COMPUTER SIMULATIONS IN CHEMISTRY

Justyna Kozłowska, Wrocław University of Science and Technology, Wrocław Centre for Networking and Supercomputing

1) Brief history of computer simulations

Computational chemistry is the use of computer simulation to predict, understand, or explain chemical phenomena. It applies methods of quantum chemistry, implemented into efficient computer programs, to compute the structures and properties of molecules and solids. Besides the fields directly associated with military projects and weather forecasting, no other scientific area more willingly explored and significantly benefited from the growing power of electronic digital computers in their earliest infancy than quantum chemistry. In fact, it is the growing power of computers that determines three main stages in the development of computational chemistry:

(i) pre-computer visions

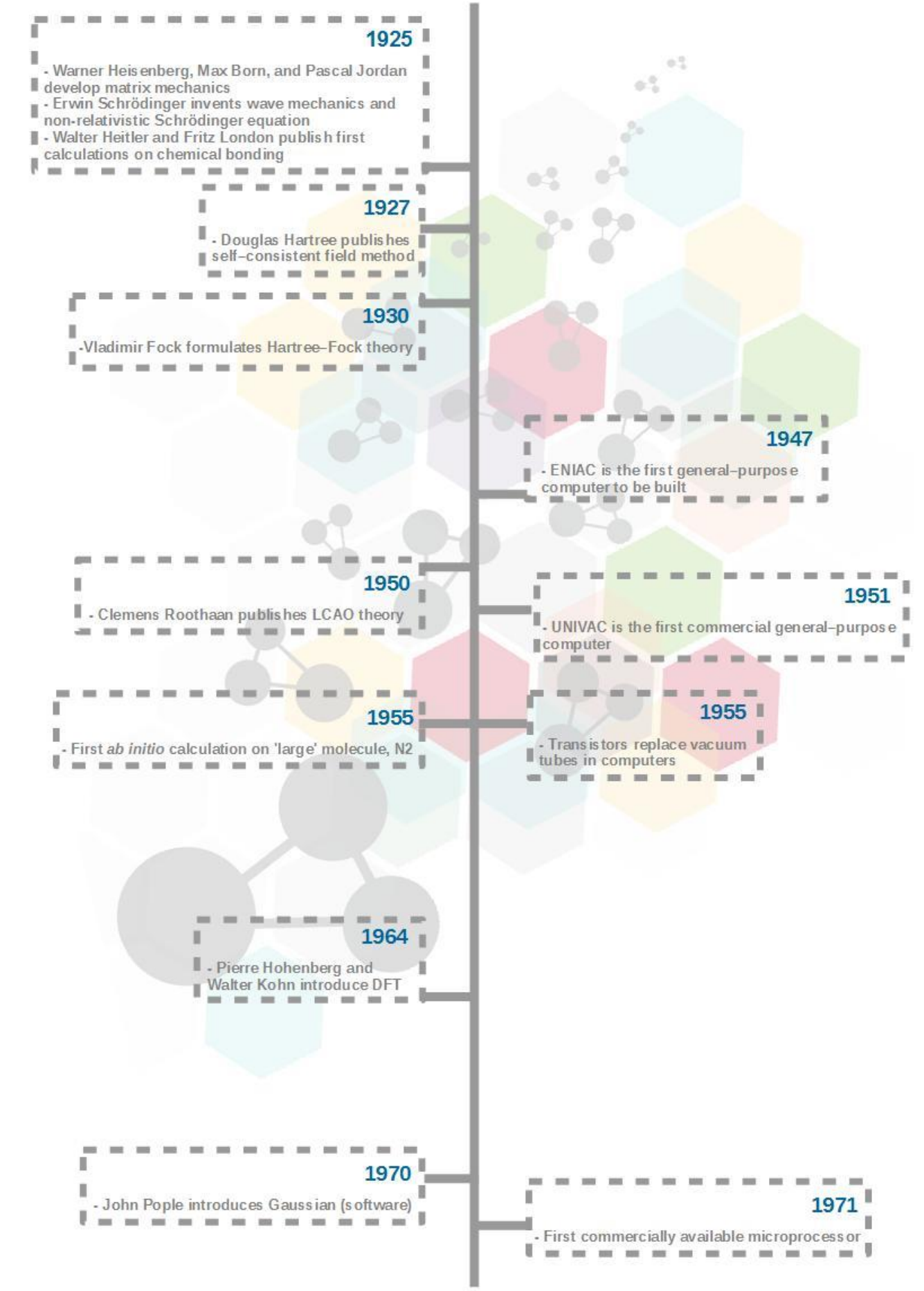
The famous starting point of quantum chemistry is the first theoretical treatment of the hydrogen molecule by Heitler and London, only a year after the appearance of highly mathematically formalized Schrödinger's equation. Among quantum chemists a new vision burst immediately that chemistry could be made computational. On the one hand, it was known that some chemical concepts could be interpreted within quantum theory. On the other, due to computational complexity, quantitative results were limited to the most simple cases. In 1929 Poul Dirac, one of the pioneers of the quantum theory, made his famous manifesto: *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.* In the quantum chemical community, this quote has become a necessary preface to any discussion about this field of science.

(ii) mainframe maturity

Application of mainframe digital electronic computers has allowed to develop in the 1950s/60s a full-fledged numerical program, that reached its maturation during the 1970s/80s. From this moment, simulation became full partner of experiment and computational chemistry an accepted (sub)discipline of chemistry. One of the earliest mentions of the term "computational chemistry" might be found in the book Computers and Their Role in the Physical Sciences (1970), where Sidney Fernbach and Abraham Haskell Taub stated: *It seems, therefore, that 'computational chemistry' can finally be more and more of a reality*.

(iii) chemistry without test tubes?

This stage begins in the 1990s and it continues to this day. The constantly growing computing power of modern computers drives the extension of computational chemistry to more complicated systems using more sophisticated quantum chemistry models.


1925
- Warner Heisenberg, Max Born, and Pascal Jordan develop matrix mechanics
- Erwin Schrödinger invents wave mechanics and non-relativistic Schrödinger equation
- Walter Heitler and Fritz London publish first calculations on chemical bonding
1927
- Douglas Hartree publishes self–consistent field method
1930
-Vladimir Fock formulates Hartree–Fock theory
1947
- ENIAC is the first general–purpose computer to be built
1950
- Clemens Roothaan publishes LCAO theory
1951
- UNIVAC is the first commercial general–purpose computer
1955
- First *ab initio* calculation on 'large' molecule, N2
1955
- Transistors replace vacuum tubes in computers
1964
- Pierre Hohenberg and Walter Kohn introduce DFT
1970
- John Pople introduces Gaussian (software)
1971
- First commercially available microprocessor

Much of modern computational chemistry calculations are conducted on Linux clusters where multiple processors can be utilized and many computations can be run in parallel. Nowadays computational chemistry provide meaningful answers to questions concerning reactivity, chemical kinetics and evolution of transition states to be obtained. Additionally, advances in computing also democratizes computation chemistry. In particular, by allowing non-specialist researchers to carry out routine investigations of simple systems using user friendly software packages on desktop computers – compared to command-line programs on mainframe or supercomputers in earlier eras.

As shown in the time-line chart above both computing technology and simulation modeling have been developed in close connection.

2) The relations between reality, mathematical models and simulations

All chemistry of atoms and molecules is determined by the interactions among the electrons and nuclei of which they are composed. This physics is accurately described by quantum mechanics and consequently fully described by the solutions of the Schrödinger equation, which in its simplest form can be written as:

$$\hat{H}\Psi = E\Psi.$$

Being written in such a form this equation might look quite harmless. Yet, within 90 years of quantum mechanic existence the only chemical systems that can be treated exactly are certain one-electron systems such as the hydrogen atom and the molecular hydrogen cation. For all of the rest chemistry we must be content within approximate solutions. The major difficulty is the treatment of the correlation between the motions of the many electrons, arising from their mutual repulsion. This is the many-body problem which lies at the heart of quantum chemistry (see Figure below). The nuclei and electrons are represented by large and small dots, respectively, and their interactions by straight lines.

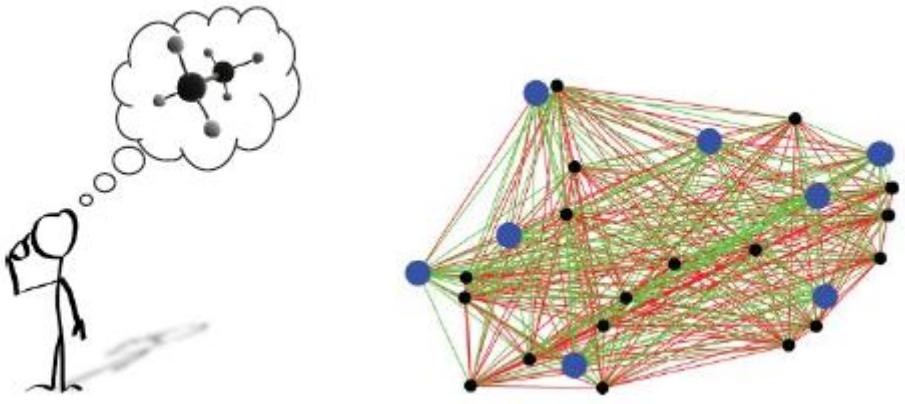


Figure from: T. Helgaker, T. A. Ruden, P. Jørgensen, J. Olsen, W. Klopper, J. Phys. Org. Chem. 17:913–933 (2004).

In order to deal with real chemical problems a number of approximations have been developed over the years. Simultaneously, thanks to the developments in computational techniques and partly as a result of spectacular advances in computer technology the practicing chemists have now a large variety of powerful techniques of different cost and accuracy, all of which can be applied to solve problems at the microscopic and molecular levels.

Among them three different methods to are used by scientists to perform calculations:

- *ab initio*, (Latin for "from the beginning") a group of methods in which molecular structures can be calculated using nothing but the Schrödinger equation, the values of the fundamental constants and the atomic numbers of the atoms present (Atkins, 1991).
- Semi-empirical approaches use approximations from empirical (experimental) measurements to provide the input data into the mathematical models.
- Molecular mechanics uses classical physics to explain and interpret the behavior of atoms and molecules of large molecular systems possessing thousands of atoms.

3) Errors, uncertainties and approximations

Errors

Reliability of computational chemistry simulations is assessed by the errors they provide.
For computational chemistry models to attain "chemical accuracy" is to provide errors smaller than:

- 1 kcal mol, (4.184 kJ) for atomization energies and reaction enthalpies
- 0.1 pm for bond distances
- 1 cm^{-1} for vibrational frequencies

More generally, reaching chemical accuracy can be crucial for the discovery of new structure−property relationships, trends, the development of new molecular materials with specific property requirements. Control over the accuracy of important physicochemical properties of molecules can be reached through application of well-established hierarchies in quantum chemistry. This allows chemists to routinely achieve chemical accuracy for *every* nonexotic and medium-sized organic molecule at substantial yet manageable computational costs.

Uncertainties

The determination of uncertainties for quantities of interest in computational chemistry comes in two forms:

- Determination of the confidence of a prediction, for instance of the property of a molecule.
- Assessment of the uncertainty by measuring the difference between properties, for instance between performance metrics of two or more computational approaches.

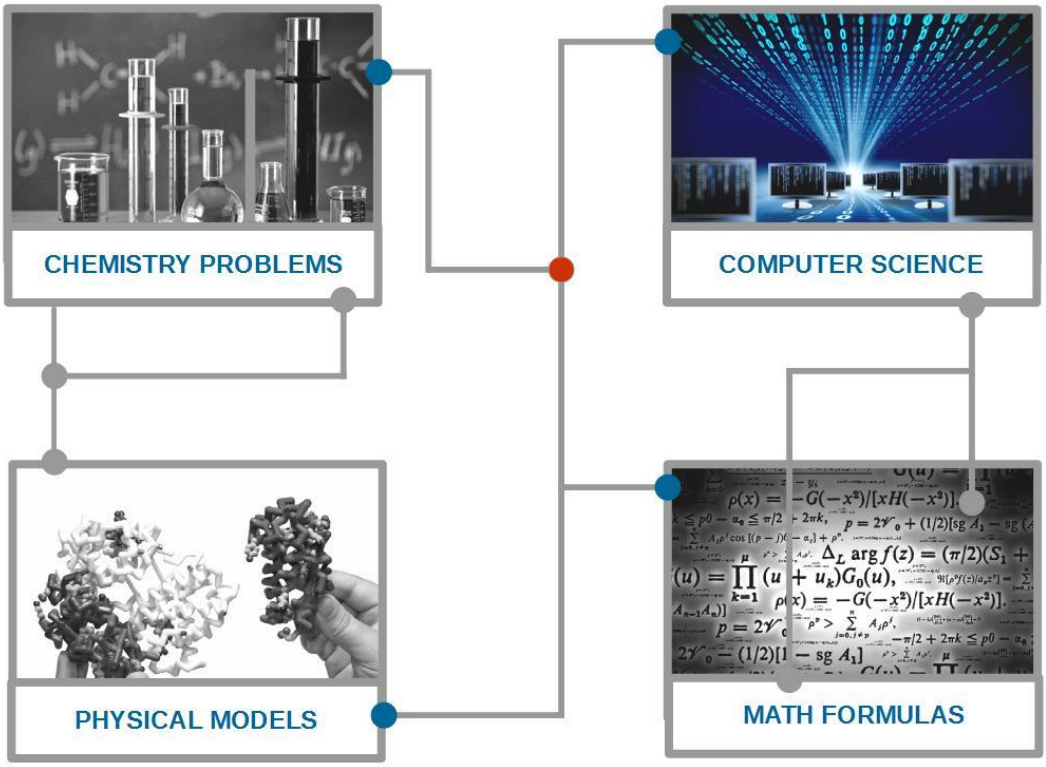
Approximations

In computational chemistry, a number of approximations need to be made in order to solve the mathematics with our current computational capabilities (as explained in the proceeding section). Among them the most important are the following:

- Born Oppenheimer approximation
- Use of nonrelativistic Hamiltonian
- Use of trial functions, molecular orbitals, in the variational procedure
- Use of single Slater determinant
- Basis set, linear combination of atomic orbitals molecular orbitals approximation (LCAO MO)

4) Computational science as an intersection of computer science, applied science and mathematics

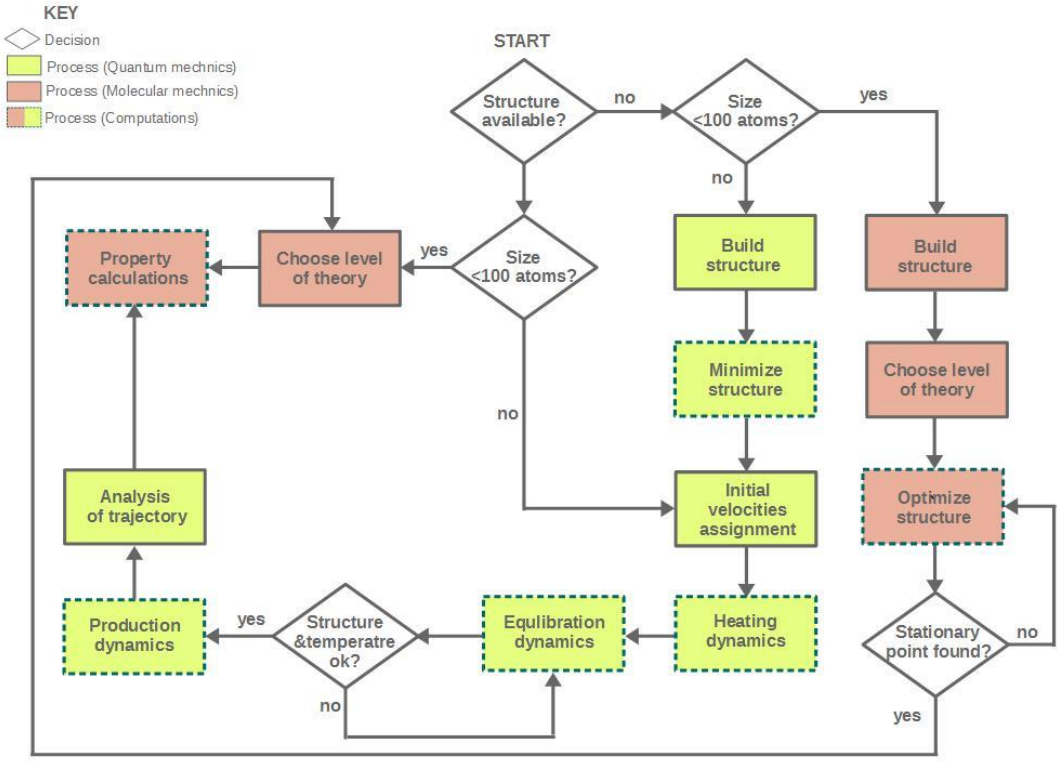
Theoretical chemistry might be defined as mathematical description of chemistry. One of its branches, computational chemistry, is considered as a mathematical method that is sufficiently well developed that it can be automated for implementation on a computer. Thus, computational chemistry involves



the application of mathematical and the theoretical principles to the solution of chemical problems. But only thanks to computer science chemists are able to investigate chemical phenomena by simulation *in silico* using computers rather than examining compounds and reactions experimentally. Computational "experiments" can be used both to predict the results of experiments and to assist in the interpretation of intriguing results.

5) Simulation steps: pre-processing, execution and post-processing

Computational chemistry investigates chemical structures and phenomena numerically, using the fundamental laws of physics. Nowadays sophisticated computer programs are employed to carry out quantum chemical computations on a wide range of molecules. Quantum chemical calculations constitute an important addendum for many experimental studies in organic, inorganic, physical chemistry, just to name a few. Due to the diversity of molecular systems, chemical processes as well as computational chemistry methods it is difficult to describe the way of conducting chemical simulations in a few words. Nevertheless, some major steps are always performed regardless the specificity of conducted research.



First and foremost, before performing any computational chemistry simulations it is necessary to define the size of the problem. This is due to the fact that the size of molecular system determines which of

the two major computational chemistry methods, namely molecular mechanics or electronic structure theory, could be used in the calculations. Secondly, one have to know how accurate results are required and simultaneously, how much time is planned for the calculations. Bearing this in mind, an appropriate level of theory (eg. quantum chemistry method, basis set or force field) should be picked up. Obviously, the choice of computational methodology depends equally on the studied chemical phenomena. Once all of these questions are answered, it remains only to select a suitable software.

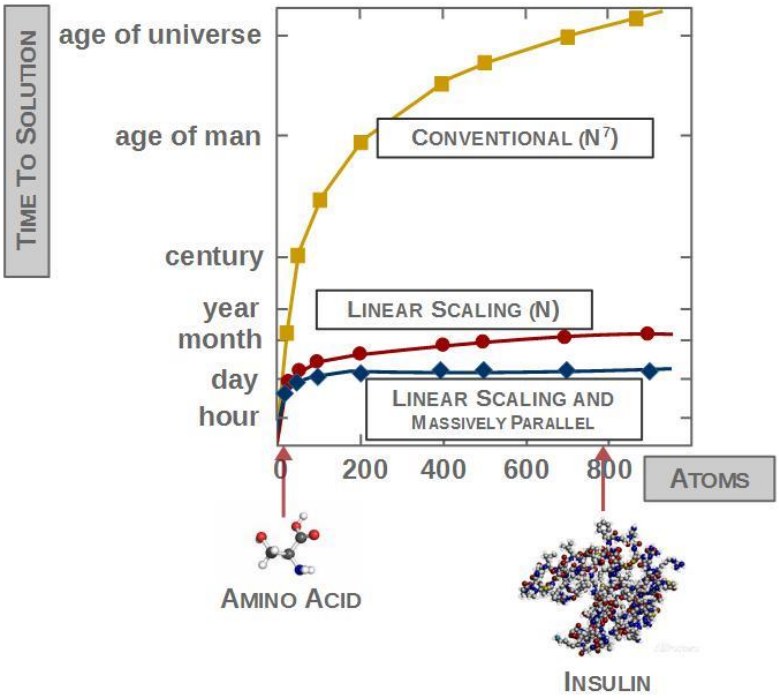
Noteworthy, two programs of the same type may calculate different properties, so one have to make sure that the chosen program computes exactly what was desired. The basic steps involved in computational chemistry experiments are showed in the flowchart below. Steps utilizing quantum chemistry are highlighted in pink, those involving molecular mechanics are highlighted in yellow.

6) Simulations' bottlenecks

One of the basics task of quantum chemistry is to fully understand the electronic structure of atoms and molecules. Over the last decades, a great progress has been made in the theoretical description of relatively simple systems. Specifically, combined with physical insights, elegant computational methods, ranging from wavefunction approaches to quantum Monte Carlo and density functional theory (DFT) have been developed to provide approximate solution of Schrödinger equation. From computational perspective all these methods involve rather lengthy and complicated program codes (ten thousands to several hundreds of thousands of lines), and have to handle a large amount of data to be stored on external devices.

The major bottleneck is that the Hilbert spaces of quantum systems grow exponentially with the size of the system. In other words, the computational cost grows dramatically whenever the size of the molecular system increases substantially (eg. proteins, nucleic acids) or an increasingly higher accuracy is requested (see Figure below). Thus the major bottlenecks of quantum computational chemistry is related to the computational cost, which has a number of aspects, i.e. computations time, memory requirements and sufficient computer power.

Note that quantum chemistry together with band structure research account for up to 30% of the computation time used at supercomputer centers [1], and *ab initio* methods constitute one of the two physics-simulation applications which prevail the use of supercomputing resources (the other one is the fusion-energy research).



In order to deal with this (unphysical) computational problem different solutions are sought. Among others, "direct" algorithms may eliminated I/O and storage bottlenecks and open the way to parallel implementations. Two major simplification schemes, known as linear scaling methods, are developed to alleviate the problem of computational expense, i.e. density fitting and local approximation. One fairly exotic solution is to use quantum simulators instead of computers based on classical mechanics. As already shown, a quantum computer of a sufficient size, like 128 logical quantum bits, would already outperform the best classical computers for *exact* chemical simulation [2].

[1] US Department of Energy 2010 National Energy Research Scientific Computing Center: 2010 Annual Report, Technical Report, Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA

[2] Aspuru-Guzik A, Dutoi A D, Love P J and Head-Gordon M 2005 Simulated quantum computation of molecular energies Science **309** 1704–7

This work was financially supported by PRACE project funded in part by the EUs Horizon 2020 research and innovation programme (2014-2020) under grant agreement 653838, and by the Polish Ministry of Science and Higher Education under grant agreement 3563/H2020/2016/2.