

Chemistry and Computer Algebra: Past, Present, Future

Michael P. Barnett
P.O. Box 322, Princeton NJ 08542-322, USA
michaelb@princeton.edu

In about 1950 Frank Boys, the founder of “computational chemistry”, used the EDSAC computer at Cambridge to differentiate polynomials represented by arrays of coefficients. This gave me the idea of applying recurrence schemes to rational expressions a few years later at MIT, which led Jean Sammett to develop the first commercial computer algebra system (FORMAC). I have used modern computer algebra software for the past decade in my work on valence theory, as just one of many chemists who use the software in varied areas of research. But with hundreds of thousands of new mathematical formulas being published by chemists each year, the scope for further proliferation of chemical computer algebra is immense. Interdisciplinary communication is the key to reaping the benefits of scientific symbiosis, and this requires the recognition of culture gaps between the disciplines and the building of appropriate bridges.

The culture gap. This involves more than differences in technical expertise and the utterly different meanings of homology, ring, field, isomorphism, toric geometry and dozens of other terms — more, even, than what constitutes linear dependence. Chemistry can be immensely precise — using a burst of laser light that lasts a few quadrillionths of a second to track vibrations inside a molecule, and using scanning tunneling microscopes to manipulate individual atoms on a surface. But it can be enormously messy — the laboratory ambience of some elegant nanotubule development is soot! Chemistry deals with colour, smell, toxicity, hardness, solubility, volatility, viscosity, luminescence, conductivity, inflammability and hundreds of other properties of millions of solid, liquid and gaseous materials, both pure and in mixtures. Chemistry seeks regularity and order and the succinct description of predictable, systematic behaviour. This implies theoretical models with mathematical properties that are smooth. But discontinuities are vividly apparent in melting and boiling and precipitation, and in the minute changes within a protein that affect life dramatically. And no matter how elegant a concept or the supporting mathematics, the behaviour of real materials in the real world is the benchmark for theoretical computations.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISSAC '03 Philadelphia PA

Copyright 2003 ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

Helping chemistry and computer algebra interrelate needs more than just the recitation of these matters and the cataloguing of published applications and suggestions for further work based on “book learning”. Computer algebra must be made commonplace in chemical education. And computer algebraists must learn to work with chemical ideas and information. Accordingly, the talk will focus on these pedagogic issues first, and then go on to explain the relevance of selected topics to chemical research under three main headings

1. the solution of sets of simultaneous multinomial equations: steady state kinetics of enzyme, gas and polymer reactions (rate constants, identifiability) — compartmental analysis — mathematical statistics — inverse kinematics of stereochemistry and molecular modeling,
2. linear algebra: unrestricted precision — interval arithmetic — error bounds — vast sparse matrices — architectural opportunities — explicit characteristic polynomials (chemical graph theory, parametrized secular equations for electronic structure calculations),
3. classical and modern analysis: ordinary and partial differential equations — dynamic theory — non-steady state kinetics — oscillatory reactions (circadian rhythms) — some mundane aspects of integration from my own research — massive tabulation of formulas — associated topics for novel algorithm development.

In applying computer algebra to any natural science, we enter uncharted waters. I will end by mentioning two topics of practical and theoretical significance, that became apparent as a result of delving into the unknown — languages for proofs and derivations, and certain types of transformation that parallel models of molecular computing.

A chemical commonplace for computer algebra. Basic physical chemistry is inescapable in the schooling of every chemist, chemical engineer and life scientist today. Word problems that relate to a wide range of well known properties of chemicals are an essential ingredient of instruction. Standard texts contain hundreds of these, and there are texts and courses that deal with Numerical Methods in Physical Chemistry explicitly. I suggest two kinds of topic that can be used in the “plots” (literary, not graphic) of multitudinous problems which require the solution of sets of simultaneous multinomial algebraic equations.

1. Many properties of solutions (substances dissolved in other substances) and binary, ternary and more complicated mixtures are multinomial in the concentrations of the components. Computing the composition of a solution or a mixture that will have prescribed

values for a selection of these properties addresses an open-ended variety of practical issues.

2. The design of capsules for drug delivery, containers for laboratory benchwork, magnetic and electrostatic force fields in scientific equipment, industrial plant for chemical reactions and separation all require the algebraic description of geometry, that gives rise to multinomial equations. So do considerations of electrostatic interactions within molecules and between molecules, and studies that involve normal and abnormal shapes of sub-cellular objects and complete cells, tissues, organs and anatomical structures.

Sets of problems that address these issues could come after the standard introduction to linear equations and Gaussian elimination in a course that was structured mathematically. And they could be dispersed in courses that were structured by chemical topic. I emphasize the chemistry and mathematics — whether the computer algebra software is used in kernel or in a worksheet or notebook mode is a separate issue. There is an abundance of material that uses the latter modes. This material often accompanies books on “Learning *subject X* using . . . *software system Y*”. Often, these books emphasize the direct production of graphical output to “build intuitive understanding”. Some of this is effective, when the illustrative material is non-trivial, *e.g.* when it proceeds from elementary differential equations to bifurcation problems and oscillatory behaviour. But a focus on the delivery mechanism should not dominate the pedagogic integration of computer algebra and natural science.

Searching the chemical literature. Many computer algebraists know that their expertise has been applied to: (a) integer programming, (b) solving van der Pol equations, (c) identifiability studies and (d) compartmental analysis. Each of these terms (and many others in the literature of computer algebra) features in numerous papers on chemistry that can be found using Chemical Abstracts, the Science Citation Index and several other on-line search tools. It should become second nature for computer algebraists to use these tools in the quest for entry points to chemical topics that provide novel opportunities for their skills.

Some basic concepts. Computer algebraists who want to join forces with research chemists can start a dialogue with their colleagues by asking for simple explanations of the following terms and ideas, which support stereochemical and kinetic studies of considerable diversity and research potential: (1) valence electrons — covalent bonding — structural formulas — single, double and triple bonds — the variety of organic compounds — acyclics — homologous series — functional groups — ring compounds — bond lengths — bond angles — isomers — stereoisomers — chirality — tautomers — steric hindrance — molecular strain — intra- and inter-molecular long range forces — solvation — molecular vibration, (2) thermodynamics of spontaneous change — law of mass action — rate constants — dynamic equilibrium — metastability — chain reactions — intermolecular energy exchange — electronic excitation — transition states.

Stereochemistry. Equipped with these concepts and the associated keywords, algebraists can ask whether the power of present molecular modeling methods can be augmented by inverse kinematic calculations of the extent to which molecules (and their tautomers) can pucker, fold or change shape in other ways could help predict the ability to permeate, diffuse, align, close and open rings, encage,

bridge or dock in processes that include molecular recognition and mimicry, replication, drug action and delivery, design of syntheses and elucidation of metabolic precursors, surface catalysis and macromolecular assembly.

Kinetics. Batch (2) of ideas also allows a distinction between (1) steady state calculations of metabolic processes — these reduce immediately to sets of simultaneous multinomial equations in the rate constants, (2) non-steady state processes — these involve simultaneous differential equations that fit into modern dynamic theory, (3) transition state studies that fall into the general methodology used to predict optimal molecular geometry and electronic energies.

And more. Once substantive discussions have started within local communities of computer algebraists and chemists on these relatively accessible topics, then the floodgates can open to joint work on the quantum theory of molecular structure, molecular modeling and dynamics of very large molecules, the statistical physics that predicts thermodynamic properties, and studies of ionic solutions, crystal structure and growth, collision processes, the theory of liquids and an ever burgeoning variety of projects directed to particular kinds of substance, property and role in life processes.

Standard texts and websites on computational chemistry and molecular modeling explain the structure of these fields, the sub-specializations, the associated catchwords and the mathematical processes. These are mostly handled by numerical methods at present and provide extensive opportunities for computer algebra. Ongoing trends can be followed by reviewing the programs of the Division of Computers in Chemistry at the national meetings of the American Chemical Society.

The technical benefits. Some of the ways in which computer algebra has helped my work can carry over to many other projects, too. It has helped me: (1) correct some general formulas that have been in the literature for many years, incorrectly stated, (2) construct new general formulas by manipulations that could not be performed by hand, (3) validate tentative hand derived formulas, by lengthy inductive proofs, (4) use very simple mathematics instead of specialized methods involving inconsistent error-prone notations, (5) carry out proofs and derivations that require the separate consideration of numerous special cases, with the resolution of conditional expressions by downward propagation of overall assumptions, (6) use recurrence schemes without numerical degradation by repeated iteration, (7) annihilate pairs of terms that are identically equal in magnitude and opposite in sign, (8) cope with numerous, awkward cusp lines in multiple integrations, (9) perform the tedious algebra in massive tabulations of formulas for individual integrals, (10) apply elaborate systems of checks, (11) transform some infinite series which have catastrophically slow convergence into adequately convergent forms, (12) keep formal square roots and trigonometric functions up to the final stage of a calculation, (13) extract common sub-expressions, to reduce storage and to expedite transmission and numerical evaluation, (14) use unrestricted precision to evaluate formulas for end use and to calibrate numerical packages, (15) explore the symbolic calculation of eigenvalues as power series in the molecular parameters, (16) interface the calculations with electronic typesetting and the web to provide validated documentation, (17) create a body of dependable machine readable material for future mechanized use.