Chemsitry, Computers, Computations, and Science Education

T. Helgaker, Department of Chemistry, University of Oslo, Norway

September 30 2005

Centre of Mathematics for Applications

University of Oslo

Chemistry: Experiment and Computation

- Chemistry is traditionally an experimental science!

  "Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational. If mathematical analysis should ever hold a prominent place in chemistry—an aberration which is happily impossible—it would occasion a rapid and widespread degradation of that science."

  August Comte, 1798–1857

  – there is a deep-seated scepticism towards computation among chemists
  – a theoretical calculation provides numbers but no understanding!

- Nevertheless:

  "The more progress sciences make, the more they tend to enter the domain of mathematics, which is a kind of center to which they all converge. We may even judge the degree of perfection to which a science has arrived by the facility with which it may be submitted to calculation."

  Adolphe Quetelet, 1796–1874

  – today, computation constitutes more than 20% of all chemical research
  – this proportion increases by about 1% every year
  – today, 13 out 34 chemistry professors in Oslo use quantum-chemical simulations
 – what are the reasons for this development?

Chemistry as a many-body problem

- Chemical systems are collections of particles of different masses and charges, bound by predominantly electrostatic forces, subject to the laws of quantum mechanics.

- Thus, the mathematical foundation of chemistry was understood about 80 years ago:

  "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

  P. A. M. Dirac, 1929

- The complications arise from the large number of particles:
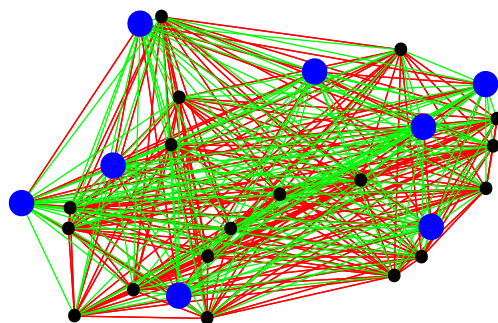
- With the development of modern computers over the last 50 years, the molecular many-body problem has become tractable.

The quantum-mechanical many-body problem

- All information about a many-particle system is contained in the wave function

$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, t)$$

- The wave function satisfies Schrödinger's equation

$$\hat{H}\Psi = \mathrm{i}\hbar\frac{\partial\Psi}{\partial t}$$

  where the Hamiltonian operator for a field-free, nonrelativistic system is given by

$$\hat{H} = -\sum_i \frac{\hbar^2}{2m_i}\left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}\right) + \frac{e^2}{4\pi\varepsilon_0}\sum_{i>j}\frac{Z_i Z_j}{r_{ij}}, \quad 1 \leq i \lesssim 10000$$

- For stationary problems, the time-dependence is trivial and all information

$$\hat{H}\Phi(\mathbf{r}) = E\Phi(\mathbf{r}), \quad \Psi(\mathbf{r}, t) = \Phi(\mathbf{r})\exp(-\mathrm{i}Et/\hbar)$$

  where E is the energy of the system

- Quantization: boundary conditions give solutions only for certain (discrete) values of E:
  - particle permutation symmetry
  - continuity of wave function and its derivatives
  - the wave function should vanish at infinity or satisfy periodic boundary conditions

The many-body problem: hierachies of approximation

- A brute-force approach is out of the question for any but the smallest systems

- The many-body problem can be solved approximately in many different ways

  - finite-element and Monte Carlo methods are little used

  - the algebraic approach: expansion in products of simple one-electron functions

- Time-independent problems are approached by solving the associated variational problem:

$$\delta E[\Phi(\mathbf{r}_i)] = 0, \quad E[\Phi] = \frac{\int \Phi^*(\mathbf{r}_i) \hat{H} \Phi(\mathbf{r}_i) \mathrm{d}\mathbf{r}_i}{\int \Phi^*(\mathbf{r}_i) \Phi(\mathbf{r}_i) \mathrm{d}\mathbf{r}_i}$$

  - $\Phi(\mathbf{r}_i)$ is expanded in functions that satisfy the boundary conditions, turning the Schrödinger equation into an (essentially) unconstrained stationary problem

  - typically, a very large number of variational parameteres (10^3–10^9) are introduced

  - all parameters are determined from first principles (rather than empirically)

  - post-variational corrections are often applied, for high accuracy (correlation)

  - accuracy can be controlled and systematically improved towards chemical accuracy

- Alternative approach: density-functional theory $\delta E[\rho(\mathbf{r})] = 0$.

  - high efficiency and flexibility, cannot be systematically improved

  - dominates solid-state chemistry and large molecular systems

Computers and computational chemistry

- Today, a large number of chemical problems have become amenable to calculation:

  - molecular structure and spectroscopic constants

  - reaction enthalpies and equilibrium constants

  - reactivity, reaction rates, and dynamics

  - interaction with applied electromagnetic fields and radiation

  - nonlinear and optical properties

- Most of these quantum-chemical calculations are carried out by nonspecialists.

- Quantum chemistry has been important in providing understanding and many qualitative models in chemistry

  - such models are good and useful

  - however, they do not constitute the bread and butter of quantum chemistry

- If quantum chemistry is to have a decisive impact on chemistry, we need to provide tools that can compete with experiment

  - ideally, our results should be as accurate as experiment: chemical accuracy

  - if we cannot consistently provide high accuracy, we will soon be out of business

- Note: much of chemistry can never be reduced to computation!

## Example: the atomization energy of CO (exp. 1071.8 kJ/mol)

- In chemistry, relative rather than total energies are important, for example:

$$D_{\mathrm{CO}} = E_{\mathrm{C}} + E_{\mathrm{O}} - E_{\mathrm{CO}} \qquad (1)$$

- Atomization energies constitute less than 1% of the total molecular energy
  - their calculation to a precision of 1 kJ/mol (chemical accuracy) is a challenging task

	E_{C}	+	E_{O}	−	E_{CO}	=	D_{CO}	err.
HF	−98964.9	−	196437.1	+	296132.2	=	730.1	−341.7
SD	−388.4	−	639.6	+	1350.2	=	322.1	−19.6
(T)	−7.7	−	11.9	+	54.2	=	34.6	15.0
vib.	0.0	+	0.0	−	12.9	=	−12.9	2.1
rel.	−40.1	−	139.0	+	177.1	=	−2.0	−0.1
tot.	−99401.1	−	197227.6	+	297700.8	=	1071.9	

The hierarchy of quantum chemistry: atomization energies (kJ/mol)
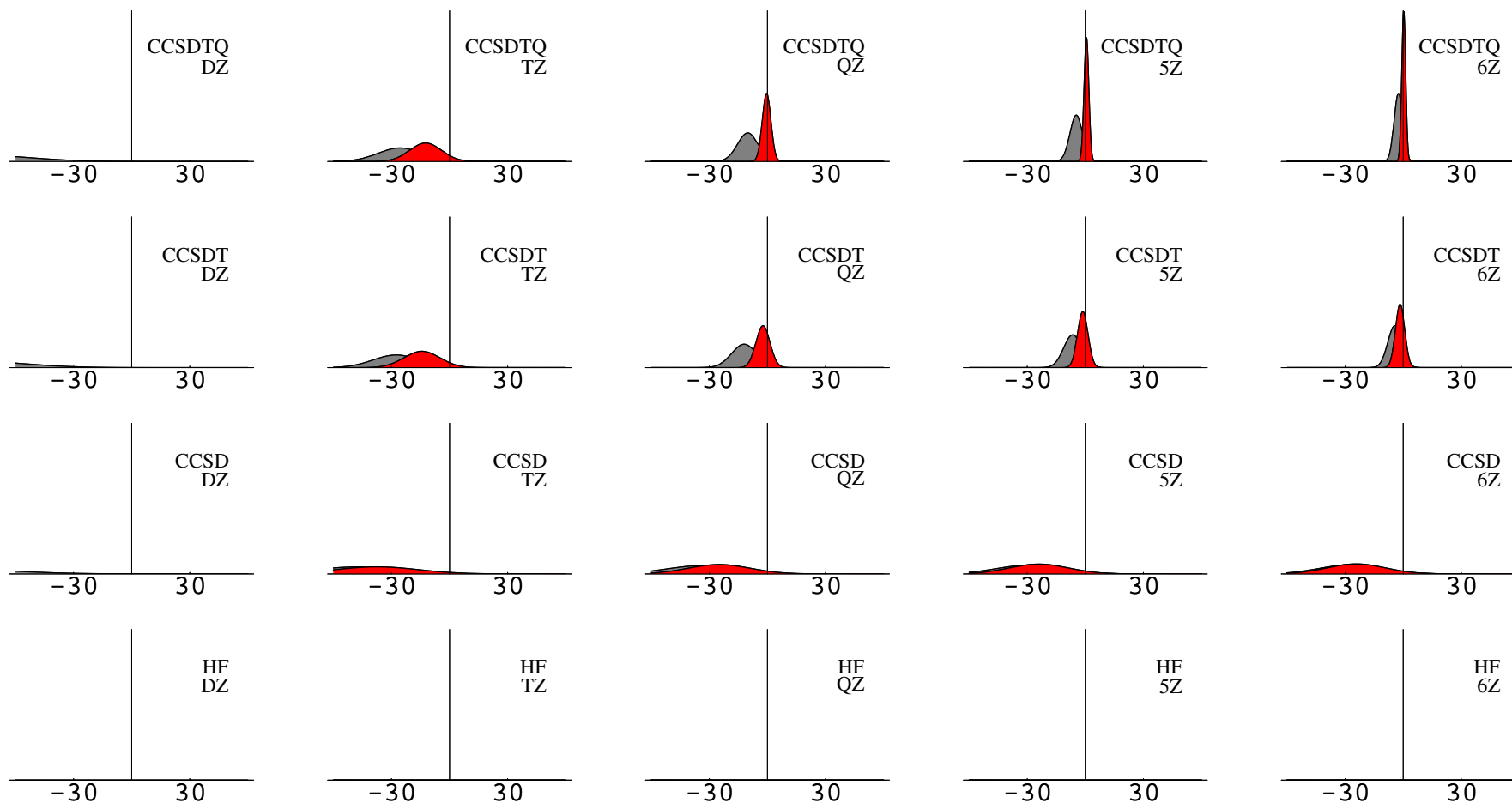
Reaction enthalpies (kJ/mol)

	DFT	CCSD(T)	exp.	vib.	rel.
$CO + H_2 \rightarrow CH_2O$	−34	−23	−21(1)	30.2	0.7
$H_2O + F_2 \rightarrow HOF + HF$	−119	−118	−129(4)	−0.6	0.5
$N_2 + 3H_2 \rightarrow 2NH_2$	−166	−165	−164(1)	86.1	1.6
$C_2H_2 + H_2 \rightarrow C_2H_4$	−208	−206	−203(2)	37.4	0.2
$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$	−211	−244	−244(1)	92.7	1.0
$CH_2O + 2H_2 \rightarrow CH_4 + H_2O$	−234	−250	−251(1)	50.3	0.5
$CO + 3H_2 \rightarrow CH_4 + H_2O$	−268	−273	−272(1)	80.5	1.2
$HCN + 3H_2 \rightarrow CH_4 + NH_2$	−320	−321	−320(3)	85.4	0.9
$HNO + 2H_2 \rightarrow H_2O + NH_2$	−429	−446	−444(1)	56.7	1.0
$C_2H_2 + 3H_2 \rightarrow 2CH_4$	−450	−447	−446(2)	85.2	0.4
$CH_2 + H_2 \rightarrow CH_4$	−543	−543	−544(2)	46.8	0.4
$F_2 + H_2 \rightarrow 2HF$	−540	−564	−563(1)	17.6	1.6
$2CH_2 \rightarrow C_2H_4$	−845	−845	−844(3)	45.8	0.6
$O_3 + 3H_2 \rightarrow 3H_2O$	−909	−946	−933(2)	72.5	2.3
mean & std. dev.	8 & 13	−1&5			
mean abs. & max. abs.	11 & 33	3 & 13			

|		HF	MP2	CCSD	CCSD(T)	exp.
H_2O	R_{OH}	94.0	95.7	95.4	95.7	95.7
NH_3	R_{NH}	99.8	100.8	100.9	101.1	101.1
C_2H_2	R_{CH}	105.4	106.0	106.0	106.2	106.2
C_2H_4	R_{CH}	107.4	107.8	107.9	108.1	108.1
N_2	R_{NN}	106.6	110.8	109.1	109.8	109.8
CH_2O	R_{CH}	109.3	109.8	109.9	110.1	110.1
CH_2	R_{CH}	109.5	110.1	110.5	110.7	110.7
CO	R_{CO}	110.2	113.2	112.2	112.9	112.8
HCN	R_{CN}	112.3	116.0	114.6	115.4	115.3
CO_2	R_{CO}	113.4	116.4	115.3	116.0	116.0
HNC	R_{CN}	114.4	117.0	116.2	116.9	116.9
C_2H_2	R_{CC}	117.9	120.5	119.7	120.4	120.3
CH_2O	R_{CO}	117.6	120.6	119.7	120.4	120.3
C_2H_4	R_{CC}	131.3	132.6	132.5	133.1	133.1

Bond distances (pm)

- Calculation of interactions among the nuclear magnetic moments in molecular systems

  - the nuclear spin–spin coupling constants are important in NMR spectroscopy

  - involves the evaluation of the response of the electrons to nuclear magnetic fields
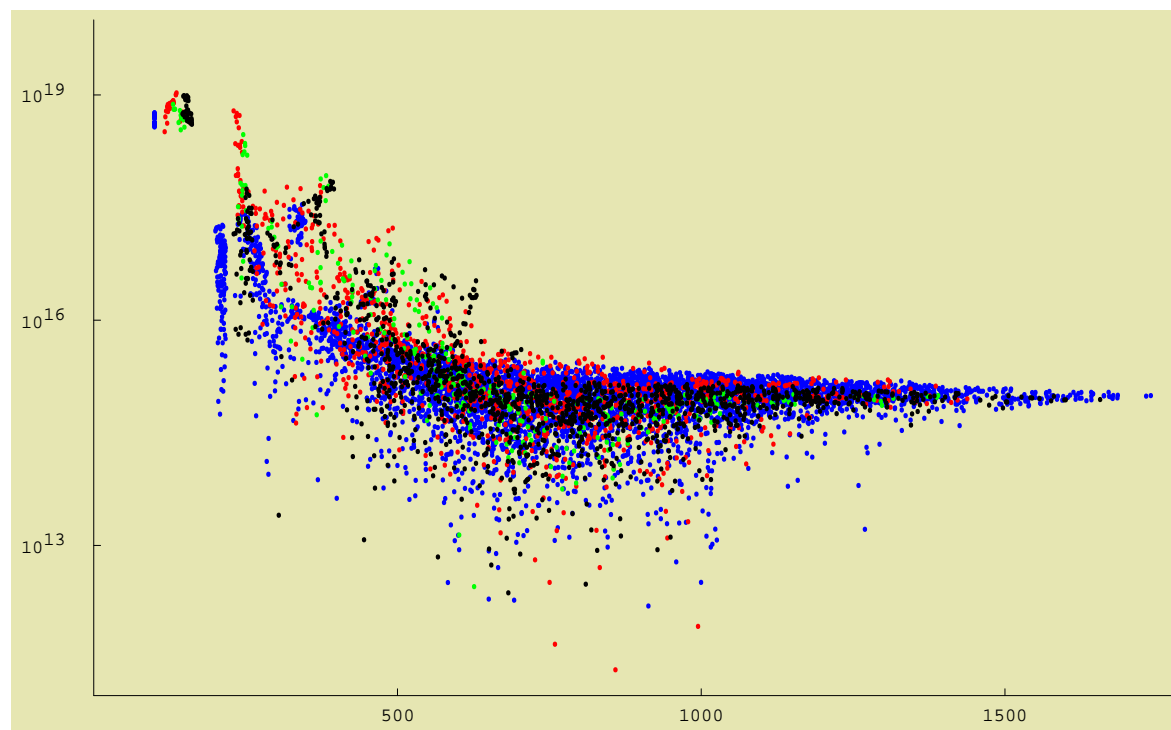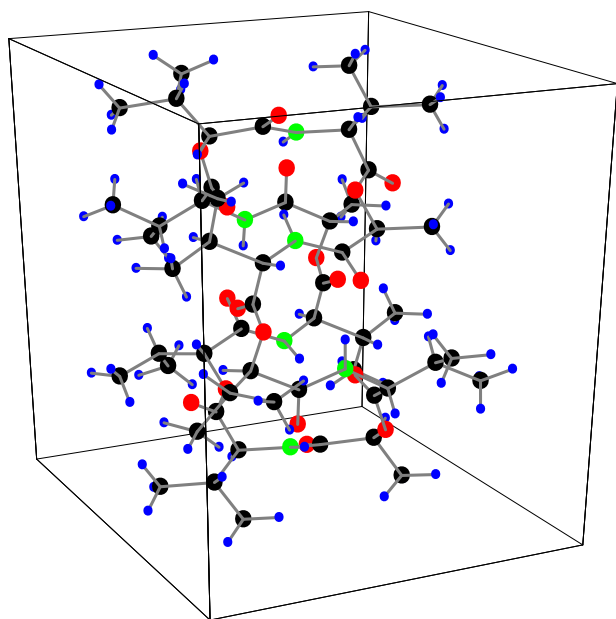
  - there are a total of 7587 spin–spin couplings to the carbon atoms in valinomycin

  - above, we have plotted the magnitude of the reduced LDA/6-31G coupling constants on a logarithmic scale, as a function of the internuclear distance

Two-photon absorption (TPA) coefficients of AF455 ($C_{137}H_{172}N_6$)

- Nonlinear optics important with development of intense lasers

- Search for molecules with special nonlinear properties

- TPA depends quadratically on photon intensity

- Second-order response of the electrons to radiation

- LDA/4-31G TPA cross sections of the ten lowest excitations (eV)
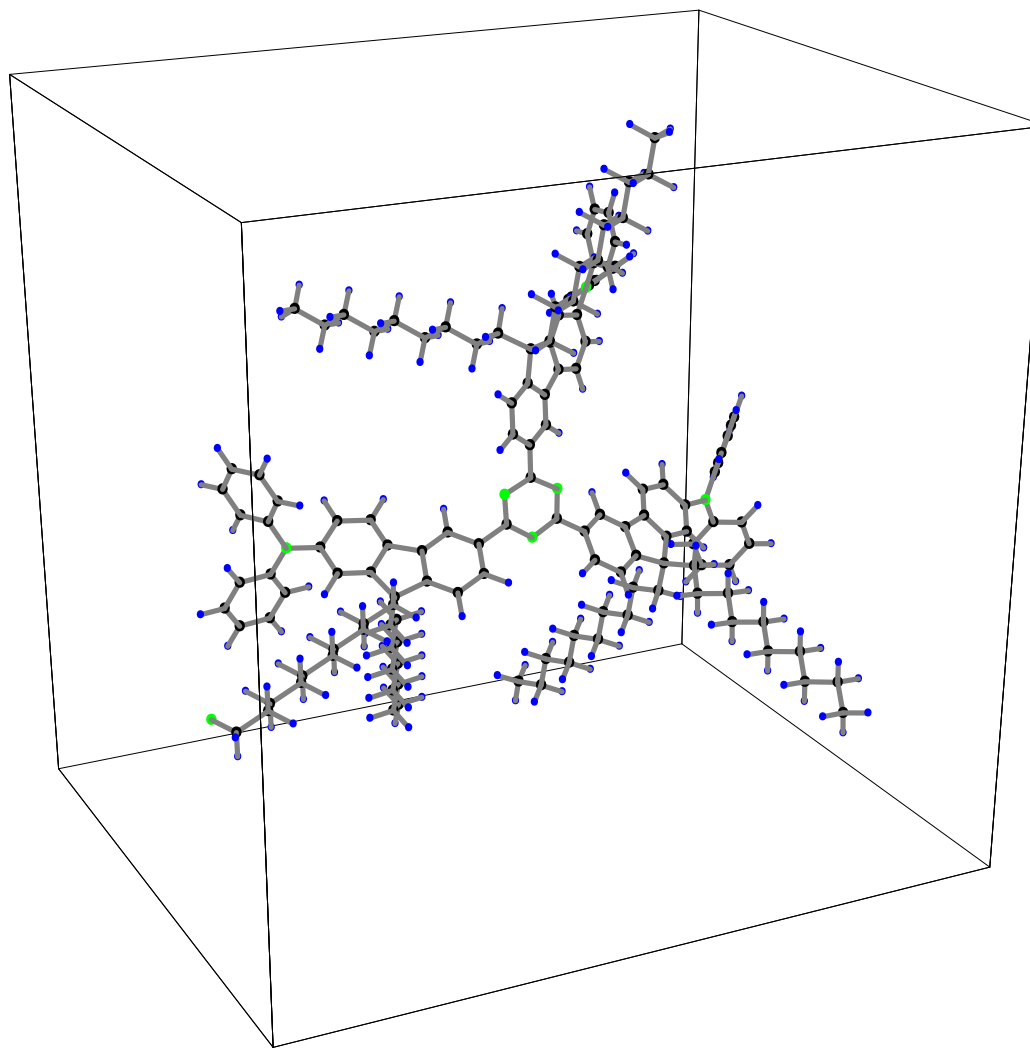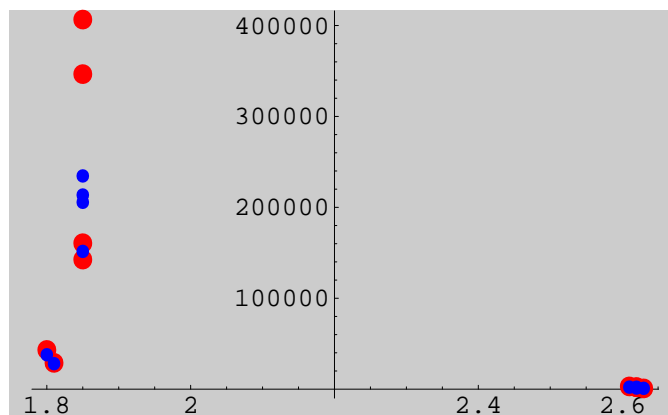
Dalton

- All calculations shown here have been carried using our local code Dalton

  – Dalton is a collaborative effort, going back nearly 25 years.

  – it is licensed freely to research groups (about 400 licenses of Dalton 2.0 since March)

  – about 740000 lines Fortran and 28000 lines of C

  – efficiency is of paramount importance

- Students are typically given the task to develop some new quantum-chemical technique, or to refine an old one within the framework of Dalton

  – students typically have a rudimentary programming experience and little physics

  – cope with a large and complicated code—initially overwhelming

  – satisfaction of developing code and methods that will be used by many people

  – introduction to the computational Dalton community

  – combination of chemistry, physics, numerical analysis and computer science

The complementary role of experiment and simulation

- Simulations play an important role in many areas of modern chemistry

  – I have used examples from my own field: molecular quantum chemistry

  – Solid-state methods are similar to those of molecular theory

    - mostly density-functional theory

    - more qualitative and less quantitative than molecular theory

    - developed by physicists

  – The two subfields are (very) slowly converging, learning from each other

- The reasons for carrying out simulations are the same:

  – experiments may be difficult to interpret or to understand

  – experiments may be difficult, impossible, dangerous, or expensive to carry out

- Experiments and simulations are complementary techniques, more and more often used in conjuction, often to great effect

  – calculations often conducted by "nonspecialists"

  – black-box methods (not black magic!)

The broader role of computation in chemistry

- Many methods of mathematics and numerical analysis are used in chemistry:
  - numerical differentation and quadrature
  - differential equations and integration
  - linear algebra and matrix theory
  - linear and nonlinear equations, optimization theory
  - curve and data fitting, finite-element method, Fourier techniques
  - feedback theory, group theory, statistics

- These methods are used for a variety of purposes:
  - Fourier techniques in spectroscopy and diffraction techniques
  - digitalization and noise filtering
  - atmospheric modeling and ozone-layer simulations
  - protein folding modeling
  - thermodynamics modeling
  - molecular dynamics and kinetics
  - modeling of ion conduction
  - transport mechanism modeling

Chemistry: a hotchpotch of subfields

- organic chemistry:
  - relatively little use of computation
  - molecular simulations are becoming widespread

- inorganic chemistry:
  - traditionally little use of computation
  - recently, large-scale solid-state simulations have become very popular

- physical chemistry:
  - large-scale computation since the 1960s
  - modelling and simulations, data analysis

- theoretical chemistry:
  - relies critically on computation
  - methods development, greedy users of computers since their invention

- analytical chemistry:
  - little use of computation
  - some data manipulation

Chemistry: the need for mathematics and computers

- In spite of the central role played by simulations in modern chemistry, its emergence is not welcomed equally enthusiastically by everyone

- In designing courses and setting up requirements, Department of Chemistry will typically opt for the lowest common denominator in mathematics and computation
  - a widespread fear that "too much mathematics" will scare students away

- This is a serious problem for chemistry in general:
  - students are turned away from chemistry, not discovering its mathematical side
  - we do not always educate the students well for future developments in our field
  - some of my colleagues have themselves experienced a similar problem, in coping with the emergence of simulations as a tool in chemistry

- It is a problem for theoretical chemistry in particular:
  - we are invisible to students with an interest in computation and simulations
  - most chemistry students do not have the background necessary for work in our field
  - it is becoming increasingly more difficult to recruit students at our own department
  - instead, we must to rely on importing PhD students from elsewhere

Some comments from my department

"Det er viktigst for en student å lære grunnleggende kjemi, fysikk og matematikk. De skal selvfølgelig ha kjennskap til numeriske metoder, men kan sette seg inn i disse underveis."

"Matematikk! Numerisk analyse i noen grad... Vi sliter med at kjemien har rykte som en myk vitenskap, vi må ha nok matte i bachelor-programmene."

"Min erfaring er at dagens studenter i liten grad behersker de enklere delene av dette—at de eksempelvis ikke er i stand til å skrive enkle datarutiner for de enkleste av disse oppgavene."

"Studenter bør oppleve beregninger og modellering som en integrert del av utdannelsen, der slike verktøy blir benyttet til å komme dypere inn i problemstillinger enn det man ellers ville. Dette stiller store krav til til undervisningform, men også til motiverte og dyktige studenter. Tviler på om dårlige studenter vil makte å se og utnytte helheten mellom teori og eksperiment. Det avanserte fremtidssamfunnet vil dog etterspørre dem som nettopp makter det."

"Det jeg savner mest i min utdannelse fra UiO er grunnleggende data- og operativsystemferdigheter og programmering. Å lære å bruke modelleringsverktøyene blir ekstra tungt fordi det er så mye annet jeg har måttet tilegne meg før jeg kommer i gang med som er virkelig moro."

Relationship to mathematics and numerical analsys

- we recognize that quantum chemistry is but one of many areas of scientific computation
  - indeed, it can be viewed as an application of mathematical and other methods to a particular field of science, in some sense secondary to the generic problems
  - however, our outstanding problems are probably real problems, not necessarily problems arising from unfamiliarity with the generic methods of mathematics and numerical analysis

- we welcome more interaction with mathematicians and computer scientists
  - applied mathematics offer a rich variety of problems and challenges
  - problems important to us may be far removed from the expertise of a mathematician, and visa versa
  - problems important to us may be irrelevant or boring to mathematicians
  - sometimes, interests may coincide, the chemistry might work and things might gel!

- the methods of quantum chemistry can only be developed within the tradition of chemistry
  - the most important contribution would be to improve education, offering courses in numerical analysis and simulation that would be of interest to students in our fields