On the generation and analysis of complex reaction networks in interstellar chemistry

Tom Lenaerts[†] and Hugues Bersini

Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle (IRIDIA), Univeristé Libre de Bruxelles, 50 Avenue Franklin Roosevelt, CP 194/6, B-1050 Brussels, Belgium.

Abstract

In Interstellar Chemistry, environmental conditions are often hard to duplicate in laboratory situations. Consequently, it is difficult to obtain a clear picture of possible molecules, reactions and overall system dynamics. Computational frameworks provide complementary mechanisms to perform this task. Here a framework is proposed that is based on the formal techniques for reaction network generation proposed by Kvasnicka and Koca (the Synthon model). Several authors have shown that these formal techniques suffer from limitations in their scaling i.e. the number of generated molecules scale exponentially with the number of atoms of the reactants. Moreover, since reactions can produce any molecular structure, isomers need to be identified. Yet, these drawbacks are in competition with the benefit that formal techniques are capable of generating new reaction mechanisms and, consequently, clarifying unresolved chemical processes. This benefit is highly relevant in Interstellar chemistry since, in that context, not all reactions are known and associated rate coefficients are often merely estimates. As a secondary benefit, this approach provides an initial framework to resolve more complex chemical engineering problems. To overcome the limitations, ideas from investigations into reaction generators and canonization algorithms are incorporated. As verification, the framework is evaluated using reactions and molecules defined in the context of interstellar chemistry. The initial results provide information about the validity of the algorithms to investigate molecular species and reactions in the different interstellar clouds. For instance, since during the generation process both dynamics (change in molecular concentrations) and the meta-dynamics (change in observed molecular types and reactions) are observed, one might gain insight in the routes of formation of the "exotic" molecular species present in interstellar clouds.

I. INTRODUCTION

Research into the structural and dynamical properties of biological and chemical networks requires, on the one hand, modeling systems that provide realistic results in terms of chemistry or biology and, on the other hand, systems that are abstract enough to draw general conclusions about observations performed on these types of properties. Motivated by these expectations and by the relevancy this kind of systems can have for different applications as for instance interstellar (Duley & Williams, 1984) and pre-biotic chemistry, a framework capable of simulating the co-evolution between the reaction network and the molecular kinetics has been developed. The framework is inspired by different realizations discussed in the context of computer chemistry or chemo-informatics (Gasteiger & Engel, 2003): molecular representation using multi-graphs with loops as analyzed in the Synthon model proposed by Kvasnicka and Koca (Koca, 1988a;Koca, 1988b;Hladka et al., 1993), isomeric molecules transformed into a unique representation using an algorithm equivalent to the CANGEN algorithm proposed by D. Weininger, A. Weininger and Weininger J. (1989), reaction networks constructed using formal and semiformal methods (Ugi et al., 1993) and the kinetics modeled using the stochastic simulation engine proposed by Gillespie (1977). The integration of all these components defines an object-oriented framework in which co-evolutionary simulations between the meta-dynamics of the reaction network and the kinetics of the different molecules can be performed (Bersini, 2000).

[†] Corresponding author, email: tlenaert@ulb.ac.be

The central core of the framework is the synthetic system that incrementally creates the reaction network. As argued by Ugi et al (1993), techniques for reaction network generation can be classified in three major categories: empirical, semiformal and formal techniques. From a theoretical and practical perspective, the latter category is the most interesting one since any form of inorganic or organic chemistry can be modeled. Moreover, as is pointed out in (Faulon & Sault, 2001), formal techniques are the only methods that can predict new reaction mechanisms and as a consequence resolve unknown chemical processes.   Although these methods have this important benefit, they are rarely exploited due to the combinatorial explosion of molecules and reactions. Especially in those chemical situations where large molecules can be formed the combinatorial explosion is extreme. In order to overcome this limitation different approaches can be taken. Depending on the particular chemistry, heuristics can be introduced that remove those reactions which are not plausible. Yet, this approach is problem dependent and therefore restricted. Recently, a problem independent approach was proposed to limit the reaction mechanism by taking into account the molecular concentrations that are produced by the kinetics of the generated reaction network (Faulon & Sault, 2001).   It is argued that this technique is general enough to be applicable to real processes and these claims were demonstrated in the context of thermal cracking experiments. Moreover, the algorithms scale in polynomial time which is an improvement over other methods that scale in exponential time. Since our system requires the combination between the molecular kinetics and the meta-dynamics of the reaction network, we mapped the approach proposed in (Faulon & Sault, 1984) onto the object-oriented Synthon framework.

The framework is developed to study kinetics and meta-dynamics in the context of interstellar chemistry (Duley & Williams, 1984). The relevance of this type of frameworks in that context is high, since conditions present in the interstellar medium are hard to duplicate in a laboratory context. An adequate computational framework can provide an experimental environment to predict the occurrence of certain molecular types and the processes that produce them as this can thus be verified on realistic data. Hence the entire method can be considered to be complementary to the existing approach of identifying the molecules through the observation of emitted radiations.

In the next section we discuss the internal details of the framework i.e. the molecular representation and the way in which the molecules are uniquely defined. Afterward the methods for kinetic and network generation are described. In this section we will show the outcome of some initial experiments that were performed in the context of interstellar chemistry. All chemical substances, reactions and reaction rates for the experiments described in this work were adopted from (Duley & Williams, 1984).


II. FRAMEWORK

The network generator is based on a mathematical chemistry proposed by Kvasnicka and Koca (Koca, 1988a;Koca, 1988b;Hladka et al., 1993).   The idea of a mathematical chemistry is to provide a model for the logical structure of chemistry. Given such a model, different chemical notions and concepts can be rationalized, formalized and unified. This kind of mathematical models is often used to formalize synthesis and retro-synthesis processes examined in organic chemistry.

The Synthon model is an extension of the Dugindji and Ugi (DU) model (Ugi et al.,1993). The general idea in the DU model is that the notion of isomerism can be extended from molecules to ensembles of molecules and, as a consequence, reactions are formal isomerizations between two appropriate ensembles. Any reaction, which is an isomerization can hence be modeled.   The DU model thus defines a formal theory for reactions, reaction mechanisms and their representations (be-matrices).

The Synthon model defines a graph-theoretical extension of the DU model. A *synthon* is informally defined by a set of atoms and virtual atoms (representing functional groups not directly active in the reaction), the chemical bonds between them, their lone-pairs and free electrons. The synthon is represented by a multi-graph with loops :

$$S(A) = \langle W, A, E \rangle$$

where the set $W=\{w_1, w_2, ...\}$ is the virtual vertex set, the set $A=\{a_1, a_2, ...\}$ is the set of remaining vertices representing the concrete atoms and $E=\{e_1, e_2, ...\}$ is the set of edges associated with vertices from the sets W and A. The edges can only be incident between two vertices of A or a vertex from the set A and a vertex from the set W. Note that certain elements from the set E are loops which define lone pairs. In Figure 1 and Figure 2 examples are shown of different molecular graphs that can be represented in the framework.
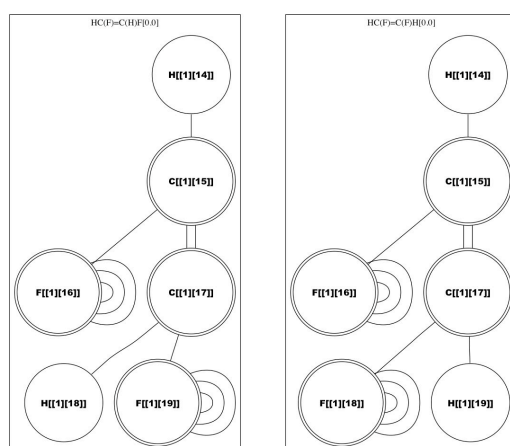
Figure 1: The molecular graph representation of two isomeric molecules: cis- and trans-difluoroethene. The difference between the molecules is defined by the difference in the order of the different nodes. The order is determined by a node identification number which is the second number between brackets in the node, e.g. for node C[[1][17]] the identity is 17. The value of this number is only relevant in terms of the ordering of the nodes. The first number, here 1, is the molecule identification number and becomes important when two molecules are joined to form a synthon. The first symbol of node [[1][17]] corresponds the atom C.

We added two extensions to the molecular representation used in the Synthon model: (1) free electrons of an atom are explicitly modeled to represent radicals and (2) ionic bonds, required for molecules like H_2^+ and H_3^+, are introduced. First, in the context of Interstellar chemistry, a lot of reactions occur between charged and uncharged molecules. To make a distinction between these different kinds we make the presence and absence of the electrons explicit. To do this a new kind of vertices is introduced and these vertices will be stored in the set R since they do not belong to the atom set A and virtual atom set W. The set $R=\{r1, r2, ...\}$ contains the electrons explicitly modeled in the molecular graph. The set of edges E is now extended so that it also contains the edges incident between a vertex from the set A and a vertex of the set R. In other words the links between electrons and atoms. Note that edges between vertices R and W or R and R are not possible. Through the introduction of this set R, reactions where for instance an electron is exchanged between a uncharged molecule and a charged one are still isomerization processes and consequently fit nicely in the existing Synthon theory. An example of this extension can be observed in Figure 2 (right). Second, molecules like H_2^+ and H_3^+ require a new type of bond. Until now, the Synthon model assumes covalent bonds where a bond is defined by a connection between two electrons. In other words the bond representation assumes a sharing of two electrons for each edge. In the context of for instance H_2^+ there is only one electron that is shared between both atoms. Moreover, this sharing cannot be represented by a simple edge $e \in E$ since no distinction

could then be made between H_2 and H_2^+ when looking at the graph. This distinction is essential and the system should also be able to recognize this automatically. The solution of the problem is straightforward. Since notion of the molecular graph that serves as a foundation in the Synthon theory does not require that the graph is connected, molecules like H_2^+ can then be represented by two disjoint parts belonging to the same molecule. In Figure 2 on the left one can see that the molecule consists two disjoint parts: i.e. H_3^+ will be represented by two disjoint parts: H^+ and H_2. The first number of each node refers to the molecular identity and this should be the same for both disjoint parts. From a chemical perspective, we assume that the molecules like H_2^+ and H_3^+ are represented by one of their resonance forms.
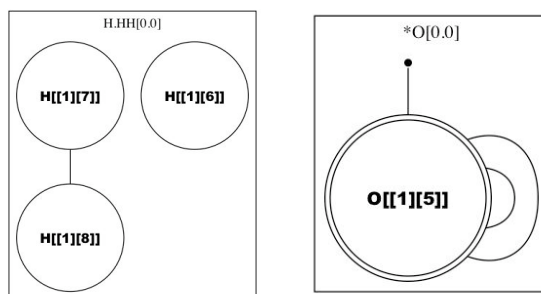

Figure 2: The molecular graph representation of H_3^+ (left) and O^+ (right). H_3^+ is represented by one of its resonance forms. The radical O^+ is represented with one free electron (the dot) and two lone pairs (loops).

In summary, the synthon is now defined as:

$$S(A) = \langle W, A, R, E \rangle$$

Using these structures and the principle of isomeric transformations, reaction networks and their properties (e.g the reaction distance between synthons) can be generated. Every reaction or chemical transformation is of the form
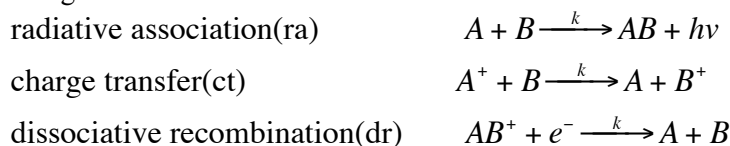
$$S(A) \Rightarrow S'(A)$$
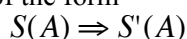
Where $S(A)$ is the educt and $S'(A)$ the product of the reaction. Since this chemical transformation creates a product which is isomeric to the educt, the process is also called an *isomerization process*. Each transformation T linking the two synthons $S(A)$ and $S'(A)$ can be represented by a reaction graph which is, in our case, the 5-tupple:

$$G_T = \langle W_T, A_T, R_T, E_T, \psi_T \rangle$$

Where the vertex and edge sets W_T, A_T, R_T and E_T contain those vertices and edges involved in the transformation and where ψ_T defines a mapping $\psi_T : E_T \rightarrow \{-1, +1\}$ assigning to each edge either $+1$ for (added edges) or -1 for (removed edges):

$$\psi(e) = \begin{cases} +1, for & e \in E_2 \ (formed \, edges) \\ -1, for & e \in E_1 \ (deleted \, edges) \end{cases}$$

Hence every concrete reaction can be specified by a reaction graph G_R describing the transformation from one synthon to another. Moreover, different reactions which are represented by the same graph but whose contents of W and A are different can be classified into the same category. These categories define abstract reaction classes, conceptually similar to the object-oriented perspective, whose structural effects are the same. For instance, consider the following interstellar reactions:

radiative association(ra) $A + B \xrightarrow{k} AB + h\nu$

charge transfer(ct) $A^+ + B \xrightarrow{k} A + B^+$

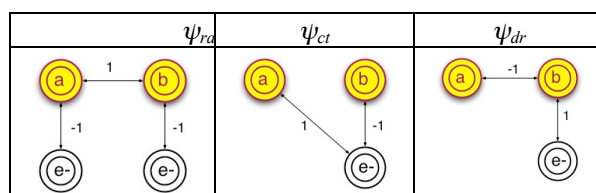dissociative recombination(dr) $AB^+ + e^- \xrightarrow{k} A + B$

Given these reaction templates, one can determine the reaction graphs. For instance, the radiative association template assumes that A and B can be linked. This requires in both A and B an atom which has at least one free electron. Assume that for A this atom is a and for B this atom is b. Furthermore, atom a has one free electron e_1 and atom b has one free electron e_2. Now the radiative association reaction describes that a covalent bond is formed by sharing the two free electrons e_1 and e_2 on atom a and b. Practically from the perspective of the molecular graph, this means that (1) the links between the two electrons and their atoms and (2) the two nodes representing the electrons are removed. Next, (3) the two atoms are connected by a new edge representing the covalent bond. The transformation scenario for radiative association is visualized in the first mapping ψ_{ra} in Figure 3. The same can also be done for the charge transfer template (ψ_{ct}) and the dissociative recombination template (ψ_{dr}).

All this information provides the necessary components to define the reaction graphs G_T:

$$G_{ra} = <W = \{\}, A = \{a,b\}, R = \{e_1, e_2\}, E = \{(a,e_1),(b,e_2),(a,b)\}, \psi_{ra} >$$

$$G_{ct} = <W = \{\}, A = \{a,b\}, R = \{e_1\}, E = \{(b,e_1),(a,e_1)\}, \psi_{ct} >$$

$$G_{dr} = <W = \{\}, A = \{a,b\}, R = \{e_1\}, E = \{(a,b),(a,e_1)\}, \psi_{dr} >$$

Figure 3: From left to right the mapping for radiative association, charge transfer and dissociative recombination are visualized. The nodes a and b refer to atoms and the node $e-$ identifies an electron.

Each G_T defines *a reaction class*. All specific chemical reactions, where a and b are replaced by real atoms, are instances of a particular class if they model the same transformation. We will refer to these instances as *reaction objects*. Hence a reaction class is defined as a set of reaction objects. A unique description of a reaction object requires the specification of not only the reactants but also the products. Following this observation the following three cosmic ray ionization reactions belong all to different classes of which they are one reaction object:

$$H_2 + c.r. \xrightarrow{k1} H_2^+ + e^-$$

$$H_2 + c.r. \xrightarrow{k2} H^+ + H + e^-$$

$$H_2 + c.r. \xrightarrow{k3} H^+ + H^-$$
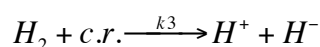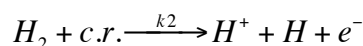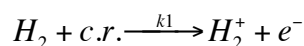
This distinction is important from both a chemical and computational perspective since, on one hand, by assuming this way of defining reaction classes and objects, a different taxonomy of reactions emerges. On the other hand, from a computational modeling perspective, objected-oriented modeling becomes the conceptually natural approach to represent the reactions classes and their objects.

Given the concept of the synthon $S(A)$ and the isomerization process G_T, reaction networks can be generated. Note that the isomeric reactions conserve mass, charge and radicals when transforming educts into products. As discussed in the literature on mathematical chemistry, the basic reaction formula then becomes:

$$(G_T)S(A) = S'(A)$$

i.e. the application of the transformation G_T on the graph $S(A)$ produces the graph $S'(A)$. As described by Ugi et al. (1993) there can now be two different types of reaction network

5

generation. The first type assumes a given *S(A)* and asks the system to find all couples $(G_T, S'(A))$ that solve the equation under the provided boundary conditions. The solutions are called b-solutions and are produced by reaction generators of the type RGB. The second type assumes a given chemical transformation G_T and the systems need to find the couples (*S(A)*, *S'(A)*) (r-solutions) that solve the equation. The reaction network generator is called a type RGR. Here, the algorithms are all of type RGB.

Before investigating the different RGB algorithms, one needs to remember that, although the chemical transformations produce new synthons, our interests is in the different molecules which they represent. One important implication is that the system needs to recognize whether created molecules are isomorphic to the molecules already present in the initial set. The algorithm used here is the same as the CANGEN algorithm (D. Weininger, A. Weininger & Weininger J., 1989) defined in the context of the generation of unique SMILES notations. The algorithm consists of two phases where first the molecular graph is canonized using an algorithm inspired by the Morgan algorithm and a second phase in which a unique notation is extracted based on the outcome of the canonization. Next to its simplicity, the strength of the algorithm is the fact that it moves the decision of how to assign unique codes to the different atoms in a molecule to the simulation phase. At that point the user provides a set of functions that produce a suitable 'hash'-code for the different nodes in the molecular graph. Concretely we used an encoding of the nodes of the molecular graph using information on the number of bonds, the directionality of the bonds, the number of lone pairs, the chirality and the atomic number. This encoding seems to provide the building blocks to distinguish between R/S and Z/E configurations (see also Figure 1). A more extensive evaluation is necessary to examine the completeness of the proposed functions.

III. NETWORK GENERATORS

Reaction generators, similar to those described in (Faulon & Sault, 2001), were implemented. The similarity is restricted to the equivalence in behavior since in this work the algorithms were constructed using a modular, object-oriented approach and the Synthon theory instead of the DU theory. We will briefly describe their workings but for an in-depth discussion we refer to the description provided by Faulon and Sault (2001). A further similarity is the fact that the reactions are here also limited to monomolecular and bimolecular reactions. Furthermore, the same constraints are provided to put bounds on the growth process. For instance, limits are introduced on the number of atoms in a molecular species (*N*), on the number of lone pairs or radicals per species (e_s) and the number of lone pairs or radicals per atom (e_p).

The *deterministic-network-generator* (DNG) generates all possible products that can be constructed from a given set of reactants and the set of chemical transformation templates. Hence the output is a network consisting of all possible reaction objects and molecules. The algorithm takes as input a set of molecular species and tries first to determine all new molecular species that can be produced using monomolecular and bimolecular reactions taking into account the constraints. This new set is constructed in such a way that it cannot contain any species from the set that is used to generate this next one. The system also maintains a set of reaction objects. This set is updated each time a reaction can be performed and the reaction is not already present in the previous reaction set. Once this new set is produced, the next iteration starts. The algorithm ends when no new reaction objects can be produced. The algorithm is guaranteed to end since species have limited size. Concretely the maximum amount of species that can be created is limited by the maximum-size constraints imposed on the elements.

Given the DNG simulation ideas, we now examine an (not completely realistic) experiment in the context of interstellar chemistry. In this experiment the set of molecular species consists

of hydrogen and oxygen. The setup is equivalent to a modeling experiment discussed by Duley and Williams (1984, chapter 7, p. 135-). The goal of such an experiment is twofold: First one wants to gain insight in the routes of molecule formation and destruction. Second, one wants to understand the physical conditions of the environment where the molecules are formed. The approach described in (Duley & Williams, 1984) identifies a number of problems when trying to reach this goal. First, the relevant chemical reactions need to be identified. Duley and Williams (1984) argue that although this might seem straightforward for few reactions, many of the associated reaction rates are still mere estimates. If the amount of reactions increases the problem becomes worse. Next to the uncertainty of the gas phase reactions, there is even a greater uncertainty on the possible contribution of the grain surface reactions. In this article, the reaction network generators are used to provide a mechanism to identify the relevant reactions. In later work we will also try to add a feature capable of reaction rate prediction. To keep things simple for now, we used the reactions and reaction rates provided in (Duley & Williams, 1984; p. 143) and repeated in Table 1.

| Class	Chemical class name	Reaction template	Rate k
G_1	Cosmic ray ionization	$A + c.r. \rightarrow A^+ + e^-$	$10^{-17}\,s^{-1}$
G_2		$AB + c.r. \rightarrow AB^+ + e^-$	$10^{-17}\,s^{-1}$
G_3		$AB + c.r. \rightarrow A^+ + B + e^-$	$10^{-19}\,s^{-1}$
G_4	Ion-molecule exchange	$A^+ + BC \rightarrow AB^+ + C$	$10^{-9}\,cm^3 . s^{-1}$
G_5	Charge rearrangement	$A^+ + B \rightarrow A + B^+$	$10^{-9}\,cm^3 . s^{-1}$
G_6	Dissociative recombination	$AB^+ + e^- \rightarrow A + B$	$10^{-6}\,cm^3 . s^{-1}$
G_7		$A^+ + e^- \rightarrow A$	$10^{-11}\,cm^3 . s^{-1}$
G_8	Neutral reactions	$A + BC \rightarrow AB + C$	$10^{-11}\,cm^3 . s^{-1}$
G_9	Photodissociation	$AB + hv \rightarrow A + B$	$10^{-11}\,s^{-1}$
G_{10}		$H_2 + hv \rightarrow 2H$	$10^{-14}\,s^{-1}$
G_{11}	Grain surface reaction	$H+H:g \rightarrow H_2+g$	$10^{-17}\,cm^3 . s^{-1}$

Table 1: Examples of reaction classes used in interstellar chemistry and their reaction rates. The column titled class refers to the different reaction graphs G_T . ($c.r.$ = cosmic ray energy, hv = electromagnetic radiation energy (photons)). Note that the classes G_{10} and G_{11} are specifically defined for hydrogen. For the photodissociation reaction G_{10} the difference is due to the difference in reaction rates. For G_{11} only the surface reaction for hydrogen is realistic and hence modeled here.

Table 1 contains a set of reaction classes that are derived from a similar list of reaction objects in (Duley & Williams, 1984; p. 143). The goal is to determine which reactions are relevant in this simple and artificial interstellar situation. The constraints that were imposed on this generator are $N=6$, $e_s=10$ and $e_p=7$. In Figure 4, the resulting network can be observed. The reaction graph is a directed graph linking reactant nodes (the molecules) to reaction nodes (the blue squares) and again to product nodes (again molecules). From the network we can retrieve the actual reaction objects by looking at the blue square nodes.   For instance, the following (reduced) table can be compiled:

Ref.	Reaction objects
1	$H + c.r. \rightarrow H^+ + e^-$
8	$H^+ + H \rightarrow H + H^+$
32	$H_2^+ + H_2 \rightarrow H_3^+ + H$
56	$H_3^+ + H \rightarrow H_2 + H + H^+$
…	…

Table 2: Reaction objects generated by the DNG algorithm. The first column refers to the number in the blue squares in the reaction network and is a uniquely defined for each object.

To allow visualization for the article, the growth of the network in Figure 1 was restricted by an observational constraint. This constraint specifies which molecules are allowed in the system. In principle the limits are normally set to the size of the molecule and other structural properties.
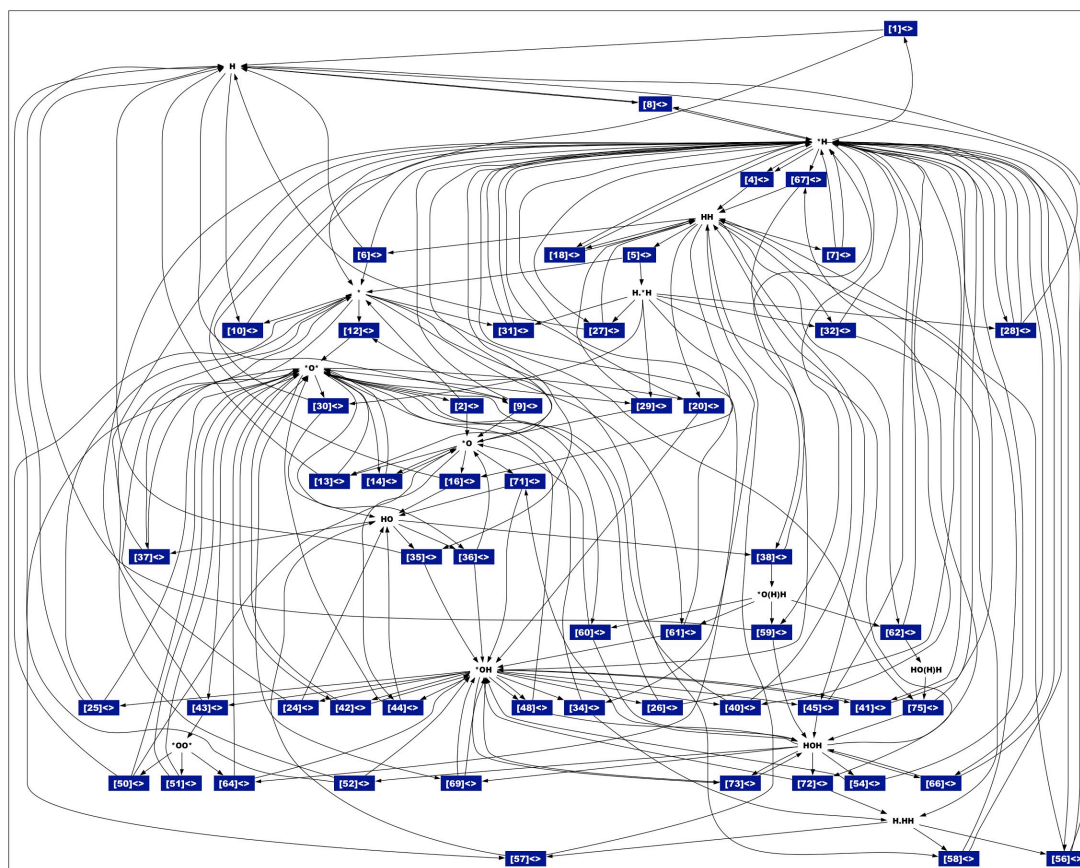
Figure 4: The reaction graph generated through DNG using the chemical transformation listed in Table 1 using initial species H and O. The blue/numbered nodes represent the reaction objects produced by the different reaction classes. The other nodes represent the molecules. Reaction objects are linked with reactant and product molecules. The relation between the molecules in this graph and real molecules is the following: H=$H+$, *H=H, HH=H_2, H.*H=H_2^+, H.HH=H_3^+, *O*=O, HO=OH^+, *OH=OH, HOH=H_2O, *OO*=O_2, *O(H)H=OH_2^+ , HO(H)H=OH_3^+ and *=e^- .

Although reduced by the visualization constraint, the network shows that even by the introduction of a few reaction classes and a few simple molecules, the reaction network can become very large. In terms of the goal of the algorithm to identify the relevant reactions, we can not really distinguish which reaction objects are relevant and which are not. From a dynamical perspective, only those reactions whose reactants are present in a certain amount are relevant. In other word, the kinetical information should allow us to reduce the complexity of the network to only those reactions which are really relevant.

As argued in the introduction, the actual explosion can be reduced using the general purpose approach proposed by Faulon and Sault (2001): exploit concentration information to decide which reactions can occur and which can't. The idea behind this is that species with low concentrations can be removed since they have negligible effect on the final distribution of products (unless these species are catalysts). The concentration information is derived using Gillespie's stochastic simulation algorithm (Gillespie, 1977). For details on the mechanism we refer to Gillespie's article. In general, the idea of the stochastic algorithm is to determine when the next reaction will occur and what reaction this will be. Hence at each iteration the algorithm returns the size of the next time step and a reaction that occurs in that time step. Given this information, the reaction is applied producing changes in the amounts of the reactants (decrease) and products (increase).

Faulon and Sault (2001) propose two network generators which make use of the concentration information: the *concentration-sampling-network-generator* (CSNG) and the *MC-sampling-network-generator* (MCNG). Although we investigated both, we will focus

here only on the second. One of the major motivations is that, in the end, we want to device an algorithm that performs the kinetic and meta-dynamical simulation at the same time which is indeed the case for MCNG. In MCNG, the set of new reaction objects is determined using the molecular species that have non-zero concentrations. We perform the same experiment as done for DNG introducing some information required for the simulation of the kinetics. Since the experiment is started with H and O, we need some initial density information on these molecules. We adopted from (Duley and Williams, 1984) the situation where the particle density n=1000 cm^{-3} and the density of O, $n(O)$=0.44 cm^{-3}. Moreover it is assumed that $n=n(H)+2n(H_2)$. This conservation equation specifies that the particle density is the same as the density of H in one cm^3. Given this equation, we can determine that H has initially a density $n(H)$=1000 cm^{-3} since there are no H$_2$ molecules. Note that the conservation equation is not completely correct since other molecules can contain H. Yet for this artificial setup we assume this simplification. Some extra parameters are also required for the kinetic simulation: (1) the number of iterations that will be performed by the Gillespie algorithm (Mc) and (2) the number of particles (Mp) that will be used to simulate the kinetics. In Figure 5 the final reaction network is shown. In Figure 6 the kinetic dynamics are shown in two different plots. To obtain the network and kinetics in figures 5, and 6 an extra constraint was introduced which removed reactions like reaction object 8 described in Table 2. The motivation for the removal was to bring the experiment as close as possible to the suggested setup in (Duley & Williams, 1894, p.143). To perform realistic interstellar simulations, other heuristics relevant to interstellar chemistry should be used.
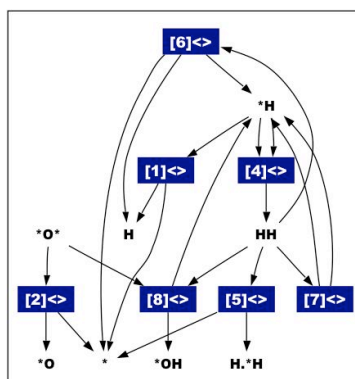

Figure 5: The reaction graph generated through MCNG using the chemical transformation listed in Table 1 using the initial species O and H with densities 0.44 cm^{-3} and 10^3 cm^{-3} respectively. The blue nodes represent the reaction objects produced by the different reaction classes. The relation between the molecules in this graph and real molecules is the same as in Figure 4.

First of all, one can observe that indeed, the introduction of the kinetics provides a drastic reduction in the size of the reaction network keeping only the most relevant reactions. In this respect the goal put forward for MCNG is fulfilled. Moreover, based on the information from the kinetics which can be seen in Figure 6, it is clear that H and H2 are the dominant species. Given this unrealistic chemical example, the simulation approach investigated in this article seems to provide good results for the study of the chemical activity in interstellar clouds. Further investigations are currently continued along this line.

IV. CONCLUSIONS

In this article, we explained the foundations of an object-oriented framework for the simulation of the dynamics and meta-dynamics of chemical reactions. The framework incorporates and extends different achievements from chemo-informatics; the Synthon model, the CANGEN algorithm, Gillespie's stochastic simulation algorithm and reaction network generation algorithms. The adaptation of the object-oriented approach is advocated for its natural modularity and scalability. The framework is currently being used in the context of interstellar chemistry and pre-biotic chemistry. The examples listed here provide a

preliminary flavor of the possibilities of the framework. Further investigation are underway to account for more realistic situations. One important extension that is currently investigated is the use of thermodynamical information (e.g. reaction enthalpy estimations) to decide which reactions will and will not be added to the network. This will also reduce the network size and will make the network more faithful to the chemistry under observation. Other extensions which are currently under investigation are the automatic calculation of the reaction rates as was proposed in (Bersini, 2000; Faulon & Sault, 2001) and some improvements of the MCNG algorithm to simulate the coevolution of kinetics and meta-dynamics.
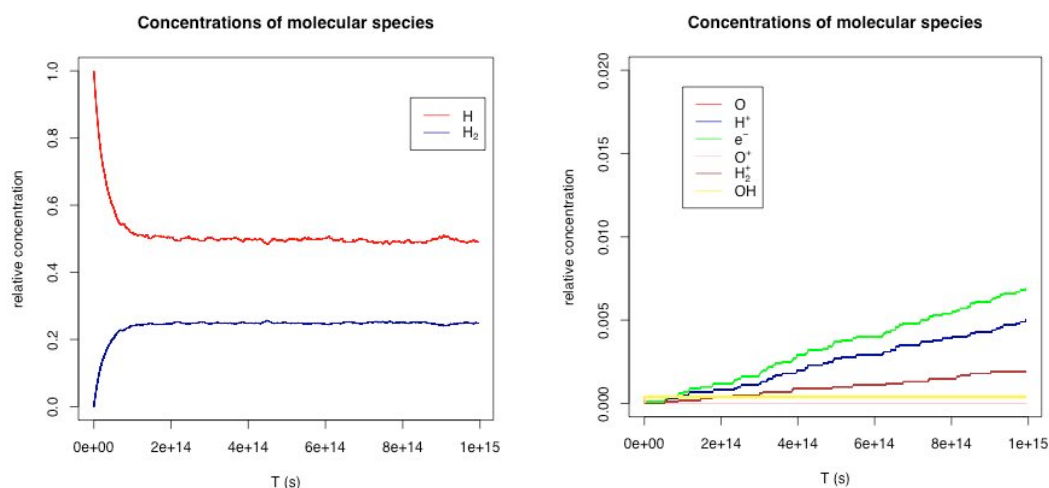
Figure 6: The kinetics produced by the Gillespie algorithm using the reaction rates shown in Table 1. On the left the concentrations of H and H2 are visualized. On the right the concentrations of the other molecules are visualized. The x-axes represent the time in seconds. The y-axes shows the relative concentrations.

REFERENCES

Bersini, H. (2000) *Reaction Mechanisms in the OO chemistry*. Proceedings of the 7th International Conference of Artificial Life, p.39-49 . Bradford books.

Duley, W.W. & D.A. Williams (1984) *Interstellar chemistry*. England, Academic Press.

Faulon, J.-L. & A.G. Sault (2001) Stochastic generator of chemical structure. 3. reaction network generator. J. Chem Inf. Comput. Sci., 41:894-908. American chemical society.

Gasteiger, J. & T. Engel (2003) *Chemoinformatics: a textbook*. WILEY-VCH Verlag Gmbh & co.

Gillespie, D.T. (1977) *Exact stochastic simulation of coupled chemical reactions*. The journal of physical chemistry, 81 (25):2340-2361.

Hladka, E, J. Koca, M. Kratochvil, V. Kvasnicka, L. Matyska, J. Pospicha, V. Potucek (1999) *The Synthon model and the program PEGAS for computer assisted organic synthesis*. Topics in Current chemistry, 66: 121-197. Heidelberg, Berlin, Springer-Verlag.

Koca, J. (1988a) *A mathematical model of the synthon*. Part XI in the series Mathematical Model of Organic Chemistry. Collection Czechoslovak Chem. Commun, 53: 1007-1017.

Koca, J. (1988b) *A graph model of the synthon*. Part XII in the series Mathematical Model of Organic Chemistry. Collection Czechoslovak Chem. Commun, 53: 3108-3117.

Ugi, I. , J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, R. Herges, M. Knauer, K. Reitsam & N. Stein (1993) *Computer assisted solution of chemical problems – The historical development and the present state of the art of a new discipline of chemistry*. Angew. Chem. Int. Ed. Engl., 32:201-227.

Weininger D., A. Weininger & J. Weininger (1989) *SMILES 2:Algorithm for generation of Unique SMILES notation*. J.Chem. Inf. Comp. Sci, 29 (2):97-101. American chemical society.