

New approaches for representing, analyzing and visualizing complex kinetic transformations

Ioannis P. Androulakis*

*Biomedical Engineering Department, and Chemical & Biochemical Engineering Department, Rutgers,
The State University of New Jersey, Piscataway, NJ 08854, USA*

Received 12 December 2005; received in revised form 16 May 2006; accepted 16 May 2006

Available online 22 August 2006

Abstract

Complex kinetic mechanisms involving thousands of reacting species and tens of thousands of reactions are currently required for the rational analysis of modern combustion systems. In order to represent, analyze and visualize effectively the ignition process, advanced computational techniques will be required. Recently, we introduced a novel approach that captured the principal elemental transformations in complex reaction mechanisms in the form of graphs. In this work, we propose new approaches in order to arrive at a compact representation of the information content of these graphs utilizing machine learning principles, such as feature selection in time series and hashing. These approaches allow the projection of the totality of the information contained in the graphs describing the chemical transformations onto a single scalar. The temporally evolving graphs are treated as streaming data and locality-preserving hashing allows the unique assignment of a scalar “motif” value to each such graph. Analysis of those motifs allows the quick identification of “clusters” of identical reaction graphs that correspond to regimes with similar kinetic characteristics. The approach is illustrated with highly complex kinetic mechanisms describing pentane autoignition. It is demonstrated how this novel representation allows the identification of regions where similar temporal history of the chemical transformations is experienced. © 2006 Elsevier Ltd. All rights reserved.

Keywords: Complex kinetic mechanisms; Combustion; Visualization

1. The complexities of combustion

Hydrocarbon oxidation, one of the leading chemical transformations in transportation applications, is a process easily stated, yet very complicated to understand in detail and even more intricate to properly model (Dean, 2001). Macroscopic models of combustion, such as the so-called Shell model, have been used for the longest time in order to predict the rate of heat release or the rate of consumption/production of the main fuel/product molecules (Crua, Kennaird, Sazhin, & Heikal, 2004; Halstead, Kirch, & Quinn, 1997). High temperature hydrocarbon oxidation can be succinctly modelled with a relatively small number of reactions indicative of rapid fuel decomposition, provided that one is mainly interested in major oxidation products. However, due to environmental concerns combustion models are revisited to address critical issues such as predictive response over

extended ranges of operation, in particular ignition at relatively low initial temperatures, and increased accuracy requirements at the ppm level. It has been realized that macroscopic characteristics are inadequate and therefore a detailed description of combustion chemistry is needed (Gardiner, 1996). A distinguishing feature of the combustion of many hydrocarbons is the “negative temperature coefficient”, NTC (Pilling & Hancock, 1997). At very low temperatures, as the temperature is raised the ignition delay of a fuel is decreased. However, for many fuels, at a certain temperature the ignition delay reverses its course and becomes longer as the temperature is raised. At some higher temperature the ignition delay decreases again. In order to capture all the intricate details of fuel-like molecules (Griffiths, 1995) mechanisms of significant size are required, potentially involving thousands of species and tens of thousands of reactions (Curran, Pitz, Westbrook, Callahan, & Dryer, 1998). One of the key complications is that different temperature regimes are known to activate different reaction pathways. Multi-point ignition in homogeneous charge compression ignition engines is a great manifestation of the impact of chemistry induced due to

* Tel.: +1 732 445 0099.

E-mail address: yannis@rci.rutgers.edu.

spatial inhomogeneities in composition, temperature and pressure. The effects of non-homogeneous fuel–air mixtures are key to understanding the nature of multi-point ignition and further conceptualize potential embodiments that enhance or control the simultaneous ignition of the mixture (Echekki & Chen, 2003).

In this paper, we discuss a novel representation of the temporal evolution of complex kinetic transformations. We extend our previous work (Androulakis, Grenda, & Bozzeli, 2004) and demonstrate how we can construct graphs representing elemental transformations over time. The graphs can be perceived as snapshots of the ensemble of chemical transformations taking place during the course of the reaction. Each reacting condition is uniquely defined by one such graph. Our present work addresses the question concerning the representation of such graphs and the possibility of identifying structural differences between them. These structural differences would indicate distinct reaction paths that are activated during the course of a reactive flow simulation. We propose a novel way of capturing the essential information of these graphs by making use of recent advances in the area of proximity-preserving hashing. We demonstrate the approach by analyzing the complex kinetics of low temperature oxidation of hydrocarbons. We will first review the concept of time-integrated element flux analysis, we then present the basic elements of the hashing procedure and then we summarize by analyzing the autoignition of *n*-pentane.

2. Element flux diagrams of complex kinetic mechanisms

The concept of element flux analysis, introduced by Revel, Boettner, Cathonnet, and Bachman (1994) allows the development of time dependent flux diagrams that identify key reaction pathways with minimal effort. The main idea is to study principal reaction pathways along the reaction process coordinate. The atomic fluxes for each atom (C, H, O, N, Cl and S) at different reaction times are calculated, based on reaction rates, and the major sources and sinks for each element are identified. According to Revel et al. (1994), the instantaneous flux of element A from species *j* to species *k* through reaction *i* is defined as:

$$\dot{A}_{ijk} = q_i \frac{n_{A,j} n_{A,k}}{N_{A,i}} \quad (1)$$

where q_i is the net production rate of reaction *i* [mol s^{-1}], $n_{A,j}$ the number of atoms A in species *j*, $n_{A,k}$ the number of atoms A in species *k* and $N_{A,i}$ is the total number of atoms of element A in reaction *i*. The total transfer of element mass for any pair of species accounting for all reactions (*i*) that these species participate as reactants/products, as a function of time *t*, can then be defined as:

$$\bar{A}_{\text{FROM,TO}}(t) = \sum_{i=1}^{N_R} A_{i,\text{FROM,TO}}(t) \quad (2)$$

The fluxes defined in Eq. (2) are functions of time. Therefore, one repeats the analysis at pre-selected time points during the reaction process. This is equivalent to taking snapshots during the reaction and trying to identify, in time, active sources and

sinks. These flow-charts can then be used to identify reaction pathways as they develop over time. The approach as described makes it cumbersome to derive global information. The fluxes have to be constructed at selected points in time. However, the time points at which these computations are performed are not always evident.

Androulakis et al. (2004) introduced the concept of a time-integrated flux indicator. The main idea is to derive (over time) an indicator for a source–sink combination based on integration of the quantity defined in Eq. (2). The quantity will be subsequently normalized as a means of representing the results. Therefore, we define the time-integrated element flux as:

$$\hat{A}_{\text{FROM,TO}} = \frac{\int_{t=0}^{\tau} \bar{A}_{\text{FROM,TO}}(t) dt}{\sum_{\text{FROM}'} \sum_{\text{TO}'} \int_{t=0}^{\tau} \bar{A}_{\text{FROM}',\text{TO}'}(t) dt} \quad (3)$$

The integrals in Eq. (6) are estimated numerically by accounting for all the time instances as the integrator is generating them. In this way, the entire reaction trajectory is taken into account, and there is no need to a priori select the location of the snapshots. The quantity defined in Banerjee and Ierapetritou (2003) allows one to assign a unique, overall, number to each source–sink pair, which is representative of the entire reaction period. The normalization defined in Banerjee and Ierapetritou (2003) insures that pathways activated at different points in time receive appropriate weighting.

Time-integrated element flux analysis allows one to characterize the main transformations that take place during the reaction period. The analysis reveals key pathways in terms of source–sink relationships. However, this analysis does not reveal any information related to the way these transformations take place. The purpose of the time-integrated element flux analysis is to establish “global” insight into the reaction pathways. It allows for a comprehensive and macroscopic view of how reactants are preferentially converted to intermediates and finally to products. By removing the temporal dependence it makes it possible to establish these types of relations. Once the skeletal mechanism has been derived a more detailed rate or sensitivity analysis could determine the most significant contributors to specific transformations. Time-integrated analysis corroborates the interpretations derived from instantaneous rate and sensitivity analysis. However, the time-integrated element flux analysis has proven to be a very powerful method for identifying skeletal mechanisms with excellent predictive capabilities and it generates significant chemical insight in very complex kinetic mechanisms. An extensive presentation of the method and its wide range of applicability was recently discussed in Androulakis et al. (2004). The essence of the approach is the generation of temporal, as well as time-integrated graphs that quantify the elemental flux from source to sink species.

3. A novel representation of temporal reaction flux diagrams in complex reaction systems

As explained in greater detail in Androulakis et al. (2004) the essence of the time-integrated element flux analysis is to provide a global view of the aggregate chemical transforma-

tions characterizing a particular reaction system. For a selected element, the algorithm traces the transformations between all reactive species and source–sink pairs are defined as a function of time. Hence, we construct a complicated graph with nodes the reactive species and arcs weighted proportionally to the rate of element flux from a source to a sink. For “simple” fuels (such as methane) or for conditions that quickly lead to equilibration (high temperature oxidation), simple kinetic representations are adequate and pathway activation is not such a critical issue. The main question is basically how far from equilibrium is the system. However, larger hydrocarbons, exhibit far more complicated oxidation characteristics. During the process of combustion alternative pathways are activated each characterized by its own distinct dynamics. Therefore, analysis of low temperature ignition/oxidation for example reveals the subsequent activation of a number of reaction pathways each described by a distinct and unique sets of characteristic pathways. Different temperature regimes are known to activate different reaction pathways. Multi-point ignition in homogeneous charge compression ignition engines is a great manifestation of the impact of chemistry induced due to spatial inhomogeneities in composition, temperature and pressure. The effects of non-homogeneous fuel–air mixtures are key to understanding the nature of multi-point ignition and further conceptualize potential embodiments that enhance

or control the simultaneous ignition of the mixture (Echekki & Chen, 2003). The active reaction pathways evolve in space and time giving rise to fundamentally different chemistries. In Androulakis et al. (2004) time-integrated flux analysis was used to decipher the intrinsic details of a complex low temperature oxidation mechanism for *n*-pentane. We demonstrated that over time, different pathways are activated depending on the initial temperature of the reactive mixture thus activating fundamentally different reaction pathways. These are succinctly summarized in Fig. 1 where the time-integrated flux diagrams are depicted under different conditions. It must be realized that different initial temperatures give rise to distinct pathways. An adiabatic plug flow reactor is used to simulated ignition of a stoichiometric mixture with the initial temperature used as a parameter. At high initial temperatures ($T_0 = 1000$ K) the fuel decomposition pathways (β -scission) dominate and the fuel readily breaks down. At low initial temperatures ($T_0 = 700$ K) two pathways involving the two successive O_2 addition steps become dominant as indicated by the activation of the pathways involving the first oxygen addition ($C_5H_{11}O_{2-2}$ and $C_5H_{11}O_{2-1}$), followed by the isomerization reactions ($C_5H_{10}OOH_{2-4}$ and $C_5H_{10}OOH_{1-3}$), followed by the second oxygen addition ($C_5H_{10}OOH_{2-4}O_2$ and $C_5H_{10}OOH_{1-3}O_2$). At intermediate temperatures ($T_0 = 850$ K) we have a combination of rapid fuel decomposition (first outlet),

$T = 1000$ K							
OUTLET	###	C5H11-2	[49.93%]				
	→	C5H11-2	→	C3H6	→	C3H5-A	→
	→	C2H3CHO	→	C2H3CO	→	C2H3	→
	→	HCCO	→	CO	→	CO2	→
OUTLET	###	C5H11-1	[25.15%]				
	→	C5H11-1	→	NC3H7	→	C2H4	→
	→	C2H2	→	HCCO	→	CO	→
OUTLET	###	C5H11-3	[24.93%]				
	→	C5H11-3	→	C4H8-1	→	C4H7	→
	→	C2H3CHO	→	C2H3CO	→	C2H3	→
	→	HCCO	→	CO	→	CO2	→
$T = 700$ K							
OUTLET	###	C5H11-2	[49.74%]				
	→	C5H11-2	→	C5H11O2-2	→	C5H10OOH2-4	→
	→	NC5KET24	→	CH3COCH2	→	CH3COCH2O2	→
	→	CH3COCH2O	→	CH3CO	→	CO	→
OUTLET	###	C5H11-1	[25.43%]				
	→	C5H11-1	→	C5H11O2-1	→	C5H10OOH1-3	→
	→	NC5KET13	→	C2H5CHO	→	C2H5CO	→
	→	C2H5O	→	CH3CHO	→	CH3CO	→
	→	CO2	→				→
OUTLET	###	C5H11-3	[24.83%]				
	→	C5H11-3	→	C5H11O2-3	→	C5H10OOH3-2	→
	→	C5H9	→	C3H6	→	C3H5-A	→
	→	C2H3CHO	→	C2H3CO	→	C2H3	→
	→	HCCO	→	CO	→	CO2	→
$T = 850$ K							
OUTLET	###	C5H11-2	[49.92%]				
	→	C5H11-2	→	C3H6	→	C3H5-A	→
	→	C2H3CHO	→	C2H3CO	→	C2H3	→
	→	HCCO	→	CO	→	CO2	→
OUTLET	###	C5H11-1	[25.23%]				
	→	C5H11-1	→	C5H11O2-1	→	C5H10OOH1-3	→
	→	C4H8-1	→	C4H7	→	C4H7O	→
	→	C2H3CO	→	C2H3	→	C2H2	→
	→	CO	→	CO2	→		→
OUTLET	###	C5H11-3	[24.84%]				
	→	C5H11-3	→	C5H11O2-3	→	C5H10OOH3-2	→
	→	C5H9	→	C3H6	→	C3H5-A	→
	→	C2H3CHO	→	C2H3CO	→	C2H3	→
	→	HCCO	→	CO	→	CO2	→

Fig. 1. Characteristic oxidation pathways for *n*-pentane (Androulakis et al., 2004).

and only the first oxygen addition (outlets 2 and 3) dominating with subsequent activation on HO₂-producing NTC pathways.

It must be emphasized that the process of identifying these graphs is in fact quite versatile and in Farrell, Johnston, and Androulakis (2004) it was demonstrated how to assess the evolution of element flux graphs in more complicated reactive simulations, such as freely propagating flames.

3.1. Temporal evolution of element flux graphs

The time-integrated element flux analysis identifies key transformations controlling the reaction process. Fig. 2 summarizes the ensemble of the most critical reaction pathways, in the form of a graph, during the oxidation of pentane. The pathways depicted in this figure correspond to the most significant species transformations as identified through the time-integrated element flux analysis (Androulakis et al., 2004) and correspond to the steps identified in Fig. 1. The coupling of the time-integrated element flux analysis and graph visualization software (Gasner, Koutsofios, & North, 2002) allow the visual inspection of the reaction pathways. If visualized in a dynamic mode the graph

changes the arc weights as a function of time and depending on the pathways that are active. The oxygen addition reaction to the fuel radicals is indicative of low temperature chemistry, NTC pathways become significant as T increases and eventually, at high T , the direct fuel decomposition reactions dominate. Each sub-graph contains reactions that become active and subsequently deactivate as the reaction progresses. Therefore, different arcs in this graph superstructure are present during the course of the simulation.

Therefore, being able to visualize the evolution of the graphs is important. The obvious reason is so that we can begin to understand the temporal evolution of the chemistry. However, the most important reason is to understand the relationship between the onset of particular pathways and reaction conditions. Graph representation of species connectivity is of course a well-established concept (Temkin, Zeigarnik, & Bonchev, 1996). The novelty of our approach is that we do not simply define a static graph representation of a chemical mechanism but rather the dynamic evolution of the reaction pathways, thus enabling the monitoring of the activation/deactivation of critical pathways as a function of time. This graph representation

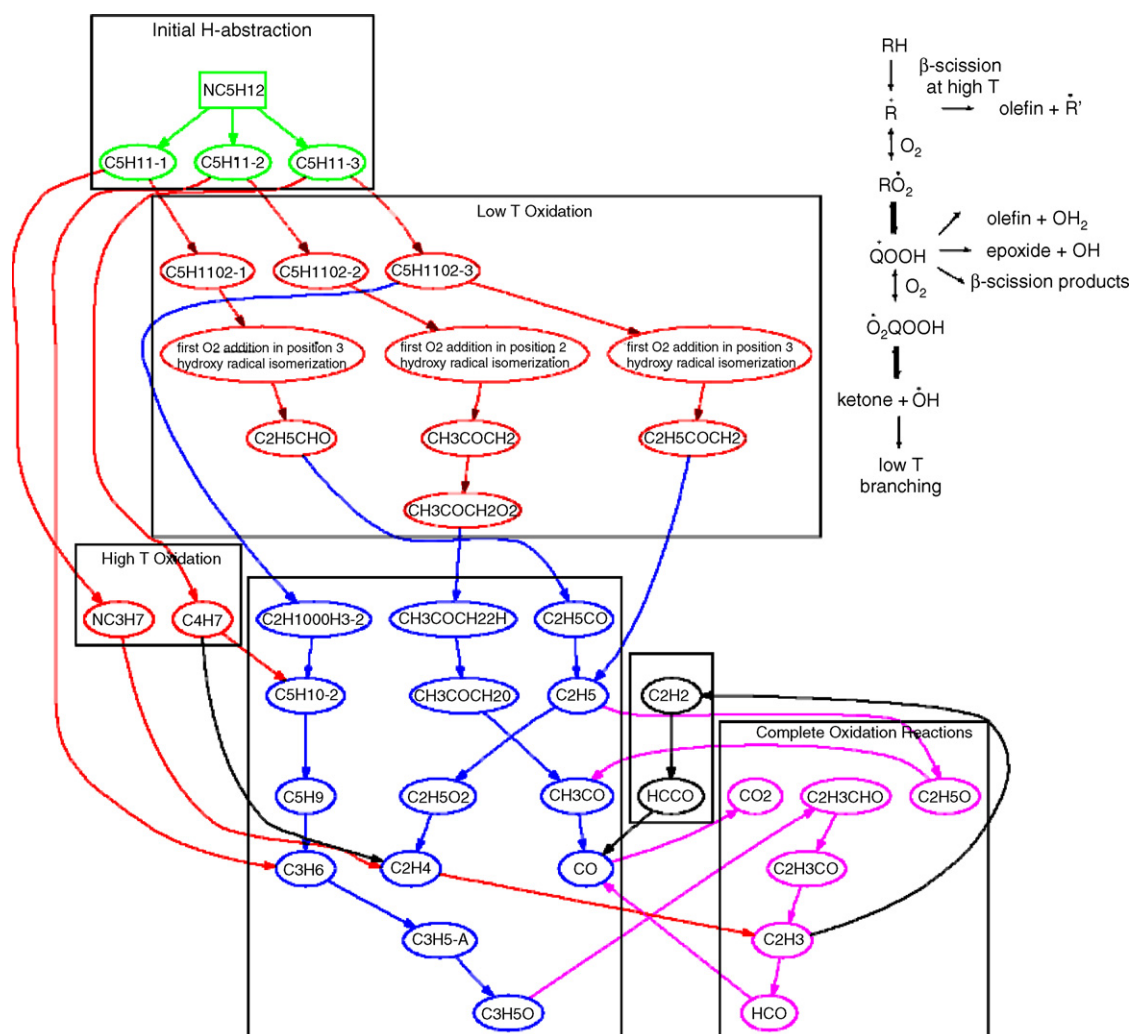


Fig. 2. *n*-Pentane oxidation flux graphs.

of the evolution of the kinetic transformation is a very powerful representation in studying the spatio-temporal sensitivities of a reaction process. Ultimately, we wish to identify the impact of local conditions (flow field, geometry) on the chemistry. Having a succinct representation of the chemistry is a very critical step. Although the flux diagrams, the way we are constructing them, do capture such dependencies, they are still too cumbersome to work with. What is therefore needed is a novel way of capturing the totality of the graph representation information, via the definition of some novel informative attribute, in order to enable the analysis of spatio-temporal emerging patterns. Visualizing graph evolution and finding frequent patterns in graphs is an area that is attracting significant interest (Abello & Korn, 2002; Brandes, Dwyer, & Schreiber, 2004; Dogrusoz, Feng, Madden, Doorley, & Frick, 2002; Gorg, Birke, Pohl, & Diehl, 2004; Kuramochi & Karypis, 2004). However, in these cases the structure and connectivity of the graph is analyzed. In studying kinetics the connectivity is always the same (superset of all physically possible source–sink species based on the mechanism) but the strength of the connectivity (weight of an arc which is “proportional” to the elemental transformation from one species to another) is evolving over time. Therefore, in order to capture the information content of the graph we need to model the spatio-temporal evolution of the connecting weights. The generated graphs can potentially have an enormous number of arcs denoting mass transformations between species. In this paper we propose a novel graph representations based on recent advances in the area of proximity-preserving hashing.

3.2. Proximity-preserving hashing for graph visualization

In order to capture the information content of the graph we need to model the spatio-temporal evolution of the connecting weights. The generated graphs can potentially have an enormous number of arcs denoting mass transformations between species. We have previously illustrated (Androulakis et al., 2004) how to identify a, still large, core of important arcs for monitoring. Therefore, at every point in space and time, during the course of a calculation, the element flux analysis will generate a number of arcs with their corresponding weights. The issue therefore becomes how to “compare” large number of such “graphs” efficiently and effectively. For a mechanism with N species the maximum number of graph entries is $N \times N$. Each graph, evaluated at time t , can be easily transformed from a matrix to a vector representation:

$$\begin{aligned} G[(\text{FROM} - 1) \times N + \text{TO}, t] \\ = \text{arc}(\text{FROM}, \text{TO}, t) = \bar{A}_{\text{FROM}, \text{TO}}(t) \end{aligned} \quad (4)$$

Therefore, each graph in time is assigned a point in a multidimensional space. Monitoring the graphs in time corresponds to the problem of analyzing multidimensional time-evolving trajectories. We view this multidimensional representation as a “pseudo time series” evolving in time. The basic question therefore becomes how to evaluate the characteristics of evolving series, how to capture their essential structure, how to identify critical events denoting transitions and how to effectively com-

pare them. In order to analyze the equivalent time series we basically follow the formalism of Keogh, Lin, and Fu (2005).

This approach (see also short discussion in Appendix A) assigns a set of signals to one of α^t bins where α is the number of discretizations per time point, and t is the number of time points, and it does it in a manner where similarly shaped signals are assigned to the same hash value, thereby essentially clustering the data into α^t categories. In order to assign the signals by shape, the signal is first normalized via the z -score given as:

$$Y_t = \frac{X_t - \mu}{\sigma}$$

$X = \{X_t, t = 1, 2, \dots\}$ denotes the raw representation of a times series, μ the time averaged value of each time series μ and σ is the corresponding standard deviation of the corresponding values of the time series. $Y = \{Y_t, t = 1, 2, \dots\}$ denotes the z -transformed representation of the series. The consequence of this transformation is two-fold, the first consequence is that this scales all of the signals to the same scale, thereby negating the effect scale upon the hash values, and secondly confers a Gaussian distribution to the values of each time point among the entire set of signals.

By transforming the signal by the z -score, and changing the distribution of values to a Gaussian distribution, we can assign symbols, $c_j \in \{a, b, c, \dots\}$, in an equi-probable manner by utilizing Gaussian breakpoints instead of uniform breakpoints. This is done essentially so that signals generated via Gaussian random noise will hash roughly uniformly within our hash space. Since the signals are normalized to $\mathcal{N}(0, 1)$ by the z -score, the Gaussian breakpoints can be obtained through tables denoting the PDF of the Gaussian, present in standard statistics textbooks.

The primary complication of this method is exponential expansion of the number of bins for every additional time point in the time series. In order to rectify piecewise averaging is utilized in order to keep the number of bins numerically tractable. The concept of piecewise averaging is similar to the concept of sub sampling, in which a signal which is normally sampled at a sampling rate of, say, S Hz is converted to a signal which is sampled at S/w Hz, and therefore we have a single value representing multiple time points. However, instead of merely dropping points we average every w time points. Therefore, the equation for piecewise averaging is given as:

$$\begin{aligned} \bar{Y}_k &= \frac{1}{w} \sum_{w(t-1)+1}^{wt} Y_t; \quad k : \text{sub-sampled time vector;} \\ t : \text{original time vector;} \quad w &= \frac{t}{k} \end{aligned} \quad (5)$$

This cuts down on the exponential expansion to $\alpha^{\text{ceil}(t/w)}$ bins, thus making it numerically tractable for longer time series. Therefore, after piecewise averaging, the signal is converted to symbols via Gaussian breakpoints.

What we have accomplished at this point is the conversion of a time signal into a set of equi-probable symbols. To convert this set of signals into a single number, we will utilize the process of converting a base α number into base 10. To do this we will

utilize the hash function proposed in Patel, Keogh, Lin, and Lonardi (2002):

$$\text{hash}(G, w, a) = 1 + \sum_{j=1}^{t/w} [\text{ord}(c_j) - 1] \times a^j \quad (6)$$

G is the vector representation of the flux graphs. This formula, then converts the symbolic representation of a signal, which is essentially a sub-sampled discretized version of the original to a base 10 hash value, which then essentially groups similar shaped signals into bins represented by the hash value. Conceptually, we propose to use the “unique” integers (hash values) as a surrogate of a new state variable, much like a reaction coordinate, indicative of the reactive state of the system. This variable captures the essential characteristics of the information contained in the detailed element flux graphs.

3.3. Analysis of a complex pentane oxidation mechanism

The representation is demonstrated with a highly complicated pentane oxidation mechanism with over 350 reacting species and over a thousand reactions based on detailed kinetic mechanisms used for the analysis of automotive fuels (Curran et al., 1998; Tenenbaum, de Silva, & Langford, 2000). The mechanism has been designed to properly account for all three regimes of hydrocarbon oxidation, namely, low and high temperature oxidation as well as the negative temperature coefficient regime. The mechanism captures the nonlinear dependence of autoignition delay as function of the initial temperature (Fig. 3). Autoignition delay is defined in the context of a rapid compression machine experiment (Westbrook, Pitz, Boercker, Curran, & Griffiths, 2003).

Computationally we define this as the time needed to achieve a certain level of oxygen or fuel conversion in a premixed homogeneous reactor. Low temperature oxidation is characterized by two distinct ignition points, a low and a high temperature (main) ignition.

Three initial conditions for stoichiometric n -pentane oxidation in air ($T_0 = 715, 840$ and 960 K) are selected for analysis because they all generate the same main ignition delay (macroscopic observable) but follow different induction pathways. The main ignition delay is a critical macroscopic observable because it really defines the core of the combustion process. Key reaction pathways for low, intermediate and high temperature reaction conditions were illustrated in Fig. 1. For each case we generate a sequence of time-dependent element flux graph, and the corresponding series representations, and use the arc weights to identify the temporal evolution of the hash values assigned to each graph.

The reduced representation using a single scalar correctly identifies that “different” induction chemistries are active in these three cases. We use the motif values to represent the totality of the information contained in the element flux diagrams. As seen in Fig. 4, the induction chemistry is quite different for an initial temperature of 840 K as opposed to an initial temperature of 715 K. The first ignition events occur at different time points for the 715 and 840 K initial temperatures. Different pathways are activated (per Fig. 2). Thus, these early ignition events “hash” to different motif values. The main ignition starts at a later time for both the initial conditions. When the initial temperature for the same stoichiometric mixture is at 960 K only the main ignition event takes place. We notice that because the underlying chemistry of high temperature (main) oxidation, irrespective of fuel

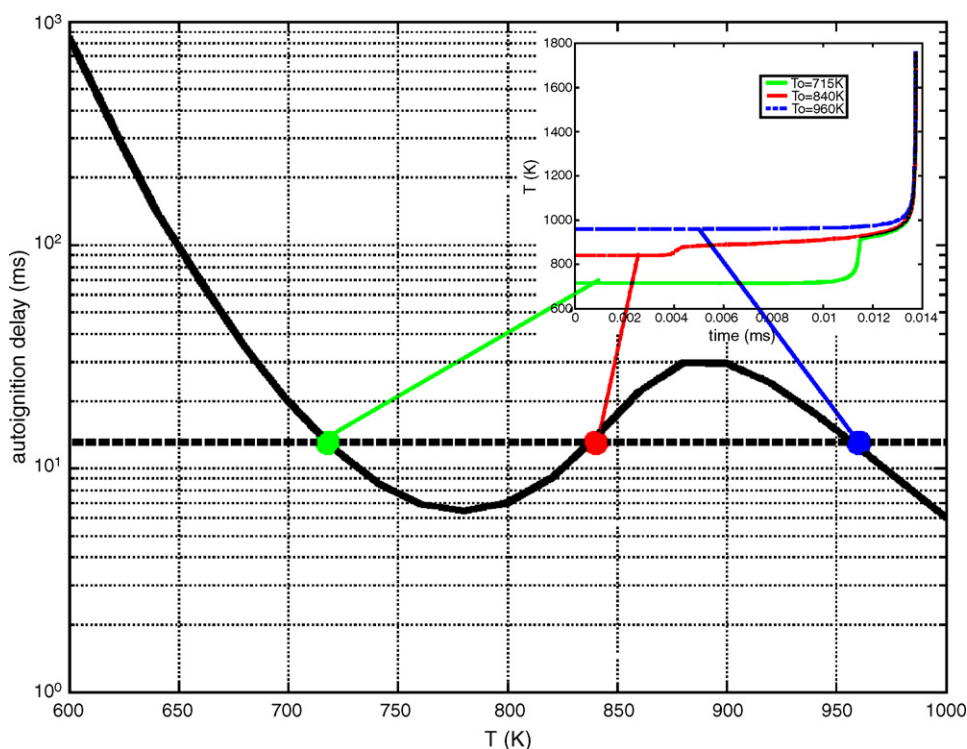


Fig. 3. n -Pentane autoignition delay.

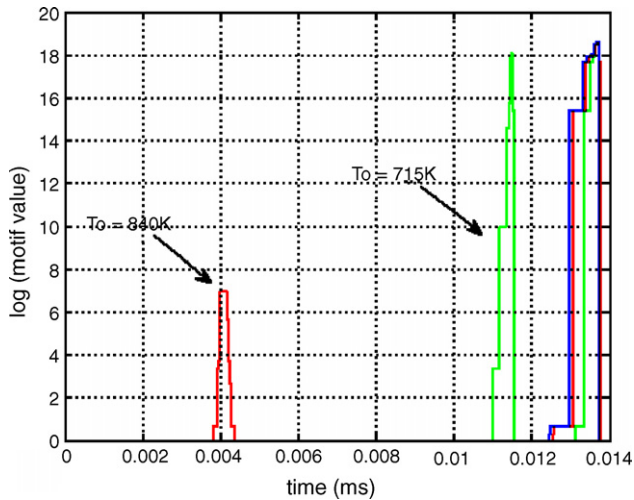
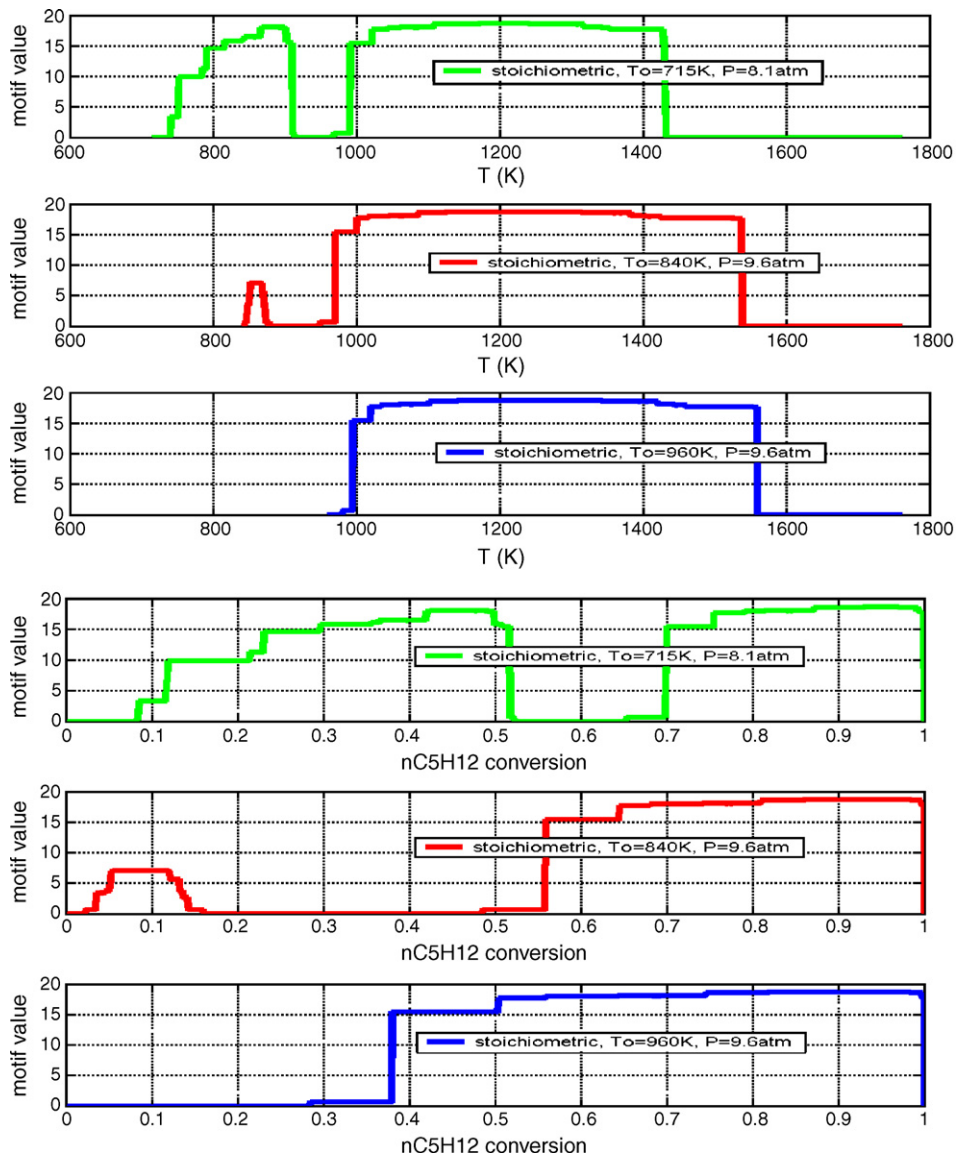


Fig. 4. Temporal evolution of motif values.

conversion is the same (in terms of active pathways, per Fig. 2), the motif values capturing the structure of the underlying flux graphs are identical only slightly shifted in time. This novel representation of the chemistry has addressed a critical issue: how to represent and compare in a very compact manner the evolution of chemistry in order to identify regimes that follow similar evolution in terms of the active reaction pathways.

Unlike previous approaches we monitor in space and in time not simply individual species (Koezler, 2001) but a metric that captures the chemistry history. It is important to realize that monitoring species concentrations would have generated a very different, and potentially incorrect, picture. As seen from the results of Fig. 5 the fuel conversion levels are quite different when the same type of “high temperature chemistry” is reached in each case. By monitoring species profiles one may have reached the conclusion that because different local conditions gave rise to ignition they may correspond to distinct mecha-

Fig. 5. Motif values (log) vs. T and fuel conversion.

nisms. In reality the local conditions are such that the same chemistry, as defined by the important reaction pathways, is activated. Eventually all initial conditions converge to the same main ignition pathways. This is independent of the level of concentration of any radical, molecular species or even temperature as can be seen in Fig. 5 when comparing the 715 and the 840 K initial temperatures. Similarly, we conclude correctly that distinct pathways drive induction. This point becomes critical when adaptive chemistry schemes are developed to describe chemical phenomena in stratified and general inhomogeneous, environments (Banerjee & Ierapetritou, 2003; Schwer, Lu, & Green, 2003). The approach proposed here defines a comprehensive metric that captures the totality of the chemical transformations in a significantly reduced domain. It is important to realize that the purpose of this analysis is not to develop a reduced representation of the chemical mechanism. Rather, the purpose is to identify, while maintaining the full complexity of the mechanism, an alternate property that characterizes the structural evolution of the active reaction pathways. Finally, in interpreting our figures it is important to realize that for illustration purposes we are reporting in all cases the log of the motif values. Thus the actual differences, particularly between the first and second

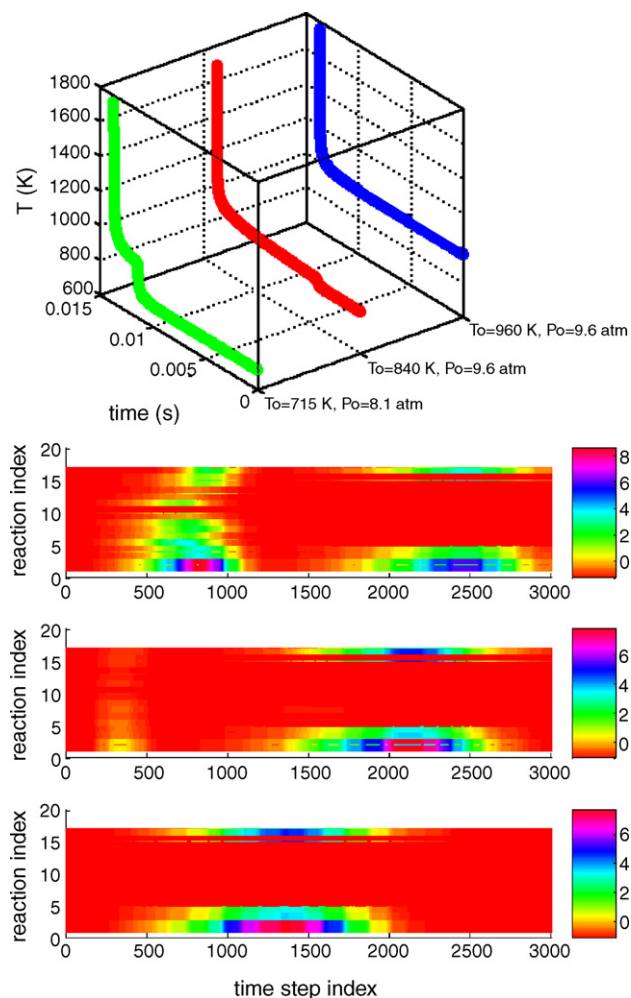


Fig. 6. Evolution of temperature profiles (top part) level of reaction activity during the course of the integration (bottom part).

ignition at low temperatures, are significantly more pronounced than Fig. 5 seems to indicate.

Fig. 6 is an alternative depiction of the chemistry evolution. The top panel presents in a comparative manner the three temperature profiles as a function of initial temperature, clearly indicating the two-stage ignition at low/intermediate initial temperature oxidation. In the bottom part we have isolated a subset of critical reactions and depict their normalized reaction rates to indicate the activity of selected pathways (Fig. 2). Normalization is arbitrary (division by largest rate) and conducted simply to enhance the visual inspection of the graph. It is important to note that the pathway activity does not depend on local conditions (T , conversion etc.) but rather on the reactive state of the system. The bottom part clearly demonstrates the onset of similar pathways at high temperatures (whether the initial temperature is high or whether the system eventually reaches that regime) irrespective of other conditions. Also, clearly we see the differences in the ignition chemistry. This fact is more clearly depicted in Fig. 7 which presents the evolution of motif values as reaction progresses in terms of local variables, such as conversion and temperature. It is important to realize that, independent

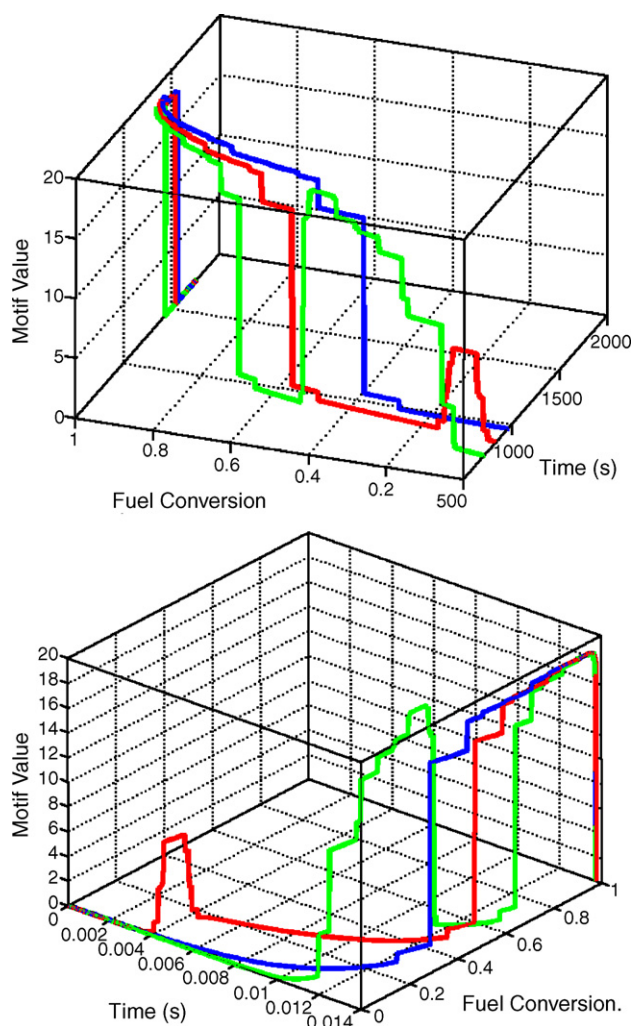


Fig. 7. Motif evolution a function of local fuel conversion and T (top) and time and local fuel conversion (bottom).

of the local conditions, the motif values correctly capture the distinct types of chemistries that are activated at each stage.

4. Conclusions and future work

Significant computational issues remain to be addressed, as we would like to improve the representation. The hash representation that was proposed is the first step in this direction. We have demonstrated that the essential information of the graphs can be captured and that strong differences and similarities can be identified. The topological properties of the hash function however, remain to be analyzed further. In particular we are currently analyzing further the actual distribution of the hash values and their implication in terms of similarities in the original space (graphs). The size of the graph describing the mass exchange among various species in the mechanism is expected to be extremely large. Alternative graph analysis, proximity-preserving hashing and soft computing methods will be explored for representing and characterizing the evolution of the kinetic transformations. Advances in nonlinear projections and self-organizing maps (Roweis & Saul, 2000; Tenenbaum et al., 2000) for reducing the dimensionality and enabling visualization and mining of time varying graphs will be investigated. We are currently exploring the ability of this representation to capture spatio-temporal variations and ultimately to be used in order to assess the interplay between complex chemistries and spatial inhomogeneities in multi-point ignition problems (Kraft, Maigaard, Mauss, Christensen, & Johansson, 2000; Maigaard, Mauss, & Kraft, 2003). Finally, we are in the process of developing post-processing analysis tools that will allow the automatic derivation of rules identifying the kinetic and rate characteristics associated with the main ranges of motif values. Our preliminary results correctly indicate that different types of chemistries are involved in each pre-ignition regime (Androulakis et al., 2004). However, in complex flow calculations where the flux information will be obtained by post-processing the species concentration information we expect to uncover far more interesting of motif values, and hence reaction pathways, spatio-temporal distributions. We should also emphasize that the computational framework for evaluating flux diagrams by post-processing reacting flow simulation results has already been implemented and has been demonstrated in the context of laminar flame calculations (Androulakis et al., 2004; Farrell et al., 2004). The ultimate goal is to identify optimal ways for the integration of flow and chemistry features for the integrated analysis of the impact of flows patterns on the evolution of chemical processes.

Appendix A

We demonstrate the steps of the symbolic transformation of an arbitrary series of the approximation of (Keogh et al., 2005) using the simple example depicted in Fig. A.1. The raw data (solid diamonds—right axis) are normalized using the standard transformation $(x - \mu)/\sigma$ (solid squares—left axis). The assumption being that, despite the small sample, the normalized values are normally distributed with a mean of 0 and S.D. = 1. The symbolic transformation (solid triangles—left axis) is based

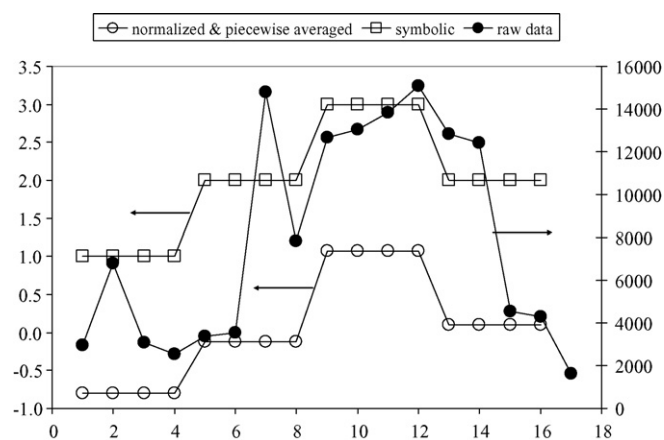


Fig. A.1. Basic elements of the HOT-SAX representation.

on the premise that the “symbols” are assigned such that the area under the curve of the normal distribution is the same. For example, in a three alphabet symbolic transformation, the symbols are assigned according to the following rule:

$$s = \begin{cases} 1 & x \leq -0.43 \\ 2 & -0.43 < x < 0.43 \\ 3 & x \geq 0.43 \end{cases}$$

The subsequent hashing assigns a distinct integer to each sequence of symbols. Signals whose symbolic transformations hash to the same value are considered to be identical. The design variables defining the transformation include the averaging window and the size of the alphabet present a sensitivity study for assessing the trade-off between accuracy of the representation (size of alphabet and discretization) and computational complexity.

References

- Abello, J., & Korn, J. (2002). MGv: A system for visualizing massive multidigraphs. *IEEE Transactions on Visualization and Computer Graphics*, 8, 21–38.
- Androulakis, I. P., Grenda, J. M., & Bozzelli, J. W. (2004). Time-integrated pointers for enabling the analysis of detailed reaction mechanisms. *AIChE Journal*, 50, 2956–2970.
- Banerjee, I., & Ierapetritou, M. G. (2003). Development of an adaptive chemistry model considering micromixing effects. *Chemical Engineering Science*, 58, 4537–4555.
- Brandes, U., Dwyer, T., & Schreiber, F. (2004). Visualizing related metabolic pathways in two and a half dimensions (Long Paper). *Graph Drawing*, 2912, 111–122.
- Crua, C., Kennaird, D., Sazhin, S., & Heikal, M. (2004). Diesel autoignition at elevated in-cylinder pressures. *International Journal of Engine Research*, 5, 1–10.
- Curran, H. J., Pitz, W. J., Westbrook, C. K., Callahan, C. V., & Dryer, F. L. (1998). Oxidation of automotive primary reference fuels at elevated pressures. In *Twenty-seventh symposium (international) on combustion* (pp. 379–387).
- Dean, A. (2001). Development and application of detailed kinetic mechanisms for free radical systems. In *International conference on the foundations of molecular modeling and simulation (FOMMS), AIChE symposium series: Vol. 97* (pp. 85–95).
- Dogrusoz, U., Feng, Q. W., Madden, B., Doorley, M., & Frick, A. (2002). Graph visualization toolkits. *IEEE Computer Graphics and Applications*, 22, 30–37.

- Echekki, T., & Chen, J. H. (2003). Direct numerical simulation of autoignition in non-homogeneous hydrogen–air mixtures. *Combustion and Flame*, 134, 169–191.
- Farrell, J. T., Johnston, R. J., & Androulakis, I. P. (2004). Molecular structure effects on laminar burning velocities at elevated temperature and pressure. In *Presented at SAE paper 2004-01-2936*.
- Gardiner, J. W. C. (1996). *Gas phase combustion chemistry*. Springer-Verlag.
- Gasner, E., Koutsofios, E., & North, G. (2002). *Drawing graphs with dot-dot user's manual*.
- Gorg, C., Birke, P., Pohl, M., & Diehl, S. (2004). Dynamic graph drawing of sequences of orthogonal and hierarchical graphs. *Graph Drawing*, 3383, 228–238.
- Griffiths, J. F. (1995). Reduced kinetic models and their application to practical combustion problems. *Progress in Energy and Combustion Science*, 21, 27–107.
- Halstead, M., Kirch, L., & Quinn, C. (1977). The autoignition of hydrocarbon fuels at high temperatures and pressures. *Combustion and Flame*, 40, 46–60.
- Keogh, E., Lin, J., & Fu, A. (2005). HOT SAX: Efficiently finding the most unusual time series subsequences. In *5th IEEE international conference on data mining* (pp. 226–233).
- Koegler, W. (2001). Case study: Application of feature tracking to analysis of autoignition simulation data. In *12th IEEE visualization 2001 conference (VIS 2001)* (pp. 461–464).
- Kraft, M., Maigaard, P., Mauss, F., Christensen, M., & Johansson, B. (2000). Investigation of combustion emissions in a homogeneous charge compression injection engine: Measurements and a new computational model. In *Proceedings of the combustion institute: Vol. 28* (pp. 1195–1201).
- Kuramochi, M., & Karypis, G. (2004). An efficient algorithm for discovering frequent subgraphs. *IEEE Transactions on Knowledge and Data Engineering*, 16, 1038–1051.
- Maigaard, P., Mauss, F., & Kraft, M. (2003). Homogeneous charge compression ignition engine: A simulation study on the effects of inhomogeneities. *Journal of Engineering for Gas Turbines and Power-Transactions of the ASME*, 125, 466–471.
- Patel, P., Keogh, E., Lin, J., & Lonardi, S. (2002). Mining motifs in massive time series databases. *ICDM, 2002*, 370–377.
- Pilling, M. J., & Hancock, A. (1997). *Low temperature combustion and autoignition*.
- Revel, J., Boettner, J. C., Cathonnet, M., & Bachman, J. S. (1994). Derivation of a global chemical kinetic mechanism for methane ignition and combustion. *Journal De Chimie Physique Et De Physico-Chimie Biologique*, 91, 365–382.
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290, 2323 pp.
- Schwer, D. A., Lu, P. S., & Green, W. H. (2003). An adaptive chemistry approach to modeling complex kinetics in reacting flows. *Combustion and Flame*, 133, 451–465.
- Temkin, O. N., Zeigarnik, A. V., & Bonchev, D. (1996). *Chemical reaction networks: A graph-theoretical approach*. CRC Press.
- Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2319 pp.
- Westbrook, C. K., Pitz, W. J., Boercker, J. E., Curran, H. J., Griffiths, J. F., et al. (2003). Detailed chemical kinetic reaction mechanisms for autoignition of isomers of heptane under rapid compression. In *Proceedings of the combustion institute: Vol. 29* (pp. 1311–1318).