

“Big Bang” in the Undergraduate Chemistry Curriculum via Symbolic Computation

Mihai Scarlete

Bishop's University, Sherbrooke, Québec

mscarlet@ubishops.ca

Gavin Heverly-Coulson¹, Amber Findleton², Starr Dostie³

Abstract – The modern delivery of concepts in physical chemistry can now take advantage of the integration of symbolic computation engines. The advancement of the friendly user-interfaces of the existent packages open to dedicated chemists the programming capacity for the creation of precise, digital definitions for most of the core notions in physical chemistry. Basic concepts such as orbitals, molecular dynamics, vibrational reaction coordinate, Stirling-compliant distribution models, thermodynamic probability and statistical entropy, etc. can now be readily calculated for medium-populated chemical systems by using the computation power of the computation engines, rather than only suggested via pictures or highly approximate calculations on the blackboard. As a result, the undergraduate curriculum can be expanded to include concepts previously introduced only in the graduate curriculum, and even subjects at the frontier of science - research objects. The impact on students is instantaneous, as they can now be equipped with tools matching the modelling/computation power utilized by high calibre researchers only a few decades ago. This paper presents the pedagogical and research results obtained by the implementation of the CHEMLOG educational system in the (under)graduate curriculum. The CHEMLOG system is based on the utilization of a symbolic computation engine interfaced with a database of chemical concepts regularly updated with the newest research results reported in the

literature in the field of physical chemistry. The analysis covers a 5-year period of classroom-delivery, as well as the analysis of the online-setup covering more than one million requests from the CHEMLOG server since 1999.

Index Terms – Self-extracting digital libraries, CHEMLOG thermodynamic database, chemical equilibrium, symbolic computation engines, server-mediated problem solving approach.

INTRODUCTION

The utilization of powerful symbolic computation engines (SCE) for the undergraduate education has resulted in a wide opening of fields previously reserved for graduate instruction.ⁱ Consequently, the analysis of chemical systems and processes, subjects usually restricted to highly specialized researchers, became a common object for the undergraduate instruction.ⁱⁱ The positive results of this approach have induced a rapid increase in the number of instructors adapting their teaching habits to the new environment. This paper presents the results obtained from the implementation of a set of internally-consistent databases comprising the numerical representations of basic concepts in the study of physical chemistry. We shall introduce the “modular template-design”, a procedure allowing graphing on an initial digital-core containing basic concepts, of successive modules of gradually increased complexity. The design of such templates not only ensures full comprehension of the correct basic concepts, but it can be taken to a higher level of utilization, which eventually results in a research-tool, allowing exploration of the limits of existent models. In order to emphasize the “Big Bang” mentioned in the title of this paper, we have selected a few examples, to prove the efficiency of the SCE in promoting research objects within the undergraduate education.

¹ Gavin Heverly-Coulson, Bishop's University, gheverly04@ubishops.ca

² Amber Findleton, Bishop's University, afindleton05@ubishops.ca

³ Starr Dostie, Bishop's University, sdostie03@ubishops.ca

1. Module for Enhanced Resolution of Oscillatory Processes

Classic analysis of mixed spectroscopic signals is now part of any basic spectroscopy course. Unavoidable part in the design of modern spectroscopic equipment, the integral transformation packages are also useful in other domains, such as the study of complex kinetic patterns of oscillatory reactions.ⁱⁱⁱ One such example is the relatively recent study of polysilane/NH₃ system, already reported elsewhere. We shall present here a version of the mentioned system of a reduced complexity down to a parameterized Lotka-system defined via the following set of equations:

```
>sys:=diff(a(t),t)=-0.1*(2-a(t))*(Xo+x(t)),diff(x(t),t)=-
1*(Xo+x(t))*(Yo+y(t)),diff(y(t),t)=-1.5*(Yo+y(t)),
diff(b(t),t)=1.5*(Yo+y(t)); fcns:={a(t), x(t), y(t), b(t)}:
```

The design of the above system allows the plotting of the concentration of one intermediate versus time, via

```
> p:=dsolve({sys, a(0)=2, y(0)=0, x(0)=0, b(0)=0}, fcns,
type=numeric); odeplot(p,[[t,y(t)]],10..300,numpoints=
50);
```

The second elementary step can be perturbed through the sequential parameter “n”:

```
>for n from 1 to 5 do sys:=diff(a(t),t)=-0.1*(2-
a(t))*(Xo+x(t)), diff(x(t),t)=-1*(Xo+x(t))*(Yo+y(t)),
diff(y(t),t)=-1.5*n*(Yo+y(t)), diff(b(t),t)=1.5*(Yo+y(t));
fcns:={a(t), x(t), y(t), b(t)}; p[n]:=dsolve({sys, a(0)=2,
y(0)=0, x(0)=0, b(0)=0}, fcns, type=numeric);
P[n]:=odeplot(p[n], [[t, y(t)]], 10..300, numpoints=50,
color=blue, axes=boxed, title='Reaction Voltéra',
color=COLOR(RGB,rand()/10^12,rand()/10^12,rand()/1
0^12)); od:
```

We shall present below two cases, in order to show increased complexity in the concentration of the intermediate, induced by the perturbation.

Fig.1 below shows the concentration of the intermediate vs. time for n=2.

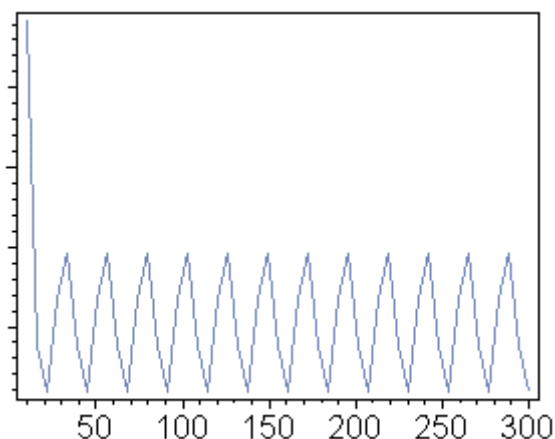


Figure 1

The deconvolution analysis of the system can now be performed from the numerical definition of the plot, and the resolution can be adjusted through the “numpoints”-counter, to any level desired by the analyst. The set of parameters chosen for the system in the case above results in a simple oscillatory behavior, and there is no difficulty associated in this case in assigning the oscillatory parameters. However, the analysis of n=3 case involved a more complicated

situation, The concentration of the intermediate in this case is presented in Fig. 2.

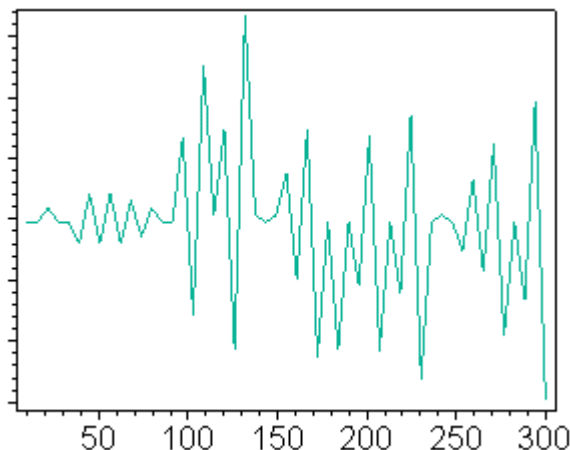


Figure 2

A simple oscillatory pattern cannot be associated to the new set of parameters, showing a more chaotic behavior. The early stages of this transition can still be analyzed by assuming interference with oscillatory-parasitic phenomena, and n=3 is one such case.

The focus on the initial stages of the perturbed oscillatory reaction has been obtained limiting the observed signal within the time frame of 0-80 units. The wave packet resulted in this crop, has been overlaid to a series of Gaussian-type signals. A degree of correlation of 95% has been accepted as sufficient for the emulation of the perturbed oscillating system, and the minimum number of Gaussian signals has been accepted as being the total number of interfering processes in the observed system.

We have started with the basic definition of a Gaussian wave packet part of any basic spectroscopy course, with the form-coefficient left as an independent variable (N.B.: however, the use of any type of signal, such as Lorentzian-response, significant for the analysis of NMR signals) e.g. through a mixing procedure schematically shown below.:

```
> g:= exp(-x^2/(FormFactor.sigma^2))*sin(2*Pi*x);
```

Two independent “FormFactor”-parameters have been associated to two such Gaussian-signals in Fig. 3, and the mixing of the two plots is presented in Fig. 3

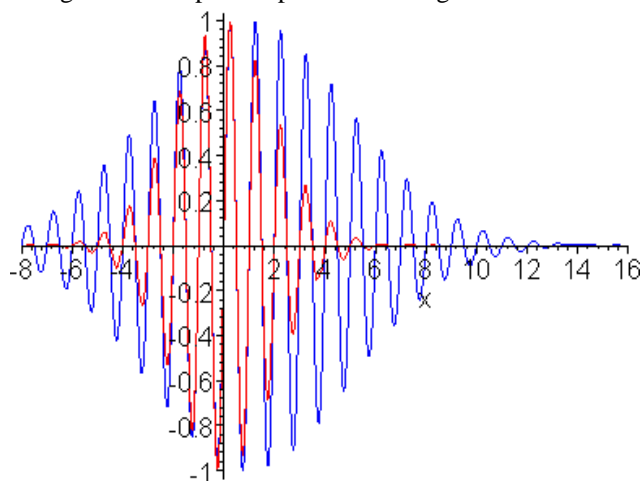


Figure 3

The overlay procedure was based on a least square method applied to the obtained numerical data on the concentration of the intermediate in the incipient chaotic stage, and on the predicted mixed Gaussian-signals. The deconvolution procedure was based on the Fourier Transform analysis of the mixed signal, transferred into the frequency domain. The integral transformation packages are part of any SCE ("intrans", in this case), so loading of the package and calling the Fourier transformation, results in the deconvolution of the mixed signal, graphically represented in Figure 4:

```
>with(intrans): h:=fourier(g,x,k); evalc(Im(h)); plot(-Im(h),k=0..10);
```

FT Spectrum of Gaussian Signals

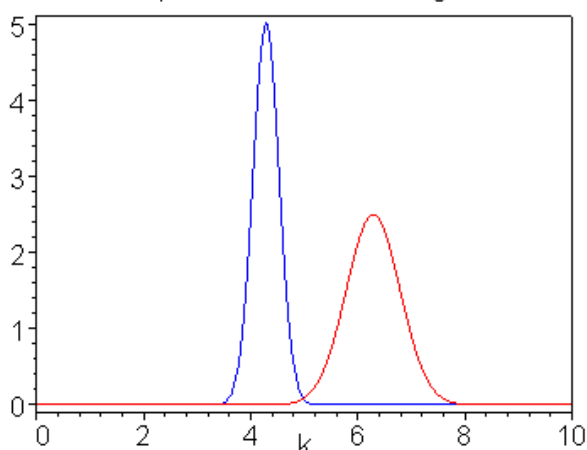


Figure 4

In the case of $n=4$ (Fig. 5), the transition to a chaotic regime is more advanced, and the analysis using 6 mixed Gaussian-signals did not result in an acceptable correlation with the

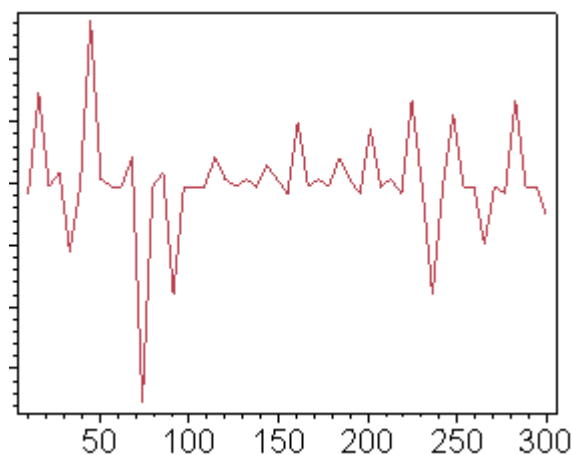


Figure 5

numerical data on the concentration of the intermediate, implying that a larger number of parasitic signals are needed to emulate the reaction.

2. From Quantum Numbers to Molecular Modeling.

The difficulty associated with the introduction of molecular modeling procedures into the undergraduate curriculum resides in the computation power required to transfer into a manageable library of the basis functions, and of the digital forms of the quantum field solutions under a

variety of boundary conditions. Moreover, even with these libraries in place, the interpretation of the results can be best achieved by the incorporation of the analytical result into a graphic user interface (GUI). Both these computation and graphical requirements are naturally incorporated in today's SCE packages. This section will present the results obtained by use of "self-extracting libraries" (a concept already presented in detail elsewhere) in the study of the parameterization of the relative weights in the Linear Combination of Atomic Orbitals (LCAO) used for molecular orbital calculations. The examples selected for this presentation show applications in the calculation of molecular orbital energy diagram (MOED) via Simple Huckel Molecular Orbital (SHMO) approximation, and the design of a GUI for hybrid subsets.

2.1. MOED VIA SHMO

Our MOED-calculation via SHMO^{iv} used the matrix algebra package. Our approach was to create a convenient template to describe molecular connectivity, illustrated below for simultaneous calculation of 4 molecules, as shown in Fig.6

```
>restart:with(linalg):SecMat:=array(1..4,1..2):SecMat[1,1]=`C2H4`:SecMat[2,1]=`C4H8`:SecMat[3,1]=`C6H12`:SecMat[4,1]=`C8H16`: for m from 2 to 8 by 2 do; SecularMatrix_Linear:=array(sparse, 1..m,1..m); for i from 1 to m do SecularMatrix_Linear[i,i]=x; od; for i from 2 to m-1 by 1 do for j from i-1 to i+1 by 2 do SecularMatrix_Linear[i,j]=1; od; od;".
```

$$\begin{array}{l}
 \text{C2H4} \\
 \text{C4H8} \\
 \text{C6H12} \\
 \text{C8H16}
 \end{array}
 \left[\begin{array}{cccccccc}
 \begin{array}{cc} x & 1 \\ 1 & x \end{array} & & & & & & & \\
 \begin{array}{cccc} x & 1 & 0 & 0 \\ 1 & x & 1 & 0 \\ 0 & 1 & x & 1 \\ 0 & 0 & 1 & x \end{array} & & & & & & & \\
 \begin{array}{cccccc} x & 1 & 0 & 0 & 0 & 0 \\ 1 & x & 1 & 0 & 0 & 0 \\ 0 & 1 & x & 1 & 0 & 0 \\ 0 & 0 & 1 & x & 1 & 0 \\ 0 & 0 & 0 & 1 & x & 1 \\ 0 & 0 & 0 & 0 & 1 & x \end{array} & & & & & & & \\
 \begin{array}{cccccccc} x & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & x & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & x & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & x & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x \end{array} & & & & & & &
 \end{array}
 \right]$$

Figure 6

Secular determinants for each case are simultaneously calculated in a loop, and embedded into a larger display structure via

```
>SecDet:=array(1..4,1..2):SecDet[1,1]:=`C2H4`:SecDet[2,1]:=`C4H8`:SecDet[3,1]:=`C6H12`:SecDet[4,1]:=`C8H16`
:for n from 1 to 4 do SecDet[1,2]:=det(matrix (C[n]
H[2*n+2])) od;
```

and displayed as in Fig. 7

$$\begin{bmatrix} \text{C2H4} & x^2 - 1 \\ \text{C4H8} & x^4 - 3x^2 + 1 \\ \text{C6H12} & -1 + 6x^2 - 5x^4 + x^6 \\ \text{C8H16} & 1 - 10x^2 + 15x^4 - 7x^6 + x^8 \end{bmatrix}$$

Figure 7

The energy levels are solved for each element of the Secular Determinant ("SecDet") matrix via

```
>Solutions:=array(1..4, 1..2): Solutions[1,1]:=`C2H4`:
Solutions[2,1]:=`C4H8`:Solutions[3,1]:=`C6H12`:Solution
s[4,1]:=`C8H16`: for jjj from 1 to 4 do Solutions[ jjj,2]
:=fsolve(SecDet[ jjj,2]=0, x); od;
```

returning the results in the table shown in Figs. 8a and 8b

C2H4		C4H8	
$\frac{\alpha - E_1}{\beta} = -1.000000000$	$\frac{\alpha - E_1}{\beta} = -1.618033989$		
$\frac{\alpha - E_2}{\beta} = 1.$	$\frac{\alpha - E_2}{\beta} = -.6180339887$		
0	$\frac{\alpha - E_3}{\beta} = .6180339887$		
0	$\frac{\alpha - E_4}{\beta} = 1.618033989$		
0	0		
0	0		
0	0		
0	0		

Figure 8a

C6H12		C8H12	
$\frac{\alpha - E_1}{\beta} = -1.801937736$	$\frac{\alpha - E_1}{\beta} = -1.879385242$		
$\frac{\alpha - E_2}{\beta} = -1.246979604$	$\frac{\alpha - E_2}{\beta} = -1.532088886$		
$\frac{\alpha - E_3}{\beta} = -.4450418679$	$\frac{\alpha - E_3}{\beta} = -1.$		
$\frac{\alpha - E_4}{\beta} = .4450418679$	$\frac{\alpha - E_4}{\beta} = -.3472963553$		
$\frac{\alpha - E_5}{\beta} = 1.246979604$	$\frac{\alpha - E_5}{\beta} = .3472963553$		
$\frac{\alpha - E_6}{\beta} = 1.801937736$	$\frac{\alpha - E_6}{\beta} = 1.$		
0	$\frac{\alpha - E_7}{\beta} = 1.532088886$		
0	$\frac{\alpha - E_8}{\beta} = 1.879385242$		

Figure 8b

2.2. PARAMETERIZED HYBRID SUBSETS

LCAO is a general procedure in molecular modeling, and access to its parameters is crucial for molecular modeling. Most of the commercial packages for molecular modeling offer only limited access to the mixing of the basis functions, usually as a limited number of prearranged modules, optimized for various classes of chemical compounds. We have created a few years ago a self-extracting library of atomic wavefunctions. In this section, we present the use of this library for parameterized hybrid subsets. Such an analysis, due to limited and highly specific applications, is not accessible in any existent commercial package, and we think that use of this simple and direct procedure may help professors, students, and researchers to understand the fluidity and the theoretical limits of the parameterized LCAO.

1. Model to be studied

In order to keep the focus on the procedure, rather than on exotic LCAO-derived systems, the example to be presented in this section is a classic hybrid system, extended to include the parameter "λ", where s, px, py represent the classic wavefunctions, embedded in the self-extracting library (not presented here).

```
>read(rtplib.m):psi[4]:= (S+lambdax)/sqrt(1+lambdax^2);psi[5]:= (lambdax+sqrt(1+lambdax^2)*py)/sqrt(
```

$2*(1+\lambda^2)$);psi[6]:=(lambda*Sp[x]sqrt(1+lambda^2)*p[y])/sqrt(2*(1+lambda^2));

A routine check of the normalization and mutual orthogonality of the subset can be performed to ensure validity of the LCAO-procedure. These tests in place, the analysis of the set, and the parameter-induced flexibility of the system can now be observed. Again, due to its highly intuitive insight, the exploration via GUI is the best option. The parameter “ λ ” is varied and the angular distribution of maximum electronic density can be studied directly in Cartesian coordinates. The continuous mixing of the atomic angular wavefunctions (AO’s) extracted from a pre-built self-extracting “rtplib.m”, returns the separate contributions of the AO’s (not displayed here), and the hybrid combination vs. λ , as an animation file. Three frames for $\lambda=0$, $\lambda=5$, and $\lambda=10$ are presented in Fig. 9 below, showing the continuous variation of the symmetry of the set from D_{3h} (9a, b) to D_{4h} (9c):

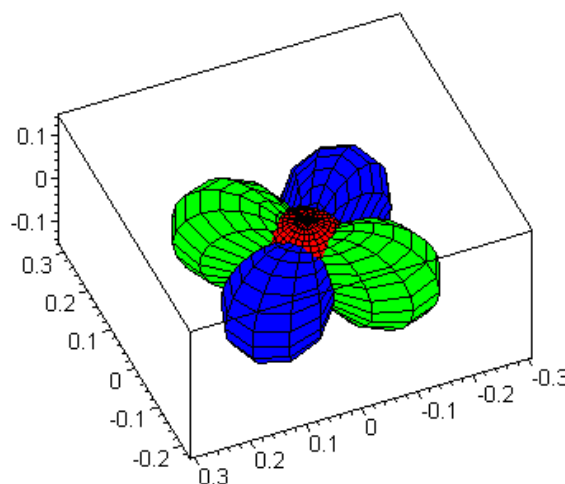


Figure 9c

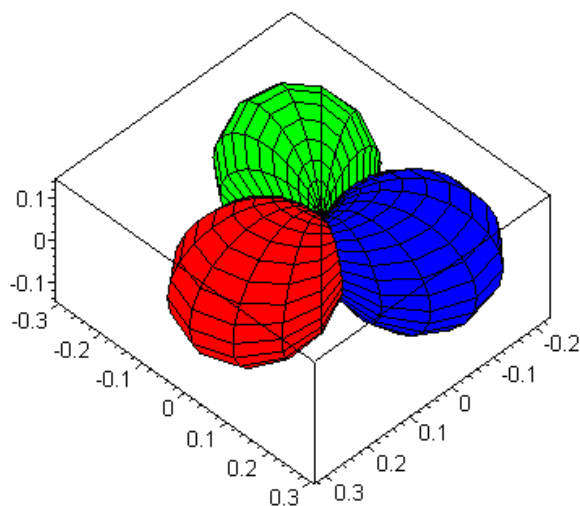


Figure 9a

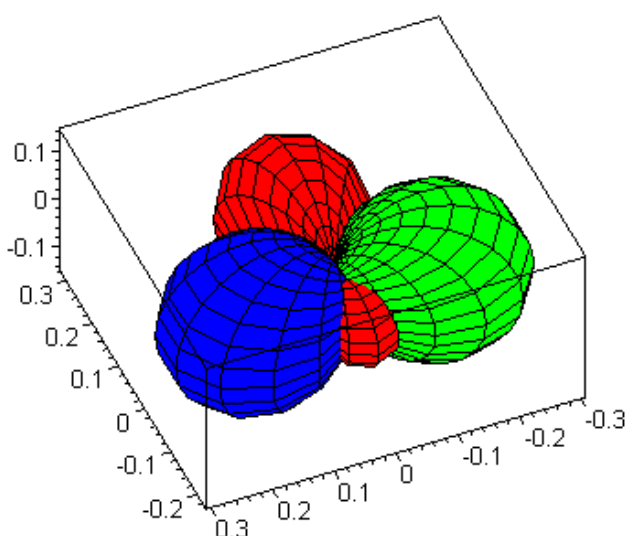


Figure 9b

3. EQUILIBRIUM COMPUTATION

One of the first applications of SCE in the undergraduate curriculum was the calculation of equilibrium states. The SCE approach becomes necessary in the case of systems of medium complexity, when the computation power required for the solving of the system becomes critical. Again, we’ll present here first a pedagogical-oriented case based on the direct use of equilibrium constants to reach the minimum total free energy of the system, in the case of automatic-calculation of pH in a solution of a triprotic acid, from the acidity constants. Then, we shall present the value of the SCE-assisted approach in the analysis of a more complex system, utilized in a recent research project.

In the first case, we have built a database with thermodynamic data (acidity constants, solubility products, instability constants for complexes, etc), in order to store in digital form and in a retrievable form this data (“TDlibrary”), while the mass balances are stored in the following procedure.

```
>read(TDlibrary); for j from 1 to 10 do c:=0.001*j;
T[j+1,1]:=c;solutions[j]:=solve({Ka1=(10^(-7)+x+y+z)*(x-
y)/(c-x), Ka2=(10^(-7)+x+y+z)*(y-z)/(x-y), Ka3(10^(-7)+x
+y+z)* z/(y-z)}, {x,y,z}); od: print(solutions);
```

The conversion of the equilibrium concentration into pH-values is readily obtained via a recurrent procedure as grouped in the table “T” in Fig. 10, as:

```
>for q from 1 to 10 do T[q+1,2]:=rhs(Solns[q +1,1])
+rhs(Solns[q+1,2])+rhs(Solns[q+1,3]); T[q+1,3]:=-log[10]
(10^(-7)+T[q+1,2]); od: print(T);
```

Concentration	H3O at Equilibrium	pH
.001	.0006277411182	3.202150245
.002	.001009805474	2.995719274
.003	.001312617082	2.881828863
.004	.001571413541	2.803681873
.005	.001801160832	2.744423395
.006	.002009881572	2.696807924
.007	.002202470385	2.657070204
.008	.002382174304	2.623008234
.009	.002551278277	2.593225146
.010	.002711462462	2.566780387

Figure 10

Extraction of the data from the table can be obtained simply by

```
>with(plots):Xvals:=[seq(T[i,1],i=2..11)];Yvals:=[seq(T[i,3],i=2..11)];
```

and transferred to a pointplot: (Fig.11)

```
>points:=pointplot(zip((x,y)->[x,y],Xvals,Yvals));
display(points);
```

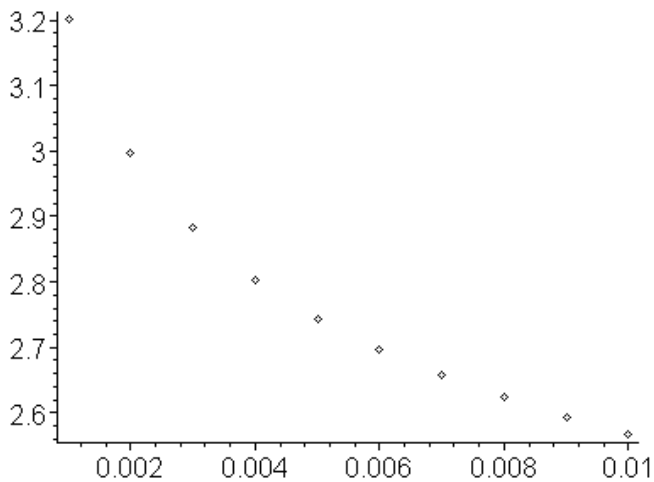


Figure 11

The interpolation of the obtained points is obtained from the preexistent least square package: obtained via

```
curve:=interp(Xvals, Yvals, z); p1:=plot(curve,
z=0.001..0.01, title='pH versus Concentration',
axes=box): display([p1, points]);
```

and returns the result presented in Fig 12.

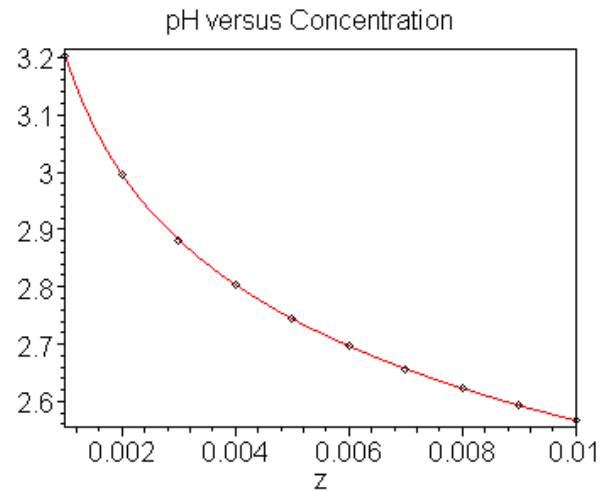


Figure 12

The analytical expression of the interpolated curve can also be extracted from the file, if needed.

A more evolved case-study is the use of more pre-structured computation modules for the calculation of equilibrium states in more complex systems, based on direct minimization of the total free energy of the system. In order to illustrate the research-utility of such modules and of the SCE-approach, we shall present below the comparative study of the complex heterogeneous systems: the $\text{Si}_{\text{liq}}/\text{SiO}_{2,\text{q}}/\text{C}_{\text{gr}}$ system, and the $\text{Si}_{\text{liq}}/\text{SiO}_x\text{N}_y/\text{SiO}_{2,\text{q}}/\text{C}_{\text{gr}}$ system. Our study has been focused on the production of CO in the gas phase, a species identified as crucial for the industrial application of these systems. The template used for the graphing of the concentration of the CO versus three important control parameters of the industrial process – pressure, temperature and initial composition of the system – has been designed to produce the 3D-graph presented in Fig. 13:

CO Volume Percentage versus T and P

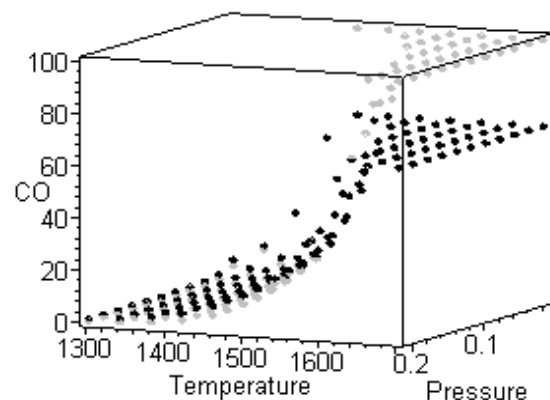


Figure 13

This way, a set of parameters could be identified, so that the equilibrium concentration of CO could be used to induce thermodynamic control of the industrial process.

Conclusions

The utilization of symbolic computation engines results in the possibility of opening of the pedagogical undergraduate curriculum to research objectives. The design and the utilization of SCE-assisted templates is consistent with a highly structured problem-solving approach, of unmatched accuracy to date in the delivery of chemical concepts, as well as in powerful computation templates for quantitative process-analysis in all major fields of physical chemistry.

REFERENCES

ⁱ Simulation of Chemical Concepts, Systems and Processes Using Symbolic Computation Engines: From Computer-Assisted Problem-Solving Approach to Advanced Tools for Research, Jonathan Rittenhouse and Mihai Scarlete, Ch. 16, Annual Report on Computational Chemistry, Elsevier, Spellmeyer Ed., **2005**

ⁱⁱ Symbolic Computation Engines and Molecular Modeling Templates: Maple-Assisted Point Group Analysis of the Vibrational Activity of Molecules., B. Vail, D. Aris, and M. Scarlete, International Journal of Quantum Chemistry, June **2005**

ⁱⁱⁱ Spectroscopic Analysis and Semiconductor Properties of Amorphous Thin Films Containing Silicon-Carbon-Nitrogen Deposited *via* Polymeric Route, M. Bercu, C. Aktik, J. Campbell, and M. Scarlete, Canadian Journal of Applied Science and Spectroscopy, **2003**, 48(1), 77

^{iv} From Quantum Numbers to Molecular Modeling. Maple-Assisted Computation, BU Press, ISBN 0-920917-16-X, **1999**