Design and Development of Computer-Aided Chemical Systems: Representation and Balance of Inorganic Chemical Reactions

I. Luque Ruiz,* C. Martínez-Pedrajas, and M. A. Gómez-Nieto[†]

Department of Computing and Numerical Analysis, University of Córdoba, Campus Universitario de
Rabanales, Edificio C2, Planta-3, E-14071 Córdoba, Spain

A model for the tracking of inorganic chemical reactions is proposed. Designed to acquire, process, and solve a great number of inorganic reactions, this model will hopefully contribute to the development of powerful computer-aided chemistry teaching systems for use within or without the environment of a virtual laboratory. Using full representation of an inorganic reaction to allow the extraction of chemical knowledge, incomplete reactions (where species are absent) may be completed by adding the necessary species, and reactions may be solved and balanced. Various types of reaction are classified, and a layer-based model is defined for the solution of different reaction types, establishing the basis for the construction of a system which, based on a wide set of production rules, is capable of solving an incomplete inorganic chemical reaction.

1. INTRODUCTION

Over the past 2 decades, advances in information technology (IT) have given rise to major changes in, for instance, the way recent generations perceive reality, and thus in the way they might grasp new information. These new cognitive models employed in the processes of capture, representation, and processing of information have become increasingly patent to teaching staff who, in turn, appreciate the need for changes to their teaching models and techniques, especially in technical/experimental subjects such as chemistry, at both high school and university levels.[1−6]

Thus, and here focusing on the field of chemistry, considerable effort has been directed toward the development of software to assist in both teaching and research; terms such as "*computer-assisted*" and "*computer-aided*" are now in everyday use when describing software tutorial systems aimed at facilitating these activities.

Software designed for computer-assisted tutorial systems for chemistry, as cited in the present reference section, is mainly centered on two areas:

(a) the proposal of problems concerning well-defined areas of chemical knowledge (these systems may or may not include subsystems aimed at teaching the user such knowledge as may be necessary for the solving of the proposed problems, with subsystems for help and correction; such systems would include tutorials on specific areas of the subject[7−12]);

(b) the development of "*virtual laboratories*" that simulate specific processes (mainly reactions), supplying the user with information about these and providing a fairly accurate software "virtual" simulation of the reaction as it would take place in a real laboratory.[12−15]

The level of interactivity of these systems varies substantially, ranging from a minimal situation where the user may only read the information supplied by the system and answer specific questions about its content, to the manipulation of working parameters or simulation in virtual laboratories, where the user may in some way manipulate the working/ learning conditions.

Perhaps, then, one of the main problems with the above-mentioned systems is their lack of true flexibility, namely, the extent to which the system may be adapted to the needs of the user during the teaching process.[10−12] Indeed, most such systems are limited to a fixed number of cases, defined in the software by static lines of code. For example, systems designed for the proposal of fixed problems are hard-coded and, in most cases, only allow the user to alter certain of their physicochemical parameters, thus affecting neither the structure of the problem nor the supplied information, just the numerical end result.[5,7,8,12,14,15]

The present authors believe that the solution to this problem may lie in a flexible architecture for software systems, and particularly in a structure based on components using procedures that can manipulate information in both directions such that, within the bounds of certain necessary rules, the user may enter information, in his/her customary form, into the system, detailing problems to be solved; the system would then solve these problems and present its solution and, more importantly, the process or reasoning behind it.[1,5,10,11,16]

In the field of chemistry, systems have been developed to facilitate man−machine communications. Translators capable of accepting an input of a character string representing a chemical species and converting it to a data structure suitable for computer processing have already been developed. Once processed, this machine representation of chemical species may then be converted into a new string or graphic and output to the user.[17−22]

* To whom correspondence should be addressed. Phone: +34 957 212082. Fax: +34 957 218630. E-mail: ma1lurui@uco.es.
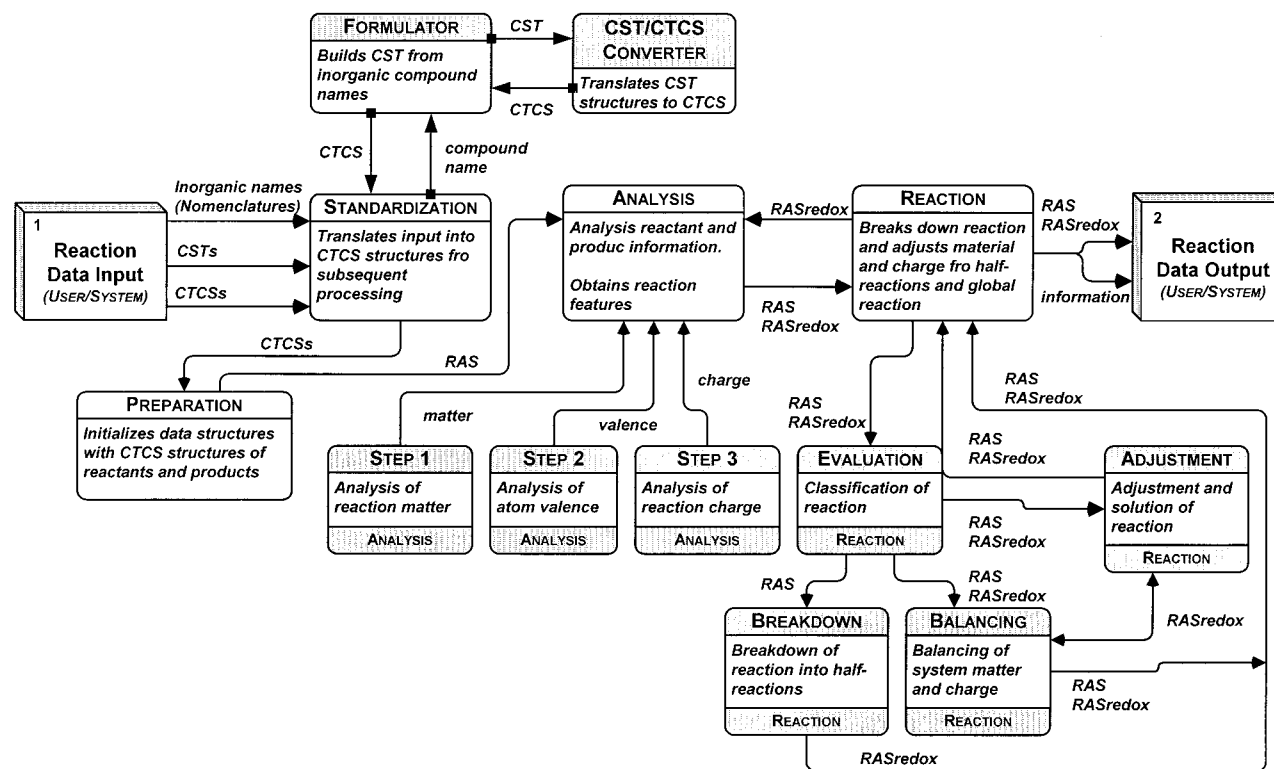   † E-mail: mangel@uco.es.

Figure 1. Architecture of the model, showing stages, information, and task flow.

Although software does exist to produce representations of chemical species based on user input, together with the calculation of certain corresponding physicochemical parameters, not many systems apply these communication models to computer-assisted learning.

The aim of the present study is to apply one such system for man−machine communication of chemical knowledge to the development of teaching systems for chemistry— systems to deal with situations ranging from the proposal of standard problems through to the performance of experiments in a virtual laboratory.

To achieve this, computing resources will be required to enable the representation and processing of reactions occurring between chemical species. These resources must be capable of representing any reaction (not just a finite, hard-coded set) and, by applying the known laws of combination, "carry out" the reaction and output the result in a form the user will understand.

This article describes the work of the authors in building a system to track reactions between inorganic species. First, there is a description of the proposed model, detailing the necessary elements for carrying out the process and the proposed abstract structures for the representation of information. The next section deals with how reactions may be classified on the basis of acquired knowledge, followed by a description of the step-by-step adjustment process for reactions (whether complete or incomplete). Finally, the useful range of the proposed model is analyzed, and future lines of study aimed at widening these limits are proposed.

2. ARCHITECTURE OF THE MODEL

When defining the architecture of the model, the man−machine communication paradigm chosen for the representation of information must be borne in mind. The proposed model is based on a series of data structures that permit the representation and transfer of chemical information, as described elsewhere.[10,17−19]

Figure 1 contains a context diagram of the architecture of the proposed model. The model supposes that data are input via two CTCS (Concise Table of Compound Structure) structure lists,[17−19] corresponding to the reactants and products involved in the reaction, and that the reaction is tracked through four steps:

If data are input using a different representation, for instance naming compounds by any of the standard nomenclatures, or using their CST (Compound Structure Table) structures, this input will be converted during the Standardization stage to the CTCS structure by the formulator subsystem.[10,18] This standardization stage lends great flexibility to data input, which may be performed directly by the user or through any other system, in addition to representing the information required for the reaction in a simplified structure (CTCS) that ensures optimal process performance.[18]

At the Preparation stage, data structures are prepared for use in the reaction. Here, two CTCS structures are input, and a set of data structures is built for use in the stages described in the next section.

These data structures enable the following activities to be performed: (1) an accurate analysis of data input, allowing the determination of the features of the reaction to be processed, with the reaction to be carried out by the system wherever possible; (2) a breakdown of the reaction (redox systems) such that it may be carried out in its basic half-reactions (this process permits the tracking of each of the component reactions occurring, allowing more information to be handled and thus reported to the user); (3) the abstract representation of the elements (compounds, subcompounds,

ions, or atoms) present in the reaction, in addition to their structural relationships at each step (start, process, and end) of the inorganic reaction.

Once the information has been suitably represented, it is examined in order to obtain the features of the process to be tracked. This is the function of the Analysis stage, during which the following processes are performed:

(1) The matter content of the system is checked, to see if there is a balance or equality on both sides of the reaction (reactants and products).

(2) Charges are checked for equality, to determine whether an ionic balance exists across both sides of the reaction.

(3) A check is made for equal valence of the same atoms on both sides of the reaction.

The end result of these three processes provides knowledge of the features of the reaction, the reaction type, and the capacity of the model for tracking and control.

The Analysis stage supplies the Reaction stage with enough information for the reaction to be carried out. At this latter stage the following steps are taken:

(1) Compounds, subcompounds, or ions are added to one or both sides of the system in order to achieve atomic consistency; that is, the set of atoms will be the same on both sides.

(2) The inorganic reaction is broken down into simpler half-reactions, where possible. Each half-reaction is analyzed and balanced for both matter and charge.

(3) Matter and charge are adjusted for the global reaction (and for redox half-reactions, where necessary). This process involves (a) balancing of matter, which may mean adding compounds, subcompounds, or ions to either or both sides of the reaction or half-reaction, and (b) balancing of ion charge, which may involve adding electrons to one side of the reaction or half-reaction.

(4) Information is gathered from each of the steps followed during the adjustment process, which will provide knowledge for the user.

(5) The solution or result obtained is checked, wherever possible, using the algebraic method, and depending on the result of this calculation, the solution to the system is decided upon.

(6) Information about the process is prepared for the output stage and subsequent presentation to the user or is transferred to another system.

Finally, the Output stage is responsible for presentation of the information corresponding to the reaction. This information may be presented to the user in a suitable form and/or transferred to another system for processing.

2.1. Structure of the Reaction. The inorganic reaction is represented by a complex data structure called the RAS (reaction abstract structure), composed of a series of other structures forming a network of information objects. The nodes of this network represent the reaction at various levels of abstraction and are interconnected using a series of pointers representing the various physicochemical relationships and structures found in the different components (compounds, subcompounds, and atoms) taking part in a reaction.

At a high level of abstraction, a chemical reaction of the form $r_1R_1 + r_2R_2 + ... \rightarrow p_1P_1 + p_2P_2 + ...$ may be represented by the structure in Chart 1.

Chart 1

> **Define RAS (Reaction Abstract Structure)**
>   *information* := **ReactionInf*,
>   /* *pointer to a structure containing information about the reaction, such as: identification of the reaction, medium in which it occurs, etc.* */
>   *reactants* := **RLS*,
>   /* *pointer to the first element on the list of reactants* */
>   *products* := **PLS*,
>   /* *pointer to the first element on the list of products* */
>   *umls* := **UMLS*,
>   /* *pointer to the first element on the list of matter imbalance* */
>   *uvls* := **UVLS*,
>   /* *pointer to the first element on the list of valence imbalance* */
>   *els* := **ELS*,
>   /* *pointer to the first element on the list of atoms taking part in the reaction* */
>   *scls* := **SCLS*,
>   /* *pointer to the first element on the list of subcompounds taking part in the reaction* */
>   *cls* := **CLS*,
>   /* *pointer to the first element on the list of compounds taking part in the reaction* */
>   *rc* := integer, /* *total charge of the reactants side* */
>   *pc* := integer, /* *total charge of the products side* */
> **End Structure RAS.**

Chart 2

> **Define Structure RLS (or PLS)** *(Reactants or Products List Structure)*
>   **ctcs** := **CTCS*,
>   /* *pointer to the CTCS structure of the compound* */
>   **next** := **RLS*,
>   /* *pointer to the next element on the list* */
>   **prior** := **RLS*,
>   /* *pointer to the previous element on the list* */
> **End Structure RLS (or PLS).**

The reactants and products taking part in the reaction are held in lists called RLS (reactants list structure) and PLS (products list structure), within which order is unimportant. These are circular lists in which each compound is represented by its CTCS structure[18] and a series of pointers designed to permit navigation through the species present in a reaction. These lists have the structure given in Chart 2 (see Figure 2).

3. ANALYSIS OF THE REACTION

The Analysis stage is responsible for determining the features of the reaction under study. Analysis is carried out of the state of the inorganic reaction, the latter being treated as an algebraic system (of n variables and m unknowns), and sufficient information is extracted for the reaction stage to "decide" upon the steps to be taken for the tracking and adjustment of the reaction.

This stage is comprised of the following steps, as illustrated in Figure 1.

Analysis of Matter. Here, the matter present in the reaction is analyzed, and one or more of the following results are obtained:

(a) Matter is balanced; that is, both sides of the equation contain the same atoms, although not necessarily in the same numbers.

(b) The atoms present in the reactants, but not in the products, are listed.

(c) The atoms present in the products, but not in the reactants, are listed.

Analysis of Charge. In this step, the charge of the species is examined and the overall charge is calculated for each side of the reaction.

Analysis of Valence. Analysis is performed for each type of atom present in both sides of the reaction, if there is a change in its coordination state. The result of this step is a
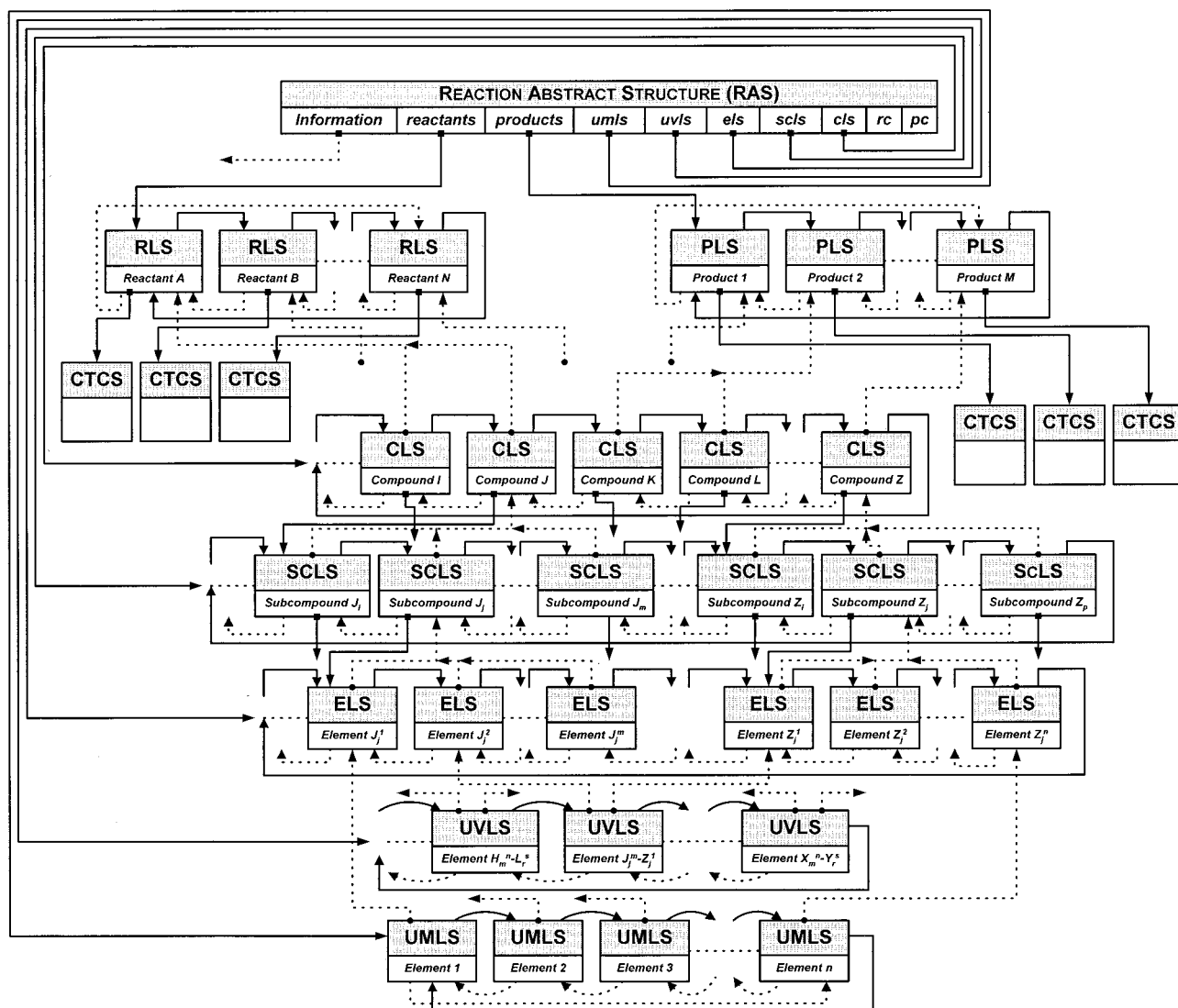
Figure 2. Context diagram of RAS and other structures used to represent inorganic reactions.

list of those atoms exhibiting such a change and the values of their coordination states for both sides of the reaction.

When carrying out this analysis, it is desirable to represent information in such a way that not only will the process be successful, but also the correct performance of the subsequent Reaction stage will be ensured.

Although the chemical reaction will take place according to the nature of the subcompounds and ions present, analysis of matter and valence requires examination of the actual atoms involved, on both sides of the reaction.

For this reason it is advisable to represent all the chemical components at different abstraction levels (compounds, subcompounds, or ions and atoms), by means of independent structures, as follows:

(1) The structures should be built by a single inspection of the CTCS structures representing the compounds taking part in the reaction.

(2) They should be completely linked to the CTCS from which they are extracted and thus connected to the RAS structure representing the reaction.

(3) The elements of these structures should be easily accessible, modifiable, and capable of interconnectivity according to their type, features, and component elements.

These structures, displayed in simplified form (for reasons of legibility) in Figure 2, are designed to represent atoms or chemical elements taking part in the reaction, and which represent the tuples of the elements of the CTCS of reactants and products (this structure may be designed as shown in Chart 3); subcompounds or ions, forming part of the compounds in the reaction, and which may be represented as shown in Chart 4; chemical compounds, or a list of the reactants and products taking part in the reaction, which may be represented as shown in Chart 5.

These three structures are interrelated via the reaction structure (RAS), by three pointers, one for each structure. Figure 2 contains a context diagram of these three structures and their relationship with the RAS, and it may be seen that while they are built by means of a simple examination of the CTCS tables of reactants and products, analysis may be carried out on the matter, valence, and charge involved in the reaction.

This analytical process gives rise to two new related data structures, as shown in Figure 2, which allow us to represent the results of the analysis; these new structures may be defined as given in Chart 6.

Chart 3

> **Define Structure ELS** *(Elements List Structure)*
> > */* Circular list linking atoms of each subcompound and like atoms of different subcompounds from both sides of the reaction. */*
> > **past :=** boolean,
> > > */* flag set if the element has been adjusted in the balancing or adjustment process */*
> > **multiply** := integer,
> > > */* value for the number of occurrences of the element after adjustment, initially set to 1 */*
> > **change** := boolean,
> > > */* flag set if valence has changed */*
> > **type** := character,
> > > */* flag indicating whether an atom is from a reactant or from a product */*
> > **category** := integer,
> > > */* category of the element in the subcompound (the **category** attribute in the **CTCS**) */*
> > **content** := integer,
> > > */* identifier for the chemical element (the **content** attribute in the **CTCS**)*/*
> > **valence** := integer,
> > > */* valence of the chemical element (the **valence** attribute in the **CTCS**) */*
> > **charge** := integer,
> > > */* charge of the chemical element (the **charge** attribute in the **CTCS**) */*
> > **occurrence** := integer,
> > > */* number of occurrences of the element in the subcompound of the chemical compound (the **occurrence** attribute in the **CTCS**) */*
> > **nextels** := *ELS,
> > > */* pointer to the next element on the list */*
> > **priorels** := *ELS,
> > > */* pointer to the previous element on the list */*
> > **nextelement** := *ELS,
> > > */* pointer to the next element, of the same subcompound, on the list */*
> > **priorelement** := *ELS,
> > > */* pointer to the prior element, of the same compound, on the list */*
> > **nextatom** := *ELS,
> > > */* pointer to the next element on the list corresponding to the same chemical element */*
> > **prioratom** := *ELS,
> > > */* pointer to the previous element on the list corresponding to the same chemical element */*
> > **subcompound** := *SCLS,
> > > */* pointer to the element, on the list of subcompounds, to which it belongs */*
> **End Structure ELS.**

Chart 4

> **Define Structure SCLS** *(Subcompounds List Structure)*
> > */* A circular list linking the subcompounds of each compound to their component atoms across both sides of the reaction */*
> > **past :=** boolean,
> > > */* flag is set if the element has been analyzed */*
> > **type** := character,
> > > */* flag to indicate whether this is a subcompound of the reactants or of the products */*
> > **multiply** := integer,
> > > */* value for the number of occurrences of the subcompound after adjustment, initially set to 1 */*
> > **category** := integer,
> > > */* category or type of chemical subcompound (the **category** attribute in the CTCS)*/*
> > **charge** := integer,
> > > */* charge of the subcompound */*
> > **occurrence** := integer,
> > > */* number of occurrences of the subcompound in the chemical compound */*
> > **nextscls** := *SCLS,
> > > */* pointer to the next element on the subcompound list */*
> > **priorscls** := *SCLS,
> > > */* pointer to the prior element on the subcompound list */*
> > **nextsubcompound** := *SCLS,
> > > */* pointer to the next subcompound of the same compound on the list */*
> > **priorsubcompound** := *SCLS,
> > > */* pointer to the prior subcompound of the same compound on the list */*
> > **firstelement** := *ELS,
> > > */* pointer to the first element on the list of atoms in the subcompound */*
> > **compound** := *CLS,
> > > */* pointer to the element to which this belongs, on the subcompound list */*
> **End Structure SCLS.**

4. REACTION STAGE

The reaction stage is responsible for carrying out the physicochemical processing of the reaction according to its features. This stage involves a series of steps, as shown in Figure 1, that are designed to evaluate, break down, balance, and solve the reaction under study.

4.1. Evaluation Step. Depending on the outcome of the *Analysis* stage, the reaction may be classified, in keeping with the present model, into one of the following categories.

Chart 5

> **Define Structure CLS** *(Compounds List Structure)*
> > */* The elements in this structure are ordered according to the category and type attributes */*
> > **past :=** boolean,
> > > */* flag to indicate whether the element has been analyzed and whether its information may be altered */*
> > **multiply** := integer,
> > > */* value for the number of occurrences of the compound after adjustment; this value is initially equal to the occurrence attribute */*
> > **type** := integer,
> > > */* flag showing whether this is a reactant or a product */*
> > **category** := integer,
> > > */* category or type of compound */*
> > **charge** := integer,
> > > */* charge of compound */*
> > **occurrence** := integer,
> > > */* number of occurrences of the chemical compound before adjustment (value entered into the system), default value 1 */*
> > **nextcls** := *CLS,
> > > */* pointer to the next element on the list of compounds */*
> > **priorcls** := *CLS,
> > > */* pointer to the prior element on the list of compounds */*
> > **firstsubcompound** := *SCLS,
> > > */* pointer to the first element on the list of subcompounds forming the compound */*
> > **CTCS** := *RLS (or *PLS),
> > > */* pointer to the element on the list of reactants or products */*
> **End Structure CLS.**

Chart 6

> **Define Structure UMLS** *(Matter Imbalance List)*
> > */* Circular list representing matter imbalance */*
> > **element :=** integer,
> > > */* identifier for the atom or chemical element producing the imbalance */*
> > **position** := character,
> > > */* flag showing whether the imbalance lies in reactants or products */*
> > **nextumls** := *UMLS,
> > > */* pointer to the next element on the list */*
> > **priorumls** := *UMLS,
> > > */* pointer to the prior element on the list */*
> > **firstelement** := *ELS
> > > */* pointer to the first element (corresponding to the same atom) on the list of elements */*
> **End Structure UMLS.**

> **Define Structure UVLS** *(Valence Imbalance List Structure)*
> > */* Circular list representing valence imbalance, obtained through examination of the **charge** attribute of the **CTCS** */*
> > **reactantvalence** := *ELS,
> > > */* pointer to the list of atoms present as reactants */*
> > **productvalence** := *ELS,
> > > */* pointer to the list of atoms present as products */*
> > **nextuvls** := *UVLS,
> > > */* pointer to the next element on the list */*
> > **prioruvls** := *UVLS,
> > > */* pointer to the prior element on the list */*
> **End Structure UVLS.**

Type 1: The reaction is complete and may be solved with no need for further contemplation of chemical compounds. Matter and charge are balanced, and there are no changes in coordination valence in any atom taking part in the reaction.

Type 2: The reaction is partially complete; matter is balanced, but there is a valence change in the atoms present.

Type 3: The reaction is partially incomplete; there is no balance of matter and/or charge, but there may be a change in the valence of the atom(s) present.

Type 4: The reaction is incomplete; matter is unbalanced, and a redox reaction is present, apparently complete in itself, revealing an oxidation change along with another of reduction. Matter imbalance is, however, due to compounds different from those considered for the previous types.

Depending on reaction type, the evaluation step will opt for one of the following courses of action (see Table 1 and Figure 1). In the following examples, compounds responsible for matter imbalance are shown in shaded boxes.

Type 1: A simple reaction in which all compounds and ions taking part are present. Control is therefore passed on to the Adjustment step to balance and solve the reaction.

MODEL FOR TRACKING INORGANIC CHEMICAL REACTIONS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 3, 2000 **749**

Table 1. Steps Associated with Reaction Types

matter balance	matter imbalance	valence balance	system	action (step)
yes	none	yes	type 1	adjustment
no	none	no	type 2	breakdown
no	H_2O, H^+, OH^-	yes	type 3.1	balancing
no	H_2O, H^+, OH^-	no	type 3.2	breakdown/balancing
no	unknown	incomplete	type 4.1	balancing
no	unknown	unknown	type 4.2	balancing

An example of this type would be

$$H_2SO_4 + NaCl \leftrightarrow Na_2SO_4 + HCl$$
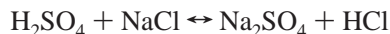
Type 2: A redox reaction (although another type of process might be present) in which all the atom types taking part are present. In this case, control is transferred to the Breakdown step to obtain and subsequently analyze half-reactions. An example of this type might be

$$Zn + HNO_3 \rightarrow Zn(NO_3)_2 + NH_4NO_3 + H_2O$$

Type 3: A reaction in which there is no balance of matter and/or charge but where this imbalance may be easily solved by considering a series of compounds or ions from a short list. Various types may occur depending on whether a redox process is present.

Type 3.1: No redox reaction is present, although the introduction of one or more of the following compounds or ions, namely, H_2O, H^+, or OH^-, may lead to a balanced reaction. In this instance, the Balancing step would be assigned to the reaction. An example would be

$$P_2O_5 + \boxed{H_2O} \rightarrow H_3PO_4$$

Type 3.2: There is a redox reaction or half-reaction where matter imbalance is due to the following compounds or ions: H_2O, H^+, or OH^-. Here, the reaction is dealt with in the same way as those of type 2, passing control to the Breakdown step or to the Balancing step in the case of a half-reaction. An example of this would be

$$HNO_3 + Cu \rightarrow Cu(NO_3)_2 + NO_2 + \boxed{H_2O}$$
$$CrI_3 + \boxed{OH^-} \rightarrow CrO_4^{2-} + IO_4^- + \boxed{H_2O}$$
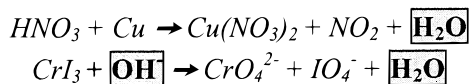
Type 4: An incomplete reaction with missing reactants, products, or both. We examine some reaction types falling within this group.

Type 4.1: There is a redox reaction, but here there is a deficit of atoms that cannot be directly obtained by adjustment of redox half-reactions; that is, it is due to species different from those described for type 3.2. There is only one atom in the reaction that exhibits changed valence, so matter imbalance is due to just one atom undergoing a redox process. An example would be

$$CoO + \boxed{H_2} \rightarrow Co + H_2O$$
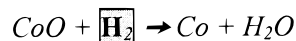
Type 4.2: In this reaction there is an imbalance of reactants and/or products that is difficult to determine directly. The reaction may or may not be redox, and where it is in fact redox, it is complete, exhibiting both oxidation and reduction processes, meaning that the imbalance is due

to compounds not taking part in the redox processes. An example of this type for when no redox process is present would be

$$AlCl_3 + \boxed{KBr} \rightarrow KCl + AlBr_3$$
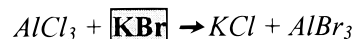
and for when a redox process is present, apparently complete

$$KNO_3 + C + \boxed{S} \rightarrow K_2S + N_2 + CO_2$$
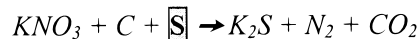
Where the reaction is classed as type 4, its information is passed on to the Balancing step, which will be briefly described later in this study, and whose purpose is to add reactants and/or products to the reaction so that it may be reclassified as one of the other types.

4.2. Breakdown Step. The purpose of this step is to break down a complete redox reaction, that is, one in which there are at least two valence changes (one by oxidation, the other by reduction) in one or more of the atoms taking part.

In this step, the redox half-reactions are built. These half-reactions contain, on both sides, the subcompounds or ions whose atoms undergo valence changes. They are represented in a data structure akin to the RAS, named the RASredox.

Once the RASredox has been built, the information is sent back to the Analysis stage to obtain the imbalance of matter, valence, and charge for each half-reaction.

4.3. Adjustment Step. This step is responsible for the balancing of valence and charge on each side of the global reaction of its redox half-reactions, where present.

The following actions may be performed, depending on reaction type: (A) for redox reactions, (A.1) adjustment of each half-reaction (oxidation and reduction), (A.2) composition of the global redox reaction, based on adjusted half-reactions, and (A.3) adjustment of the reaction based on information extracted from the adjusted global redox reaction, and (B) for non-redox reactions, (B.1) adjustment of the reaction.

The various adjustment processes mentioned above are performed using two different techniques: analytical, where the reaction is represented by a system of linear equations to be solved by a simple matrix calculation; manual, where the system is adjusted by making use of the abstract representations of the reaction and following the traditional manual or empirical method.

4.3.1. Analytical Adjustment. Analytical adjustment is performed using the widely known matrix method,[23-26] which is based on the consideration that the species present in the reaction such as entities of the form $X_a Y_b^c$, where X and Y are atomic elements, a and b, are the subscripts or occurrences with which they are present in the species, and c is the charge of the species in question. Using this representation, a reaction of the form

$$o_1 \text{Reactant}_1 + o_2 \text{Reactant}_2 + \cdots = o_j \text{Product}_j + o_k \text{Product}_k + \cdots$$
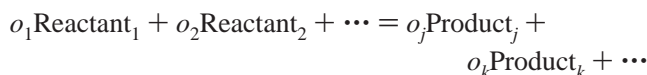
and taking the principle of atom balance (AB) into account, it may be written as $SV = 0$, where S is the stoichiometric matrix $A \times C$, in which A is the number of different elements in the reaction and C the number of species, and V is the coefficient matrix (o_i) that ensures the reaction is balanced.

Table 2. Percentage of Different Reaction Types

relationships[a]	presence, %	example reaction
$E = A; E = C - 1$	62.7	$NH_3 + O_2 \rightarrow NO_2 + H_2O$
$E = A - 1; E = C - 1$	21.6	$KClO_3 \rightarrow KCl + O_2$
$E = A - 2; E = C - 1$	9.7	$Fe_2(SO_4)_3 + Ba(OH)_2 \rightarrow Fe(OH)_3 + BaSO_4$
$E = A; E = C - 2$	4.5	$KMnO_4 + H_2O_2 + H_2SO_4 \rightarrow K_2SO_4 + MnSO_4 + H_2O + O_2$
$E = A - 3; E = C - 1$	1.5	$(PNCl_2)_3 + KSO_2F \rightarrow (PNF_2)_3 + KCl + SO_2$

[a] **A**, number of elements; **C**, number of compounds (reactants + products); **E**, number of linearly independent equations.

The solution to this reaction using an algebraic method will depend on the relationship between the range of matrix **S** and the number of linearly independent equations. Thus, if we let E be the number of linearly independent equations in **S**, we have the following:

(1) If range(S) = $E - 1$, then it is a determinate compatible system and has one simple solution.

(2) If range(S) = $E - 1$, then it is an indeterminate compatible system and has infinite solutions depending on the values of a coefficient o_i, and these solutions are a linear combination.

(3) If range(S) < $E - 1$, then the reaction has n families of solutions (where $n > 1$), depending on the number n of o_i coefficients.

Subramaniam et al.[26] has studied a wide range of inorganic reactions, the results of which are shown in Table 2. It is clear that most of these reactions may be easily solved by the algebraic method, since there are enough linearly independent equations for their solution; only a small percentage require consideration of more than one parameter or coefficient (o_i) to find a solution.

This study by Subramaniam et al. is incomplete, since there exist more complex equations in which E-range(S) > 1, thus calling for new equations, as proposed by Olson.[24] These equations may be as follows:

(1) A new equation represents the principle of charge balance (CB).

(2) In the case of redox reactions, a series of equations represents the principle of electron balance (EB). Depending on the reaction, up to six equations may be obtained for the global reaction, shown diagrammatically as[24]

$$\sum_{OX.} \text{reactants} = \sum_{RED.} \text{reactants};$$
$$\sum_{OX.} \text{products} = \sum_{RED.} \text{products};$$
$$\sum_{OX.} \text{reactants} = \sum_{RED.} \text{products}$$
$$\sum_{OX.} \text{products} = \sum_{RED.} \text{reactants};$$
$$\sum_{RED.} \text{reactants} = \sum_{RED.} \text{products};$$
$$\sum_{OX.} \text{reactants} = \sum_{OX.} \text{products}$$

or just four equations, in the case of redox half-reactions; two for the oxidation half-reaction

$$\sum_{OX.} \text{reactants} = n; \quad \sum_{RED.} \text{products} = n$$

and another two for the reduction half-reaction

$$\sum_{RED.} \text{reactants} = n; \quad \sum_{OX.} \text{products} = n$$

where $\sum$ is the summation of the products of the occurrences of the elements by the electrons they contribute to the oxidation (OX.) or reduction (RED.) process and n is the number of electrons transferred in the half-reactions.

While in some cases consideration of the CB and EB principles only yields equations that are linear combinations of those obtained by the AB principle, on other occasions it will provide equations that are easily solved by simple algebraic techniques.

Analytical adjustment performed by the present system is based on the following:

(1) A series of equations is obtained from the application of the AB, CB, and EB principles, the latter being applied to redox systems.

(2) The system **SV** = 0 is built and the range of the **S** matrix determined, leading to solution of the reaction using the Gaussian technique.[27] Where E < range(S), a coefficient, V_i, is chosen, solving the reaction accordingly.

(3) An analytical solution is obtained for the global reaction and for each of the half-reactions, in the case of redox reactions.

4.3.2. Manual Adjustment. Analytical adjustment will successfully perform the adjustment of reactions, and is fast and computationally accurate. Nevertheless, for the development of systems such as that proposed above, it is highly recommended that the user have access, when desired, to all available information concerning the chemical−mathematical process of adjustment, since in many chemistry courses the student may not have a high enough level to fully understand the algebraic methods employed in the analytical adjustment process. Furthermore, these methods are indispensable for the adjustment of the global reaction after obtaining a partial result from the sum of the adjusted redox half-reactions using either of the two techniques (manual or analytical).

Manual adjustment, known as the ping-pong or inspection method, is the technique most commonly used in the majority of chemistry courses, whenever a computer system is not employed for this purpose, so its coverage will favor the familiarization of users with software systems that embrace it.[23,24]

This method, while usually slower than the analytical method, almost always produces the same results as the analytical method, but provides the user with more chemical information. However, if it includes a suitable representation of the chemical reaction, as proposed above, its speed disadvantage is considerably offset, particularly when carried

MODEL FOR TRACKING INORGANIC CHEMICAL REACTIONS

J. Chem. Inf. Comput. Sci., Vol. 40, No. 3, 2000 **751**

out in conjunction with the application of a series of basic rules, such as those described below.

Manual adjustment should be performed making use of the ELS, SCLS, and CLS, using the ELS to identify the elements that only appear in one species in the reaction, and starting the process from that point. Where there is more than one such element, the process should begin, where possible, with the element that is present in the species with the highest number of different atoms.[23]

During the adjustment process, values for the past attributes of these structures are continuously updated, reporting that the element/subcompound/compound has been fixed or balanced, and similarly, values of the multiply attribute are updated with the corresponding multiplier value that ensures the balancing of the reaction.

4.4. Balancing Step. This step is where matter in the reaction is balanced. Information is fed in from the evaluation step, represented by RAS and RASredox structures; the reaction has now been classified by one of the preestablished types, and knowledge has been extracted regarding material imbalance. The balancing step, when required by the evaluation step, works very closely with the adjustment process.

Incoming information is processed according to a series of rules that determine the species to be considered on one or both sides of the reaction in order to balance the material.

The balancing of charge (electrons) is carried out directly in the Adjustment step by examination of the electron species (e^-), which may be represented in the same way as any other chemical species using the CTCS.[18] Balancing is performed according to the values for the rc and pc attributes of the RASredox of redox reactions or half-reactions.

Balancing the material is equally simple when dealing with Type 2 or 3 reactions, since here only the H_2O, H^+, and OH^- species need to be examined. In such cases the following simple rules are applied, taking into account the presence of ion species in the reaction, the species from the above list present on both sides of the reaction, information on the reaction medium (covered in the information attribute of the RAS), etc.

For type 4 reactions, somewhat more complicated rules must be applied. These rules must take into account, in addition to information supplied by the RAS, chemical information about the reaction in question and accordingly describe a knowledge frame that will allow the determination of the species or set of species to be dealt with in the reaction. This knowledge frame and the balancing process for type 4 reactions will be discussed in a forthcoming article in this Journal.[28]

5. DISCUSSION

Powerful software for use in computer-aided teaching of chemistry must satisfy not only a series of technological and efficiency requirements but also a number of demands imposed by this field, such as (a) the system must have the ability to represent chemical information exactly as it is depicted in the real world, (b) it should be open and flexible enough to support information supplied by the user, and (c) it must be capable of representing knowledge and passing this on to the user.

The first step toward satisfying these requirements is that the software should include resources to represent any species and any chemical reaction, whether defined by the developer or entered by the user.

The present article has presented an architectural and structural model that allows the representation of inorganic reactions, employing abstract data structures which enable a complete, computationally efficient representation of both reactants and their intra- and intermolecular relationships within a reaction. The system can balance complete reactions as well as some incomplete ones (type 3), producing the solution, a description of the steps to be taken, and the species needed to balance the reaction.

Balancing of the reaction is performed in every case, both analytically and manually, on the global reaction and on redox half-reactions, thus ensuring the verification of the solution without incurring significant additional computational costs. Manual adjustment by the inspection method enables the user to be shown a step-by-step description of the process just as it would generally be carried out, as well as completing the process when analytical analysis is performed on redox half-reactions.

In the case of totally incomplete reactions (type 4), the model is open to the consultation of a knowledge frame by means of which solutions may be proposed with a certain degree of accuracy to ensure the completion of the reaction. This extensive knowledge frame, to be described in a future study, is based on the type of chemical process and features of the reaction and also on the species present.

REFERENCES AND NOTES

(1) Lagowski, J. J. Chemical Education: Past, Present, and Future. *J. Chem. Educ.* **1998**, *75* (4), 425−436.

(2) Johnson, J. F.; Morris, K. M. Courseware is the Key: Bridging Lecture and Laboratory in General Chemistry. *J. Comput. High. Educ.* **1997**, *8* (2), 109−125.

(3) Tissue, B. M. The Costs of Incorporating Information Technology in Education. *http://www.chem.vy.edu/archive/chemconf97/paper04.html*.

(4) Sculley, J. *Interactive Multimedia. Visions of Multimedia for Developers, Educators & Information Providers*; Microsoft Press: 1988.

(5) Hippe, Z. *Artificial Intelligence in Chemistry*; Elsevier: Amsterdam. 1991.

(6) Lawson, A. E. *Science Teaching and the Development of Thinking*; Wadsworth: Belmont, CA, 1995.

(7) Immirzi, A. Using Symbolic Problems in Teaching Chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 780−784.

(8) Karizo, R. W. Mathematica-Assisted Learning in Physical Chemistry. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 96−103.

(9) Hippe, Z. S. Design and Application of an Intelligent Information System SCANKEE for Solving Selected Chemical Problems. *Comput. Chem.* **1998**, *22* (1), 133−140.

(10) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Error Detection, Recovery, and Repair in the Translation of Inorganic Nomenclatures. 1. A Study of the Problem. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 7−15; 2. A Proposed Strategy. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 16−24; 3. An Error Handler. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 483−490.

(11) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. A Tutorial System for Inorganic Chemical Formulation and Reactions. *Proceedings of the 1994 ACM Symposium on Applied Computing*; ACM Press: 1994; pp 583−587.

(12) (a) *Enzyme Lab.*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_D/5D1/prog1−5d1.html. (b) *Chemistry Comes Alives!*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_SP/SP18/prog1-SP18.html. (c) *ChemDemos*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_SP/SP8/abs-sp8.html, (c) GC Instrument Simulator, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_99/9901W/abs-9901.html. (d) *Equilibrium Calculator*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_D/1D1prog3−1D1.html. (e) *Frost Diagram*, http://jchemed.chem.wisc.edu/JCEsoft/Issues/Series_C/4C1/prog2−4C1.html.

(13) Rzepa, H. S.; Tonge, A. P. VchemLab: A Virtual Laboratory. The Storage, Retrieval, and Display of Chemical Information Using Standard Internet Tools. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1048−1053.

(14) Bell, J. T.; Fogler, H. S. Vicher: A Virtual Reality Based Educational Module for Chemical Reaction Engineering. *Comput. Appl. Eng. Educ.* **1996**, *4* (4), 285−296.

(15) Yarger, J. L.; Chizmeshya, A.; Yarger, B. K. Using Java to Animate the Vibrations of Molecules: Calculation and Visualization of Molecular Vibrations in (NSF)3. *http://journals.springer-ny-com/sam-bin/swilma/cmp.840562295.html.*

(16) Scudder, P. H. Database vs. Expert System Teaching Paradigms: Using Organic Reactions Mechanisms To Teach Chemical Intuition. *J. Chem. Educ*. **1997**, *74* (7), 777−781.

(17) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Inorganic Chemical Knowledge Representation Using Dynamic Data Structures. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 378−384.

(18) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Computer Translation of Inorganic Chemical Nomenclature to a Dynamic Abstract Data Structure. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 526−533.

(19) Luque Ruiz, I.; Cruz Soto, J. L.; Gómez-Nieto, M. A. Knowledge Representation through Objects in the Development of Expert Systems for Chemical Synthesis and Reactions. In *Advanced Methods in Artificial Intelligence*; Lecture Notes in Computer Science 682; Springer-Verlag: Berlin, 1993.

(20) Cooke-Fox, D. I.; Kirby, G. H.; Rayner, J. D. Computer Translation of IUPAC Systematic Organic Chemical Nomenclature. 4. Connection Tables to Structure Diagrams. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 122−127.

(21) Eggert, A. A.; Jacob, A. T.; Middlecamp, C. H. Converting Chemical Formulas to Names. As Expert Startegy. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 223−237.

(22) Brecher, J. S. Name=Struct: A Practical Approach to the Sorry State of Real-Life Chemical Nomenclature. *J. Chem. Inf. Comput. Sci.,* in press.

(23) Herndon, W. C. On Balancing Chemical Equations: Past and Present. A Critical Review and Annotated Bibliography. *J. Chem. Educ.* **1997**, *74* (11), 1359−1362.

(24) Olson, J. A. An Analysis of the Algebraic Method for Balancing Chemical Reactions. *J. Chem. Educ.* **1997**, *74* (5), 538−542.

(25) Philips, J. C. Algebraic Constructs for the Graphical and Computational Solution to Balancing Chemical Equations. *Comput. Chem.* **1998**, *22* (4), 295−308.

(26) Subramaniam, R.; Goh, N. K.; Chia, S. The Relationship between the Number of Elements and the Number of Independent Equations of Elemental Balance in Inorganic Chemical Equations. *J. Chem. Educ.* **1995**, *72* (7), 894−895.

(27) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical recipes in C. The art of scientific computing*, 2nd ed.; Cambridge University Press: 1993 *(http://cfata2.harvard.edu/numer-ical-recipes/).*

(28) Luque Ruiz, I.; Gómez-Nieto, M. A. Solving Incomplete Inorganic Chemical Systems through a Fuzzy Knowledge Frame. *J. Chem. Inf. Comput. Sci.,* submitted for publication.