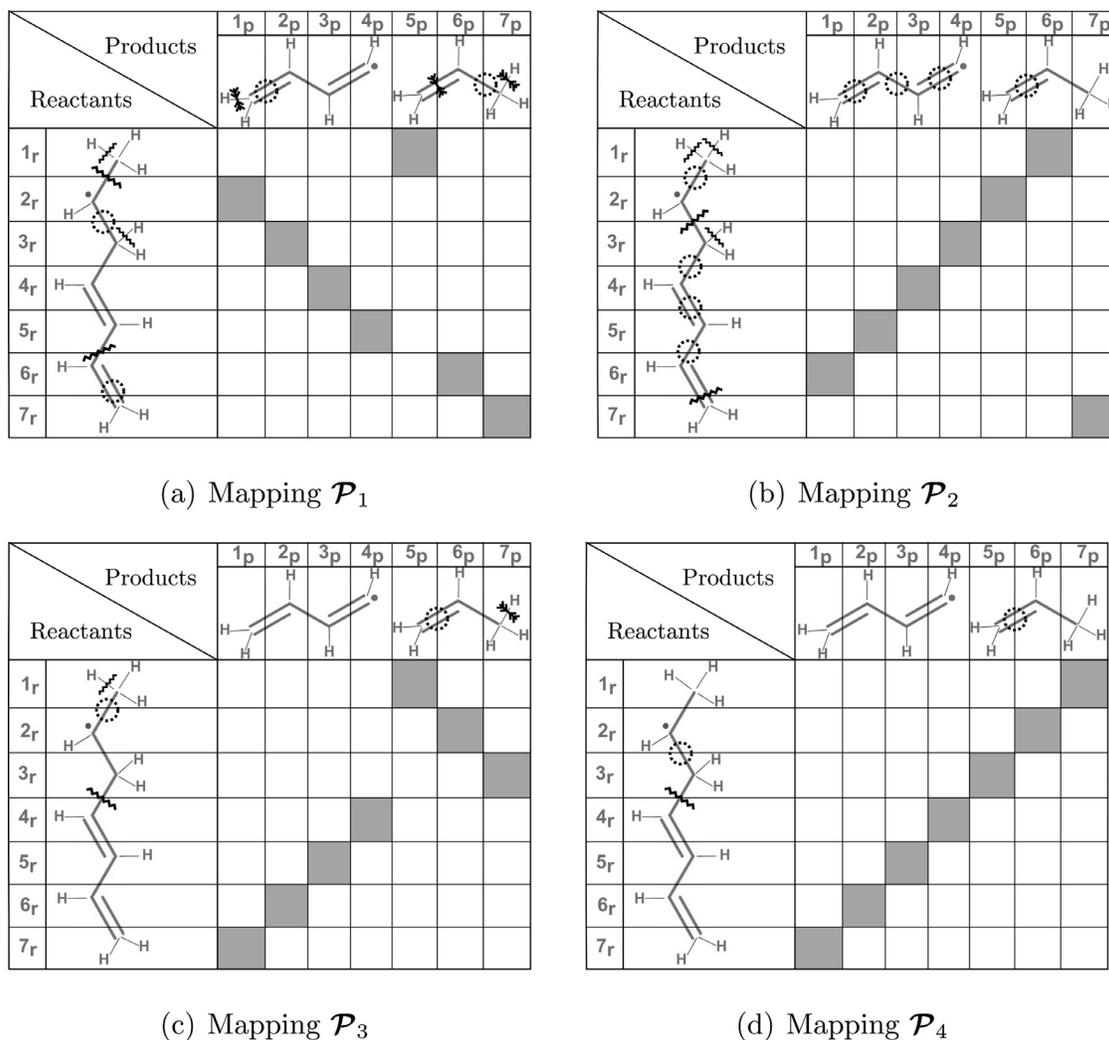
Fig. 9. Four of the mappings output by the hierarchical pattern matching for the reaction x13C$_7$H$_{11}$6 → N-C$_4$H$_5$ + C$_3$H$_6$. An index i_r (resp. j_p) refers to the carbon atom of the reactants (resp. products) located in the same row (resp. column). The shaded cells indicate the correspondence between atoms i_r from the reactants and atoms j_p of the products. The bond changes induced by the mappings are directly indicated on the molecules: broken lines correspond to broken bonds in the reactant, arrows to bonds being created in the products, and dotted circles indicate bonds that are transferred, but change type between reactant and product ("transformed" bonds). Bonds without specific marking are transferred unchanged from reactants to products.
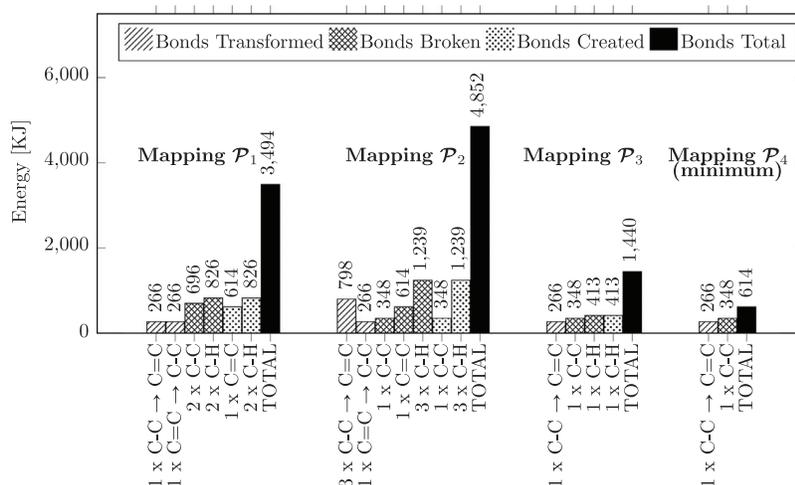
Fig. 10. Detailed enthalpy calculations showing bonds created, broken, and transformed for the four distinct sets of mapping matrices shown if Fig. 9. Mapping $\mathcal{P}_4$ yields a minimal reconfiguration enthalpy, and is therefore deemed realizable. The corresponding bond breaking pattern, shown in Fig. 9(d) is indeed the simplest and expected one.
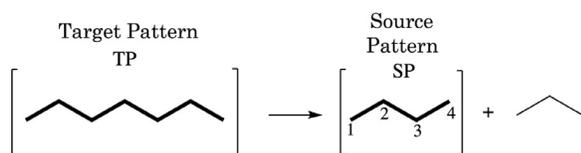
Fig. 11. Skeleton representation of Reaction 8, indicating the initial Target and Source patterns as the reactant and the second largest molecule in the reaction, respectively.

Fig. 12. Fitting the source pattern SP into the target pattern TP shown in Fig. 11 using a sub-graph isomorphism algorithm. A total of 8 combinations are identified, accounting for all symmetries.

Fig. 13. Two of the 8 possible pattern reactions at the start of the second iteration. In the first case, the C_4 source pattern has been removed from the middle of the C_7 target pattern, leaving on the reactant side an isolated atom and a C_2 pattern. In the second case, C_4 source pattern was removed from the side of the C_7 target pattern. Applying Step 1 to them identifies in the first case, the C_3 chain on the product side as target pattern, and the C_2 chain on the reactant side as the source pattern, while on the second case, SP and TP are identical, and SP is by default chosen to be on the reactant side.

reactions in a kinetic mechanism. However, a brute-force application of the criterion requires the examination of every single permutation between the skeleton atoms on the reactants and products side in order to decide if it is realizable or not. As the number of permutations is equal to the factorial of the number of atoms on either side, the brute-force approach becomes quickly prohibitive from the computational point of view when applied to the large hydrocarbon molecules typically found in combustion. For example, any decomposition reaction involving dodecane, a C_{12} molecule, would require evaluating Eq. (23) nearly half a billion times. To dramatically accelerate the computation, we introduce a pre-processing step using pattern matching arguments, the Hierarchical Pattern Matching Algorithm or HPMA, to reliably identify a much smaller set of permutations that have the potential of being realizable, and automatically eliminate a large fraction of the high reconfiguration enthalpy mapping scenarios.

4.4.1. General algorithm

We introduce *skeleton patterns* (or simply *patterns*), which designate any substructure in the skeleton representation of a molecule or combination thereof. A pattern can be as small as a single skeleton atom, correspond to a molecule, or be as big as the entire set of molecules appearing on the product side of a reaction. Any subset of a pattern is still considered a pattern, as is any set of several patterns. Our objective is to identify permutations that retain large reactant patterns in the products, or vice-versa, since those will be associated with the smallest $\mathscr{H}_\mathcal{P}$ values. This is done using a recursive procedure similar to that one would use to solve a jigsaw puzzle. Mainly, we constrain atom transfer between reactants and products for the largest common pattern first, and try to fit the smaller remaining patterns afterward. The steps involved at each level of the recursive algorithm are described next, with Reaction 8 being used as illustration. Here, the algorithm is applied successively to each of the n_R reactions contained in the kinetic mechanism. Note that reactions will be treated as a collection of reactant patterns and product patterns throughout the algorithm, and will therefore be referred to as *pattern reactions*.

- **Step 1.** Identify the *second* largest pattern contained within a single molecule on either reactant and product side of the reaction. This pattern is labeled *Source Pattern* (SP). We label *Target Pattern* (TP) the collection of patterns found on the side of the reaction opposite to SP. In the first iteration, SP is simply the second largest molecule, while TP is either the set of reactants (if SP is a product) or the set of products (if SP is a reactant). An example is provided in Fig. 11.

- **Step 2.** Establish an exhaustive list of all possible ways to fit the source pattern SP into the target pattern TP without breaking any bonds within those two patterns. This pattern matching step is done in practice using a sub-graph isomorphism algorithm (here, the VF2 algorithm [31] from the Boost Graph Library [32]), directly compatible with the graph skeleton representation used to describe the molecules. The iso-morphism algorithm automatically takes into account the multiple possibilities to fit source patterns into target patterns due to molecular

symmetries. A total of 8 possible matches can be identified for the SP/TP pair shown in Fig. 11. Those fits are listed in Fig. 12.

- **Step 3.** For each of the solutions returned by the isomorphism algorithm, first record the connection between the skeleton atoms in SP and the matching part in TP to be part of the appropriate permutation matrix. Then, remove the source pattern and corresponding part in the target pattern from the reaction, thereby generating a new pattern reaction with a smaller list of atoms to transfer from reactants to products. This step is illustrated in Fig. 13.

Steps 1 to 3 are then applied to each of the newly generated pattern reactions created at Step 3. The recursive algorithm stops when the pattern reaction at Step 3 is empty, with no more atoms to match between the reactant and product side. At that point, the single permutation matrix corresponding to the various SP/TP matches made throughout the iterations leading to that empty pattern reaction is returned and stored as a potential candidate for the reconfiguration enthalpy minimization. The hierarchical nature of the algorithm is evident through the choice of the source and target patterns at each iteration, since the largest patterns are matched first, guaranteeing a small number of broken bonds, and therefore, a small $\mathscr{H}_\mathcal{P}$ value later on. Applying HPMA on Reaction 8 reduces the number of permutations on which to apply Eq. (23) from 7! = 5040 to just 24, among which Mapping $\mathcal{P}_4$ from Fig. 9 can be found.

4.4.2. Special treatments

While the algorithm, as described above, covers most of the pattern matching configurations encountered in practice, some specific situations require some special treatment: the presence of different types of skeleton atoms (in our case, the presence of oxygen), and cases for which Step 2 cannot return any matching solutions.

4.4.2.1. Reactions containing oxygen.
The skeleton representation of the molecule accounts for the nature of the skeleton atoms through labels. The isomorphism algorithm being applied on the

Fig. 14. Initial selection of source and target patterns for a oxygen-containing reaction.
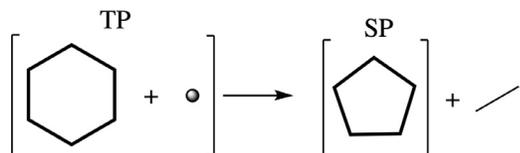
Fig. 15. Conversion of a 6-carbon ring to a 5-carbon ring. Due to the presence of cycles, the iso-morphism algorithm is unable to fit the 5-carbon cycle into the 6-carbon cycle.

labeled graphs will therefore ensure that each pattern matching solution matches oxygen atoms to oxygen atoms and carbons to carbons. For reasons of efficiency, the selection of SP and TP from Step 1 is slightly modified when the reaction contains both carbon and oxygen to prioritize patterns that contain oxygen. In contrast to the main algorithm, patterns are ranked first on the number of oxygen they contain, and then, on their size. Patterns containing carbon only are ignored. Also in contrast to the main algorithm, the Source Pattern SP is chosen as the *smallest* oxygen containing pattern in the pattern reaction. An illustration of the oxygen-specific selection of source and target patterns is provided in Fig. 14.

4.4.2.2. No-match cases. In some situations, the iso-morphism algorithm is unable to find solutions with a given SP/TP pair, as identified through Step 1, due to extensive changes in the geometry of the molecule. A typical example is the conversion of a 6-carbon ring to a 5-carbon ring, as illustrated in Fig. 15.

In these cases, a systematic procedure is used. Here, the constraints on the pattern matching algorithm are relaxed by replacing SP by modified source patterns SP*, obtained by arbitrarily breaking one bond in SP. The pattern reaction is therefore replaced by *n* pattern reactions involving SP* instead of SP, *n* being the number of bonds involved in SP. HPMA then proceeds as usual on each of the modified pattern reactions. For example, the pattern reaction depicted in Fig. 15 is replaced in Fig. 16 by 5 modified pattern reactions where SP, initially a 5-carbon cycle, has been replaced by a list of a 5-carbon linear patterns, allowing it to be fitted in the 6-carbon ring by the isomorphism algorithm. Note that this procedure can be repeated when needed by breaking more than one bond in the initial source pattern.

The Hierachical Pattern Matching Algorithm, combined with our criterion of minimum reconfiguration enthalpy is called the *Constrained Hierachical Pattern Matching Algorithm*, or CHPMA. The main steps in CHPMA are summarized in the flowchart in Fig. 17.

5. Applications and examples

The numerical tracking framework, composed of CHPMA and the transfer equations formulated in Section 3, has been implemented in an in-house kinetic code, and used to simulate homogeneous isochor reactors with a variety of initial conditions. The kinetic mechanism used in this section is the high-temperature part of the mechanism developed by Narayanaswamy et al. [33–36], composed of 235 species and 2187 reactions. This mechanism is especially well-suited to demonstrate the benefits of the tracking approach, since it contains consistent submodules describing the oxidation of a variety of hydrocarbon fuels. The tracking equations
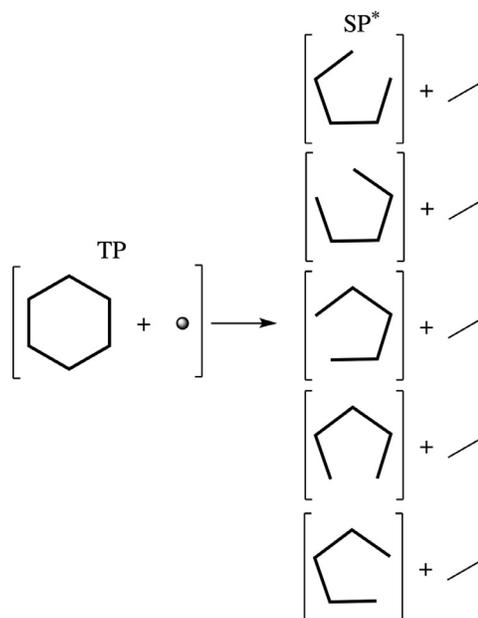
Fig. 16. Conversion of a 6-carbon ring to a 5-carbon ring: the 5-carbon cycle source pattern is replaced by 5 distinct linear patterns which can now be matched to the 6-carbon cycle target pattern by the iso-morphism algorithm.

are integrated in time, along with the species concentration equations, using the stiff ODE solver DVODE [37]. Conservation properties of labeled quantities, as described in Eqs. (20) and (21) are checked and satisfied at all times. Symmetry properties of the labeled fuel molecules are also properly recovered, with atoms structurally identical in the fuel being found in identical amounts in all molecules in the system at all time.

In the following, we present several practical examples displaying some of the new capabilities provided by the tracking approach described above. The first one illustrates in a simple case how the tracking algorithm complements conventional pathway analysis, while the others focus on the formation of soot precursors, especially large aromatic species such as naphthalene, and how it relates to the initial structure of the fuel molecules. Whenever necessary, specific notations for post-processing are introduced to simplify the analysis of the results. We conclude the section with a discussion of the limitations and challenges of the method.

To obtain a consistent picture, fixed conditions, summarized in Table 2, are used throughout the examples. Three different hydrocarbons are used as fuel: *n*-heptane (NC7), toluene (TOL), and methyl-cyclohexane (MCH). The fuel/oxygen mixture is initialized for all cases at a temperature $T = 1200$ K and a pressure of 1 bar. A fixed 3% molar carbon content and a molar C/O ratio of 2 are used, yielding slightly different equivalence ratios for each case, from 5.14 to 6.28. Except when specified otherwise, all simulations are run for a total time of 0.5 s.

As mentioned above, for each label tracked in the chemical kinetic network, an additional $\sum_{i=1}^{n^S} n_i^A$ equations need to be solved as well. For the chemical mechanism used in these examples, this represents a total of 1530 equations, in addition to the 236 equations needed to the species concentrations and temperature. While not prohibitive provided that the number of labels tracked per simulation remain small, the large number of time-evolving variables emphasize the need for targeted post-processing approaches, directly extracting relevant and meaningful information from the large volume of data generated. This process is illustrated in the examples below.
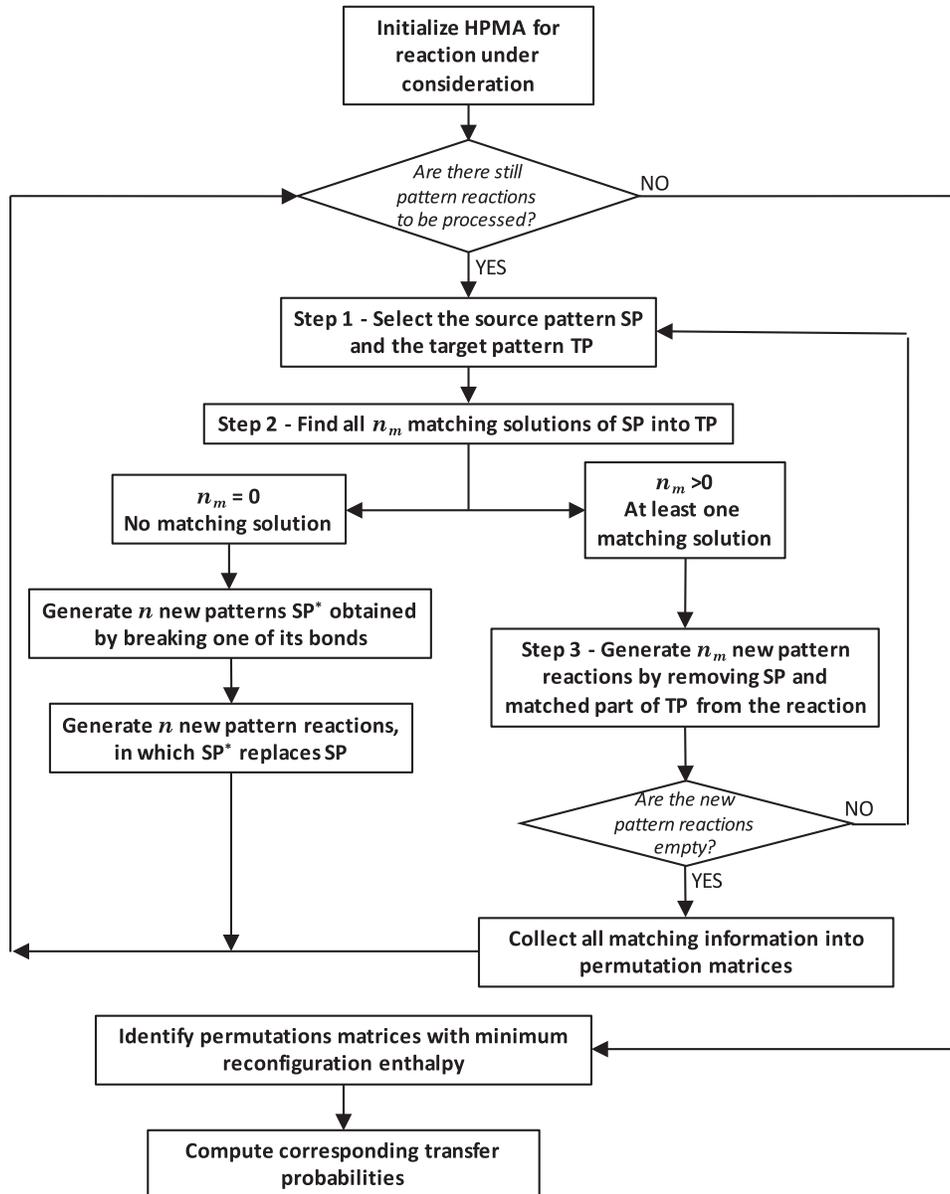
Fig. 17. *Constrained Hierarchical Pattern Matching Algorithm* flowchart.

Table 2
Parameters for the homogeneous, isochor reactor simulations used in Sections 5.2 to 5.3.

Simulation index		C1	C2	C3	C4
Fuel		NC7	TOL	MCH	NC7 (80%mol) TOL (20%mol)
Initial composition [% mol]	NC7	0.0043			0.0034
	TOL		0.0043		
	MCH			0.0043	
	O_2			0.0075	
	N_2			0.0098	
Equivalence ratio		6.28	6	5.14	6.05
Pressure				1 bar	
Temperature				1200 K	
Carbon molar content				3%	
C/O ratio				2	
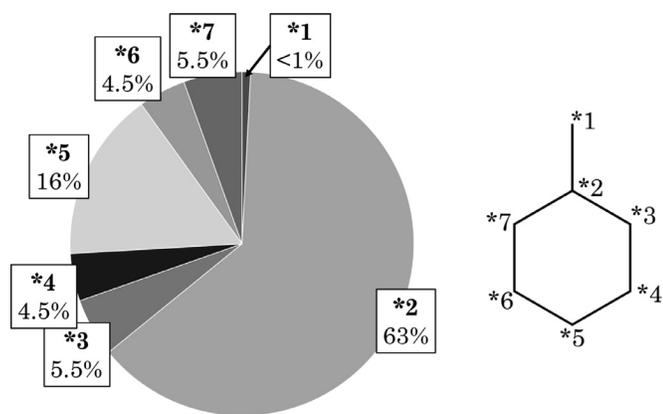
Fig. 18. Early CO formation during MCH rich oxidation. The chart displays the relative amount of each label in CO, showing that most of the carbon in CO originated from the tertiary carbon in MCH. The initial labeled fuel molecule, identical to Fig. 2(b), is reproduced on the right for reference.

5.1. Early CO production in methyl-cyclohexane rich oxidation

In the first example, we consider the early production of carbon monoxide, CO, during the rich oxidation of methyl-cyclohexane, with the objective to identify where, in the initial MCH molecule, the sole carbon in CO is coming from. Case 3 (Table 2) is run for a total time of 1μs, each atom in MCH carrying its own individual label, from *1 to *7. At the end of this very short simulation, the mole fraction of CO reaches 3.2×10^{-8}, and the relative amount of each initial label in CO is recorded. The results are shown in Fig. 18.

The fact that most CO in these early stages of oxidation originates from the tertiary carbon in MCH (labeled *2 in Fig. 18), while the methyl carbon (labeled *1) contributes very little, may seem counter-intuitive at first. However, it can be explained qualitatively, if not quantitatively, using reaction flux analysis to identify the main reaction pathway leading from the fuel to CO. CO is formed at more than 50% by the reaction:

$$T - C_3H_5 + O_2 \rightarrow CH_3 + CO + CH_2O \qquad (26)$$

In turn, $T - C_3H_5$ is formed mostly exclusively through successive β-scissions of the fuel radical obtained by H-abstraction from the tertiary carbon of MCH, named MCHR1 in the kinetic mechanism, as shown in Fig. 19:

A brief examination of Reaction 26 shows that the carbon in CO naturally comes from the secondary carbon in $T - C_3H_5$, which can be traced back directly to the tertiary carbon in MCH from Fig. 19. This direct pathway does explain the high percentage of *2 labels in CO. Interestingly, the H-abstraction to form MCHR1 only accounts for 7% of the consumption of MCH. MCH mostly decomposes to CH_3 and phenyl (63%), but at those rich conditions, the methyl radical primarily recomposes into larger molecules, such as C_2H_5 or C_2H_6, and is not readily oxidized, and therefore, does contribute very little to the formation of CO. This example has been chosen for its simplicity (the existence of a single major pathway) rather than its scientific relevance. Yet, it illustrates how a global "origin-destination" picture for the fuel molecule can generate new insights about the dynamics of a chemical system, which would otherwise be very complicated to identify or obtain via conventional kinetic network analysis.

5.2. Atom selectivity in combustion products

In direct continuation of the simple analysis above, our second example investigates how the initial structure of a fuel molecule affects the formation of combustion products, and in this specific case, the formation of naphthalene (A_2). This is done by assigning a distinct label to each atom in the initial fuel molecule. At any time during the simulation, the amount of each label in a species S_i can be used to quantitatively trace back the origin of the species all the way to the fuel structure. In practice, those amounts are directly provided by the variables α_i^{*k}, defined in Eq. (15), which are vectors containing the amount of label *k at the various skeleton atom locations in species S_i. Here, cases C1, C2, and C3 (Table 2) are run successively. We record the vector $\alpha_{A_2}^{*k}$ for each *k at the end of the simulations ($t = 500$ ms). Results are provided in Fig. 20.

Our objective here is to identify and quantify preferential links between the fuel structural features and the naphthalene produced during combustion. If there is no preferential link or contribution, the quantity of each label at every atom location in the product should be proportional to its initial amount. Since the fuel molecules selected here each contain 7 skeleton atoms (in this case, carbon only), α_{A_2} in the absence of preferential contribution would be:

$$\alpha_{A_2; j}^{*k} = \frac{1}{7} = 0.1428, \text{ or } 14.28\% \text{ for all } *k \text{ and } j. \qquad (27)$$

Figure 20 shows that for both n-heptane and methyl-cyclohexane, a close-to-equal contribution of all fuel atoms to each location in A_2 is obtained, very likely indicating that naphthalene is formed by the recombination of small hydrocarbon molecules indiscriminately of their origins. This means there is no preferential pathway between atoms in the fuel and atoms in the product. In the case of toluene, however, a strong variability is observed. For example, atoms at central location 1 in naphthalene are predominantly labeled *2 and *3, that is, coming from the tertiary carbon in toluene, and those adjacent to it in the aromatic cycle. This points towards the existence of a significant, more direct pathway between toluene and A_2, for instance, the addition of a C_3 species on the methyl branch of toluene to form the second ring, while pathways going through small hydrocarbon pieces contribute to a lesser extent.

To consolidate the information contained in the α values into one single, better conditioned quantity, we define the *selectivity* $\sigma_i^{\mathcal{F}}$ of species S_i with respect to the skeleton atoms in a fuel molecule $\mathcal{F}$ as:

$$\sigma_i^{\mathcal{F}} = \frac{1}{\sqrt{n_{\mathcal{F}}^A - 1} \, \frac{n_i^A C_i}{n_{\mathcal{F}}^A}} \left[\frac{1}{n_{\mathcal{F}}^A} \sum_{*k \in \mathcal{F}} \left(K_i^{*k} - \frac{n_i^A C_i}{n_{\mathcal{F}}^A} \right)^2 \right]^{1/2} \qquad (28)$$

where :

- K_i^{*k} is defined in Eq. (6),
- The term $\frac{n_i^A C_i}{n_{\mathcal{F}}^A}$ is the mean value of K_i^{*k} over all labels *k in the fuel molecule $\mathcal{F}$.
- The numerator corresponds to the standard deviation of the set of data $\{K_i^{*k}\}_{*k \in \mathcal{F}}$.
- $\sqrt{n_{\mathcal{F}}^A - 1}$ is a normalization term enforcing that $\sigma_i^{\mathcal{F}}$ be bounded between 0 and 1.

The selectivity $\sigma_i^{\mathcal{F}}$ can be interpreted as the relative standard deviation of the set of data $\{K_i^{*k}\}_{*k \in \mathcal{F}}$. In effect, it measures the *dispersion*, or equivalently, the inhomogeneity, of the contributions of each atom of a fuel $\mathcal{F}$ to a species of interest, S_i. Thus, a value $\sigma_i^{\mathcal{F}} = 1$ means that S_i is exclusively formed from one, and one only, atom in the fuel molecule $\mathcal{F}$. On the contrary, a value $\sigma_i^{\mathcal{F}} = 0$, means that all atoms initially in $\mathcal{F}$ contribute equally to the formation of S_i. This quantity can be used to identify the emergence of potential competing pathways in the kinetic network.
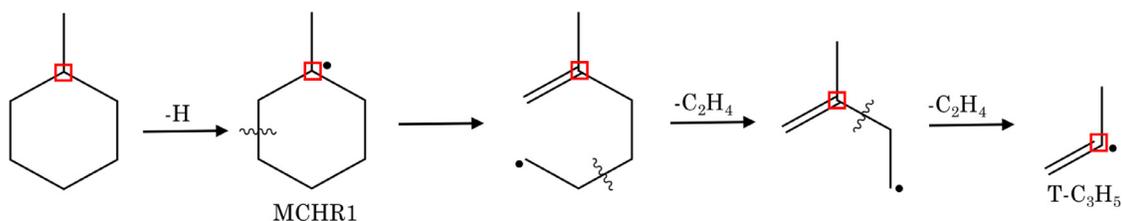
Fig. 19. Main pathway from MCH to T-C_3H_5. The location of what is initially the tertiary carbon in MCH has been highlighted throughout the reaction pathway. Note that by symmetry, β-scission in the second step can occur on both sides of the MCHR1 molecules, only one is shown for simplicity.
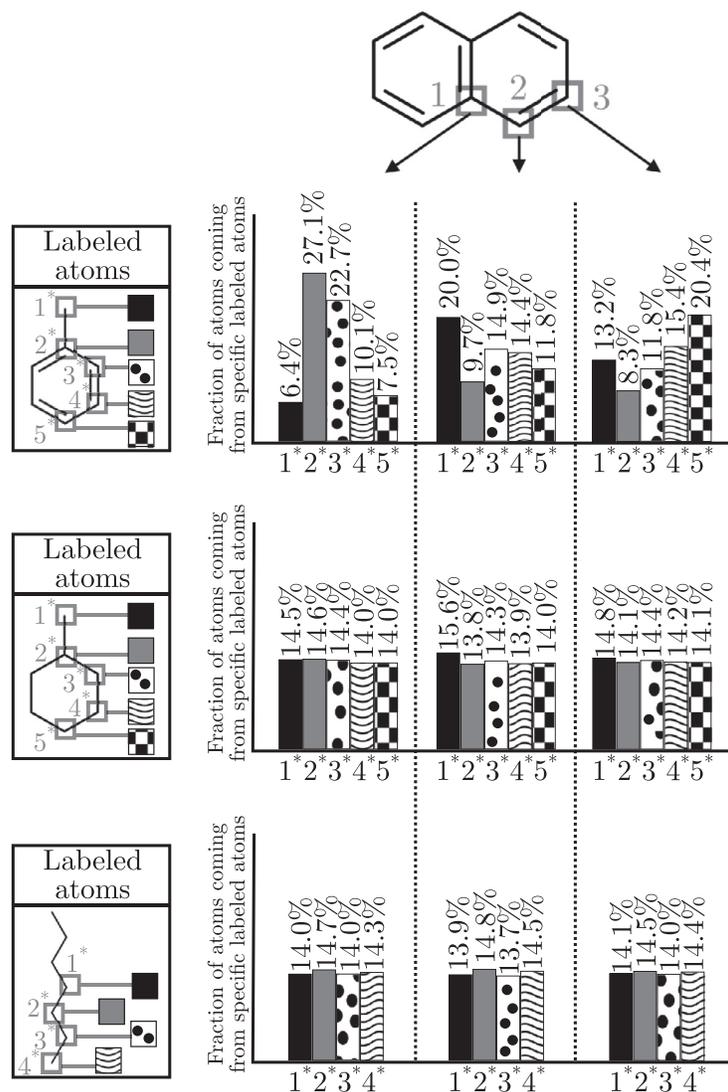
Fig. 20. Origin of atoms ($\alpha_{i,j}^{\cdot k}$) at different locations in naphthalene, for cases C1, C2, and C3 (Table 2), showing much larger variability for toluene than for heptane or methyl-cyclohexane. All atoms of each fuel molecule have been labeled individually (note that the non-labeled atoms in the fuels and A_2 would yield identical results as their symmetrical counterparts, and are therefore not tracked, and their results not shown), and all results are taken at $t = 500$ ms.

$\sigma_i^{\mathcal{F}}$ was computed at $t = 500$ ms for cases C_1, C_2, and C_3, each case considering a different fuel $\mathcal{F}$ (NC7, TOL, MCH). Here, i refers to the following set of intermediate or product species:

$$S_i \in \{CH_4, C_2H_2, C_2H_4, C_2H_6, C_3H_8, A_2\}. \tag{29}$$

Results are shown in Fig. 21. Consistent with the observation on naphthalene in Fig. 20, we observe that all species considered do exhibit preferential concentrations of certain atoms in the fuel, but to various extent. For example, toluene is associated with the highest selectivity, with a 15% value obtained in CH_4 and C_2H_4. The

intuitive explanation is that small hydrocarbon fragments are likely to come not from the aromatic ring, but rather from the loss of the methyl branch, or, when the ring opens up, from the extremities of the resulting linear molecule. Note that for heptane and methyl-cyclohexane, the product species can be split into two groups. The first one, containing CH_4, C_2H_6, and C_3H_8, shows higher selectivities than the other species, C_2H_2, C_2H_4, and A_2. This indicates that those species are obtained initially in significant amount through direct break-up of the fuel, retaining specificities from the fuel break-up pattern, even after a long residence time. In contrast, the
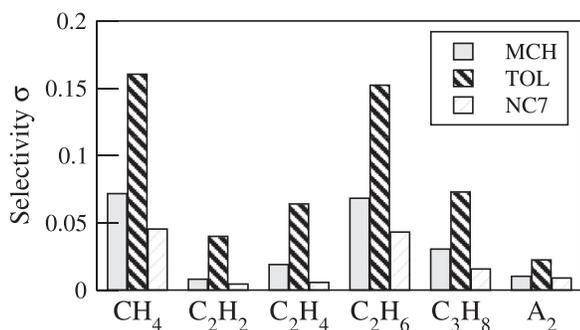
Fig. 21. Atom selectivity $\sigma_i^{\mathcal{F}}$ for the three fuel molecules considered in this study. A value of 0 indicates equal contribution of all atoms in the fuel to the species, while a value of 1 would indicate that only one atom in the fuel participates in the formation of the species.
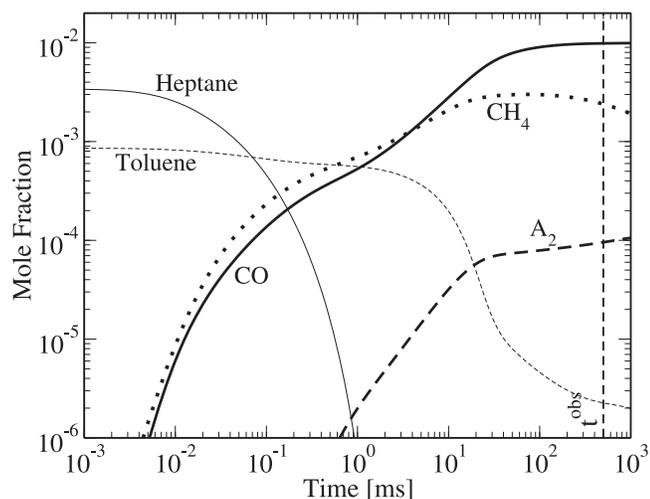
Fig. 22. Time evolution of some key species concentrations for case C_4. Note the logarithmic scale used for time, in order to better show the various characteristic times involved.

second group of species appears to be secondary species, whose formation involves a larger number of pathways, thereby mixing better all atoms available.

5.3. Atom tracking in multi-component mixtures

The tracking algorithm is then applied to a mixture of two hydrocarbon fuels (case C_4), containing heptane (80% in volume) and toluene (20% in volume). The time evolution of some key species concentrations is displayed in Fig. 22, showing two very different characteristic times for fuel consumption, heptane being consumed quickly (1 ms) and entirely, while the toluene mole fraction remains roughly constant, equal to its initial value, over a period 10 times as long. Due to the high equivalence ratio, significant amounts of hydrocarbon species, including toluene, remain after oxidation.

This configuration is used to investigate the contribution of individual fuel components in the formation of combustion intermediates and products. For this purpose, at $t = 0$, all atoms in toluene are labeled $\star$ (only one label being used, the integer is dropped for simplicity), while the atoms forming heptane and molecular oxygen are left untouched. This corresponds to the configuration depicted in Fig. 3. At $t = 500$ ms, the value $K_i^\star$ for each species S_i in the system is recorded, and normalized by the local concentration of skeleton atoms in S_i, yielding the fraction of $C^\star$ in each
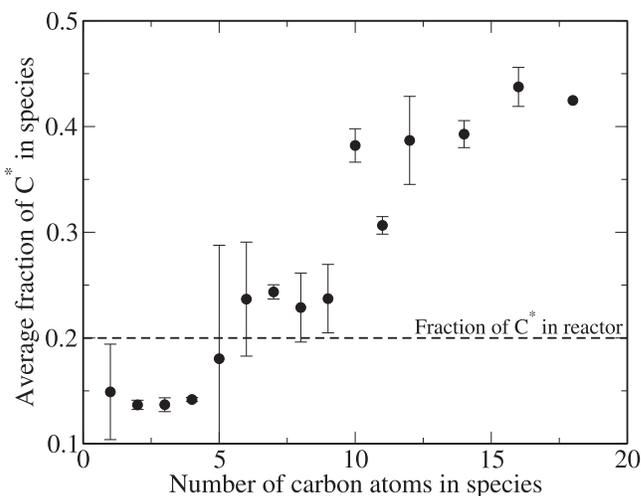
Fig. 23. Average fraction of $C^\star$ in species present at $t = 500$ ms for Case C_4, conditioned on the number of carbon in the species. Error bars indicate for each number of carbon atoms the minimum and maximum values observed. The dashed line indicates the average fraction of labeled atoms, that is, initially coming from toluene molecules, in the system.

species:

$$\beta_i = \frac{K_i^\star}{n_i^A C_i} \tag{30}$$

The values β_i, $i = 1$ to n^S, are then averaged conditioned on the the number of carbon in the corresponding species. The resulting conditional averages are plotted in Fig. 23.

A very clear trend can be observed, with β nearly monotonically increasing with the number of carbons in the species. Small hydrocarbon species (less than 5 carbons) are primarily formed by atoms initially contained in n-heptane, while large molecules contain a disproportionate amount of atoms originating from the initial toluene molecules. Large gas-phase species are precursors to soot particulates. Therefore, this result is very consistent with numerous experimental observations, for instance the work of Homan et al. [22], who showed that ring-forming carbons have a much higher probability to end up in soot in rich combustion.

5.4. Discussion

The examples above illustrate how the tracking methodology developed in this work can positively complement existing kinetic analysis and model development tools, especially in the realm of multi-component fuel combustion. There are however a number of assumptions that need to be kept in mind when interpreting and analyzing the results.

- First and foremost, the methodology is devised as non-intrusive. The labeled atoms can be seen as ideal isotopes, in the sense that they do not modify in any way the physical and chemical properties of their carrier species, such as heat capacities, enthalpies, or entropies, nor do they modify the rates of the reactions they are involved in. In contrast, experiments will be impacted by the presence of the isotopes, to varying degree depending on the actual isotopes. The underlying assumption in all experimental works cited in the introduction is that, especially for carbon isotopes, those effects are small. However, in the absence of thorough uncertainty quantification, comparisons between the present tracking algorithm results and experimental data should be kept qualitative in nature.
- The current tracking algorithm tracks individual atoms only, not bonds. Therefore, it does not provide statistics on whether or

not two atoms sharing a covalent bond in the fuel molecule remain bonded together in the products. This capability is left for future development of the method.

- CHPMA may fail to determine accurate transfer probabilities for reactions with non-trivial transition states, since this information is typically not provided in standard Chemkin mechanism files. Note that to add reliability to the automatic procedure, the algorithm can automatically identify reactions involving a large number of structural re-arrangements (which often correlates with complex transition states), leaving the possibility for the user to directly input the transfer probabilities manually. Transfer probabilities are also affected by the existence of isomerization reactions that may not be explicitly included in kinetic mechanisms, and as mentioned in Section 4.3, ambiguities in the structure of the molecules or the presence of global, rather than elementary, reactions. The resulting uncertainties are difficult to quantify at this stage. However, the two-tier molecular representation (skeleton and enhanced) and the conservative choice to use the skeleton representation as default whenever structural uncertainty exists, decrease the risk for "spurious selectivity", where pathways or products are wrongly attributed to specific structures of the fuel molecules. Any observed preferential behavior (*e.g.* toluene disproportionally forming high molecular weight products in the heptane/toluene mixture studied in Section 5.3) can be confidently considered qualitatively legitimate and significant.

6. Conclusion

A novel tracking algorithm has been developed and implemented in simple constant-volume homogeneous reactor simulations. The algorithm contains three main ingredients: a systematic and tailored treatment of the structural features of all species involved in the kinetic model, an enthalpy-based criterion to identify, in any given reaction, how atoms are transferred from reactants to products in a manner compatible with the expected transition states, and a pattern matching algorithm relying on graph isomorphisms to efficiently apply the enthalpy criterion. The method has been applied to several cases involving one or several fuels, where it demonstrated new capabilities compared to conventional kinetic mechanism analysis. While no direct comparison with experimental data was done, the results are qualitatively consistent with existing experimental work. The method is ideally suited to address some recurrent questions in combustion chemistry, including, but not limited to, how to best define surrogate fuels for experimental and numerical investigations of realistic combustion devices, as well as soot mitigation strategies using oxygenated molecules, which will be explored in future work.

References

[1] Chevron 2014 annual report, Technical Report, 2014. Chevron Corporation URL: https://www.chevron.com/annualreport/2014/
[2] A.K. Agarwal, Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines, Prog. Energy Combust. Sci. 33 (3) (2007) 233–271.
[3] A. Demirbas, Importance of biodiesel as transportation fuel, Energy Policy 35 (9) (2007) 4661–4670.
[4] A. Janssen, M. Jakob, M. Müther, S. Pischinger, Tailor-made fuels from biomass potential of biogenic fuels for reducing emissions, MTZ Worldw. 71 (2010) 54–60.
[5] G. Knothe, C.A. Sharp, T.W. Ryan, Exhaust emissions of biodiesel, petrodiesel, neat methyl esters, and alkanes in a new technology engine †, Energy Fuels 20 (1) (2006) 403–408.
[6] Y. Xu, I. Keresztes, A.M. Condo Jr., D. Phillips, P. Pepiot, T.C. Avedisian, Spherically symmetric droplet burning characteristics of algae-derived renewable diesel, conventional #2 diesel, and their mixtures, Fuel 167 (2016) 295–305.
[7] M.E. Myers Jr, J. Stollsteimer, A.M. Wims, Determination of gasoline octane numbers from chemical composition, Anal. Chem. 47 (13) (1975) 2301–2304.
[8] R.J. Gill, D.B. Olson, Estimation of soot thresholds for fuel mixtures, Combust. Sci. Technol. 40 (5–6) (2007) 307–315.
[9] C. McEnally, L. Pfefferle, Improved sooting tendency measurements for aromatic hydrocarbons and their implications for naphthalene formation pathways, Combust. Flame 148 (4) (2007) 210–222.
[10] C.S. McEnally, L.D. Pfefferle, Sooting tendencies of nonvolatile aromatic hydrocarbons, Proc. Combust. Inst. 32 (1) (2009) 673–679.
[11] C.S. McEnally, L.D. Pfefferle, Sooting tendencies of oxygenated hydrocarbons in laboratory-scale flames, Environ. Sci. Technol. 45 (6) (2011) 2498–2503.
[12] S.W. Benson, J.H. Buss, Additivity rules for the estimation of molecular properties. thermodynamic properties, J. Chem. Phys. 29 (3) (1958). 546–28
[13] S.W. Benson, Thermochemical kinetics, Wiley, 1976.
[14] Y. Xuan, G. Blanquart, Numerical modeling of sooting tendencies in a laminar co-flow diffusion flame, Combust. Flame 160 (2013) 1657–1666.
[15] P. Pepiot, H. Pitsch, R. Malhotra, S. Kirby, A. Boehman, Structural group analysis for soot reduction tendency of oxygenated fuels, Combust. Flame 154 (1–2) (2008) 191–205.
[16] C.K. Westbrook, W.J. Pitz, H.J. Curran, Chemical kinetic modeling study of the effects of oxygenated hydrocarbons on soot emissions from diesel engines †, J. Phys. Chem. A 110 (21) (2006) 6912–6922.
[17] K.G. Joback, R.C. Reid, Estimation of pure-component properties from group–contributions, Chem. Eng. Commun. 57 (1–6) (2007) 233–243.
[18] T. Turányi, Sensitivity analysis of complex kinetic systems. tools and applications, J. Math. Chem. 5 (3) (1990) 203–248, doi:10.1007/BF01166355.
[19] A.E. Lutz, R.J. Kee, J.A. Miller, SENKIN: A Fortran program for predicting homogeneous gas phase chemical kinetics with sensitivity analysis, Technical report SAND-87-8248, Sandia National Labs., Livermore, CA (USA), 1988.
[20] W. Sun, Z. Chen, X. Gou, Y. Ju, A path flux analysis method for the reduction of detailed chemical kinetic mechanisms, Combust. Flame 157 (7) (2010) 1298–1307, doi:10.1016/j.combustflame.2010.03.006.
[21] E.M. Fisher, W.J. Pitz, C.K. Westbrook, H.J. Curran, Detailed chemical kinetic mechanisms for combustion of oxygenated fuels, Proc. Combust. Inst. 28 (2) (2000) 1579–1586.
[22] H.S. Homan, W.K. Robbins, A carbon-14 tracer study of the relative fractions of various fuel carbons in soot, Combust. Flame 63 (1-2) (1986) 177–190.
[23] B.A. Buchholz, C.J. Mueller, A. Upatnieks, G.C. Martin, W.J. Pitz, C.K. Westbrook, Using carbon-14 isotope tracing to investigate molecular structure effects of the oxygenate dibutyl maleate on soot emissions from a DI diesel engine, SAE Trans.: J. Fuels Lubr. (2005) 1–15.
[24] A. Eveleigh, N. Ladommatos, R. Balachandran, A. Marca, Conversion of oxygenated and hydrocarbon molecules to particulate matter using stable isotopes as tracers, Combust. Flame 161 (11) (2014) 2966–2974.
[25] V.A. Bunev, Tracer method in numerical simulation of combustion processes, Combust. Explos. Shock Waves (2007) 1–9.
[26] V.A. Bunev, Numerical simulation of the effect of the addition of NO and NO2 on a rich hydrogen flame using the tracer method, Combust. Explos. Shock Waves (2009) 1–7.
[27] I. Mayer, On bond orders and valences in the ab initio quantum chemical theory, Int. J. Quant. Chem. 29 (1) (1986) 73–84, doi:10.1002/qua.560290108.
[28] A.C. Society, Cas registry, https://www.cas.org/content/chemical-substances.
[29] A. Dalby, J.G. Nourse, W.D. Hounshell, A.K.I. Gushurst, D.L. Grier, B.A. Leland, J. Laufer, Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited, J. Chem. Inf. Comput. Sci. 32 (3) (1992) 244–255.
[30] B. deB. Darwent, Bond dissociation Energies in simple molecules, Technical report NSRDS-NBS 31, National Bureau of Standards, 1970. United States Department of Commerce, 1970.
[31] L.P. Cordella, P. Foggia, C. Sansone, M. Vento, An improved algorithm for matching large graphs., Proceedings of the third IAPR-TC-15 International Workshop Graph-Based Representations, 2001, pp. 149–159.
[32] J.G. Siek, L.Q. Lee, A. Lumsdaine, The boost graph library: user guide and reference manual, Addison-Wesley Professional, 2001.
[33] G. Blanquart, P. Pepiot, H. Pitsch, Chemical mechanism for high temperature combustion of engine relevant fuels with emphasis on soot precursors, Combust. Flame 156 (2009) 588–607.
[34] K. Narayanaswamy, G. Blanquart, H. Pitsch, A consistent chemical mechanism for oxidation of substituted aromatic species, Combust. Flame 157 (2010) 1879–1898.
[35] K. Narayanaswamy, H. Pitsch, P. Pepiot, A chemical mechanism for low to high temperature oxidation of methylcyclohexane as a component of transportation fuel surrogates, Combust. Flame 162 (2015) 1193–1213.
[36] K. Narayanaswamy, P. Pepiot, H. Pitsch, A chemical mechanism for low to high temperature oxidation of n-dodecane as a component of transportation fuel surrogates, Combust. Flame 161 (2014) 866–884.
[37] P.N. Brown, G.D. Byrne, A.C. Hindmarsh, VODE: A Variable Coefficient ODE Solver, SIAM J. Sci. Stat. Comput. 10 (1989) 1038.