

COMPUTER SIMULATIONS FOR THE NANO-SCALE**I. Štich¹**

*Center for Computational Materials Science (CCMS), Department of Physics
Faculty of Electrical Engineering and Information Technology
Slovak University of Technology, Ilkovičova 3, SK-812 19 Bratislava, Slovakia*

Received 10 January 2007, accepted 17 January 2007

A review of methods for computations for the nano-scale is presented. The paper should provide a convenient starting point into computations for the nano-scale as well as a more in depth presentation for those already working in the field of atomic/molecular-scale modeling. The argument is divided in chapters covering the methods for description of the (i) electrons, (ii) ions, and (iii) techniques for efficient solving of the underlying equations. A fairly broad view is taken covering the Hartree-Fock approximation, density functional techniques and quantum Monte-Carlo techniques for electrons. The customary quantum chemistry methods, such as post Hartree-Fock techniques, are only briefly mentioned. Description of both classical and quantum ions is presented. The techniques cover Ehrenfest, Born-Oppenheimer, and Car-Parrinello dynamics. The strong and weak points of both principal and technical nature are analyzed. In the second part we introduce a number of applications to demonstrate the different approximations and techniques introduced in the first part. They cover a wide range of applications such as non-simple liquids, surfaces, molecule-surface interactions, applications in nanotechnology, etc. These more in depth presentations, while certainly not exhaustive, should provide information on technical aspects of the simulations, typical parameters used, and ways of analysis of the huge amounts of data generated in these large-scale supercomputer simulations.

PACS: 07.05Tp, 71.15Pd, 71.15Mb, 02.70Ss

KEYWORDS: Computer simulation, Condensed systems, Electronic structure,
Computational statistical mechanics, Supercomputing

¹E-mail address: stich@elf.stuba.sk

Contents

1	Introduction	4
2	The model and basic equations	6
3	Reformulation of basic equations	10
3.1	Electronic structure	11
3.1.1	Hartree-Fock theory	12
3.1.1.1	Hartree-Fock equation	14
3.1.1.2	Exchange and correlation in the Hartree-Fock approximation	17
3.1.2	Cusp conditions: short-range correlations	18
3.1.3	Non-dynamical long-range correlations	21
3.1.4	Post Hartree-Fock methods	22
3.1.5	Density functional theory	24
3.1.5.1	Thomas-Fermi approximation	25
3.1.5.2	Basic theorems of the DFT	26
3.1.5.3	Kohn-Sham equation	28
3.1.5.4	The LDA approximation	31
3.1.5.5	Exchange-correlation energy, E_{xc}	33
3.1.5.6	Accuracy issues	35
3.1.5.7	Beyond the LDA/LSD approximation: GGA functionals	38
3.1.5.8	Beyond the GGA approximation: meta and orbital-dep. funct.	41
3.1.5.9	Generalization to finite temperatures: Free Energy functional	43
3.1.5.10	Excited states and gaps	44
3.1.5.11	Discontinuity of E_{xc}	45
3.1.5.12	Symmetry adapted wave functions	46
3.1.5.13	Time-dependent DFT	49
3.1.6	Quantum Monte-Carlo	51
3.1.6.1	Variational Monte-Carlo	51
3.1.6.2	Diffusion Monte-Carlo	53
3.1.6.3	Importance sampling	57
3.1.6.4	Excited states	59
3.1.7	Basis sets, pseudopotentials, extended systems and ionic forces	60
3.1.7.1	Basis sets	61
3.1.7.1.1	Gaussians and Slater functions	61
3.1.7.1.2	Plane waves	62
3.1.7.1.3	Generalized plane waves	63
3.1.7.1.4	Mixed Basis Sets	63
3.1.7.1.5	Real Space Grids	64
3.1.7.2	Pseudopotentials	64
3.1.7.2.1	Norm-conserving pseudopotentials	65
3.1.7.2.2	Vanderbilt ultrasoft pseudopotentials	67
3.1.7.3	Extended systems	67
3.1.7.4	Ionic forces	70

3.2	Ionic dynamics	72
3.2.1	Classical ions	72
3.2.2	Quantum ions	74
3.2.2.1	Path integral approach	75
3.2.2.2	Discrete formulation of PI: high-temperature expansion	76
3.2.2.3	Primitive approximation	77
3.2.2.4	Symmetry of the wave function	80
3.2.2.5	Technical issues	81
3.3	Solving the basic equations	82
3.3.1	Ehrenfest dynamics	83
3.3.2	Born-Oppenheimer dynamics	84
3.3.2.1	Diagonalization of \hat{H}_e	84
3.3.2.2	Accuracy issues	90
3.3.3	Car-Parrinello dynamics	91
3.3.4	From dynamics to optimization	97
3.3.5	Parallelization and $O(N)$ methods	98
4	Applications: From the silicon crystal to technological processes	101
4.1	Structurally disordered solids and liquids	101
4.1.1	Liquid and amorphous silicon	101
4.2	Solid surfaces	108
4.2.1	Takayanagi reconstruction of the Si(111) surface	108
4.2.2	Dynamics of the Takayanagi reconstructed Si(111)- 7×7 surface	111
4.3	Interaction of molecules with surfaces	115
4.3.1	Explicit determination of PES: Dissociation of H_2 on Mg(0001) surface	117
4.3.2	Statistical sampling of PES: Dissociation of Cl_2 on Si(111)- 2×1 surface	119
4.4	Simulation of surface probe microscopy (SPM)	123
4.4.1	Simulation of imaging with dynamic AFM apparatus $\{v_1\}$	124
4.4.2	Simulation of nanomanipulation with dynamic AFM apparatus $\{v_2, v_3\}$	129
4.5	Simulation of chemical reactions: the importance of entropy	132
4.5.1	Simulation of the methanol-to-gasoline process	133
4.6	Simulations in nanotechnology	139
4.6.1	Organometallic nanojunctions $\{v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}\}$	140
4.7	Quantum Monte-Carlo	155
4.7.1	Zero-temperature QMC	155
4.7.1.1	Homogeneous electron gas	155
4.7.1.2	Odd-numbered, positively charged H-clusters	157
4.7.1.3	Toward QMC-based <i>ab-initio</i> molecular dynamics	160
4.7.2	Finite-temperature QMC	161
4.7.2.1	Odd-numbered, positively charged H-clusters	162
5	Outlooks	166
	References	166

1 Introduction

Computer modeling of matter is a standard tool used in materials science on all length-scales from macroscopic all the way down to *atomic/molecular-scale*. In the last decades computational physics became the third pillar of physics in addition to the two more traditional experimental and theoretical physics. Computational physics is in many respects akin to experimental physics as what it often amounts to is a “computer experiment”. Typical example being computational statistical mechanics, where a given property of interest is literally “measured” on the computer in a way analogous to how it is measured in the laboratory experiment. On the other hand, the approach the computational physics uses has a lot in common with theoretical physics. For example, it often starts with Hamiltonians, equations-of-motion, etc., which form the basis of the simulations, which are an integral part of the theoretical physics. One could see it as a bridge between theory and experiment. The simulation tools obviously depend on the length- and time-scales modeled in the simulation. However, all scales are mutually interconnected and hence, ultimately one needs to go to the atomic/molecular-scale, which is subject of the present review.

Why do we need the atomic/molecular-scale simulations? Part of the information one needs, can be measured, but experiments are difficult, expensive, and limited. This is true on all length-scales, and particularly on the atomic/molecular scale. Simple example is for instance femtosecond processes in condensed systems where it is often easier to set up a credible computer experiment than to perform an accurate enough real laboratory experiment. Sometimes the properties of interest cannot be measured in a laboratory experiment at all. For example, the properties of matter under the pressures and temperatures deep in the earth mantle cannot be directly measured in a laboratory experiment. All that can be done is to set up an accurate simulation able to extract the properties required. The other advantage of the computational physics is that it deals with well-defined systems. If the question is what the system response to a probe will be under ideal circumstances, then computational approach can often deliver the answer more easily than a real laboratory experiment. The “on-line” research makes exploratory research very easy. It is so simple to shuffle atoms and design new nanostructures at will on the computer and “watch” the system response to such modifications. Very often experiments on the nano-scale can relatively easily be performed with a high degree of reproducibility. However, the underlying atomic-scale processes typically remain totally murky. For example there are numerous surface probe microscopy experiments, which have demonstrated the ability to manipulate atoms, molecules, bonds, etc. Most of them can only be understood with the assistance of a computer simulation of the tip-sample interactions. There are also cost-related issues. While the cost of the computer hardware is steadily decreasing and power increasing, no such trend can be seen in the cost of the experimental equipment. Moreover, experimental equipment is almost always highly specialized equipment able to measure only one or a small group of properties of matter, whereas computers are very versatile equipments, which can be used to solve a variety of tasks. These factors increasingly contribute to the excellent cost/knowledge gained ratio in computational physics. Hence, on the atomic/molecular scale the complete picture can often only be obtained from the results of *simulations* right down to the *quantum mechanical atomic/molecular level*, where the *many-particle Schrödinger equation* at last provides a concise and universal description of matter.

This review will cover only a fairly limited range of length- and time-scales. Physics studies matter at all length-scales from astrophysics all the way down to subnuclear. As Fig. 1 shows

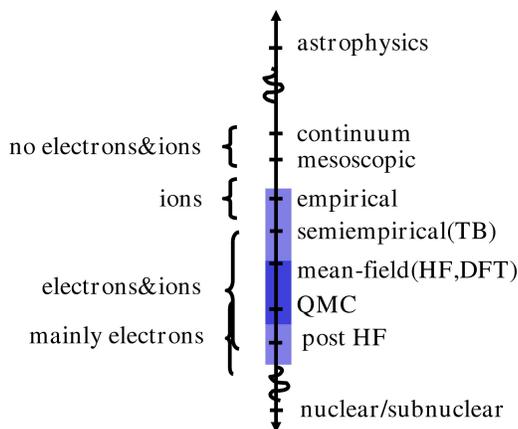


Fig. 1. Outline of the different simulation techniques used in simulation of matter. The techniques, which are of interest in this review, are indicated in the light blue box and those which will explicitly be dealt with are indicated in a dark blue box. TB means tight-binding, HF Hartree-Fock, DFT Density functional theory, post HF means methods such as MP2, configuration interaction (CI), coupled cluster (CC), and QMC stands for the quantum Monte-Carlo method. The respective degrees-of-freedom modeled are shown on the left. The two opposite limits: astrophysics and subnuclear physics are also shown for reference.

each length-scale deals with different degrees-of-freedom and requires a different sophistication in description of the matter at that length- and time-scale. This paper will deal exclusively with the *angstrom/nanometer scale*. Our model shall explicitly and simultaneously include *two degrees-of-freedom: electrons and ions*.

In practice there are two key parameters, which will determine the method of choice. Those are:

- the maximum allowed absolute error ε ,
- the maximum allowed time T , needed to achieve the specified accuracy.

Most processes of interest, such as chemistry and nanotechnology take place at room temperature. Current “*chemical accuracy*” is considered to be ≈ 1 kcal/mol. However, often a considerably higher accuracy is required. For example to calculate energy differences requires an error of $\varepsilon \leq 100$ K ≈ 0.1 kcal/mol ≈ 0.01 eV. Many phenomena, e.g. superconductivity, require even higher accuracy. This level of accuracy is extremely hard to achieve across the board of applications and only a few of those discussed here (QMC and some post HF methods) have the aspiration to challenge the chemical accuracy. This review should help the interested reader to strike the right balance between the required accuracy ε and the allowed computer time T to achieve it.

The aim of this review is to guide the reader through the world of atomic/molecular-scale simulations. The first part will provide a self-contained description of the methods of *ab-initio* simulation and the basic approximations. First the basic equations will be formulated. Solutions of those equations will be sought. The line of argument will be divided in three broad parts: electrons, ions, and techniques for solving the equations. In the part describing the methods for modeling the electrons the Hartree-Fock, density functional techniques and the quantum Monte-

Carlo techniques are addressed more in detail. The choice is dictated by practical considerations: the first, while usually not sufficiently accurate for real applications, is of great pedagogical value. The density functional techniques are usually the method of choice striking often an ideal balance between accuracy and feasibility. On the other hand, quantum Monte-Carlo is the method of choice if correlation effects are important in systems beyond very small molecules. The description of methods for description of ions addresses in addition to classical ions also quantum ions by the path integral method. Some technical aspects, such as the importance and role of parallel computers will briefly be addressed. In the second part the application of those methods will be demonstrated on selected systems. The applications will be mainly taken from the author's own work and the work of his group. Where available, the presentation will be accompanied by computer graphics animations of the simulations denoted as $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}\}$. The applications will be augmented by a few selected applications taken from the work of other groups. Incidentally, the presentation to a certain degree also follows the evolution of the field in time: from non-simple liquids, disordered systems, surfaces, interaction of atoms and molecules with surfaces, simulation of surface probe techniques, chemical reactions to applications relevant to nanotechnology. Contrary to the brief demonstrations of results obtained by the presented theory, which appear in the first part, the projects presented in the second part are more in depth presentations aiming at providing information on the typical parameters used in simulations, practically achievable length- and time-scales etc. These projects are also showing how the results, i.e. huge amounts of numerical data from large-scale simulations, can be interpreted in order to obtain a coherent picture of the system or process under consideration. The account is by no means complete. The paper only briefly touches on the methods of quantum chemistry; no methods for quantum transport are discussed, etc. No account of magnetic properties, biological applications, etc. is given. Hence, the presented angle of view will be a very personal one with no aspiration to cover all the facets of the atomic/molecular-scale modeling.

2 The model and basic equations

We start by description of the underlying model for most simulations at the atomic/molecular scale. A typical example being a molecule, atomic or molecular cluster, crystal, solid surface, structurally disordered solid, liquid, etc. The model explicitly includes *ions* with their *electrons* constituting the condensed matter under study. Hence, in the most general case, one is seeking a solution to a quantum mechanical problem for N_I ions defined by their masses M_I , nuclear charges eZ_I , and position vectors $\{\vec{R}_I\}$ and N_e electrons defined by electron mass m_e , electron charge e , located at positions $\{\vec{r}_i\}$. The corresponding quantum-mechanical Hamiltonian for the coupled electron-ion system

$$\begin{aligned}
 \hat{H} &= - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 + \sum_{i < j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} - \sum_{I,i} \frac{e^2 Z_I}{|\vec{R}_I - \vec{r}_i|} + \sum_{I < J} \frac{e^2 Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \\
 &= - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_i \frac{\hbar^2}{2m_i} \nabla_i^2 + V_{n-e}(\{\vec{r}_i\}, \{\vec{R}_I\}) \\
 &= - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + \hat{H}_e(\{\vec{r}_i\}, \{\vec{R}_I\})
 \end{aligned} \tag{1}$$

consists of the sum of kinetic energies of the ions, electrons, the electron-electron, electron-ion, and ion-ion interaction. For simplicity a non-relativistic quantum-mechanical formalism is assumed and energy is measured in general units. The more convenient atomic units (a.u.) will be introduced and used throughout later. The state of this coupled electron-ion system is described by time-dependent wave function $\Phi(\{\vec{r}_i\}, \{\vec{R}_I\}; t)$, which is solution to the time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \Phi(\{\vec{r}_i\}, \{\vec{R}_I\}; t) = \hat{H} \Phi(\{\vec{r}_i, \vec{R}_I\}; t). \quad (2)$$

This most general equation is too complicated to be solved directly. Even if it were technically possible to find the solution to equation (2), for the sake of clarity and insight, it is desirable to separate the ionic and electronic degrees of freedom as their mass ratio $M_I/m_e = 1836$. In other words, the electronic dynamics is always much faster than the ionic, and hence, the electronic degrees of freedom will always experience “static ions”. The goal here is to separate the electronic and ionic degrees of freedom in the Schrödinger’s wave equation [1–3]². The simplest way to achieve this objective is to consider single-configuration ansatz [4, 5] for the total wave function $\Phi(\{\vec{r}_i\}, \{\vec{R}_I\}; t)$ in the form

$$\Phi(\{\vec{r}_i\}, \{\vec{R}_I\}; t) \approx \Psi(\{\vec{r}_i\}, t) \chi(\{\vec{R}_I\}, t) \exp \left[\frac{i}{\hbar} \int_{t_0}^t dt' \tilde{E}_e(t') \right], \quad (3)$$

with both electronic and ionic wave functions, respectively, separately normalized at every time, i.e., $\langle \chi(\{\vec{R}_I\}; t) | \chi(\{\vec{R}_I\}; t) \rangle = 1$ and $\langle \Psi(\{\vec{r}_i\}; t) | \Psi(\{\vec{r}_i\}; t) \rangle = 1$. The phase factor is taken in a convenient form of mean value of the H_e

$$\tilde{E}_e = \int d\vec{r} d\vec{R} \Psi^*(\{\vec{r}_i\}; t) \chi^*(\{\vec{R}_I\}; t) \hat{H}_e \Psi(\{\vec{r}_i\}; t) \chi(\{\vec{R}_I\}; t). \quad (4)$$

The single-configuration ansatz (3) clearly leads to a mean-field description of the coupled dynamics. Using ansatz (3) and (4) in Schrödinger’s equation (2) (after multiplication from left by $\langle \Psi |$ and $\langle \chi |$ and imposition of energy conservation $d\langle H \rangle / dt \equiv 0$) yields

$$i\hbar \frac{\partial \Psi}{\partial t} = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + \left\{ \int d\vec{R} \chi^*(\{\vec{R}_I\}; t) V_{n-e}(\{\vec{r}_i\}, \{\vec{R}_I\}) \chi(\{\vec{R}_I\}; t) \right\} \Psi, \quad (5)$$

$$i\hbar \frac{\partial \chi}{\partial t} = - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 \chi + \left\{ \int d\vec{r} \Psi^*(\{\vec{r}_i\}; t) \hat{H}_e(\{\vec{r}_i\}, \{\vec{R}_I\}) \Psi(\{\vec{r}_i\}; t) \right\} \chi. \quad (6)$$

This set of equations defines the so-called time-dependent self-consistent field method (TDSCF) introduced in the early days of quantum mechanics by Dirac [6, 7]. Equations (5), (6) define quantum-mechanical motion of both ionic and electronic degrees of freedom. The motion proceeds in time-dependent effective potentials obtained from quantum-mechanical expectation values $\int d\vec{R} \dots$, $\int d\vec{r} \dots$, respectively. The single determinant ansatz (Eq. (3)) yields a mean-field description of the coupled electron-nuclear quantum dynamics.

² The following part is little technical. The readers who accept that the electronic and ionic dynamics may be separated and that the ions are usually sufficiently massive to behave classically, may proceed directly to Eq. (11) describing the Ehrenfest dynamics.

The ions by being massive, compared to electrons will in most cases behave classically. The reason is, that the quantum fluctuations, which are of the order of $\hbar\omega$ (where ω is a typical frequency associated with ionic dynamics) are, under normal circumstances (massive ions, room temperatures) much smaller than thermal fluctuations, which are of the order of $k_B T$ (with k_B the Planck's constant and T temperature). The case, where the assumption of negligible quantum fluctuations is not met and $\hbar\omega \approx k_B T$ will be dealt with later. Hence, it is desirable to reformulate the wave equation (6) for ionic dynamics in terms of classical equation of motion. This can be done in classical limit ($\hbar \rightarrow 0$) to quantum mechanics by writing the nuclear wave function as a product of an amplitude factor A and a phase S in the form [3, 6, 8]

$$\chi(\{\vec{R}_I\}; t) = A(\{\vec{R}_I\}; t) \exp \left[iS(\{\vec{R}_I\}; t) / \hbar \right]. \quad (7)$$

After insertion of ansatz (7) into Eq. (6) and taking the classical limit, the coupled electron-nuclear equations take the following form

$$i\hbar \frac{\partial \Psi}{\partial t} = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + \left\{ \int d\vec{R} \chi^*(\{\vec{R}_I\}; t) V_{n-e}(\{\vec{r}_i\}, \{\vec{R}_I\}) \chi(\{\vec{R}_I\}; t) \right\} \Psi, \quad (8)$$

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I \int d\vec{r} \Psi^* \hat{H}_e \Psi = -\nabla_I V_e^E(\{\vec{R}_I(t)\}). \quad (9)$$

The problem with the equations (8), (9) is that Eq. (8) still contains explicitly the nuclear wave function. This can be remedied by taking the classical limit also in Eq. (8) in which case the $|\chi(\{\vec{R}_I\}; t)|^2$ in classical limit can be replaced by $\prod_I \delta(\vec{R}_I - \vec{R}_I(t))$. This yields

$$\int d\vec{R} \chi^*(\{\vec{R}_I\}; t) \vec{R}_I \chi(\{\vec{R}_I\}; t) \xrightarrow{\hbar \rightarrow 0} \vec{R}_I(t), \quad (10)$$

and the coupled electron-nuclear equations become

$$\begin{aligned} i\hbar \frac{\partial \Psi}{\partial t} &= - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + V_{n-e}(\{\vec{r}_i\}, \{\vec{R}_I\}) \Psi \\ &= - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 \Psi + V_{n-e}(\{\vec{r}_i\}, \{\vec{R}_I\}) \Psi, \end{aligned} \quad (11)$$

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I \int d\vec{r} \Psi^* \hat{H}_e \Psi = -\nabla_I V_e^E(\{\vec{R}_I(t)\}). \quad (12)$$

These equations are often referred to as “*Ehrenfest molecular dynamics*” (EMD) [9]. Equations (11), (12) are coupled, as the potential energy in the time-dependent electronic equation depends on the ionic positions and vice versa, the potential energy from which the forces are derived in Eq. (12) require the knowledge of the electronic wave functions.

Both TDSCF and EMD are clearly mean-field theories. However, they incorporate transitions between different electronic states, which can be shown by expanding the time-dependent electronic wave function $\Psi(\{\vec{r}_i\}, \{\vec{R}_I\}; t)$ in terms of eigenstates of the electronic Hamiltonian H_e

$$\hat{H}_e(\{\vec{r}_i\}, \{\vec{R}_I\}) \Psi_k(\{\vec{r}_i\}, \{\vec{R}_I\}; t) = E_k(\{\vec{R}_I\}) \Psi_k(\{\vec{r}_i\}, \{\vec{R}_I\}; t) \quad (13)$$

as

$$\Psi(\{\vec{r}_i\}, \{\vec{R}_I\}; t) = \sum_{k=0}^{\infty} c_k(t) \Psi(\{\vec{r}_i\}, \{\vec{R}_I\}; t) \quad (14)$$

where $\{c_k\}$ are complex numbers giving the probabilities ($\equiv |c_k|^2$) of finding the system in the given eigenstate k and $\{\vec{R}_I\}$ are instantaneous nuclear positions at time t (Eq. (12)).

The wave function $\Psi(\{\vec{r}_i\}, \{\vec{R}_I\}; t)$ will often be dominated by the ground-state wave function Ψ_0 so that at each time t , $|c_0(t)|^2 \equiv 1$. Such an approximation will hold if the energy gap between the ground-state Ψ_0 and first excited state Ψ_1 is large compared to the thermal energy $k_B T$. Typical situation where the above approximation is well satisfied are wide-gap insulators. The opposing limit is a case of a metal, where there is no gap whatsoever. Practical implications of this approximation will be discussed in chapters 3.1.5.9, 3.1.5.12, and 3.1.5.13. The simplification brought about by this approximation is large, as in this case the nuclei will move on a single potential energy surface (PES) following Eq. (12) with

$$V_e^E = \int d\vec{r} \Psi_0^* \hat{H}_e \Psi_0 \equiv E_0(\{\vec{R}_I\}). \quad (15)$$

The E_0 can be computed for each configuration of the ions $\{\vec{R}_I\}$ by solving the time-independent Schrödinger's equation $H_e \Psi_0 = E_0 \Psi_0$ for only the ground-state. Equation (15) means that electronic structure is included via static, time-independent Schrödinger equation for a set of fixed nuclear positions encountered in each molecular dynamics step. This is the basis of the so-called *Born-Oppenheimer molecular dynamics* (BOMD), which can be summarized as follows

$$\hat{H}_e \Psi_0 = E_0 \Psi_0, \quad (16)$$

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I \min_{\Psi_0} \{\langle \Psi_0 | H_e | \Psi_0 \rangle\}. \quad (17)$$

A comment is due here. The BOMD was formulated for the ground-state wave function, which is the most commonly used approximation which will also be widely used in this paper. However, any other single electronic state Ψ_k can be used to describe excited state dynamics of a system. Typical examples being for instance photochemistry or laser excitations in solid-state systems. From the discussion leading to the BOMD it may seem that this formulation excludes its use to metal. The use of the BOMD to excited state dynamics and to metals will be discussed later (chapters 3.1.5.12, 3.1.5.13).

So far two types of dynamics were introduced. The EMD (Eqs. (11 – 12)) and BOMD (Eqs. (16 – 17)). The two schemes are identical in that both are mean-field approaches starting from a single-reference wave function but differ in two major respects: (1) the EMD allows for transitions between different electronic states, whereas the BOMD is strictly limited to one state, e.g. the ground-state wave function and (2) there is no real electronic dynamics associated with BOMD as the ground-state electronic wave function is obtained only in each ionic dynamics step, whereas in EMD, the wave function, which minimized $\langle H_e \rangle$ initially will stay in its minimum along with the moving ions.

In principle, there is yet another type of dynamics, often referred to as *classical trajectory generation* [10]. The basis of this approach is construction of an entire PES $E_0(\{\vec{R}_I\})$ defined

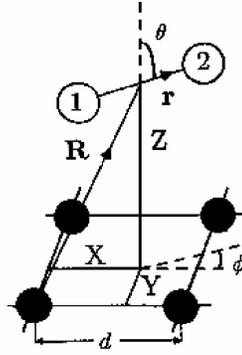


Fig. 2. Definition of the geometrical variables characterizing a diatomic molecule approaching a surface.

in Eq. (15) for each ionic configuration $\{\vec{R}_I\}$. From the knowledge of $E_0(\{\vec{R}_I\})$ one could generate the gradients $\nabla_I E_0(\{\vec{R}_I\})$. With this knowledge, one could run “classical trajectories” containing in principle identical information as the BOMD. The problem here is the dimensionality of the underlying configuration space $\{\vec{R}_I\}$. Imagine a very simple system, for instance a diatomic molecule over a static surface. Already this very simple system has 6 degrees of freedom $(x, y, z, r, \theta, \phi)$: x, y -coordinate of the molecule defining the position of the molecule in the surface unit cell, z distance of the molecule from the surface, r the molecular stretch, and θ, ϕ the molecular orientation; see Fig. 2. If each degree of freedom were sampled with just 10 points 10^6 independent energies would be required to construct the corresponding PES. Most likely, most of these points would be calculated for parts of the configuration space which the molecule at any given temperature would not visit. The situation would be even more computationally involved if surface response, such as surface relaxation due to molecule-surface interaction and/or surface vibrations were taken into account. This is the reason why the coupled electron-ion dynamics defined in Eqs. (11–12) and (16–17) is very economic. Evaluation of the electronic wave function is typically much more time consuming than classical propagation of the ions. Hence, computation of the PES *on-the-fly*, as is the case in both EMD and BOMD greatly reduces the computational load as only those points on the PES are evaluated, which are actually visited by the ionic degrees of freedom in a given thermodynamic state. The situation is akin to computation of integrals in very high-dimensional space. Here too, evaluation of an integral on a regular grid of points is the least efficient way to evaluate the integral and the alternative stochastic methods based on Metropolis Monte-Carlo techniques offer a much more computationally economic approach. The reason is, that similarly to EMD and BOMD, only those integrand points will be visited, which make a “significant” contribution to the integral value.

3 Reformulation of basic equations

We will now seek the solutions of the EMD and BOMD coupled electron-nuclear dynamics equations. In both schemes the equations for the electronic and ionic degrees-of-freedom are very different: the former are solutions to either time-dependent (EMD) or time-independent

(BOMD) Schrödinger's equation, whereas the latter are simple, classical Newton's equations. The two degrees-of-freedom will be discussed separately.

3.1 Electronic structure

Perhaps the simplest approximation for the total many-body wave function $\Psi(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_j, \dots, \xi_{N_e})$ with $\xi_i = (\vec{r}_i, \sigma_i)$ where \vec{r}_i, σ_i are position vector and projection of the spin of the i^{th} electron, respectively was made by Hartree [11] in writing $\Psi(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_j, \dots, \xi_{N_e})$ as a product of non-interacting single-particle molecular orbitals (MO) $\{\varphi_i(\xi)\}$.

$$\Psi_H(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_j, \dots, \xi_{N_e}) = \varphi_1(\xi_1)\varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}). \quad (18)$$

Ansatz (18), also called *Hartree approximation*, represents a very poor approximation to the true many-body wave function. The reason is that the wave function in this approximation does not satisfy the basic symmetry under particle exchange. Electrons are indistinguishable Fermions and hence, the total electronic wave function, including its spin part, must be antisymmetric under spatial exchange³

$$\Psi_i(\xi_1, \xi_2, \dots, \xi_i, \dots, \xi_j, \dots, \xi_{N_e}) = -\Psi_i(\xi_1, \xi_2, \dots, \xi_j, \dots, \xi_i, \dots, \xi_{N_e}). \quad (19)$$

The simplest way to satisfy the antisymmetry of the electronic wave function (19) is to take it in the form of a (Slater) determinant or sum of (Slater) determinants of independent, non-interacting electrons described by single-particle orbitals $\{\varphi_i^L(\xi)\}$

$$\Psi_i = \sum_L c_{iL} \Phi_L = \sum_L c_{iL} \begin{vmatrix} \varphi_1^L(\xi_1) & \varphi_1^L(\xi_2) & \cdots & \varphi_1^L(\xi_{N_e}) \\ \varphi_2^L(\xi_1) & \varphi_2^L(\xi_2) & \cdots & \varphi_2^L(\xi_{N_e}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{N_e}^L(\xi_1) & \varphi_{N_e}^L(\xi_2) & \cdots & \varphi_{N_e}^L(\xi_{N_e}) \end{vmatrix}. \quad (20)$$

The familiar *Pauli exclusion principle* follows directly from the wave function in the form (20). What happens if we attempt to doubly occupy a spinorbital $\varphi(\xi)$ by putting electron i and j in it? Two rows in the Slater determinant will be equal and, hence, $\Psi \equiv 0$! The simplest MO ansatz includes only a single term in the expansion (20)

$$\Psi_{HF} = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \varphi_1(\xi_1) & \varphi_1(\xi_2) & \cdots & \varphi_1(\xi_{N_e}) \\ \varphi_2(\xi_1) & \varphi_2(\xi_2) & \cdots & \varphi_2(\xi_{N_e}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{N_e}(\xi_1) & \varphi_{N_e}(\xi_2) & \cdots & \varphi_{N_e}(\xi_{N_e}) \end{vmatrix}. \quad (21)$$

This approximation, called also *Hartree-Fock approximation* [12, 13] can only be expected to yield reasonable results if the MOs in Ψ_{HF} are variationally optimized: the so-called *Self-Consistent Field* (SCF) approach. How this can be done is shown below.

³ Note that the antisymmetry of the many-body wave function implies that there is a node (position(s) in space where the wave function has a value equal zero) of the wave function inbetween the two particles positions. The nodal surfaces are a very important ingredient of the many-body wave function.

3.1.1 Hartree-Fock theory

The Hartree-Fock theory is described in an excellent manner in book by Szabo and Ostlund [14], and hence we will limit ourselves to just the outline of the method. Let's assume a wave function of the Slater determinant form (21) and find an expression for the expectation value of the energy

$$E_{HF} = \langle \Psi_{HF} | \hat{H}_e | \Psi_{HF} \rangle, \quad (22)$$

$$\hat{H}_e = \sum_i \hat{h}(i) + \sum_{i < j} \frac{1}{r_{ij}} = \hat{H}_1^{core} + \hat{H}_2. \quad (23)$$

The above Hamiltonian is written in atomic units (a.u.); $\hbar = 1$, $m_e = 1$, $e = 1$. In these units, energy is measured in units of Rydbergs and lengths in units of Bohr radii, a_B . Atomic units will be used throughout, unless stated otherwise. First examine the core Hamiltonian \hat{H}_1^{core}

$$\begin{aligned} \langle \Psi_{HF} | \hat{H}_1^{core} | \Psi_{HF} \rangle &= \sum_i \langle \Psi_{HF} | \hat{h}(i) | \Psi_{HF} \rangle = \\ &= \sum_i \langle \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) | \hat{h}(i) | \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) \rangle. \end{aligned} \quad (24)$$

The nature of the approximation is best evidenced by considering a very simple case $N_e = 2$, e.g. a Helium atom

$$\Psi_{HF} = \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(\xi_1) & \varphi_1(\xi_2) \\ \varphi_2(\xi_1) & \varphi_2(\xi_2) \end{vmatrix} = |\varphi_1(\xi_1) \varphi_2(\xi_2)\rangle. \quad (25)$$

Look at one term in the above sum, for instance $h(1)$

$$\begin{aligned} \langle \varphi_1(\xi_1) \varphi_2(\xi_2) | \hat{h}(1) | \varphi_1(\xi_1) \varphi_2(\xi_2) \rangle &= \\ &= \frac{1}{2} \int [\varphi_1^*(\xi_1) \varphi_2^*(\xi_2) - \varphi_1^*(\xi_2) \varphi_2^*(\xi_1)] \hat{h}(1) [\varphi_1(\xi_1) \varphi_2(\xi_2) - \varphi_1(\xi_2) \varphi_2(\xi_1)] d\xi_1 d\xi_2 = \\ &= \frac{1}{2} \int \varphi_1^*(\xi_1) \varphi_2^*(\xi_2) \hat{h}(1) \varphi_1(\xi_1) \varphi_2(\xi_2) d\xi_1 d\xi_2 \\ &+ \frac{1}{2} \int \varphi_1^*(\xi_2) \varphi_2^*(\xi_1) \hat{h}(1) \varphi_1(\xi_2) \varphi_2(\xi_1) d\xi_1 d\xi_2 \\ &- \frac{1}{2} \int \varphi_1^*(\xi_2) \varphi_2^*(\xi_1) \hat{h}(1) \varphi_1(\xi_1) \varphi_2(\xi_2) d\xi_1 d\xi_2 \\ &+ \frac{1}{2} \int \varphi_1^*(\xi_1) \varphi_2^*(\xi_2) \hat{h}(1) \varphi_1(\xi_2) \varphi_2(\xi_1) d\xi_1 d\xi_2 \\ &= \frac{1}{2} \left(\langle \varphi_1 | \hat{h} | \varphi_1 \rangle + \langle \varphi_2 | \hat{h} | \varphi_2 \rangle \right). \end{aligned} \quad (26)$$

Now look at \hat{H}_2

$$\begin{aligned} \langle \varphi_1(\xi_1) \varphi_2(\xi_2) | \frac{1}{r_{12}} | \varphi_1(\xi_1) \varphi_2(\xi_2) \rangle &= \\ &= \frac{1}{2} \int [\varphi_1^*(\xi_1) \varphi_2^*(\xi_2) - \varphi_1^*(\xi_2) \varphi_2^*(\xi_1)] \frac{1}{r_{12}} [\varphi_1(\xi_1) \varphi_2(\xi_2) - \varphi_1(\xi_2) \varphi_2(\xi_1)] d\xi_1 d\xi_2 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2} \int \varphi_1^*(\xi_1) \varphi_2^*(\xi_2) \frac{1}{r_{12}} \varphi_1(\xi_1) \varphi_2(\xi_2) d\xi_1 d\xi_2 \\
 &+ \frac{1}{2} \int \varphi_1^*(\xi_2) \varphi_2^*(\xi_1) \frac{1}{r_{12}} \varphi_1(\xi_2) \varphi_2(\xi_1) d\xi_1 d\xi_2 \\
 &- \frac{1}{2} \int \varphi_1^*(\xi_2) \varphi_2^*(\xi_1) \frac{1}{r_{12}} \varphi_1(\xi_1) \varphi_2(\xi_2) d\xi_1 d\xi_2 \\
 &+ \frac{1}{2} \int \varphi_1^*(\xi_1) \varphi_2^*(\xi_2) \frac{1}{r_{12}} \varphi_1(\xi_2) \varphi_2(\xi_1) d\xi_1 d\xi_2.
 \end{aligned}$$

The $\frac{1}{r_{12}}$ operator prevents separation over electronic coordinations of electron 1 and electron 2. However, since ξ_1 and ξ_2 are dummy variables, the first and second terms are equal, as are the last two. Thus for the two-electron operator $\frac{1}{r_{12}}$,

$$\langle \varphi_1(\xi_1) \varphi_2(\xi_2) | \frac{1}{r_{12}} | \varphi_1(\xi_1) \varphi_2(\xi_2) \rangle = \langle 12|12 \rangle - \langle 12|21 \rangle = \langle 12|12 \rangle, \quad (27)$$

where

$$\langle ij|kl \rangle = \int d\xi_1 d\xi_2 \varphi_i^*(\xi_1) \varphi_j^*(\xi_2) \frac{1}{r_{12}} \varphi_k(\xi_1) \varphi_l(\xi_2), \quad (28)$$

$$\langle ij||kl \rangle = \langle ij|kl \rangle - \langle ij|lk \rangle. \quad (29)$$

Generalizing the above results to the general case, we can write down the HF energy for a given set of occupied spinorbitals

$$\begin{aligned}
 E_{HF} &= \langle \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) | \hat{H}_1^{core} | \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) \rangle \\
 &+ \langle \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) | \hat{H}_2 | \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) \rangle = \\
 &= \sum_i \langle \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) | \hat{h}(i) | \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) \rangle \\
 &+ \sum_{i<j} \langle \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) | \frac{1}{r_{is}} | \varphi_1(\xi_1) \varphi_2(\xi_2) \cdots \varphi_{N_e}(\xi_{N_e}) \rangle \\
 &= \sum_i \langle \varphi_i | \hat{h} | \varphi_i \rangle + \sum_{i<j} \langle ij||ij \rangle. \quad (30)
 \end{aligned}$$

Now consider a more general case, where some orbitals needn't be doubly occupied. We will now explicitly decompose the spinorbitals into special and spin component

$$\varphi_i(\xi) = \psi_i(\vec{r}) \chi_i(\sigma). \quad (31)$$

For an electron $\chi(\sigma)$ can be written as a column/row vector of either spin-up or spin-down and the expressions for multiplication of the spin wave functions of the type (\cdot) , (\cdots) are easy to perform. After a little algebra one obtains

$$E_{HF} = 2 \sum_i^{occ} f_i h_{ii} + \sum_{i,j}^{occ} \{ \alpha_{ij} (ii|jj) + \beta_{ij} (ij|ij) \}$$

$$= 2 \sum_i^{occ} f_i h_{ii} + \sum_{i,j}^{occ} \{ \alpha_{ij} J_{ij} + \beta_{ij} K_{ij} \} \quad (32)$$

$$= 2 \sum_i^{occ} f_i h_{ii} + \sum_{i,j}^{occ} \left\{ \alpha_{ij} \int d\vec{r}_1 d\vec{r}_2 \varphi_i^*(\vec{r}_1) \varphi_j^*(\vec{r}_2) \frac{1}{r_{12}} \varphi_i(\vec{r}_1) \varphi_j(\vec{r}_2) \right\} \\ + \sum_{i,j}^{occ} \left\{ \beta_{ij} \int d\vec{r}_1 d\vec{r}_2 \varphi_i^*(\vec{r}_1) \varphi_j^*(\vec{r}_2) \frac{1}{r_{12}} \varphi_i(\vec{r}_2) \varphi_j(\vec{r}_1) \right\}, \quad (33)$$

with

$$\begin{aligned} f_i &= \begin{array}{ll} 1 & \text{if } i \text{ doubly occupied,} \\ 1/2 & \text{if } i \text{ singly occupied,} \end{array} \\ \alpha_{ij} &= \begin{array}{ll} 2 & \text{if } i \text{ and } j \text{ doubly occupied,} \\ 1 & \text{if } i \text{ or } j \text{ doubly occupied and the other singly occupied,} \\ 1/2 & \text{if } i \text{ and } j \text{ singly occupied,} \end{array} \\ \beta_{ij} &= \begin{array}{ll} -1 & \text{if } i \text{ and } j \text{ doubly occupied,} \\ -1/2 & \text{if } i \text{ or } j \text{ singly occupied and the other doubly occupied,} \\ -1/2 & \text{if } i \text{ and } j \text{ singly occupied with parallel spins,} \\ 1/2 & \text{if } i \text{ and } j \text{ singly occupied with opposite spins.} \end{array} \end{aligned}$$

In Eq. (32) J_{ij} is termed Coulomb integral and has the physically reassuring interpretation of accounting for *electronic repulsion* between electrons in molecular orbital i , and molecular orbital j . K_{ij} , the *exchange integral* has no classical analog and no true physical interpretation. It is the heritage of having written the many-body wave function in the form of a Slater determinant. In loose terms we can say that it ‘‘correlates the motion of electrons i and j when they have parallel spins, lowering the energy since those electrons avoid each other better’’.

3.1.1.1 Hatree-Fock equation

Now let us find the optimized molecular orbitals which from the Slater determinant $\Psi_{HF} = |\varphi_1 \varphi_2 \cdots \varphi_{N_e}\rangle$. In other words, we want to find the ‘‘best’’ wave function Ψ_{HF} subject to the constraint that the spin orbitals all remain orthonormal one to another. The best wave function is the one, which minimizes the electronic energy expression (32) will be determined by the method of Lagrange multipliers, which is a general method for finding extremes of functions or functionals subject to constraints. We now return to the spinorbital notation. The orthonormality constraints can be stated as

$$\langle a|b\rangle - \delta_{ab} = 0. \quad (34)$$

For N_e spinorbitals there are $N_e(N_e + 1)/2$ independent constraints, so we need that many multipliers in our constraint minimization. The function, which will be minimized, is the sum of the energy function E_{HF} and the constraints

$$\mathfrak{S}[\{\varphi_a\}] = E_{HF}[\{\varphi_a\}] - \sum_a^{N_e} \sum_b^a \varepsilon_{ba} (\langle b|a\rangle - \delta_{ab}).$$

Since ε_{ab} is Hermitian, $\varepsilon_{ab}^* = \varepsilon_{ba}$, the above equation can be rewritten as

$$\mathfrak{S}[\{\varphi_a\}] = E_{HF}[\{\varphi_a\}] - \sum_a^{N_e} \sum_b^{N_e} \varepsilon_{ba} (\langle b|a\rangle - \delta_{ab}), \quad (35)$$

without introducing any new undetermined multipliers. The best set of orbitals $\{\varphi_i\}$ can be determined by setting the *variation* of $\mathfrak{S}[\{\varphi_a\}]$ to zero

$$\delta\mathfrak{S}[\{\varphi_a\}] = \delta E_{HF}[\{\varphi_a\}] - \sum_a^{N_e} \sum_b^{N_e} \varepsilon_{ba} \delta \langle b|a\rangle \stackrel{!}{=} 0. \quad (36)$$

We have an electronic energy in terms of spinorbitals

$$\begin{aligned} E_{HF} &= \sum_a^{N_e} \langle a|H_1|a\rangle + \frac{1}{2} \sum_{ab}^{N_e} \langle ab|ab\rangle \\ &= \sum_a^{N_e} \langle a|H_1|a\rangle + \frac{1}{2} \sum_{ab}^{N_e} [\langle ab|ab\rangle - \langle ab|ba\rangle], \end{aligned}$$

and we can write the variation of the energy δE_{HF} as

$$\begin{aligned} \delta E_{HF} &= \sum_a^{N_e} \left(\langle \delta a|\hat{H}_1|a\rangle + \langle a|\hat{H}_1|\delta a\rangle \right) + \\ &+ \frac{1}{2} \sum_{ab}^{N_e} \left[\begin{aligned} &\langle \delta ab|ab\rangle + \langle a\delta b|ab\rangle + \langle ab|\delta ab\rangle + \langle ab|a\delta b\rangle \\ &- \langle \delta ab|ba\rangle - \langle a\delta b|ba\rangle - \langle ab|\delta ba\rangle - \langle ab|b\delta a\rangle \end{aligned} \right]. \end{aligned} \quad (37)$$

After some thought one will realize that there are only two unique two-electron integrals in (37), and that it can be written

$$\delta E_{HF} = \sum_a^{N_e} \langle \delta a|\hat{H}_1|a\rangle + \sum_{ab}^{N_e} (\langle \delta ab|ab\rangle - \langle \delta ab|ba\rangle) + \text{complex conjugate}. \quad (38)$$

The other variation we need is $\varepsilon_{ab} \delta \langle a|b\rangle$, which can be written as

$$\begin{aligned} \sum_{ab} \varepsilon_{ab} \langle \delta a|b\rangle + \sum_{ab} \varepsilon_{ba} \langle a|\delta b\rangle &= \sum_{ab} \varepsilon_{ba} \langle \delta a|b\rangle + \sum_{ab} \varepsilon_{ab} \langle b|\delta a\rangle \\ &= \sum_{ab} (\varepsilon_{ba} \langle \delta a|b\rangle + \varepsilon_{ba}^* \langle \delta a|b\rangle). \end{aligned} \quad (39)$$

So the whole expression boils down to

$$\delta\mathfrak{S} = \sum_a^{N_e} \langle \delta a|\hat{H}_1|a\rangle + \sum_{ab}^{N_e} (\langle \delta ab|ab\rangle - \varepsilon_{ba} \langle \delta a|b\rangle) + \text{complex conjugate} = 0. \quad (40)$$

If we now define a Coulomb operator $J_b(1)$ and exchange operator $K_b(1)$ as

$$J_b(1) = \int d\vec{r}_2 |\varphi_b(2)|^2 r_{12}^{-1}, \quad (41)$$

$$K_b(1)\varphi_a(2) = \int d\vec{r}_2 \varphi_b^*(2) r_{12}^{-1} \varphi_a(2) \varphi_b(1). \quad (42)$$

This allows a more compact notation to be employed in writing the variation in \mathfrak{S}

$$\begin{aligned} \delta\mathfrak{S} = \sum_a \int d\vec{r}_1 \delta\varphi_a^*(1) \left[\hat{h}(1)\varphi_a(1) + \sum_b \{(J_b(1) - K_b(1))\varphi_a(1) - \varepsilon_{ba}\varphi_b(1)\} \right] \\ + \text{complex conjugate} = 0. \end{aligned} \quad (43)$$

Equation (43) must be valid for an arbitrary variation $\delta\varphi_a^*(1)$, and hence, after rearranging the resulting equation to look like some sort of eigenvalue equation, yields,

$$\left[\hat{h}(1) + \sum_b (J_b(1) - K_b(1)) \right] \varphi_a(1) = \sum_b \varepsilon_{ba} \varphi_b(1), \quad (44)$$

$$\hat{f}(1)\varphi_a(1) = \sum_b \varepsilon_{ba} \varphi_b(1), \quad (45)$$

where \hat{f} is the so-called *Fock operator*. As the above equation shows, its solution will converge to a linear combination of Hartree-Fock canonical orbitals. However, the linear combination and the canonical orbitals are related by a unitary rotation described by a transformation matrix U_{ba} , which is unitary, meaning that the sets of wave functions related by unitary transformations will differ only a phase factor, affecting nothing observable. Hence, we are free to choose U_{ba} to whatever we please, and if we chose it to make the ε_{ba} matrix diagonal, we can rewrite the *Hartree-Fock equation* as

$$\hat{f}|\varphi_a\rangle = \varepsilon_a |\varphi_a\rangle. \quad (46)$$

When this is done, the resulting orbitals will converge to the Hartree-Fock canonical orbitals. Equations (46) are coupled integro-differential equations as the solution of orbital a depends on all other orbitals b , which form a static mean-field in which electron a moves. Hence, the solutions will have to be obtained by *iterative solution*. One starts with a guess for the set on one-particle solutions $\{\varphi_i^0\}$, from which the Fock operator \hat{f} is constructed and a first iteration on the solutions $\{\varphi_i^1\}$ obtained. This procedure is repeated until the input wave functions $\{\varphi_i^n\}$ do not differ from the input ones $\{\varphi_i^{n+1}\}$.

Though the Hartree-Fock theory has been formulated in the early 1930ths, for a long time numerical calculations could only be performed for atoms. The breakthrough came in 1951 when Hall [15] and Roothaan [16] independently proposed to expand the Hartree-Fock orbitals into a set of atom-centered basis functions, $\varphi_i = \sum_{\nu=1}^M c_{i\nu} \chi_\nu$. Obviously, $M \propto N_e$, as $N_e \propto N_I$. In this form the solution of Eq. (46) can be cast in a matrix diagonalization problem. This point

will be discussed in more detail in chapters 3.1.7.1 and 3.3.2.1. The numerical effort is easy to estimate: There are M^4 integrals $\langle ij|kl\rangle$ and the solution of the Hartree-Fock equation (46) requires diagonalization of a M -dimensional matrix, which is a $O(M^3)$ step. The numerical cost is dominated by the diagonalization step, and hence the overall numerical cost is $O(M^3)$.

The eigenvalues ε_a of the Hartree-Fock equation are called “orbital energies”. In general, they are negative for occupied (bound) orbitals and positive for virtual (continuum) orbitals. The orbital energies of the occupied orbitals can be identified with the *ionization potentials* of the system

$$\varepsilon_a = -IP(a) \quad (47)$$

via *Koopmans’ theorem* [17]. Equation (47) states that the orbital energy ε_a of the a -th occupied orbital is equal to the negative value of the energy necessary to remove an electron from the orbital φ_a . However, this is only an approximation because in deriving the theorem it has been assumed that the orbitals do not change after ionization. The accuracy of the Koopmans’ theorem is of the order of 0.5–2.0eV for valence orbitals and much larger for the core orbitals.

Despite its simplicity, HF approach is a very useful approximation. For instance, equilibrium structure constants of most main group systems are obtained with errors of about 2% in bond distances and a few degrees in bond angles. Simple examples are discussed in [14]. Exceptions are mainly found for weak bonds like those in F_2 and for systems containing transition metals. The HF approach is useless for description of metals or binding energies but reaction energies can be computed to within about 10 kJ/mol for a wide range of reactions. In addition, HF approach, which features an exact exchange, is an ideal starting point for more accurate treatment of electronic structure by the so-called post HF methods (chapter 3.1.4) or by the quantum Monte-Carlo method (chapter 3.1.6).

3.1.1.2 Exchange and correlation in the Hartree-Fock approximation

Let us now talk about the errors introduced by the SCF approach. Hartree-Fock is a mean-field theory, in which each electron has its own wave function, which in turn obeys an effective 1-electron Schrödinger equation. The effective Hamiltonian (Fock operator) contains the average field (Coulomb and exchange) of all other electrons in the system. The total electronic wave function for the system, ignoring the complications introduced by the Pauli exclusion principle, is a simple product of the orbitals. Following the Born interpretation of wave functions, this implies that if $P(\vec{r}_1, \vec{r}_2)$ is the probability density for finding electrons labeled 1 and 2 in regions of space around \vec{r}_1 and \vec{r}_2 , respectively,

$$P(\vec{r}_1, \vec{r}_2) = P(\vec{r}_1)P(\vec{r}_2), \quad (48)$$

i.e. the probability density for a given electron is independent of the positions of all others.

In reality, however, the motions of electrons are more intimately correlated. Because of the direct Coulomb repulsion of electrons, the instantaneous position of electron 2 forms a center of a region in space which electron 1 will avoid. This avoidance is more than that caused by the mean-field, and is local; if electron 2 changes position, the Coulomb hole for electron 1 moves with it. In contrast, in the mean-field theory, electron 1 has no knowledge of the instantaneous position of electron 2, only its average value, and thus motions are uncorrelated, and there is no depletion in

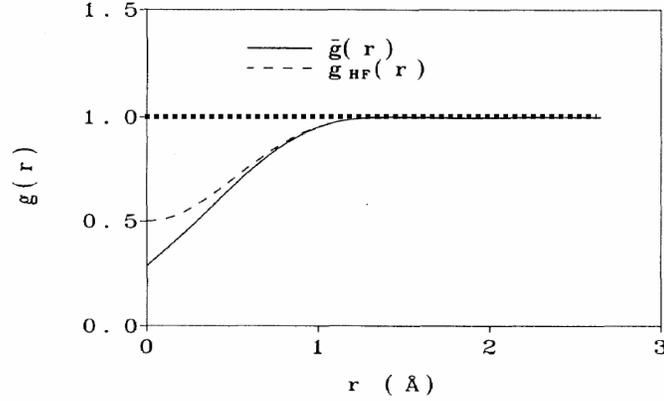


Fig. 3. Pair-correlation function in unpolarized homogeneous electron gas system [18]. Short-dash line: Hartree-treatment; dashed line: Hartree-Fock treatment; full line: the result after correlation is introduced.

$P(\vec{r}_1, \vec{r}_2)$ near $\vec{r}_1 = \vec{r}_2$. This deficiency can be nicely seen in Fig. 3, which shows the electronic pair correlation function of an electron in homogeneous electron gas (HEG) [18]. HEG is a model system in which the ions are smeared out and substituted by a positive background with the same electronic density. As will become evident later, this much studied system is of paramount importance in condensed-matter physics. Three types of pair correlation functions are shown in Fig. 3: Hartree-type of description does not form any hole around an electron; Hartree-Fock description improves the failure of the Hartree description quite significantly by introduction of the exchange-hole. However, as mentioned above, inclusion of the true electronic correlation still significantly modifies the exchange-hole by forming the correlation-hole in addition to the exchange-hole.

3.1.2 Cusp conditions: short-range correlations

The spurious effects described above, the *short-range correlations*, can be traced back to the r_{ij}^{-1} operator and the corresponding behavior of the wave function in regions close to $r_{ij} = 0$. Consider first the Schrödinger equation for the ground-state of the hydrogen atom,

$$\hat{H}\Psi_{1s} = -\frac{1}{2}\nabla^2\Psi_{1s} - \frac{1}{r}\Psi_{1s} = E_{1s}\Psi_{1s}, \quad (49)$$

the potential energy diverges as $r \rightarrow 0$ but $H\Psi_{1s}$ remains finite because there is a canceling divergence in the kinetic energy, which is reflected in the shape of the $1s$ wave function, which has a sharp cusp at the origin.

The potential energy diverges whenever an electron approaches a nucleus or another electron. Many-electron wave functions therefore contain similar cusps. Any necessary electron-nuclear cusps are built into the single-particle orbitals obtained by solving the mean-field HF equations. But these orbitals do not contain the electron-electron cusps [19, 20]. More generally, cusp conditions apply to any system with Coulomb interactions [19].

The many-electron wave function $\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e})$ satisfies the Schrödinger equation,

$$\hat{H}\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e}) = E\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e}), \quad (50)$$

where

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^{N_e} \nabla_i^2 + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e}). \quad (51)$$

The paradox here is that close to $r_{ij} = 0$ the left hand side of (50) becomes infinite because of the r_{ij}^{-1} Coulomb singularity, whereas E is constant, and so the right hand side is well behaved. The local energy $H\Psi/\Psi$ cannot have singularities since it is constant, and the only conclusion is that there must be an additional singularity or cusp, in the left hand side of (50) which exactly cancels r_{ij}^{-1} close to $r_{ij} = 0$. Since the electrons are not necessarily close to a nucleus, the only candidate for this canceling term is again the kinetic energy.

If we pick a pair of electrons i and j and introduce the center-of-mass, difference coordinates $\vec{r} = \vec{r}_i - \vec{r}_j$ and $r_{c.m.} = (\vec{r}_i + \vec{r}_j)/2$, \hat{H} may be rewritten as

$$\hat{H} = -\nabla_r^2 - \frac{1}{4} \nabla_{r_{c.m.}}^2 - \frac{1}{2} \sum_{\substack{k=1 \\ (k \neq i, j)}}^{N_e} \nabla_k^2 + V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e}). \quad (52)$$

Since Ψ is an exact eigenfunction, the corresponding exact local energy $\hat{H}\Psi/\Psi$ is everywhere equal to the eigenenergy E and does not diverge as $r \rightarrow 0$.

As we know, the Hartree-Fock wave function Ψ_{HF} is not an eigenfunction, so the local energy $\hat{H}\Psi_{HF}/\Psi_{HF}$ is not constant and may diverge as $r \rightarrow 0$. To study this possibility we write an approximation to the true wave function Ψ in the form

$$\tilde{\Psi} = e^{-u_{\sigma_i \sigma_j}(r)} \Psi_{HF}. \quad (53)$$

Consider the behavior of the local energy

$$\hat{H}\tilde{\Psi}/\tilde{\Psi} = \frac{1}{e^{-u_{\sigma_i \sigma_j}(r)} \Psi_{HF}} \hat{H} e^{-u_{\sigma_i \sigma_j}(r)} \Psi_{HF}, \quad (54)$$

as $r \rightarrow 0$ while $\vec{r}_{c.m.}$ and all other electron positions remain held fixed. Any divergent terms in the kinetic energy must arise from the action of the ∇_r^2 operator on $e^{-u_{\sigma_i \sigma_j}(r)}$. We first treat the case of antiparallel spins. The local energy remains finite as long as

$$\begin{aligned} & \frac{1}{e^{-u_{\uparrow\downarrow}(r)} \Psi_{HF}} \left(-\nabla^2 + \frac{1}{r} \right) e^{-u_{\uparrow\downarrow}(r)} \Psi_{HF} = \\ & = u''_{\uparrow\downarrow} + \frac{2u'_{\uparrow\downarrow}}{r} - \left(u'_{\uparrow\downarrow} \right)^2 + 2u'_{\uparrow\downarrow} \hat{r} \cdot \frac{\nabla \Psi_{HF}}{\Psi_{HF}} - \frac{\nabla^2 \Psi_{HF}}{\Psi_{HF}} + \frac{1}{r} \end{aligned} \quad (55)$$

remains finite, where \hat{r} is a unit vector in the \vec{r} direction, and the primes denote differentiation with respect to r . If electrons i and j have opposite spins the value of $\Psi_{HF}(\vec{r} = 0)$ is generally

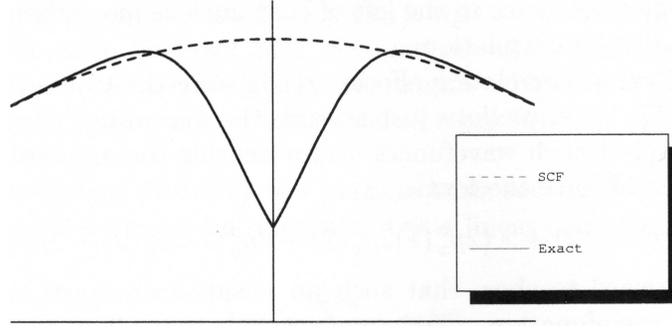


Fig. 4. Schematic shape of the exact wave function (full line) and Ψ_{HF} (dashed line) around the point of coalescence. The wave functions are plotted against $z = z_i - z_j$ with the two electrons having identical x and y coordinates.

nonzero. We insist that u' and u'' tend to finite values as $r \rightarrow 0$, and so only the second and the final terms in (55) diverge. These two divergences cancel out if we impose the *opposite spin cusp condition*

$$\frac{du_{\uparrow\downarrow}}{dr}\Big|_{r=0} = -\frac{1}{2}, \quad (56)$$

and hence also

$$\frac{\partial\Psi}{\partial r_{ij}}\Big|_{r_{\uparrow\downarrow} \rightarrow 0} = -\frac{1}{2}\Psi\Big|_{r_{\uparrow\downarrow} \rightarrow 0} \rightarrow 0. \quad (57)$$

If electrons i and j have parallel spins, The Pauli principle guarantees, that Ψ_{HF} is an odd function of \vec{r} and

$$\Psi_{HF} = \nabla\Psi_{HF}\Big|_{r=0}\cdot\vec{r} + O(r^3). \quad (58)$$

It follows, that the Laplacian of Ψ_{HF} tends to zero as $r \rightarrow 0$, but the second, fourth, and last term in Eq. (55) all diverge. The sum of the divergent contributions is

$$\frac{2u'}{r} + \frac{2u'}{\nabla\Psi_{HF}\Big|_{r=0}\cdot\vec{r}}\nabla\Psi_{HF}\Big|_{r=0}\cdot\hat{r} + \frac{1}{r} = \frac{4u'}{r}, \quad (59)$$

and so the *parallel-spin cusp condition* is

$$\frac{du_{\uparrow\uparrow}}{dr}\Big|_{r=0} = -\frac{1}{4}, \quad (60)$$

and hence also

$$\frac{\partial\Psi}{\partial r_{ij}}\Big|_{r_{\uparrow\uparrow} \rightarrow 0} = -\frac{1}{4}\Psi\Big|_{r_{\uparrow\uparrow} \rightarrow 0} \rightarrow 0. \quad (61)$$

These are the well-known cusp conditions [19, 20], which show that in whatever direction one moves from $r = 0$, the wave function increases linearly. The exact wave function must have the shape depicted in Fig. 4, showing the existence of a *Coulomb hole* around the point of coalescence.

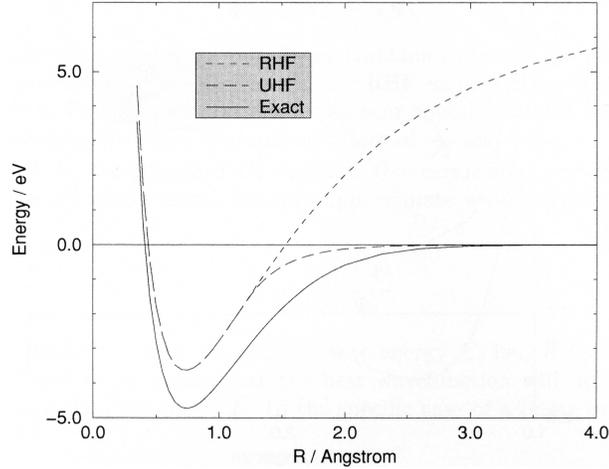


Fig. 5. Potential energy surfaces for H_2 molecule. Full line very accurate model, long dashes: UHF model, short dashes RHF model.

3.1.3 Non-dynamical long-range correlations

Let's take a look at the opposite limit of the so-called *long-range* or *non-dynamical correlations*. The effects of neglecting electron correlation in Hartree-Fock are spectacularly illustrated when one attempts to compute complete PESs curves for diatomic molecules using SCF. Fig. 5 shows potential energy curves for H_2 from both an accurate calculation and from Hartree-Fock. It is seen that the spin-restricted Hartree-Fock (RHF) approximation breaks down as dissociation is reached predicting energies which are much too high, and a PES curve characteristic of the interaction of ions rather than neutral atoms. The RHF wave function for the $X^1\Sigma_g^+$ ground-state of H_2 takes the form

$$\Psi_X = \hat{A}\sigma_g^\alpha(1)\sigma_g^\beta(2), \quad (62)$$

where \hat{A} is the antisymmetrizing operator, α and β are the usual one-electron spin functions, and the bonding orbital $\sigma_g = Z_{\sigma_g}(\varphi_A + \varphi_B)$, with φ_A and s -like orbital centered on atom A, and Z_{σ_g} a normalization constant. As the atoms become infinitely separated, $\varphi_A \sim 1s_A$, $Z_{\sigma_g} \sim \frac{1}{\sqrt{2}}$ and thus

$$\Psi_X \sim \frac{1}{2}\hat{A}\left(1s_A^\alpha 1s_B^\beta + 1s_B^\alpha 1s_A^\beta + 1s_A^\alpha 1s_A^\beta + 1s_B^\alpha 1s_B^\beta\right). \quad (63)$$

The first two terms are direct products of neutral 2S hydrogen atom wave functions on the two atoms A and B, as desired. However, the last two terms describe a spurious $H^+ \cdots H^-$ pair. The overall energy of this unphysical wave function exceeds the energy of two hydrogen atoms by half the difference of the ionization energy and electron affinity of H, i.e. 6.4 eV, and the long-range potential energy curve has an unphysical ionic R^{-1} behavior.

The failure of RHF can be easily understood in terms of electron correlation. At long internuclear distances, if one electron is located near atom A, the other will on physical grounds

be close to atom B. This correlation is reflected in the exact wave function, which is asymptotically the product of hydrogenic orbitals on the two nuclei. In contrast, within the Hartree-Fock framework, each electron is made to experience only the average effect of the other. Since in RHF, the two electrons are constrained to be in the same special orbital, this σ_g orbital will be symmetrical between the atoms, and thus each electron has an equal probability of being on A or B, irrespective of the position of the other electron. The possibility of both electrons being on the same atom is not excluded, as reflected in the RHF wave function (63).

In the case of H_2 , and in fact for a number of other dissociating molecules, Hartree - Fock theory can give correct behavior provided the restriction to identical special orbitals for α and β spin is relaxed. The unrestricted HF (UHF) wave function is identical to RHF at short bond lengths, but when the two atoms are separated, it becomes more advantageous for the α and β spin orbitals to localize on different hydrogen atoms. In this way, a correct asymptotic energy is obtained, as seen in Fig. 5. However, the wave function can never be identical to the exact wave function. Asymptotically, the UHF wave function is either $\hat{A}1s_A^\alpha 1s_B^\beta$ or $\hat{A}1s_B^\alpha 1s_A^\beta$, whereas the true wave function is the sum of these two degenerate determinants. Moreover, although the energy is unaffected, the UHF wave function is not an eigenfunction of the spin-squared operator \hat{S}^2 , being an unphysical mixture of singlet and triplet states. The spin contamination has very serious effects, such as in the case of F_2 , where the UHF does not repair the RHF failure.

3.1.4 Post Hartree-Fock methods

Let's now return to the above example of the H_2 molecule and show how the inability of the HF method to describe correctly dissociation and long-range correlation can be rectified. We first examine an excited state $^1\Sigma_g^+$ of H_2 . The RHF wave function is of the form

$$\Psi_E = \hat{A}\sigma_u^\alpha(1)\sigma_u^\beta(2). \quad (64)$$

We now have two electrons in the antibonding orbital $\sigma_u = Z_{\sigma_u}(\varphi_A - \varphi_B)$ and asymptotically

$$\Psi_E \sim \frac{1}{2}\hat{A}\left(1s_A^\alpha 1s_B^\beta + 1s_B^\alpha 1s_A^\beta - 1s_A^\alpha 1s_A^\beta - 1s_B^\alpha 1s_B^\beta\right). \quad (65)$$

This wave function, similarly to (63) also contains an unphysical mixture of covalent and ionic terms. Notice, however, that it is now possible to construct purely ionic or purely covalent wave functions by taking a linear combination of Ψ_X and Ψ_E . In $\Psi_X - \Psi_E = \sigma_g^2 - \sigma_u^2$, the ionic terms cancel exactly, and the correct asymptotic wave function is obtained. This is an example of *configuration interaction* (CI), where the wave function is considered a mixture of several Slater determinants, as suggested in Eq. (20). For H_2 at general internuclear separation, the form of the CI wave function is

$$\Psi = c_X\Psi_X + c_E\Psi_E. \quad (66)$$

It is clear, that the coefficients c must be treated as variational parameters, as near equilibrium the RHF wave function is already a good approximation, and hence $c_X \cong 1$ and $c_E \cong 0$. In contrast, in the dissociation limit their values must be $\frac{1}{\sqrt{2}}$ and $-\frac{1}{\sqrt{2}}$.

In general, variational principle is used in the CI method to determine the CI coefficients. For any approximate wave function, the so-called Rayleigh quotient

$$\bar{E} = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \quad (67)$$

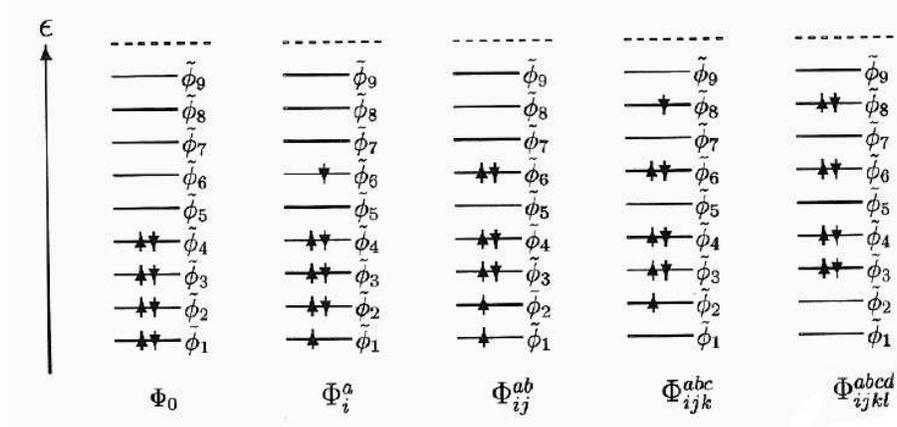


Fig. 6. Schematic construction of closed-shell ground-state Φ_0 , singly Φ_i^a , doubly Φ_{ij}^{ab} , triply Φ_{ijk}^{abc} and quadruply Φ_{ijkl}^{abcd} excited determinants.

is an upper bound to the exact ground-state energy E , i.e., $\bar{E} \geq E_G$. Variational methods assume that the best wave function is the one at the minimum of \bar{E} . The linear ansatz (20), (66) boils down to finding eigensolutions of a symmetric matrix, which is a straightforward and well studied problem.

The CI method is the conceptually simplest method for accounting for correlation effects (both short- and long-range). In general, all effects beyond the Hartree-Fock level are termed as “*correlation*” effects and the energy difference between the exact energy eigenvalue of a N_e -electron system and the SCF (HF) energy as *correlation energy*

$$E_{corr} = E - E_{SCF}. \quad (68)$$

The main idea is based on the expansion theorem. Let $\{\varphi_i\}$, $i = 1, 2, \dots, \infty$ be a complete basis of the one-particle Hilbert space (e.g. HF). Then it can be proven that all N_e -electron Slater determinants Φ_L in Eq. (20), which can be constructed by placing N_e electrons into these Slater determinants in all possible ways, span the full antisymmetric N_e -particle space. This holds, of course, also for the eigenfunctions of the N_e -electron Hamiltonian. This is schematically shown in Fig. 6.

Such a procedure has a prohibitive computational cost. Consider, for example, a system of N_e electrons occupying states chosen from a set of $2N_e$ single-particle orbitals. The total number of N_e -electron Slater determinants for large N_e is

$$\frac{(2N_e)!}{N!N!} \underset{N \rightarrow \infty}{\approx} \frac{e^{2N_e \ln 2N_e}}{e^{2N_e \ln N_e}} = e^{2N_e \ln 2}. \quad (69)$$

Hence, the number of determinants rises exponentially with the number of electrons, N_e . The full CI approach in conjunction with a good basis set can be applied only to small systems. The current practical limit is in the region of small molecules, such as H_2O .

The computational load can be reduced by including only the most important determinants. This is normally done by considering low-energy excitations from a reference determinant, normally the HF ground-state. One can, for example, include only single and double excitations (see, Fig. 6), with a computational load which scales as $O(N_e^6)$. Unfortunately, such a truncated method is not size-consistent as the energy does not scale linearly with the number of electrons. For that reason, for example, calculation of binding energies is not possible. This is intuitively clear, as for example, a double excitation on the atoms results to quartet excitations from the bonded system. The size-consistency problem can be overcome via *coupled-cluster* (CC) expansion [21]. The CC method implicitly includes all excitations from the reference determinant with approximated coefficients in the expansion. The method is not variational. The CC methods do provide highly accurate results but the size of the system is limited by the $O(N_e^6)$ scaling.

A different approach is followed in the so-called *Møller-Plesset* (MP) perturbation correction, which adds dominant effects (first-, second- and higher order) of electron correlation to the SCF treatment. However, there is no way to establish or estimate the convergence properties of this series in general. The simplest method of this family is the MP2, taking into account the first term of the perturbation expansion. MP2 typically improves SCF if the latter is already a useful approximation.

The post-HF methods are described in a straightforward manner in Ref. [14]. Given the generally prohibitive cost of all the post-HF methods in connection with the coupled electron-ion dynamics outlined in part 2.1, we will not pursue these methods any further here. Instead, when highly accurate results are required, and/or the system is highly correlated, we will follow another path, based on quantum Monte-Carlo (QMC) methods, which have comparable accuracy to the best post-HF methods and yet much more favorable scaling with the system size.

3.1.5 Density functional theory

Density functional theory (DFT) is, in addition to the Hartree-Fock, another theory limiting the expansion of the wave function (20) to just a single determinant. However, as will become clear later, it makes attempt to include electronic correlations in an approximate way. Intuitively, an integrated quantity, such as the electronic density

$$n(\vec{r}) = N_e \int d\vec{r}_2 \cdots \int d\vec{r}_{N_e} \Psi^*(\vec{r}, \vec{r}_2, \cdots, \vec{r}_{N_e}) \Psi(\vec{r}, \vec{r}_2, \cdots, \vec{r}_{N_e}) \quad , \quad (70)$$

should be easier to describe than the much more complicated object, such as the many-body wave function $\Psi(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_{N_e})$.

On one hand, the dimensionality of the problem is drastically reduced from $3N_e$ to just 3, on the other hand, though, the electronic density is a rather featureless object which in a molecule or solid shows only relatively small departures from the overlapped densities of its atomic constituents. In order to appreciate some qualitative features of the density in a real system, we show in Fig. 7 the spherically averaged density in the ground-state of the carbon atom. The density falls monotonically from the nucleus. Hence, it may be difficult to capture the details contained in the many-body wave function in a featureless object, as the charge density is. The difficulty of the task we are up to is appreciated if we consider the energy differences which govern the stability of the different concurrent structures, such as the different phases of the same system, concurrent cluster structures or different isomers of a molecular system, which

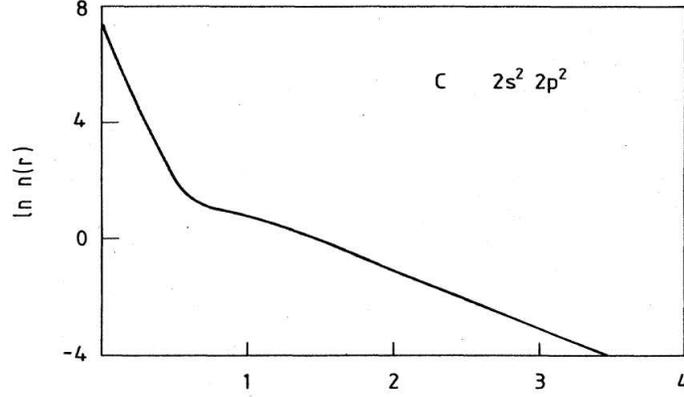


Fig. 7. Spherical average of electronic charge density in ground-state of C atom [25].

often are just a tiny fraction of the correlation energy which itself is just a few percent of the total energy.

3.1.5.1 Thomas-Fermi approximation

A scheme based on the idea of substituting wave function by the corresponding electronic charge density dates back to the days of Thomas [22] and Fermi [23]. The *Thomas-Fermi (TF) scheme* assumes that the motions of the electrons are uncorrelated, and that the corresponding kinetic energy of the electrons can be described by a local approximation based on results for free electrons. The electrons are treated as independent particles, and the electron-electron interaction energy arises solely from the electrostatic energy,

$$E_{es}[n] = \frac{e^2}{2} \int d\vec{r} \int d\vec{r}' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|}. \quad (71)$$

The kinetic energy functional

$$T[n] = \int d\vec{r} t[n(\vec{r})] \quad (72)$$

can be expressed through kinetic energy density of a system of noninteracting electrons with density n

$$\begin{aligned} t[n] &= \frac{2}{(2\pi)^3} \int_{|k| \leq k_F} d\vec{k} \frac{\hbar^2 k^2}{2m_e} = \\ &= \frac{2\hbar^2}{(2\pi)^3 2m_e} \int_{|k| \leq k_F} d\vec{k} k^2 = \frac{2\hbar^2}{(2\pi)^3 2m_e} \int_0^{k_F} \int_0^\pi \int_0^{2\pi} k^2 k^2 \sin\theta d\theta d\phi dk = \frac{\hbar^2}{10\pi^2 m_e} k_F^5. \end{aligned} \quad (74)$$

The density is related to the Fermi wave vector via

$$n = \frac{2}{(2\pi)^3} \left(\frac{4}{3} \pi k_F^3 \right) \Rightarrow k_F = \sqrt[3]{3\pi^2 n}. \quad (75)$$

This gives,

$$T_0 [n] = c_k n^{\frac{5}{3}}, \quad (76)$$

where $c_k = 3\hbar^2(3\pi^2)^{\frac{2}{3}}/10m$. The energy functional in the TF approximation reads

$$E [n] = T [n] + E_{es} [n] + \int d\vec{r} n(\vec{r}) V_{ext}(\vec{r}), \quad (77)$$

where $V_{ext}(\vec{r})$ is an external potential in which the electrons move, such as the potential generated by the ions, or external field. We will seek the minimum of the functional (77) subject to the condition of charge conservation,

$$\int d\vec{r} n(\vec{r}) = N_e. \quad (78)$$

Applying the method of Lagrange multipliers yields the following functional to minimize

$$\mathfrak{S} [n] = T [n] + E_{es} [n] + \int d\vec{r} n(\vec{r}) \left\{ V_{ext}(\vec{r}) + \lambda \right\}, \quad (79)$$

where λ is the Lagrange multiplier imposing the normalization constraint (78). The optimum electronic density is solution to the equation

$$\frac{\delta \mathfrak{S} [n]}{\delta n} = \frac{5}{3} c_k n^{\frac{2}{3}} + e^2 \int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + V_{ext}(r) + \lambda \stackrel{!}{=} 0. \quad (80)$$

Results of the studies with the TF model have shown that the TF model has severe deficiencies. The charge density is infinite at the nucleus, and it does not decay exponentially far from the nucleus of an atom, but as r^{-6} . It has also been shown that TF theory does not result in atoms binding to form molecules or solids. The absence of shell structure in the TF atom means that the observed periodic variation of properties with atomic number cannot be reproduced. In fact, the atoms shrink with increasing atomic number Z as $Z^{1/3}$ [24].

At this stage it might seem hopeless to consider the electronic density as the basic variable any further. Yet, this conclusion is wrong as the most successful modern electronic structure methods take density as the basic variable. The failure of the TF model was due to the fact that, in addition to a very approximate treatment of the kinetic energy, it almost entirely ignored that fact that electrons are quantum objects. In fact, TF is a special case of the *density functional theory (DFT)*, which makes away with most of the shortcomings of the TF model.

3.1.5.2 Basic theorems of the DFT

There are a number of excellent review papers [3, 25, 26] which cover different aspects of the DFT theory and applications. We will cover here mainly those aspects of the DFT theory,

which are needed to solve the coupled Born-Oppenheimer (Eqs. (16)-(17)) or Ehrenfest dynamics (Eqs. (11)-(12)).

The DFT expresses ground-state (GS) properties of a system, including the energy E , in terms of functionals of the GS electron density, i.e. they are expressed by the knowledge of the density alone. The basic theorems of the DFT formalism were derived by Hohenberg and Kohn [27], who have shown that $E[n]$ satisfies a variational principle. Levy [28] provided a simpler and more general derivation. We consider N_e electrons moving in an external potential, $V_{ext}(\vec{r})$, i.e. the Hamiltonian is

$$\tilde{H} = \hat{T} + V_{ee} + \sum_{i=1}^{N_e} V_{ext}(\vec{r}_i), \quad (81)$$

where \hat{T} and V_{ee} are the kinetic and electron-electron interaction operators, respectively. We consider “**N**-representable” densities $n(\vec{r})$ ⁴, i.e. those obtainable from some antisymmetric wave function $\Psi(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e})$. Levy defined the functional,

$$F[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + V_{ee} | \Psi \rangle, \quad (82)$$

where the minimum is taken over all Ψ that give the density n . $F[n]$ is totally universal as it does not refer to any specific system. Let us denote the ground-state energy, wave function, and density by E_{GS} , Ψ_{GS} , and n_{GS} , respectively. The basic two theorems of the DFT are

$$E[n] \equiv \int d\vec{r} V_{ext}(\vec{r}) n(\vec{r}) + F[n] \geq E_{GS}, \quad (83)$$

which is the *variational property*, and for all N -representable densities $n(\vec{r})$ we have the *minimum property*

$$\int d\vec{r} V_{ext}(\vec{r}) n_{GS}(\vec{r}) + F[n_{GS}] = E_{GS}. \quad (84)$$

To prove the variational principle (83), we introduce the notation $\Psi_{\min}^n(\vec{r})$ for a wave function that minimizes $F[n]$ in Eq. (82),

$$F[n] = \langle \Psi_{\min}^n | \hat{T} + V_{ee} | \Psi_{\min}^n \rangle \quad (85)$$

and setting $V = \sum_i V_{ext}(\vec{r}_i)$, we have

$$\int d\vec{r} V_{ext}(\vec{r}) n(\vec{r}) + F[n] = \langle \Psi_{\min}^n | V + \hat{T} + V_{ee} | \Psi_{\min}^n \rangle \geq E_{GS} \quad (86)$$

according to the minimum property of the GS. This proves the inequality (83). Using the minimum property once more, we find

$$E_{GS} = \langle \Psi_{GS} | V + \hat{T} + V_{ee} | \Psi_{GS} \rangle \leq \langle \Psi_{\min}^{n_{GS}} | V + \hat{T} + V_{ee} | \Psi_{\min}^{n_{GS}} \rangle. \quad (87)$$

⁴Hohenberg and Kohn [21] worked in the space of V -representable densities, i.e. those that can be realized for some external potential, V_{ext} . This space is a subspace of N -representable densities.

We subtract the interaction with the external potential and obtain

$$\langle \Psi_{GS} | \hat{T} + V_{ee} | \Psi_{GS} \rangle \leq \langle \Psi_{\min}^{n_{GS}} | \hat{T} + V_{ee} | \Psi_{\min}^{n_{GS}} \rangle. \quad (88)$$

On the other hand, the definition of $\Psi_{\min}^{n_{GS}}$ yields the reverse relation between the two sides of Eq. (88). This is possible only if

$$\langle \Psi_{GS} | \hat{T} + V_{ee} | \Psi_{GS} \rangle = \langle \Psi_{\min}^{n_{GS}} | \hat{T} + V_{ee} | \Psi_{\min}^{n_{GS}} \rangle. \quad (89)$$

Hence, we have

$$\begin{aligned} E_{GS} &= \int d\vec{r} V_{ext}(\vec{r}) n_{GS}(\vec{r}) + \langle \Psi_{GS} | \hat{T} + V_{ee} | \Psi_{GS} \rangle \\ &= \int d\vec{r} V_{ext}(\vec{r}) n_{GS}(\vec{r}) + \langle \Psi_{\min}^{n_{GS}} | \hat{T} + V_{ee} | \Psi_{\min}^{n_{GS}} \rangle \\ &= \int d\vec{r} V_{ext}(\vec{r}) n_{GS}(\vec{r}) + F[n_{GS}]. \end{aligned} \quad (90)$$

These theorems show, that the GS charge density determines the GS wave function, from which all GS properties can be calculated which themselves, are functionals of the GS density. The problem remains, how to find the functional $F[n]$ or an approximation to it. As the $F[n]$ functional contains the electron-electron interaction potential V_{ee} , it must contain all the many-body physics, exchange-correlation effects, etc. Clearly finding a valid approximation to $F[n]$ will not be an easy task. The simplest approximation, completely ignoring the quantum nature of the electrons is represented by the TF model, see chapter 3.1.5.1 above.

3.1.5.3 Kohn-Sham equation

Many of the drawbacks of the TF approach can be traced to the approximate treatment of the kinetic energy. The task of finding good approximations to the energy functional is greatly simplified by using a different separation, first introduced by Kohn and Sham [29]

$$E[n] = T_0[n] + \int d\vec{r} n(\vec{r}) \left(V_{ext}(\vec{r}) + \frac{1}{2} \Phi(\vec{r}) \right) + E_{xc}[n]. \quad (91)$$

T_0 is the kinetic energy of a system with density n for a noninteracting system, Φ is the classical Coulomb interaction for electrons, and $E_{xc}[n]$ defines the exchange-correlation energy. T_0 differs from the true kinetic energy T , but it is of comparable magnitude and is treated exactly in this approach. This removes many of the deficiencies of the TF model, such as the lack of shell structure of atoms or the absence of chemical bonding in molecules and solids. All terms in Eq. (91), except for $E_{xc}[n]$, can be evaluated exactly. Hence, at this stage all many-body effects have been factored into the exchange-correlation term $E_{xc}[n]$. The approximations to $E_{xc}[n]$ play a central role in the DFT.

The minimization applied to Eq. (91) subject to the normalization constraint

$$\int d\vec{r} n(\vec{r}) = N_e \quad (92)$$

leads to the following constraint functional to be minimized

$$\mathfrak{S}[n] = T_0[n] + \int d\vec{r} n(\vec{r}) \left(V_{ext}(\vec{r}) + \frac{1}{2} \Phi(\vec{r}) \right) + E_{xc}[n] + \lambda \left(\int d\vec{r} n(\vec{r}) - N_e \right) \quad (93)$$

with λ being the Lagrange multiplier imposing the constraint (92). The optimum density $n(\vec{r})$ can be found by variation of $\mathfrak{S}[n(\vec{r})]$,

$$\frac{\delta \mathfrak{S}[n]}{\delta n(\vec{r})} = \frac{\delta T_0[n]}{\delta n(\vec{r})} + V_{ext}(\vec{r}) + \frac{1}{2} \Phi(\vec{r}) + \frac{\delta E_{xc}[n]}{\delta n(\vec{r})} + \lambda \stackrel{!}{=} 0, \quad (94)$$

where the last term defines the exchange-correlation potential V_{xc} . This equation is formally very similar to the corresponding TF equation (80), the only differences being the presence of the V_{xc} potential and the form, the kinetic energy is incorporated in the model. While several approximations to V_{xc} exist, see chapters 3.1.5.4, 3.1.5.7, and 3.1.5.8, despite considerable effort, no reliable approximation to the kinetic energy functional exists. Kohn and Sham made at this stage a step backward and reintroduced the orbitals back into their theory [29]. They conveniently defined the T_0 functional through the independent-particle wave functions $\{\varphi_i(\vec{r})\}$ as

$$T_0 = \sum_i^{occ} \langle \varphi_i | -\frac{1}{2} \nabla^2 | \varphi_i \rangle, \quad (95)$$

and the Φ classic Coulomb energy of the electrons as

$$\frac{1}{2} \int d\vec{r} \Phi(r) n(\vec{r}) = \frac{1}{2} \int d\vec{r} \int d\vec{r}' \underbrace{\frac{n(\vec{r}')}{|\vec{r}' - \vec{r}|}}_{V_H(\vec{r})} n(\vec{r}). \quad (96)$$

The $V_H(\vec{r})$ defines the so-called Hartree potential. With these definitions the equation defining the optimum set of the so-called *Kohn-Sham orbitals* becomes

$$\begin{aligned} \mathfrak{S}[n] &= \sum_i^{occ} \langle \varphi_i | -\frac{1}{2} \nabla^2 | \varphi_i \rangle + \int d\vec{r} n(\vec{r}) \left(V_{ext}(\vec{r}) + \int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r}' - \vec{r}|} \right) \\ &+ E_{xc}[n] - \sum_{ij}^{occ} \varepsilon_{ij} (\langle \varphi_i | \varphi_j \rangle - \delta_{ij}), \end{aligned} \quad (97)$$

where ε_{ij} is the matrix of Lagrange multipliers imposing the mutual orthonormality of the Kohn-Sham orbitals $\int d\vec{r} \varphi_i^*(\vec{r}) \varphi_j(\vec{r}) = \delta_{ij}$. The optimum orbitals can be found by constraint optimization

$$\frac{\delta \mathfrak{S}[n]}{\delta \varphi_i^*} = \left(-\frac{1}{2} \nabla^2 + \frac{1}{2} \int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r}' - \vec{r}|} + V_{ext}(\vec{r}) + \underbrace{\frac{\delta E_{xc}}{\delta n(\vec{r})}}_{V_{xc}(\vec{r})} \right) \varphi_i - \sum_{ij}^{occ} \varepsilon_{ij} \varphi_j \stackrel{!}{=} 0. \quad (98)$$

Note that similarly to the case of the SCF equations (44–45), also equation (98) converges to linear combination of canonical *Kohn-Sham equation*. In the same way as with the SCF equation one can manipulate by unitary transformations the matrix ε_{ij} to be diagonal⁵. In such a case

⁵The form of the Kohn-Sham orbitals to which the equation (98) converges depends on the orthonormalization procedure used; see also chapters 3.3.2.1 and 3.3.3.

Eq. (98) takes the form

$$-\frac{1}{2}\nabla^2\varphi_i + \underbrace{\left(\frac{1}{2}\int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r}-\vec{r}'|} + V_{xc}(\vec{r}) + V_{ext}(\vec{r})\right)}_{V_{eff}(\vec{r})} \varphi_i = \varepsilon_i\varphi_i. \quad (99)$$

This is a mean-field equation where the effect of the other electrons enters via charge density

$$n(\vec{r}) = \sum_{i=1}^{occ} |\varphi_i|^2. \quad (100)$$

The corresponding total energy Eq. (91), can be expressed as

$$E^{KS} = \sum_i^{occ} \varepsilon_i - \frac{1}{2} \int d\vec{r} V_H(\vec{r})n(\vec{r}) + E_{xc}[n] - \int d\vec{r} V_{xc}(\vec{r})n(\vec{r}). \quad (101)$$

In order to solve Eqs. (99), (100), we still need an approximation to $V_{xc}(\vec{r})$. The quality of this approximation will crucially determine the quality of the resulting single-particle orbitals $\{\varphi_i\}$, and the GS properties. A very simple model for $V_{xc}(\vec{r})$ was proposed by Kohn and Sham [29] who assumed that the inhomogeneous electron gas Eqs. (99), (100) behaves locally as the homogeneous electron gas with the same density

$$E_{xc}^{LDA} = \int d\vec{r} n(\vec{r}) \varepsilon_{xc}^{HEG}[n(\vec{r})]. \quad (102)$$

Here $\varepsilon_{xc}^{HEG}[n(\vec{r})]$ is the exchange and correlation energy per particle of a homogeneous electron gas with density n . The approximation (102) is called *local density approximation* (LDA). Numerous approximations as well as very accurate calculations exist for $\varepsilon_{xc}^{HEG}[n]$. Kohn and Sham [29] also showed that approximation (102) is exact in the limiting cases of slowly varying density and very high densities. Equations (99 – 102) are the famous *Kohn-Sham equations* which is the most widely used model for realistic calculations. Note, that Eq. (99) is nonlinear, as it depends on all occupied orbitals via the electronic density (100). This means, that similarly to the Hartree-Fock equation (46), its solutions have to be found in a self-consistent manner. One starts with an initial guess for the set of Kohn-Sham orbitals, $\{\varphi_i^0\}$, from which the charge-density (100) is computed and effective potential $V_{eff}(\vec{r})$ determined. The Kohn-Sham equation (99) is solved which yields the first iteration of $\{\varphi_i^1\}$. This procedure is repeated until the input wave functions $\{\varphi_i^n\}$ and/or potentials do not differ from the output wave functions $\{\varphi_i^{n+1}\}$ and/or potentials. The numerical cost is easily determined as $O(N_e^3)$, as the dominant step is the diagonalization of the Kohn-Sham Hamiltonian matrix, or a similar procedure with the same asymptotic scaling; see also chapter 3.3.2.1.

The limiting cases noted above are not realized in most calculations, such as application to atoms, molecules, surfaces, or solids. So why is it, that despite these facts, the Kohn-Sham equations are still, after more than 40 years, in use and yield results which are generally of surprising accuracy? Kohn and Sham themselves did not expect their equations to provide an accurate description of chemical bonding [29]. A decade passed before the first attempts were made to test the ability of the DFT in LDA approximation to describe the bonds in molecules,

and remarkably, they showed that they generally reproduce ground-state geometries and vibration and phonon frequencies. The reasons for that will be further analyzed in chapter 3.1.5.5. The DFT remains the basis of most of the parameter-free calculations in extended systems and more recently also in applications to molecules, clusters, and even bioapplications.

In the above discussion we have tacitly neglected the spin and assumed that the orbitals are all doubly occupied. The *generalization to spin systems* requires consideration of spin indices, in which case the analogy of Eqs. (99 – 102) reads

$$-\frac{1}{2}\nabla^2\varphi_{i\sigma} + \underbrace{\left(\int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r}-\vec{r}'|} + V_{xc}^\sigma(\vec{r}) + V_{ext}(\vec{r})\right)}_{V_{eff}^\sigma(\vec{r})} \varphi_{i\sigma} = \varepsilon_{i\sigma}\varphi_{i\sigma}, \quad (103)$$

$$n_\sigma(\vec{r}) = \sum_{i=1}^{occ} |\varphi_{i\sigma}|^2, \quad (104)$$

$$V_{xc}^\sigma = \frac{\delta E_{xc}[n_\uparrow n_\downarrow]}{\delta n_\sigma(\vec{r})}, \quad (105)$$

$$E_{xc}^{LSD}[n_\uparrow n_\downarrow] = \int d\vec{r} n(\vec{r}) \varepsilon_{xc}^{HEG}[n_\uparrow(\vec{r}), n_\downarrow(\vec{r})]. \quad (106)$$

This straightforward extension of the DFT can be shown to lead to analytical problems, as the ground-state wave function $\Psi_{GS}(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e})$ does not uniquely define the effective potential $V_{eff}^\sigma(\vec{r})$ [30, 31]. Nevertheless the practical consequences for spin DFT calculations are not severe [32], and indeed, they are routinely done.

3.1.5.4 The LDA approximation

In both formulations of the DFT theory the approximation to the exchange-correlation energy and potential (102) and (106) play a central role and will determine the accuracy of the results. Let us analyze this point in more detail. The first approximation of the exchange effects in HEG was given by Dirac [33] in the form

$$\varepsilon_x^{Dirac} = -\frac{3}{4} \left(\frac{3}{\pi}\right) n^{\frac{1}{3}} = \frac{0.458}{r_s} (a.u.), \quad (107)$$

where r_s is a measure of the interelectronic distance ($n = 1/(4/3)\pi r_s^3 \Rightarrow r_s = (3/(4\pi n))^{1/3}$). The exchange interaction means, that an electron of a given spin, say spin \uparrow , will be surrounded by a region where the density of electrons with the same spin is reduced. Qualitatively, the exchange hole of form (107) can be easily inferred from the following argument. Replace an electron by a uniform density inside a sphere of radius r_0 , and zero elsewhere. Since the exchange hole contains a single electron [34], $r_0 = (3/(4\pi n_\uparrow))^{1/3}$, and we obtain an exchange energy with the dependence as that of (107). Later, Gaspar adopted the Dirac approximation and derived a slightly different result using variational approach [35]. This method was shown to reproduce well the HF results for a positively charged atom, showing that an approximation

based on homogeneous gas can give realistic results even if the density is far from uniform. The Gáspár's form of $\varepsilon_x^{Gaspard}$ is customarily used in the LDA approximation (102) for exchange. The much more complicated correlation term, ε_c , is commonly based on the results of very accurate quantum Monte-Carlo results for a homogeneous electron gas of various densities [36, 37]; see also chapters 3.1.6 and 4.7.1.1. The LSD approximation (Eq. 106) requires also results for spin-polarized system and a means for interpolation for partial polarizations. Perdew and Zunger [38] proposed the formula

$$\varepsilon_c^{PZ} = \begin{cases} A \ln r_s + B + C r_s \ln r_s + D r_s, & r_s \leq 1, \\ \gamma / (1 + \beta_1 \sqrt{r_s} + \beta_2 r_s), & r_s > 1. \end{cases} \quad (108)$$

For dense electron densities ($r_s \leq 1$) the random phase approximation provides the parameters. Parameters for a fully polarized gas can be obtained from scaling arguments. Other parameters were obtained by fitting the results of Ceperley and Alder [36]. Another form that has found widespread use is due to Vosko, Wilk, and Nusair [39],

$$\begin{aligned} \frac{\varepsilon_c^{VWN}}{A} &= \ln \left(\frac{r_s}{F(\sqrt{r_s})} \right) + \frac{2b}{\sqrt{4c-b^2}} \arctan \left(\frac{\sqrt{4c-b^2}}{2\sqrt{r_s}+b} \right) \\ &- \frac{bx_0}{F(x_0)} \left[\ln \left(\frac{\sqrt{r_s}-x_0}{F(x_0)} \right) + \frac{2(b-2x_0)}{\sqrt{4c-b^2}} \arctan \left(\frac{\sqrt{4c-b^2}}{2\sqrt{r_s}+b} \right) \right], \end{aligned} \quad (109)$$

where $F(x) = x^2 + bx + c$ and the fitting parameters, which differ for the polarized (P) and unpolarized (U) cases, are obtained from the data of Ceperley and Alder [36]. Interpolation between U and P results are usually based on a formula involving spin polarization ζ

$$\varepsilon_{xc}(n, \zeta) = f(\zeta) \varepsilon_{xc}^U(n) + (1 - f(\zeta)) \varepsilon_{xc}^P(n), \quad (110)$$

$$\zeta = \frac{n_\uparrow(\vec{r}) - n_\downarrow(\vec{r})}{n_\uparrow(\vec{r}) + n_\downarrow(\vec{r})}. \quad (111)$$

Functional forms of $f(\zeta)$ based on the Hartree-Fock and random phase approximations have been suggested by Bath and Hedin [40] and Vosko *et al.* [39].

Few comments are due here. Note that the effective potential, $V_{eff}(\vec{r})$, is local, and Eqs. (99–102) are no more complicated than Hartree-Fock's. The kinetic energy, the electrostatic interaction between core and valence electrons are treated exactly. Only the exchange energy, E_x , and the even smaller correlation contribution require approximation. This is in marked contrast to the TF, where the large kinetic energy term is approximated. Note also the problem of “self-interaction”. In Eq. (99) the Coulomb interaction term includes also interaction with itself. The same is true also for the HF equations (46), where, however, the nonlocal exchange term of the form $-\frac{1}{2} \sum_{ij} \int d\vec{r}' \int d\vec{r}'' \frac{\varphi_i^*(\vec{r}) \varphi_j^*(\vec{r}') \varphi_j(\vec{r}) \varphi_i(\vec{r}'')}{|\vec{r}-\vec{r}''|}$ exactly cancels this spurious self-interaction. In contrast, in the DFT in LDA approximation for exchange (and correlation) the corresponding $V_{xc}(\vec{r})$ is local and depends only on the charge density $n(\vec{r})$.

The other comment regards the form of Eq. (99), which at the first glance looks like an ordinary stationary Schrödinger equation. However, we should keep in mind, that this is not quite true, as the orbital energies ε_i are in reality merely Lagrange multipliers, having little to

do with the excitation energies. For more on this point see also chapters 3.1.5.10-13. This is at variance with the HF treatment, where the Koopmans' theorem [17] gives some justification for equating the orbital energies with the excitation energies.

3.1.5.5 Exchange-correlation energy, E_{xc}

The crucial simplification in the DFT scheme was achieved by mapping the interacting system, whose energy we seek, onto the fictitious, non-interacting system of which we solve Eqs. (99–101). The relationship between these two systems can be studied by considering the scaled electron-electron interaction $\lambda/|\vec{r} - \vec{r}'|$ with λ varying from 0 (non-interacting system) to 1 (physical system). This is done in the presence of an external potential, V_λ [41], such that the ground-state of the Hamiltonian

$$\hat{H}_\lambda = -\frac{1}{2}\nabla^2 + V_{ext}(\vec{r}) + V_\lambda + \lambda V_{ee} \quad (112)$$

has the density $n(\vec{r})$ for all λ . The exchange-correlation energy of the interacting system can be expressed in terms of an integral over the coupling constant λ [42, 43],

$$E_{xc} = \frac{1}{2} \int d\vec{r} n(\vec{r}) \int d\vec{r}' \frac{1}{|\vec{r} - \vec{r}'|} n_{xc}(\vec{r}, \vec{r}' - \vec{r}), \quad (113)$$

$$n_{xc}(\vec{r}, \vec{r}' - \vec{r}) \equiv n(\vec{r}') \int_0^1 d\lambda \left(g(\vec{r}, \vec{r}', \lambda) - 1 \right), \quad (114)$$

where $g(\vec{r}, \vec{r}', \lambda)$ is the pair-correlation function of the system with density $n(\vec{r})$ and Coulomb interaction λV_{ee} . The exchange-correlation hole, n_{xc} , describes the effect of the interelectronic repulsions, i.e. the fact that an electron present at point \vec{r} reduces the probability of finding another one at \vec{r}' . The exchange-correlation energy is viewed as the energy resulting from the interaction between an electron and its exchange-correlation hole (Eqs. (113), (114)).

We note that since $g(\vec{r}, \vec{r}') \rightarrow 1$ as $|\vec{r} - \vec{r}'| \rightarrow \infty$, the above separation into electrostatic and exchange-correlation energies can be viewed as an approximate separation of the consequences of long-range and short-range effects, respectively, of the Coulomb interaction. Take into account the isotropic nature of the Coulomb interaction V_{ee} , a variable substitution $\vec{R} = \vec{r}' - \vec{r}$ in Eq. (113) yields

$$E_{xc} = \frac{1}{2} \int d\vec{r} n(\vec{r}) \int_0^\infty dR R^2 \frac{1}{R} \int d\Omega n_{xc}(\vec{r}, R). \quad (115)$$

This equation shows that the exchange-correlation energy depends only on the spherical average of $n_{xc}(\vec{r}, R)$, so that the approximations to E_{xc} can give excellent values, even if the description of nonspherical parts $n_{xc}(\vec{r}, R)$ are wrong. There is an important sum-rule that requires that the xc-hole contains one electron for all \vec{r}

$$\int d\vec{r}' n_{xc}(\vec{r}, \vec{r}' - \vec{r}) = -1. \quad (116)$$

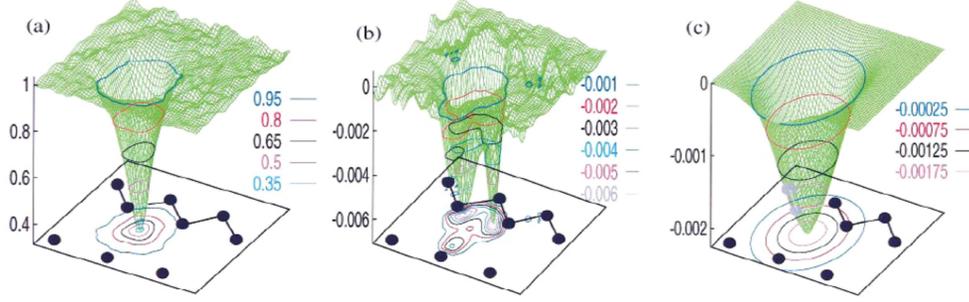


Fig. 8. Exchange-correlation hole in Si crystal [44]. (a) spin averaged pair correlation function from variational quantum Monte-Carlo (VMC), (b) exchange-correlation hole (VMC), and (c) exchange-correlation hole (LDA), with one electron fixed at the tetrahedral interstitial site in the (110) crystal plane.

This means that we can consider $-n_{xc}(\vec{r}, \vec{r}' - \vec{r})$ as a normalized weight factor, and define locally the radius of the xc-hole

$$\left\langle \frac{1}{\bar{R}} \right\rangle_{\vec{r}} = - \int d\vec{r}' \frac{n_{xc}(\vec{r}, R)}{|\bar{R}|}, \quad (117)$$

leading to

$$E_{xc} = -\frac{1}{2} \int d\vec{r} n(\vec{r}) \left\langle \frac{1}{\bar{R}} \right\rangle_{\vec{r}}. \quad (118)$$

This shows that, *provided the sum-rule (116) is satisfied, the exchange-correlation energy depends only weakly on the details of the n_{xc}* . This property is behind the success of the LDA/LSD approximation.

The expression for E_{xc} in Eqs. (113-114) can be generalized for the spin case

$$E_{xc} = \frac{1}{2} \sum_{\alpha\beta} \int d\vec{r} n_{\alpha}(\vec{r}) \int d\vec{r}' \frac{n_{\beta}(\vec{r}')}{|\vec{r} - \vec{r}'|} \int_0^1 d\lambda (g_{\alpha\beta}(\vec{r}, \vec{r}', \lambda) - 1). \quad (119)$$

The insensitivity of the xc-energy to the details of the xc-hole, provided the sum-rule (116) is satisfied is illustrated in Figs. 8, 9. Figure 8 shows the spin-averaged pair-correlation function $\bar{g} = \frac{1}{2} \sum_{\alpha\beta} g_{\alpha\beta}(\vec{r}, \vec{r}')$ and exchange-correlation hole $\bar{n}_{xc} = \sum_{\alpha\beta} n_{xc}^{\alpha\beta}$ from a very accurate calculation (quantum Monte-Carlo; see also chapter 3.16) and in the LDA approximation for a real system, Si crystal [44]. It is evident, that the approximate LDA xc-hole indeed has a form significantly different from the accurate one, as the former is a single minimum function, whereas the latter has two minima. However, as shown in Fig. 9 for nitrogen atom, most of the differences in the xc-hole are removed after the spherical average (Eq. 115) is taken [45]. The figure shows just the dominant part of the xc-hole, the exchange hole, n_x from an exact calculation, namely HF (remember, that exchange is treated exactly with the HF approach), and in the LDA approximation (Eq. (102)) for two representative values of \vec{r} . Indeed, the approximate and exact holes are

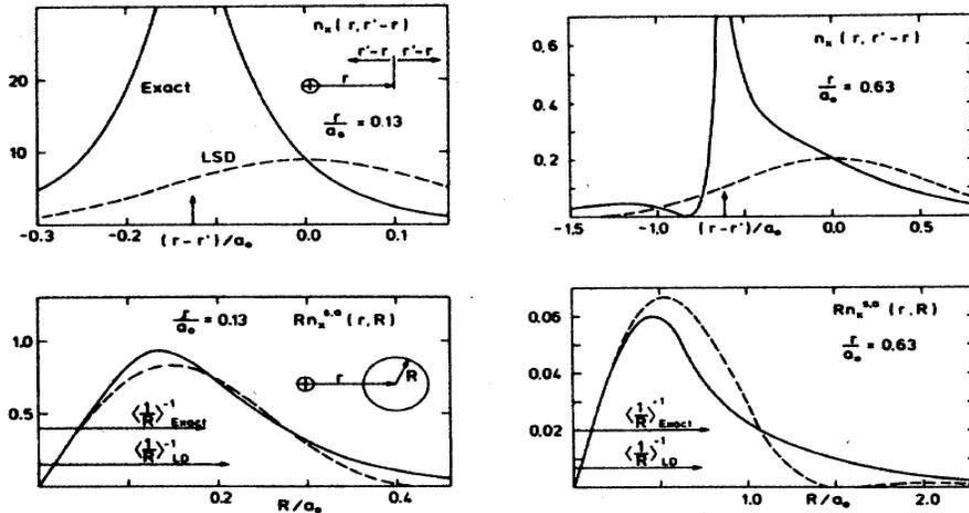


Fig. 9. Exchange-hole in a nitrogen atom [45]. Solid line: exact result, dashed line: LDA result for a spin up electron for $r=0.13$ (left panel) and $r=0.63$ (right panel). The top panels show the hole along a line through the nucleus and the electron. The arrow indicates the nuclear position and $\vec{r}' - \vec{r}' = 0$ gives the electron position. Note, that the exact hole has a large weight at the nucleus, while the approximate hole is centered on the electron. The lower panels show the spherical average of the hole around the electron. The area under the curve is proportional to the exchange energy.

again qualitatively different: the LDA hole is very symmetric around its center \vec{r}' . The spherical averages, however, are remarkably similar, and the values of the exchange energy differ by only a few percent. Hence, the large differences in the exchange hole arise almost entirely from the non-spherical parts, which contribute nothing to the exchange energy.

3.1.5.6 Accuracy issues

Above we have presented arguments why DFT even in very simple approximations, such as the local approximations LDA and LSD, should provide not only simple but also reasonably accurate description of a wide range of realistic systems. Now we want to analyze the accuracy of the DFT and compare it with the concurrent HF approach. The methods have one common feature, namely that both are single determinant approaches. However, while the HF has an exact exchange and no correlation whatsoever, the DFT includes both exchange and correlation but both at an approximate level. This is immediately clear by comparing the non-local HF exchange operator (42) with the local LDA exchange contribution (102) and (106). Moreover, there is yet another error arising from the DFT in the LDA/LSD approximation, namely the degree by which the system deviates from the homogeneous electron gas by which it is (locally) approximated. Hence, we expect the largest errors for atoms, smaller for molecules and clusters, and smallest for solids, especially those, which have quasi free-electron-like behavior, such as, e.g. aluminum. By the same token, free-electron-like solids exhibit the largest errors in the HF

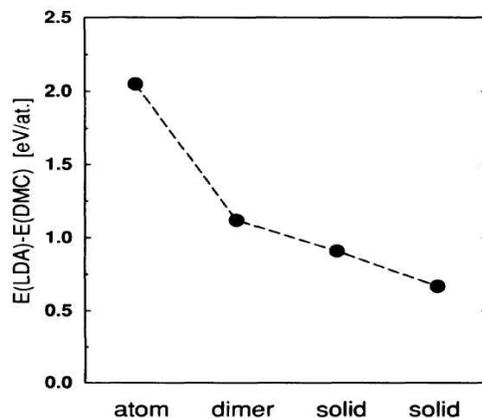


Fig. 10. Total energy difference between LDA and quantum Monte-Carlo results for nitrogen atom, dimer, molecular (Pa3) and atomic solids (I2₁3) [46].

treatment. The system-size dependence of the LDA error is shown in Fig. 10, where the energy difference between LDA and very accurate quantum Monte-Carlo results are shown [46]. We note that the nitrogen system represents a very stringent test, especially for the dimer, which is often used as a test case for quantum chemical methods as it has very large correlation energy due to the triple bond and an extremely small interatomic distance. The errors are indeed appreciable and even if the error for solids is significantly smaller, there is a small difference between the atomic and molecular structures, presumably caused by the different character of the correlation in molecular and atomic solids. The differences in the accuracy between the two solids is quite understandable, as one is a molecular solid, whereas the other an atomic solid; the latter being more accurately modeled by LDA than the former.

To set the stage, let us first take a look at the simple example of an H₂ molecule, which we have discussed above in chapters 3.1.3, 3.1.4 on HF and post HF approaches. The DFT result is shown in Fig. 11. As can be seen, inclusion of correlation in the model has a dramatic effect on the quality of the result. The calculated DFT points, obtained with an approximate (PBE) XC-functional, see chapter 3.1.5.7, result in a curve almost indistinguishable from the more accurate (CI) results. A dramatic improvement of the DFT results over the corresponding HF results resulting from inclusion of an approximate correlation is quite general.

Let us now take a closer look at the systems of intermediate size, namely molecules and clusters. In Fig. 12 we compare binding energies ($E_b = \sum_i E_{at}^i - E_{bond}$, i.e. the difference between the energy of constituent atoms and the bonded system) of two very different medium-size systems [47]. LDA, HF, and very accurate Quantum Monte-Carlo results (see also chapter 3.1.6) are compared. The first is a system of Si_n clusters of different size [48] and the other is a system of hydrocarbons (CH₄, C₂H₂, C₂H₄, C₂H₆, C₆H₆, CH₄, C₃H₄ both propyne and allene), which cover a variety of carbon-carbon bonds (e.g. C₂H₂, C₂H₄, C₂H₆, and C₆H₆ correspond to triple, double, single, and 1.5 C-C bonds, respectively) [49]. These systems represent a crucial test of the accuracy of the LDA approximation as the systems compared are very far from

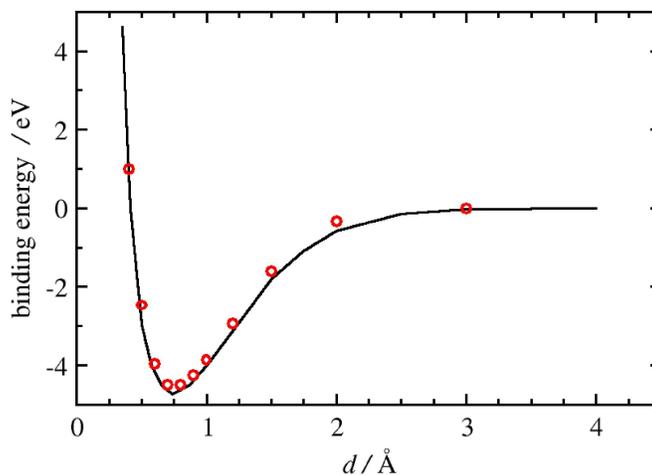


Fig. 11. Potential energy surfaces for H_2 molecule. Full line: very accurate model (CI calculation), red points: DFT calculation with PBE XC-functional (see chapter 3.1.5.7 below).

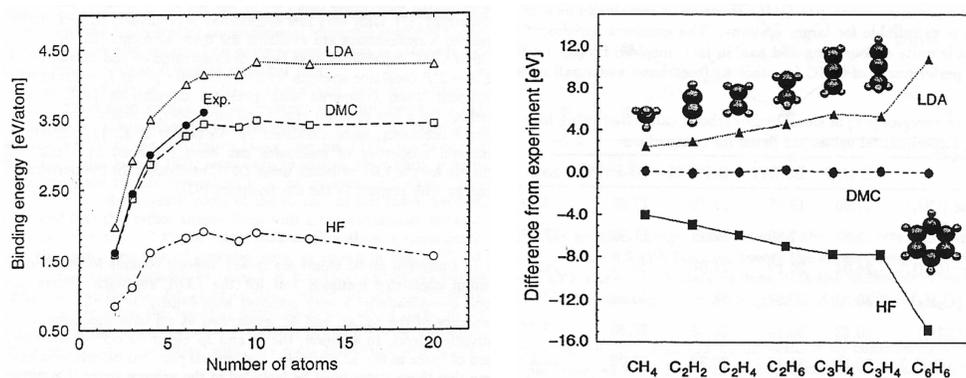


Fig. 12. Comparisons of binding energies in LDA, HF and a very accurate quantum Monte-Carlo method. Left panel: small and medium-size Si_n clusters [48]; right panel: hydrocarbon molecules [49].

the limit of a homogeneous system, where LDA provides an accurate system description. For both systems we find, as is typically the case, that HF is about 30% underbound and LDA about 20% overbound. The deviations from the true results may be quite considerable (for example, for benzene: +12eV in the LDA and -15eV in the HF treatment), as the right panel of Fig. 12 indicates. The feature of the true result being bracketed by HF and LDA results from below and above, respectively, is quite a general feature. This reflects the fact that LDA by the nature of the approximation tends to homogenize the system whereas HF does just the opposite.

As these examples show, the LDA results are almost always appreciably better than the HF results. This is especially true of the extended systems, where LDA often provides reasonably

accurate results. This was one of the main reasons why DFT techniques enjoyed decade-long popularity in the solid-state community. On the other hand, the DFT techniques, in the LDA approximation, were long ignored by the quantum chemistry community.

The question arises about whether the accuracy of the DFT techniques could be further improved beyond the LDA/LSD accuracy. While no simple cure exists for the HF techniques, surprisingly simple improvements can be adopted for the DFT techniques.

3.1.5.7 Beyond the LDA/LSD approximation: GGA functionals

The improvements, which would alleviate the limitations imposed by the local density approximation, are based on abandoning the strict locality of the approximation. For a while we will now consider just the dominant term in the exchange-correlation energy, namely the exchange part. It is clear from the nature of the exchange, that this term cannot be local. A simple-minded extension is based on a *gradient expansion*. This takes into account not only the value of the density at each point \vec{r} but also its curvature, etc. Up to fourth order for the exchange-only energy functional $E_x[n_\uparrow, n_\downarrow]$ for a spin compensated systems ($n/2 = n_\uparrow = n_\downarrow$) Herman, van Dyke, and Ortenburger [50] found the following gradient expansion,

$$E_x^{GE4}[n] = \int d\vec{r} \varepsilon_x^{LDA}(n) \{1 + c_{x2}\xi + c_{x4}[\eta^2 + c_{\eta\xi}^x \eta\xi + c_{\xi^2}^x \xi^2]\}, \quad (120)$$

where

$$e_x^{LDA}(n) = -\frac{3k_F(\vec{r})}{4\pi} n(\vec{r}), \quad (121)$$

$$k_F(\vec{r}) \equiv [3\pi^2 n(\vec{r})]^{1/3}, \quad (122)$$

$$\xi(\vec{r}) \equiv \left(\frac{\nabla n(\vec{r})}{2k_F(\vec{r})n(\vec{r})} \right)^2, \quad (123)$$

$$\eta(\vec{r}) \equiv \frac{\nabla^2 n(\vec{r})}{4k_F^2(\vec{r})n(\vec{r})}. \quad (124)$$

A simpler version of a gradient expansion, based just on simple density difference, $n(\vec{r}) - n(\vec{r}')$, was proposed as early as 1964 by Hohenberg and Kohn [27]. In fact, keeping just the zero- and first-order contributions in expansion (120) yields

$$E_x[n] = A_x \int d\vec{r} n^{4/3} + C_x \int d\vec{r} |\nabla n|^2 / n^{4/3} + \dots, \quad (125)$$

where $A_x = -\frac{3}{4}(3/\pi)^{1/3}$ and $C_x = -7/432\pi(3\pi^2)^{1/3}$ (a.u.). However, all these early attempts to go beyond LDA were unsuccessful. The reason for this is in the sum rules the exact exchange hole must satisfy

$$n_x(\vec{r}, \vec{r}') = -\frac{n(\vec{r})}{2}, \quad (126)$$

$$n_x(\vec{r}, \vec{r} + \vec{R}) \leq 0, \quad (127)$$

$$\int d\vec{R} n_x(\vec{r}, \vec{r} + \vec{R}) = -1. \quad (128)$$

It is well known [42] that the LDA approximation, which retains only the first term on the right in Eq. (125), satisfies the exact conditions (126-128). It was soon discovered that the *gradient expansion approximation* (GEA), which retains both the first and second terms, violates conditions (127) and (128) [51]. This observation explains the failure of GEA to provide the accurate correction to LDA that might have been expected. It was also found, that a cut-off GEA exchange hole, with the cut-offs chosen to satisfy sum-rules (127) and (128), yields an accurate functional for the exchange energy. In addition to the exchange-hole sum-rules outlined above, there are other conditions, which the correlation part must satisfy, such as, e.g. behavior in slow varying/rapidly varying density limit etc. In addition, it has been shown, that the second-order expansion (125) is generally insufficient and a fourth-order contribution containing $\nabla^2 n$ can make equal contributions as the second-order [52]. The so constructed functionals are called *generalized gradient approximation*, or GGA functionals

$$E_{xc}^{GGA} = \int d\vec{r} (n_\uparrow, n_\downarrow, |\nabla n_\uparrow|, |\nabla n_\downarrow|, \nabla^2 n_\uparrow, \nabla^2 n_\downarrow). \quad (129)$$

Many forms of such functionals have been suggested. Some of the most important are given in [53]. One should note that some functionals may not be entirely general as they often have asymptotic behavior valid only for finite systems [54]. Generally, the functionals fall either in the category of semiempirical, which come up with a general functional form with parameters fitted to some data base or first-principles, which fix all the parameters without any need for fitting. The semiempirical GGA's can be remarkably successful for small molecules. For example the semiempirical Becke exchange [54] with either VWN [39] or LYP [55] correlation outperformed correlated *ab initio* methods (MP2 and CISD) in calculations of atomization energies of 32 molecules [56]. From this standpoint, DFT is then of semi-empirical nature. For example, for the well-known BLYP and B3LYP semiempirical functionals, experimental data from up to 407 atomic and molecular systems have been used in order to find the best values for the 15 adjustable parameters [57]. The *ab initio* approach builds on the positive features of the LSD approximation by incorporating exact constraints with the quest to improve the description by the added features. A representative example of this type of functional is the so-called PW91 functional [58]. This functional uses second-order expansion in terms of density gradients (up to ∇n). The form of the functional was later slightly reformulated by Perdew, Burke, and Ernzerhof [59], and is known under the abbreviation PBE. The PBE form has several strong points: (1) it incorporates correctly both the slow and rapidly varying limit of $n(r)$, (2) it cancels the logarithmic singularities of ϵ_c^{LDA} in the high density limit. The PBE functional for atoms and molecules has been compared with results of LSD, BLYP, and B3LYP functionals and found to perform equally well [60] for the properties considered. In Tab. I, we show the results with a range of different GGA functionals for dissociation energies of the same molecules considered above [47, 48]. Although not quite as accurate as the best correlated calculations considered above, the GGAs show an enormous improvement over the LDA, with mean error of about 0.3 eV and maximum error of 2%. However, given the way the GGA functionals are constructed, one must be cautious

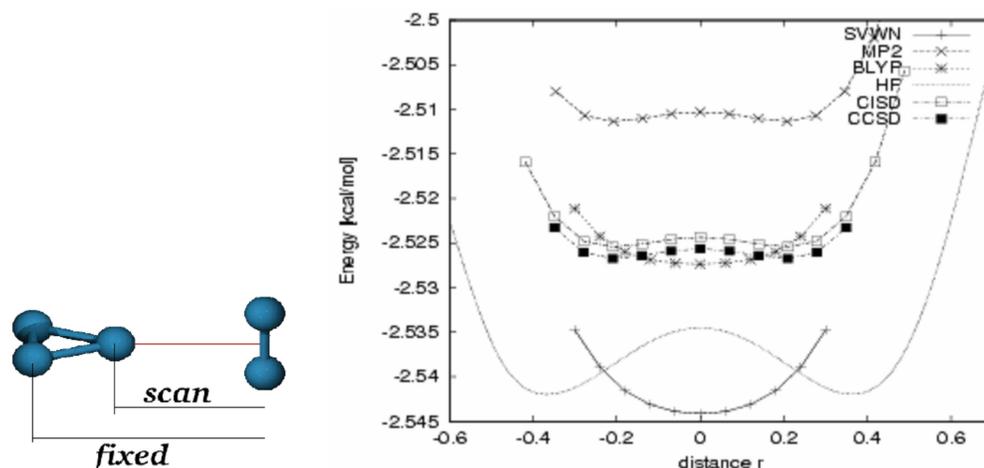


Fig. 13. PES curves for proton movement between the two minima in the H_5^+ cluster.

in regarding these results as a benchmark of the accuracy of the GGAs in general. In particular, while in most cases the GGA energies do represent a definite improvement over the LDA/LSD results, the GGA XC-hole is unlikely to be any better than in the LDA/LSD approximation.

This last point is illustrated in Fig. 13, which shows results for one of the simplest forms of a molecular ionic cluster, namely H_5^+ [61]. The cluster consists of a H_3^+ core ion with one molecule attached to one of its vertices, $H_3^+ \cdots H_2$. As shown in Fig. 13, this structure is qualitatively correctly described not only by high-quality correlated calculations (MP2, CISD, CCSD), but even by HF. The failure of the LDA in the VWN form [39] is not rectified by inclusion of the GGA semilocal corrections, as both methods predict a qualitatively incorrect structure, namely two H_2 molecules with the proton in between, $H_2 \cdots H^+ \cdots H_2$. Fig 12 also highlights the importance of the correlation. As can be seen, the effect of the correlation is to lower the barrier for the proton

Tab. I. Binding energies of small hydrocarbon molecules calculated with three different GGA functionals [47, 48]. The GGA results are compared to HF, LDA, and experimental results.

Binding energies (eV)	HF	LDA	BLYP	B3LYP	PW91	Exp.
methane (CH_4)	14.20	20.59	17.90	18.05	17.78	18.19
acetylene (C_2H_2)	12.70	20.49	17.31	17.20	17.38	17.59
Ethylene (C_2H_4)	18.54	28.19	24.12	24.12	24.07	24.41
ethane (C_2H_6)	23.87	35.37	30.30	30.57	30.29	30.85
allene (C_3H_4)	22.63	35.87	30.23	30.18	30.43	30.36
propyne (C_3H_4)	22.70	35.70	30.12	30.12	30.32	30.45
benzene (C_6H_6)	44.44	70.01	58.38	58.48	59.26	59.24

diffusion quite considerably ~ 0.2 eV, compared to the results of correlated methods.

In Eqs. (112 – 114) we have separated the short-range correlations from the long-range correlations and dealt mainly with the short-range once. The overall spectacular success of the DFT theory has mainly to do with situations where the short-range correlations dominate, such in extended atomic systems. As the example above shows, it often fails to describe correlations at long-range, such as van der Waals interactions. Nevertheless, van der Waals interactions, resulting from density fluctuations in regions separated in space, are quite common in numerous condensed systems. The most widely known system is graphite, where carbon atoms in one plane are held together by very strong covalent bonds, while the planes interact through only much weaker, van der Waals, interactions. Other examples are molecular crystals and clusters, where the LDA approximation is entirely inappropriate and the GGAs may significantly improve the description of the weak bonds [62]. However, in general, the van der Waals interactions up to now have not been properly incorporated in the DFT theory [63]

In general, the GGA functionals lead to improved bond angles, lengths, and energies. In particular, the strengths of hydrogen bonds and other weak bonds between closed-shell systems are, often but not always (see above), significantly improved over LDA results. This is especially true for finite systems, even though in some cases the GGA functionals overcorrect [64]. When applied to extended systems, which typically are much better described by the local functionals, the GGA functionals in most cases tend to more severely overcorrect.

3.1.5.8 Beyond the GGA approximation: meta and orbital-dependent functionals

The next step in the development of gradient approximations is to incorporate the kinetic energy density. Version, based on the PBE functional adds a term of the form [65]

$$\tau_{\sigma}(\vec{r}) = \sum_i^{\text{occ}} \frac{1}{2} |\nabla \psi_{i\sigma}|^2. \quad (130)$$

The $\tau_{\sigma}(\vec{r})$ is the kinetic energy density for the occupied Kohn-Sham orbitals. However, this functional, as some other early functionals of the same form, included parameters determined by fitting to experimental data. This feature is avoided in the recent work, where a new, so-called TPSS functional was constructed, whose form satisfies the requirement that the exchange potential be finite at the nucleus for ground-state one- and two-electron densities [66, 67]. This constraint is satisfied by LDA, but violated by GGAs. Extensive tests indicate results of generally very high quality.

Other functionals are based on relaxing the requirement that the exchange potential be determined within the DFT and rely on combination of exact treatment of exchange, as is the case in the HF, with added approximately treated correlation. These are called hybrid functionals [68], of which the widely used B3LYP functional [57], is an example.

Let's now pursue this idea little further. Similarly as Kohn and Sham [29] did with the orbital-dependent kinetic energy functional setting $T_0[n] \stackrel{!}{=} \sum_{\sigma=\uparrow\downarrow} \sum_i^{\text{occ}} \langle \varphi_{i\sigma} | -\frac{1}{2} \nabla^2 | \varphi_{i\sigma} \rangle$ in deriving their equation, one can consider a framework based on *orbital-dependent functional* E_{xc} . If one performs a power series expansion of E_{xc} in terms of interaction strength, e^2 (e being the elementary

charge), one obtains the exact exchange as leading term

$$E_x^{exact} = -\frac{1}{2} \sum_{\sigma=\uparrow\downarrow} \sum_{i,j=1}^{N_{e\sigma}} \int d\vec{r} \int d\vec{r}' \frac{\varphi_{i\sigma}^*(\vec{r}) \varphi_{j\sigma}^*(\vec{r}') \varphi_{j\sigma}(\vec{r}) \varphi_{i\sigma}(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (131)$$

which is nothing but the Hartree-Fock exchange energy but evaluated with KS orbitals. Once we accept to use an expression for E_{xc} which explicitly depends on the KS orbitals, the question is how to compute the corresponding exchange-correlation potential (98), (105). The way to accomplish this task is indicated by the Hohenberg-Kohn theorem [27]. Application of this theorem to the KS system shows that there is a one-to-one correspondence between $n(\vec{r})$ and $V_{eff}(\vec{r})$. The latter, in turn, determines the KS orbitals from which the density can be obtained. Thus, we can formally write any of these quantities as functionals of the other quantities and an explicit dependence on one of them introduces an explicit dependence on the others. If we use this observation in order to rewrite Eqs. (98), (105) using the chain rule of functional differentiation we have

$$V_{xc}(\vec{r}) = \sum_{\sigma=\uparrow\downarrow} \sum_{i=1}^{N_{e\sigma}} \int d\vec{r}' \frac{\delta E_{xc}}{\delta \varphi_{i\sigma}(\vec{r}')} \frac{\delta \varphi_{i\sigma}(\vec{r}')}{\delta n(\vec{r})} + \text{complex conjugate.} \quad (132)$$

The functional derivative with respect to the orbitals can easily be calculated from its explicit functional form. In order to calculate the functional derivative of the orbitals with respect to the density, we now view the orbitals as functional of $V_{eff}(\vec{r})$ and use, for a second time, the chain rule for functional differentiation to obtain

$$V_{xc}(\vec{r}) = \sum_{\sigma=\uparrow\downarrow} \sum_{i=1}^{N_{e\sigma}} \int d\vec{r}' \int d\vec{r}'' \frac{\delta E_{xc}}{\delta \varphi_{i\sigma}(\vec{r}')} \frac{\delta \varphi_{i\sigma}(\vec{r}')}{\delta V_{eff}(\vec{r}'')} \frac{V_{eff}(\vec{r}'')}{\delta n(\vec{r})} + \text{complex conjugate.} \quad (133)$$

The third functional derivative on the right-hand of Eq. (133) is the inverse of the static density response function of the KS system

$$\chi_s^{-1}(\vec{r}, \vec{r}') = \frac{\delta V_{eff}(\vec{r}')}{\delta n(\vec{r})}. \quad (134)$$

Multiplying Eq. (133) from right one obtains

$$\int d\vec{r}' V_{xc}(\vec{r}') \chi_s(\vec{r}', \vec{r}) = \sum_{\sigma=\uparrow\downarrow} \sum_{i=1}^{N_{e\sigma}} \int d\vec{r}'' \frac{\delta E_{xc}}{\delta \varphi_{i\sigma}(\vec{r}'')} \frac{\delta \varphi_{i\sigma}(\vec{r}'')}{\delta V_{eff}(\vec{r})} + \text{complex conjugate.} \quad (135)$$

We are not going to pursue this avenue any further as the resulting equations are too cumbersome and are hardly used when solving the coupled Born-Oppenheimer (16), (17) or Ehrenfest dynamics (11), (12). However, as shown in the next paragraph, they are conceptually important for the excited-state dynamics.

It is clear from the present discussion that a wide variety of functional approximations to DFT are available. This, on one hand, makes the possibility of finding an appropriate functional for any given system under consideration more probable, but on the other hand, makes also testing the different functionals available a necessity as there is no way to guarantee that a given

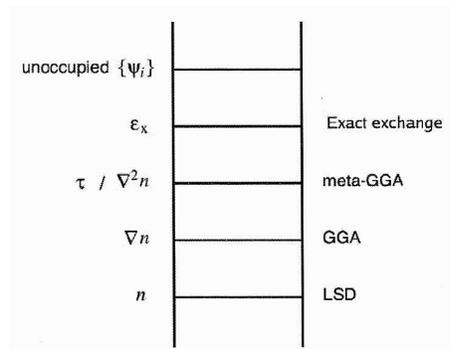


Fig. 14. The ladder of DFT schemes (after Perdew).

functional will provide a good description of the system under consideration. This makes the DFT theory to a semiempirical method. Nevertheless, the development of approximations to the exchange-correlation functionals over the last 20 years has improved the performance of the DFT techniques considerably and many people believe, that the progress up the ladder (Fig. 14) will continue until the chemical accuracy of ~ 1 kcal/mol will be achieved. There are numerous methods, which aim to improve various aspects of the DFT techniques, examples being, for example the GW [69], LDA+U [70] techniques. We will not pursue these improvements. We believe that, given the simplicity, the DFT techniques are a perfect choice for systems, where they provide a decent description. On the other hand for those systems or for those properties where this is not the case, we believe that rather than trying to improve the shortcomings of the DFT techniques it is better to use a true many-body approach, such as quantum Monte-Carlo, rather than to incorporate some many-body aspects into the DFT, which, by its nature, is a mean-field single-body approach.

3.1.5.9 Generalization to finite temperatures: Free Energy functional

Imagine we deal with systems at high temperatures compared with the Fermi temperature. In such a case the total energy functional which corresponds to zero temperature has to be generalized to the *free energy functional*. This is especially needed if we deal with metals, where even very small temperatures will lead to electronic excitations around the Fermi level. The free energy approach is a mean-field approach in the spirit similar to Ehrenfest molecular dynamics (11), (12). In the free-energy approach we assume that the electrons equilibrate more rapidly than the time-scale of the nuclear motion. The electrons do not follow the ions adiabatically. The incoherent electronic transitions due to nuclear motion are included in an averaged way by considering the thermal population of the electronic levels. This is evident from the fact, that free energy is an averaged macroscopic thermodynamic potential. DFT maps the interacting system into a fictitious non-interacting system, i.e. into an ideal Fermionic gas. Statistical mechanics of ideal gasses is well established.

The grand free energy can be written as

$$\Omega^{KS}(\mu, V, T) = -2k_B T \sum_i \ln \{1 + \exp[-\beta(\varepsilon_i - \mu)]\} \quad (136)$$

where $\beta = 1/k_B T$ and $\{\varepsilon_i\}$ are single-particle eigenvalues of the Kohn-Sham Hamiltonian,

$$H^{KS} = -\frac{1}{2}\nabla^2 - V_H(\vec{r}) - \sum_I \frac{Z_I}{|\vec{R}_I - \vec{r}|} + \underbrace{\frac{\delta\Omega_{xc}[n]}{\delta n(\vec{r})}}_{V_{xc}(\vec{r})}, \quad (137)$$

$$H^{KS} \varphi_i(\vec{r}) = \varepsilon_i \varphi_i(\vec{r}), \quad (138)$$

where Ω_{xc} is the exchange-correlation functional at finite temperature. In order to obtain the free energy analogy of the total energy (Eq. 101), the corresponding terms properly generalized to finite temperatures have to be included. In the framework of the Kohn-Sham-like formulation the Helmholtz free energy functional for the interacting case becomes

$$\begin{aligned} F^{KS}(N, V, T) &= \Omega^{KS}(\mu, V, T) + \mu \int d\vec{r} n(\vec{r}) - \frac{1}{2} \int d\vec{r} V_H(\vec{r}) n(\vec{r}) \\ &+ \Omega_{xc} - \int d\vec{r} V_{xc}(\vec{r}) n(\vec{r}). \end{aligned} \quad (139)$$

The corresponding single-particle density, which is needed to solve the equations (137)-(139) is given by

$$n(\vec{r}) = \sum_i f_i(\beta) |\varphi_i|^2, \quad (140)$$

$$f_i(\beta) = \frac{1}{1 + \exp[\beta(\varepsilon_i - \mu)]}, \quad (141)$$

where the fractional occupation numbers $\{f_i\}$ are obtained from the Fermi-Dirac distribution at temperature T in terms of the Kohn-sham eigenvalues $\{\varepsilon_i\}$.

The first generalization of the Kohn-Sham theory to finite temperature is due to Mermin [71]. Stability analysis of the free-energy functional (139) shows that this functional, similarly to the Mermin functional [71], has a stationary point which is a saddle point and not an extremum [72]. This means that direct minimization techniques, see below, cannot be applied. We will return to this point in chapter 3.3.2.1, when we deal with techniques to solving the Kohn-Sham equations.

3.1.5.10 Excited states and gaps

The KS equations in its usual form are ground-state theory and there is generally no rigorous way to calculate for example electronic excitation energies due to photoabsorption. In particular, in usual approximations to E_{xc} , such as LDA, GGA, all the band gaps are severely underestimated. This poses severe problems if one tries to solve the coupled Born-Oppenheimer (16), (17) or Ehrenfest dynamics (11), (12) in excited state. Such processes are of paramount importance in photochemistry, photoswitching etc. In the previous chapter we have briefly addressed the

situation where there are many closely-spaced excited electronic states, such as, for instance, in a metal. Here we focus on the opposite limit, where a single electronic state, or a bunch of them, is completely decoupled from the rest.

3.1.5.11 Discontinuity of E_{xc}

The deficiency of the DFT to describe the band gaps and the HOMO (highest occupied molecular orbital)-LUMO (lowest unoccupied molecular orbital) gaps has to do with the discontinuity of the exchange-correlation functional E_{xc} as a function of particle number, which occurs at integer particle numbers N_e [73, 74]. In order to discuss this property one has to generalize the definition (82), (83) of the Hohenberg-Kohn functional to *non-integer particle numbers*, i.e. to densities of the form

$$\int d\vec{r}n(\vec{r}) = N_e + \varepsilon \quad (142)$$

where N is an integer and $\varepsilon \in \langle 0, 1 \rangle$. The generalization of the Hohenberg-Kohn functional can be given as

$$F_{frac}[n] = \min_{\hat{D} \rightarrow n} Tr \left[\hat{D} \left(\hat{T} + \hat{V}_{ee} \right) \right] \quad (143)$$

where the search runs over all statistical mixtures

$$\hat{D} = (1 - \nu) |\Psi_{N_e}\rangle \langle \Psi_{N_e}| + \nu |\Psi_{N_e+1}\rangle \langle \Psi_{N_e+1}| \quad (144)$$

of an N_e -particle state $|\Psi_{N_e}\rangle$ and an $(N_e + 1)$ -particle state $|\Psi_{N_e+1}\rangle$ which yield the given density $n(\vec{r})$. The density and energy of the ensemble described by \hat{D} are given by

$$n(\vec{r}) = Tr \left[\hat{D}n(\vec{r}) \right] = (1 - \nu) n_{N_e}(\vec{r}) + \nu n_{N_e+1}(\vec{r}), \quad (145)$$

$$E[n] = F_{frac}[n] + \int d\vec{r}V_{ext}(\vec{r})n(\vec{r}) = (1 - \nu)E_{N_e} + \nu E_{N_e+1}, \quad (146)$$

where $n_{N_e}(\vec{r})$ and $E_{N_e}(n_{N_e+1}(\vec{r})$ and $E_{N_e+1})$ are the density and energy corresponding to the state $|\Psi_{N_e}\rangle$ ($|\Psi_{N_e+1}\rangle$). Since the functional (146) is defined on the domain of densities with non-integer particle numbers, its minimum has to be found under the condition of integer particle number which is enforced by a Lagrange multiplier μ (= to the chemical potential). At the minimizing density

$$\mu(N_e) = \left. \frac{\delta E[n]}{\delta n(\vec{r})} \right|_{n_N} = \frac{\partial E_{N_e}}{\partial N_e}, \quad (147)$$

Eqs. (146) and (147) show that the chemical potential $\mu(N)$ jumps discontinuously as N passes through an integer if E_{N_e} and E_{N_e+1} are separated by a finite energy gap. The discontinuity is

$$\Delta(N) = \lim_{\nu \rightarrow 0} \left(\left. \frac{\delta E[n]}{\delta n(\vec{r})} \right|_{n_{N_e+\nu}} - \left. \frac{\delta E[n]}{\delta n(\vec{r})} \right|_{n_{N_e-\nu}} \right). \quad (148)$$

Tab. II. Calculated Kohn-Sham and experimental energy gaps in semiconductors [75]. xcLDA is the LDA exchange-correlation, xEXACT+cLDA stands for exact exchange and LDA correlation, Δ_x means exact exchange only and Δ_{exp} the experimental value.

Solid/eV	xcLDA	xEXACT+cLDA	Δ_x	Δ_{exp}
Si	0.52	1.43	5.84	1.17
C	4.15	5.06	8.70	5.46
GaN(Γ)	1.90	3.46	7.63	3.30
InN(Γ)	-018	1.40	6.14	1.95

This expression can be rewritten in terms of the ionization potential $I(N)$ and affinity $A(N)$ of the N -particle system

$$\Delta(N_e) = I(N_e) - A(N_e) = E(N_e + 1) + E(N_e - 1) - 2E(N_e), \quad (149)$$

where $E(N_e)$ is the ground-state energy of the N_e -particle system. Insertion of Eq. (91) into Eq. (148) yields

$$\Delta(N_e) = \Delta_{KS}(N_e) + \Delta_{xc}(N_e) \quad (150)$$

where

$$\Delta_{KS}(N_e) = \lim_{\nu \rightarrow 0} \left(\left. \frac{\delta T_0[n]}{\delta n(\vec{r})} \right|_{n_{N_e+\nu}} - \left. \frac{\delta T_0[n]}{\delta n(\vec{r})} \right|_{n_{N_e-\nu}} \right) = \varepsilon_{N_e+1}(N_e) - \varepsilon_{N_e}(N_e) \quad (151)$$

is the *KS gap* defined as the difference between $N_e + 1$ -th and N_e -th KS eigenvalue of the N_e -particle system and

$$\Delta_{xc}(N_e) = \lim_{\nu \rightarrow 0} \left(\left. \frac{\delta E_{xc}[n]}{\delta n(\vec{r})} \right|_{n_{N_e+\nu}} - \left. \frac{\delta E_{xc}[n]}{\delta n(\vec{r})} \right|_{n_{N_e-\nu}} \right) \quad (152)$$

is the derivative discontinuity of the exchange-correlation functional E_{xc} .

This *last property is not satisfied by standard LDA and GGA functionals*. In order to recover the discontinuity in E_{xc} one would have to go to *orbital-dependent functionals* briefly mentioned above (Eqs. (131) – (135)). The effect on the band gaps in solids is shown in Tab. II. One can see that LDA severely underestimates the band gaps for all cases listed. In some cases, such as InN, the band gap disappears altogether in the LDA description. Adding exact exchange improves results considerably. This is necessarily due to the discontinuity in the Δ_{xc} term, absent in the LDA treatment. On the other hand, adding just the exact exchange without any correlation overdoes the correction and leads to too wide band gaps.

3.1.5.12 Symmetry adapted wave functions

Apart from the subtleties of the discontinuities in the E_{xc} , one can attempt to construct excited states with correct spin symmetry. This is what the so-called restricted open-shell Hartree-Fock (ROHF) [76] and Kohn-Sham (ROKS) [77] attempt to do. Most important types of photochemical reactions are those that take place in the first excited singlet state S_1 or in the lowest triplet

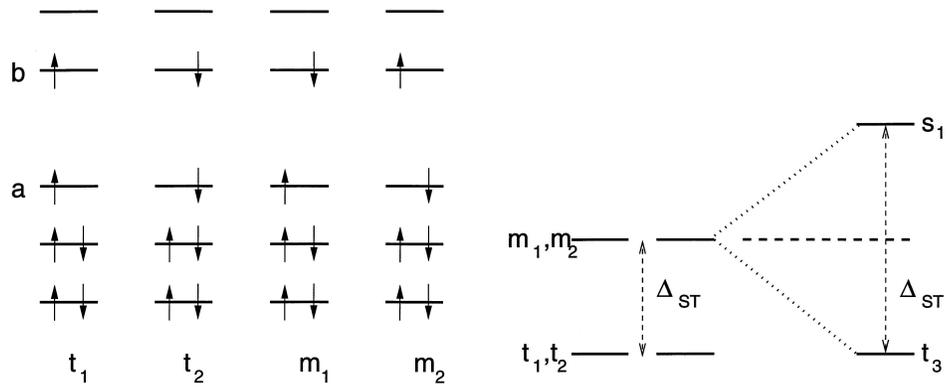


Fig. 15. Symmetry adapted wave functions [77]. Left: the four possible determinants $|t_1\rangle$, $|t_2\rangle$, $|m_1\rangle$, and $|m_2\rangle$ as a result of the promotion of a single electron from the HOMO to the LUMO of a closed-shell system. Right: The energy scheme relating the mixed states $|m_1\rangle$ and $|m_2\rangle$ and the spin-adapted states $|t_3\rangle$ and $|s_1\rangle$.

state T_1 , the other states being in general too short-lived to be of interest. Since T_1 is the lowest state with triplet symmetry, the standard Kohn-Sham theory can be used. However, S_1 is not amenable to the ground-state Kohn-Sham theory, as due to the spin symmetry a minimum of two determinants are necessary for the description of the wave function of non-degenerate excited singlet states of a closed-shell system.

This can easily be seen. Consider a closed shell system with $2n$ electrons in n doubly occupied orbitals. The orbitals are taken to be spin-restricted which means that the spatial parts of the wave functions are the same for spin-up (α) and spin-down (β) orbitals. If a single electron is transferred from the HOMO orbital a to the LUMO orbital b , four different determinants can be formed in accordance with the Pauli principle. Two of these determinants are degenerate with an expectation value of the total spin momentum $\langle \hat{S}^2 \rangle = 2$: $t_1 = |1\bar{1} \cdots ab\rangle$, $t_2 = |1\bar{1} \cdots \bar{a}\bar{b}\rangle$. The other two degenerate determinants, m_1 and m_2 , are not eigenfunctions of \hat{S}^2 . The $\langle \hat{S}^2 \rangle = 0$, and $\langle \hat{S}^2 \rangle = 2$ can be isolated by Clebsh-Gordon decomposition:

$$t_3 = \frac{1}{\sqrt{2}}(m_1 - m_2) = \frac{1}{\sqrt{2}} (|1\bar{1} \cdots a\bar{b}\rangle - |1\bar{1} \cdots \bar{a}b\rangle),$$

$$s_1 = \frac{1}{\sqrt{2}}(m_1 + m_2) = \frac{1}{\sqrt{2}} (|1\bar{1} \cdots a\bar{b}\rangle + |1\bar{1} \cdots \bar{a}b\rangle).$$

As can be seen from the energy diagram in Fig. 15, the energy of the singlet state can be calculated from the energies of m and t states

$$E_{S_1}[\{\phi_i\}] = 2E_m^{KS}[\{\phi_i\}] - E_t^{KS}[\{\phi_i\}]. \quad (153)$$

A generalized version of this expression can be found in the so-called Ziegler's sum method [78]. The energies E_m^{KS} and E_t^{KS} of the mixed and triplet determinants

$$E_m^{KS} [\{\phi_i\}] = T_s [n] + \int d\vec{r} V_{ext}(\vec{r}) n(\vec{r}) + \frac{1}{2} \int d\vec{r} V_H(\vec{r}) n(\vec{r}) + E_{xc} \{n_m^\alpha, n_m^\beta\}, \quad (154)$$

$$E_t^{KS} [\{\phi_i\}] = T_s [n] + \int d\vec{r} V_{ext}(\vec{r}) n(\vec{r}) + \frac{1}{2} \int d\vec{r} V_H(\vec{r}) n(\vec{r}) + E_{xc} \{n_t^\alpha, n_t^\beta\}, \quad (155)$$

are expressed in terms of the Kohn-Sham spin-density functionals constructed from the wave functions $\{\phi_i\}$. The state S_1 is given by

$$|S_1 [\{\phi_i\}]\rangle = \frac{1}{\sqrt{2}} |m [\{\phi_i\}]\rangle - |t [\{\phi_i\}]\rangle. \quad (156)$$

The Kohn-Sham equations for the first excited singlet state are obtained by varying Eq. (153) using (154) and (155) subject to orthonormality constraints $\sum_{i,j=1}^{n-1} \varepsilon_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij})$. The equation for the doubly occupied orbitals $i = 1, 2, \dots, n-1$ reads

$$\left\{ \begin{array}{l} -\frac{1}{2} \nabla^2 + V_H(\vec{r}) + V_{ext}(\vec{r}) \\ + V_{xc}^\alpha [n_m^\alpha, n_m^\beta] + V_{xc}^\alpha [n_m^\alpha, n_m^\beta] - \frac{1}{2} V_{xc}^\alpha [n_t^\alpha, n_t^\beta] - V_{xc}^\beta [n_t^\alpha, n_t^\beta] \end{array} \right\} \phi_i(\vec{r}) = \sum_{j=1}^{n-1} \varepsilon_{ij} \phi_j(\vec{r}), \quad (157)$$

whereas

$$\left\{ \frac{1}{2} \left[-\frac{1}{2} \nabla^2 + V_H(\vec{r}) + V_{ext}(\vec{r}) \right] + V_{xc}^\alpha [n_m^\alpha, n_m^\beta] - \frac{1}{2} V_{xc}^\alpha [n_t^\alpha, n_t^\beta] \right\} \phi_a(\vec{r}) = \sum_{j=1}^{n-1} \varepsilon_{aj} \phi_j(\vec{r}) \quad (158)$$

and

$$\left\{ \frac{1}{2} \left[-\frac{1}{2} \nabla^2 + V_H(\vec{r}) + V_{ext}(\vec{r}) \right] + V_{xc}^\beta [n_m^\alpha, n_m^\beta] - \frac{1}{2} V_{xc}^\alpha [n_t^\alpha, n_t^\beta] \right\} \phi_b(\vec{r}) = \sum_{j=1}^{n-1} \varepsilon_{bj} \phi_j(\vec{r}) \quad (159)$$

are the two different equations for the singly-occupied open-shell orbitals a and b , respectively (Fig. 15). Note, that these Kohn-Sham-like equations feature *orbital-dependent exchange-correlation potentials*, which was above identified as a necessity for obtaining the required discontinuities of E_{xc} as function of the particle number.

Examples of the results obtained for the vertical and adiabatic excitation energies calculated for some compounds [77] are shown in Tab. III. As can be seen, due to the orbital-dependent exchange-correlation potential, the method of symmetry-adapted wave function yields reasonable results for the excitation energies, which could not be achieved if the differences in Kohn-Sham energies $\Delta_{KS}(N_e) = \varepsilon_{N_e+1}(N_e) - \varepsilon_{N_e}(N_e)$ were used to estimate the excitation energies.

Tab. III. The calculated excitation energies for the lowest singlet excitations of three molecules [77]. The energies are in eV. Experimental and theoretical values (CCSD) [79] are given for comparison. LDA (BLYP) stands for LDA (BLYP) prescription for E_{xc} in Eqs. (150) – (153).

Compound	Transition	LDA	BLYP	Theor./CCSD	Exp.
Formaldehyde	vertical	3.81	3.51	3.95	4.07
	adiabatic	3.55	3.16		3.50
Acetaldehyde	vertical	4.22	3.83	4.26	4.28
	adiabatic	3.81	3.35		3.69
Acetone	vertical	4.33	3.93		4.48
	adiabatic	3.85	3.41		3.77

3.1.5.13 Time-dependent DFT

A somewhat more general way to deal with excitations and excited states is provided by the *time-dependent DFT method* (TDDFT). The method is based on a generalization of the Hohenberg-Kohn theorem to time-dependent densities and potentials [80]. Runge and Gross proved that there is a one-to-one correspondence between the time-dependent external potential $V_{ext}(\vec{r}, t)$, and the time-dependent electron density, $\rho(\vec{r}, t)$, for a fixed initial state [80]. Similar to the static case, one can cast the many-electron problem into the Kohn-Sham non-interacting electrons form assuming non-interacting v -representability. This assumption means that the density of the interacting system can be reproduced by the non-interacting effective potential $V_{eff}[\rho](\vec{r}, t)$, i.e.,

$$\rho(\vec{r}, t) = \sum_i^{occ} |\phi_i(\vec{r}, t)|^2, \quad (160)$$

where the orbitals $\{\phi_i(\vec{r}, t)\}$ satisfy the time-dependent Kohn-Sham equations

$$i \frac{\partial}{\partial t} \phi_i(\vec{r}, t) = \left(-\frac{1}{2} \nabla^2 + V_{ext}(\vec{r}, t) + \underbrace{\int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|}}_{V_{eff}[n](\vec{r}, t)} + V_{xc}[n](\vec{r}, t) \right) \phi_i(\vec{r}, t). \quad (161)$$

The $V_{eff}[\rho](\vec{r}, t)$ defines the exchange-correlation potential $V_{xc}[\rho](\vec{r}, t)$. In the usual adiabatic approximation [81], the exchange-correlation potential is taken to be simply the derivative of the ground-state exchange-correlation energy E_{xc} , with respect to the density, $V_{xc}[n](\vec{r}, t) \approx \delta E_{xc}[n]/\delta n$.

Let us now suppose a time-dependent perturbing potential $V_1(\vec{r}, t)$, for instance an oscillating electric field $V_1(\vec{r}, t) = Ez \cos(\omega t)$, is switched at time $t = t_0$

$$V_{ext}(\vec{r}, t) = V_0(\vec{r}) + V_1(\vec{r}, t) = \begin{cases} V_0(\vec{r}), & t \leq t_0, \\ V_0(\vec{r}) + Ez \cos(\omega t), & t > t_0, \end{cases} \quad (162)$$

where $V_0(\vec{r})$ is the attractive Coulomb potential between electrons and nuclei. The first-order density response for interacting particles to the perturbation may be obtained from

$$n(\vec{r}, t) - n_0(\vec{r}) \approx n_1(\vec{r}, t) = \int dt' \int d\vec{r}' \chi(\vec{r}, t, \vec{r}', t') V_1(\vec{r}', t') \quad (163)$$

with the interacting response function

$$\chi(\vec{r}, t, \vec{r}', t') = \left. \frac{\delta n(\vec{r}, t)}{\delta V_{ext}(\vec{r}', t')} \right|_{V_0}. \quad (164)$$

Expressing the right-hand-side of Eq. (163) in terms of Kohn-Sham response function of non-interacting particles

$$\chi_{non-int}(\vec{r}, t, \vec{r}', t') = \left. \frac{\delta n(\vec{r}, t)}{\delta V_{eff}(\vec{r}', t')} \right|_{V_{eff}[n_0]} \quad (165)$$

we have

$$n_1(\vec{r}, t) = \int dt' \int d\vec{r}' \chi_{non-int}(\vec{r}, t, \vec{r}', t') V_{eff,1}(\vec{r}', t'), \quad (166)$$

where

$$V_{eff,1}(\vec{r}, t) = V_1(\vec{r}, t) + \int d\vec{r}' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + \int d\vec{r}' \int dt' f_{xc}[n_0](\vec{r}, t, \vec{r}', t') n_1(\vec{r}', t') \quad (167)$$

with the exchange-correlation kernel

$$f_{xc}[n_0](\vec{r}, t, \vec{r}', t') = \frac{\delta V_{xc}[n](\vec{r}, t)}{\delta n(\vec{r}', t')}. \quad (168)$$

The frequency-dependent linear density response is obtained by Fourier transform

$$n_1(\vec{r}, \omega) = \int d\vec{r}' \chi_{non-int}(\vec{r}, \vec{r}', \omega) V_1(\vec{r}', \omega) + \int d\vec{r}'' \int d\vec{r}''' \chi_{non-int}(\vec{r}, \vec{r}'', \omega) \left(\frac{1}{|\vec{r}'' - \vec{r}'''}| + f_{xc}[n_0](\vec{r}'', \vec{r}''', \omega) \right) n_1(\vec{r}''', \omega) \quad (169)$$

and

$$\chi_{non-int}(\vec{r}, \vec{r}', \omega) = \sum_{ph\sigma} \left(\frac{\phi_p(\vec{r}) \phi_h^*(\vec{r}) \phi_p^*(\vec{r}') \phi_h(\vec{r}')}{\omega - (\varepsilon_p - \varepsilon_h)} - \frac{\phi_p^*(\vec{r}) \phi_h(\vec{r}) \phi_p(\vec{r}') \phi_h^*(\vec{r}')}{\omega + (\varepsilon_p - \varepsilon_h)} \right). \quad (170)$$

ϕ_p and ϕ_h are the Kohn-Sham particle (occupied) and hole (unoccupied) orbitals corresponding to the Kohn-Sham energies ε_p and ε_h , respectively.

Tab. IV. Comparison of TDDFT excitation energies in eV in H₂CO [86] with experiment [87]. Active space contained 6 *h* and 70 *p* states. LDA exchange-correlation potential is used.

Transition	¹ A ₂	³ A ₂	³ A ₁	¹ B ₁	³ B ₂	¹ B ₂	³ A ₁	¹ A ₁
	n → π*	n → π*	π → π*	σ → π*	n → 3s	n → 3s	n → 3p	n → 3p
TDDF(LDA)	3.68	3.08	6.39	8.94	6.56	6.89	7.59	7.83
Exp.	3.94	3.50	5.53	8.68	6.83	7.09	7.79	7.97

Note, that the density response, $n_1(\vec{r}, \omega)$, appears on both sides of Eq. (169), and hence, Eqs. (169), (170) have to be solved self-consistently. This is most conveniently done by casting Eq. (169) into a matrix eigenvalue problem [82]. We will not follow this procedure here. One complication in solving Eqs. (169), (170) is that the sum over hole states are in principle infinite. However, the excitation energies exhibit typically only minor dependence on the size of the active space included. Another complication is that TDDFT merely provides excitation energies, but excited state wave functions are not properly defined. Only recently first excited state geometry optimization was presented [83–85].

In Tab. IV we show the computed excitation energies for an extended system using the TDDFT method. Similarly to the spin-adapted wave function approach, there is a reasonable agreement between the calculated and experimental excitation energies. However, inspecting the data for the transitions to the 3*s* and 3*p* Rydberg states, the deficiency of the LDA exchange-correlation potential at long-range becomes apparent. Switching to GGA’s changes little.

3.1.6 Quantum Monte-Carlo

In order to achieve chemical accuracy (1 kcal/mol) or to treat strongly correlated systems, one needs to take into account the many-body nature of the wave function of the electronic subsystem (13). In chapter 3.1.4 we have outlined methods to incorporate the many body effects via multideterminant expansion of the electronic wave function. While in principle these methods can be made arbitrarily accurate, their current practical applicability is limited to very small systems, such as diatomic molecules. The quantum Monte-Carlo (QMC) method [88, 89] attacks the problem from an entirely different standpoint. Instead of constructing the electronic wave function (20) with all the possible excitations built in, as is done in the quantum chemistry methods (chapter 3.1.4), the QMC method attempts just to sample the many-body wave function, subject, in principle, to only statistical errors.

3.1.6.1 Variational Monte-Carlo

Imagine, for a moment, that we “somehow” know a many-body model wave function, which we will call trial wave function, $\Psi_T(\vec{R})$, with a $3N_e$ -dimensional vector of electron position vectors $\vec{R} \equiv (\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{N_e})$. Then the energy of the system can be estimated as

$$E \approx \frac{\int \Psi_T^*(\vec{R}) \hat{H}_e \Psi_T(\vec{R}) d\vec{R}}{\int \Psi_T^*(\vec{R}) \Psi_T(\vec{R}) d\vec{R}}. \quad (171)$$

In order to evaluate integral (171) we need to perform integration in $3N_e$ -dimensional space. Even for systems of medium size, this results in integration in extremely highly dimensional spaces. The worst way to evaluate the integral (171) would be to start by computing the values of the function to be integrated on M mesh points \bar{R} and applying some form of numerical integration, such as the Simpson's rule. The reason is that in a d -dimensional space the error ε scales as $O(M^{-4/d})$. Therefore, as the dimension d increases, the error falls off increasingly slowly. In such a situation it is much more efficient to use statistical methods, such as Monte-Carlo, where $\varepsilon = O(M^{-1/2})$. Notice, that the error is independent of the dimension of the integration space. In order to use Monte-Carlo methods to evaluate the integral, Eq. (171) can be rewritten as follows

$$E_{\text{VMC}} = \frac{\int |\Psi_T(\bar{R})|^2 \left[\Psi_T^{-1}(\bar{R}) \hat{H}_e \Psi_T(\bar{R}) \right] d\bar{R}}{\int \Psi_T^*(\bar{R}) \Psi_T(\bar{R}) d\bar{R}} = \int p(\bar{R}) g(\bar{R}) d\bar{R} \quad (172)$$

with probability density $p(\bar{R}) = |\Psi_T(\bar{R})|^2 / \int \Psi_T^*(\bar{R}) \Psi_T(\bar{R}) d\bar{R}$ ($p(\bar{R}) \geq 0$, $\int d\bar{R} p(\bar{R}) = 1$). Metropolis algorithm⁶ [90] is used to sample a set of points $\{\bar{R}_m; m = 1, \dots, M\}$ from the configuration-space probability density $p(\bar{R})$. At each of these points the ‘‘local energy’’

$$E_L(\bar{R}) = \Psi_T^{-1}(\bar{R}) \hat{H}_e \Psi_T(\bar{R}) \quad (173)$$

is evaluated and the average energy accumulated

$$E_{\text{VMC}} = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M E_L(\bar{R}_m). \quad (174)$$

The trial moves are usually sampled from a Gaussian centered on the current position of the walker, the variance of the Gaussian being chosen such that the average acceptance probability is $\approx 50\%$. Expression (174) is the basis of the *variational Monte-Carlo (VMC) method*.

However, in order to evaluate E_{VMC} we need to specify the *model trial many-body wave function* $\Psi_T(\bar{R})$. Since we don't know the exact many-body wave function we have to guess it. A very successful ansatz is taking $\Psi_T(\bar{R})$ in a Slater-Jastrow type

$$\Psi_T(\bar{R}) = \sum_{\alpha} D_{\alpha}^{\uparrow} D_{\alpha}^{\downarrow} J(\bar{R}) = \sum_{\alpha} D_{\alpha}^{\uparrow} D_{\alpha}^{\downarrow} \exp \left[\sum_i \chi(\vec{r}_i) - \sum_{j>i} u(\vec{r}_i, \vec{r}_j) + \dots \right] \quad (175)$$

where $J(\bar{R})$ is the so-called *Slater-Jastrow factor*, $D_{\alpha}^{\uparrow}, D_{\alpha}^{\downarrow}$ are Slater determinants of spin-up and spin-down single-particle orbitals (HF, DFT or similar), $\chi(\vec{r}_i)$ is a one-body function describing the electron-nuclear correlation, while $u(\vec{r}_i, \vec{r}_j)$ describes the electron-electron correlations.

⁶The Metropolis algorithm generates the sequence of sampling points \bar{R}_m by moving a single walker according to the following rules: (1) Start the walker at a random position \bar{R} ; (2) make a trial move to a new position \bar{R}' chosen from some probability density function $T(\bar{R}' \leftarrow \bar{R})$. After the trial move the probability that the walker initially at \bar{R} is now in the volume element $d\bar{R}'$ is $d\bar{R}' \times T(\bar{R}' \leftarrow \bar{R})$; (3) Accept the trial move to \bar{R}' with probability $A(\bar{R}' \leftarrow \bar{R}) = \min \left(1, \frac{T(\bar{R} \leftarrow \bar{R}') P(\bar{R}')}{T(\bar{R}' \leftarrow \bar{R}) P(\bar{R})} \right)$. If the trial move is accepted, the point \bar{R}' becomes the next point of the walk; if the trial move is rejected, the point \bar{R} becomes the next point of the walk. If $P(\bar{R})$ is high, most trial moves away from \bar{R} will be rejected and point \bar{R} may occur many times in the set of points making up the random walk. (4) return to step (2) and repeat.

In particular, $u(\vec{r}_i, \vec{r}_j)$ is constructed so that $\Psi_T(\bar{R})$ satisfies exactly the cusp-conditions (56) and (60). Note that wavefunction (175) is constructed in terms of orbitals, instead of spinorbitals and hence is not antisymmetric on exchange of electrons with opposite spins. It can be shown [88] that it gives the same expectation value for any spin-independent operator. The role of the Slater-Jastrow factor is quite straightforward to understand. The antisymmetry of the wave function creates an exchange hole that keeps parallel-spin electrons apart, but there are no correlations between the positions of electrons with antiparallel spins. There is a significant probability of finding two antiparallel-spin electrons very close to each other (c.f. Fig. 3) and electron-electron Coulomb repulsion is high. The purpose of the two-body term $u(\vec{r}_i, \vec{r}_j)$ is to reduce the magnitude of the many-electron wave function whenever two electrons come close to one another and decreases the electron-electron interaction energy. The unwelcome effect of the two-body term is that it pushes electrons away from regions of high charge density. Hence, it overdoes the correction effect, as the charge densities are normally quite well described even at the single-particle level (HF, DFT). The one-body term $\chi(\vec{r}_i)$ in (175) alters the charge density without disturbing the pair-correlation function, which the two-body term was fixing. The Slater-Jastrow factor has typically around 20 adjustable parameters, which represents a reasonable compromise between variational freedom and computational complexity. At the first glance the trial wave function (175) looks quite similar to the CI wave function (20). The main difference between the two approaches lies in the fact that the cusp conditions are explicitly built in the wave function of the form (175). This very significantly reduces the number of determinants needed in the expansion. The reason is that in the CI expansion (20) the singularity in the many-body wave function due to the cusp conditions is expanded in terms of the multiple determinants, while they are handled explicitly in the wave function (175). In practical terms, good quality CI calculations may use $10^6 - 10^9$ Slater determinants, while often a single determinant is all that is needed in the QMC trial wave function.

The variational parameters of the trial wave function (175) are optimized by minimizing either the variance of the local energy

$$\sigma^2 = \frac{\int |\Psi_T|^2 \left[\frac{\hat{H}_e \Psi_T}{\Psi_T} - E_{\text{VMC}} \right]^2 d\bar{R}}{\int |\Psi_T|^2 d\bar{R}} \quad (176)$$

or total energy E_{VMC} [88, 89]. In practice one generates M configurations using some guess for the variational parameters in the wavefunction (175) and minimizes σ^2 or E_{VMC} by modifying the variational parameters in (175). This process is repeated until the “best” wave function is determined within the constraints of the functional form dictated by the Slater-Jastrow factor. Depending on the sophistication of the parameters of the trial wave function, typically up 85–95% of the correlation energy is recovered.

3.1.6.2 Diffusion Monte-Carlo

Further improvement can be achieved either by improving the quality of the trial wave function Ψ_T , or by the so-called *diffusion Monte-Carlo* technique (DMC) [88, 89]. While the former route is certainly a valid option, in practical terms it may often be difficult to improve Ψ_T to accuracy comparable to the best quantum chemistry methods. Such accuracy can be achieved using the DMC technique.

DMC is a stochastic projection method for solving the many-body Schrödinger equation in imaginary time

$$-\frac{\partial}{\partial t}\Phi(\bar{R}, t) = (\hat{H}_e - E_T)\Phi(\bar{R}, t), \quad (177)$$

where E_T is an energy offset whose importance will become clear later. Equation (177) may be rewritten in integral form,

$$\Phi(\bar{R}, t + \tau) = \int G(\bar{R} \leftarrow \bar{R}', \tau) \Phi(\bar{R}') d\bar{R}', \quad (178)$$

where

$$G(\bar{R} \leftarrow \bar{R}', \tau) = \langle \bar{R} | \exp[-\tau(\hat{H}_e - E_T)] | \bar{R}' \rangle \quad (179)$$

is a Green's function obeying the same equation as the wave function

$$-\frac{\partial}{\partial t}G(\bar{R} \leftarrow \bar{R}', t) = (\hat{H}_e - E_T)G(\bar{R} \leftarrow \bar{R}', t) \quad (180)$$

with the initial condition $G(\bar{R} \leftarrow \bar{R}', 0) = \delta(\bar{R} - \bar{R}')$. It is easy to show that the operator $\exp[-\tau(\hat{H}_e - E_T)]$ in the $\tau \rightarrow \infty$ limit projects out the lowest eigenstate $|\Psi_0\rangle$ with non-zero overlap with the chosen $|\Phi(t=0)\rangle = |\Phi_{init}\rangle$. Using the spectral expansion of the propagator

$$\exp(-\tau\hat{H}_e) = \sum_i |\Psi_i\rangle \exp(-\tau E_i) \langle \Psi_i | \quad (181)$$

where $\{\Psi_i\}$, $\{E_i\}$ denote a complete set of eigenfunctions and eigenvalues of H_e , respectively. In the $\tau \rightarrow \infty$ limit we have

$$\begin{aligned} & \lim_{\tau \rightarrow \infty} \langle \bar{R} | \exp[-\tau(\hat{H}_e - E_T)] | \Phi_{init} \rangle = \\ &= \lim_{\tau \rightarrow \infty} \int G(\bar{R} \leftarrow \bar{R}') \Phi_{init}(\bar{R}') d\bar{R}' = \\ &= \lim_{\tau \rightarrow \infty} \sum_i \Psi_i(\bar{R}) \exp[-\tau(E_i - E_T)] \langle \Psi_i | \Phi_{init} \rangle = \\ &= \lim_{\tau \rightarrow \infty} \Psi_0(\bar{R}) \exp[-\tau(E_0 - E_T)] \langle \Psi_0 | \Phi_{init} \rangle. \end{aligned} \quad (182)$$

By adjusting E_T to be equal to E_0 in (182) causes an exponential damping of all higher states in energy Ψ_i leaving only the ground-state Ψ_0 to survive. This property of the projector $\exp[-\tau(\hat{H}_e - E_T)]$ is the basis of the DMC method.

Let's now take a closer look at Eq. (177) by explicitly inserting the Hamiltonian \hat{H}_e into the imaginary-time Schrödinger equation (177)

$$-\frac{\partial}{\partial t}\Phi(\bar{R}, t) = \underbrace{-\frac{1}{2} \sum_i \nabla_i^2 \Phi(\bar{R}, t)}_{diffusion} + \underbrace{(V(\bar{R}) - E_T) \Phi(\bar{R}, t)}_{rate}. \quad (183)$$

Equation (183) with just the first term on the right-hand side included represents the *diffusion equation* in a $3N_e$ -dimensional space with diffusion constant $1/2$. $\Phi(\bar{R}, t)$ plays the role of the density of diffusing particles. This equation can be simulated by a random walk of particles through configuration space: the well-known “drunkard’s walk”. If instead just the second term is included we have the so-called *rate equation*, describing branching process, such as radioactive decay or exponential birth and death process in population. Thus the entire equation (183) can be simulated as a combination of diffusion and a branching process, in which the number of diffusers increases or decreases at a given point proportional to the density of diffusers already there. This branching process serves to decrease the probability density in regions where $V(\bar{R})$ is large and enhance it in regions of favorable potential energy.

The Green’s function for a diffusion process in a $3N_e$ -dimensional space is a $3N_e$ -dimensional Gaussian with variance τ in each dimension

$$G(\bar{R} \leftarrow \bar{R}', \tau) = (2\pi\tau)^{-3N_e/2} \exp\left[-\frac{|\bar{R} - \bar{R}'|^2}{2\tau}\right]. \quad (184)$$

The Green’s function for the entire process can be obtained using the Trotter-Suzuki formula

$$e^{-\tau(\hat{A}+\hat{B})} = e^{-\tau\hat{B}/2}e^{-\tau\hat{A}}e^{-\tau\hat{B}/2} + O(\tau^3) \quad (185)$$

and $\hat{A} = \hat{T}$ and $\hat{B} = \hat{V} - E_T$ we have

$$\begin{aligned} G(\bar{R} - \bar{R}', \tau) &= \langle \bar{R} | e^{-\tau(\hat{T}+\hat{V}-E_T)} | \bar{R}' \rangle \\ &= e^{-\tau[V(\bar{R})-E_T]/2} \langle \bar{R} | e^{-\tau\hat{T}} | \bar{R}' \rangle e^{-\tau[V(\bar{R})-E_T]/2} + O(\tau^3). \end{aligned} \quad (186)$$

The approximate Green’s function for small τ is therefore given by [91]

$$\begin{aligned} G(\bar{R} - \bar{R}', \tau) &\approx (2\pi\tau)^{-3N/2} \exp\left[-\frac{(\bar{R} - \bar{R}')^2}{2\tau}\right] \\ &\times \exp[-\tau(V(\bar{R}) + V(\bar{R}') - 2E_T)/2]. \end{aligned} \quad (187)$$

In Eq. (187) the first part of $G(\bar{R} - \bar{R}', \tau)$ clearly accounts for the diffusion process whereas the term

$$P = \exp[-\tau(V(\bar{R}) + V(\bar{R}') - 2E_T)/2] \quad (188)$$

accounts for the branching or birth/death process, with P being the number of walkers that survive to the next step [91]. If $P < 1$, the walker continues evolution with probability P , and if $P \geq 1$, the walker continues and in addition, at the same position a new walker is created with probability $P - 1$. These processes can be incorporated via a single command $M_{new} = INT(P + \eta)$, where M_{new} is the number of walkers evolving to the next step at a given position and η is a random number from a uniform distribution $\eta \in \langle 0, 1 \rangle$.

The energy offset E_T , which determines the renormalization (Eq. 187), is used to control the total population of walkers. The E_T is adjusted from time to time during the simulation so that the total number of walkers fluctuates around a preset number, typically between 10^2 and 10^3 . In Fig. 16 we depict the DMC algorithm for the simplest case, a single electron moving in a parabolic, one-dimensional potential. This figure also illustrates the difficulties that arise if one

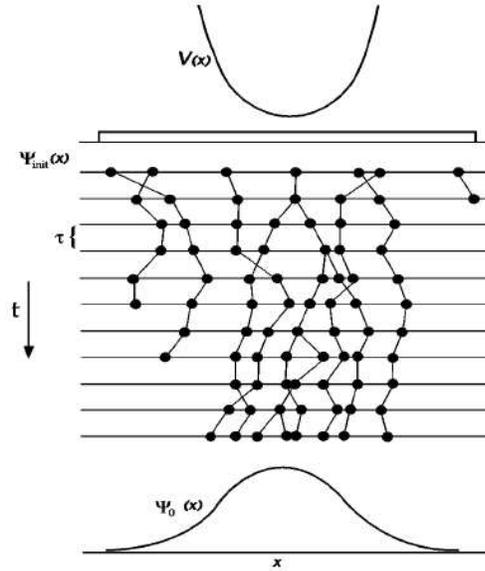


Fig. 16. Illustration of the DMC algorithm for the case of a one-dimensional potential [88]. The initial walker distribution samples a uniform Ψ_{init} . As the imaginary time progresses, the distribution samples the ground-state wave function Ψ_0 . Note the disappearance of the walkers in the regions of high potential energy and birth in regions of low values.

wants to work with the DMC wave function represented by the density of walkers. Note also that the walkers, which take the form of a Dirac delta function, $\delta(\bar{R} - \bar{R}')$, effectively play the role of the basis set. This is a very important feature in a highest accuracy approach, as it avoids the severe problem associated with the choice of basis sets without compromising accuracy; see also chapter 3.1.7.1.

In the discussion above we have identified the many-body wave function $\Phi(\bar{R}, t)$ as the density of the walkers and, hence, tacitly assumed that the many-body wave function is non-negative. In reality, of course, this is not valid for the wave function $\Phi(\bar{R}, t)$, but for its modulus squared $|\Phi(\bar{R}, t)|^2$. Any fermionic many-body wave function must have nodes, as it must satisfy the antisymmetry condition (19). Probabilistic methods, of which DMC is one incarnation, can handle only positive distributions, and therefore are not suitable to solving the quantum many-body problem. Straightforward generalizations, which have been applied with good results to some very small problems [92, 93], generally lead to the well-known Fermionic sign problem.

The way out of this deadlock is provided by the *fixed node approximation*⁷ [91, 94-97]. The fixed node approximation uses the fact that the imaginary-time Schrödinger equation is a linear equation which allows a general wave function $\Phi(\bar{R}, t)$ with nodes to be decomposed into sep-

⁷We assume here that the system has time-reversal symmetry, i.e. that the Hamiltonian is real. This is a special case of a more general fixed-phase approximation [93,97]. Such a generalization is necessary for studying many-body systems in magnetic field or states with non-zero angular momentum.

arate domains of positive and negative contributions $\Phi_+(\bar{R}, t)$, $\Phi_-(\bar{R}, t)$, respectively. Both can be taken as positive and the total wave function constructed from the solutions in the positive/negative subregions as $\Phi(\bar{R}, t) = \Phi_+(\bar{R}, t) - \Phi_-(\bar{R}, t)$. The problem with this approach is that the antisymmetric wave function $\Phi(\bar{R}, t)$ exponentially decays as $\exp(-\Delta E t)$ for $t \rightarrow \infty$ where ΔE is the energy difference between antisymmetric and symmetric ground states. The solution to this problem consists in introducing absorbing barriers placed on the nodal surface, dividing the configuration space into a set of nodal pockets. The DMC walkers are initially scattered throughout both regions. The simulations in both pockets proceed independently. In the $t \rightarrow \infty$ limit the walker densities in the two regions are proportional to the lowest-energy eigenfunctions satisfying zero boundary conditions at the absorbing barrier. The antisymmetric eigenfunction of the problem solved satisfies the same boundary conditions, which determine the eigenfunctions within each nodal bounded region. The $t \rightarrow \infty$ walker densities within each region are therefore proportional to the antisymmetric eigenfunction in that region.

The only remaining problem is that the $dN_e - 1$ -dimensional nodal surfaces (d being the dimensionality of the physical space) are almost never known. This difficulty is circumvented by using an approximate nodal surface, which is normally obtained from a VMC trial wave function. The fixed-node DMC energies are variational [88, 91, 96]

$$E_G \leq E_{\text{DMC}} \leq E_{\text{VMC}}. \quad (189)$$

The implementation of the fixed-node DMC is straightforward. One scatters the DMC walkers throughout the configuration space and moves them in the usual way. After each DMC move the sign of the trial wave function is checked and when the walker crosses the nodal surface it is deleted. Within each pocket, the fixed-node DMC projects out the lowest-energy nodeless wave function satisfying zero boundary conditions on the nodal surface. The typical fixed-node DMC errors are typically $\approx 5\%$ of the correlation energy. This surprisingly high accuracy is due to the fact that it is relatively easy to guess the approximate nodal surface⁸ and the fact that we are making errors in parts of the space, where the weight of the wave function is small and hence are rarely visited by the walkers and hence their contribution to the energy is small.

3.1.6.3 Importance sampling

The DMC algorithm outlined above is extremely inefficient. One reason is that the renormalization factor P in Eq. (188) typically wildly fluctuates from one MC step to the next. The efficiency of the configuration space sampling can be highly improved if one uses the information which has already been gathered in the VMC calculation. This is the basis of the so-called *importance-sampling transformation* [91, 99]. Instead of sampling directly the desired wave function $\Phi(\bar{R}, t)$, let us consider sampling of a different function $f(\bar{R}, t)$, which is a product of $\Phi(\bar{R}, t)$ and the trial wave function $\Psi_T(\bar{R})$,

$$f(\bar{R}, t) = \Phi(\bar{R}, t) \Psi(\bar{R}). \quad (190)$$

After multiplying Eq. (177) by $\Psi_T(\bar{R})$ we obtain

$$-\frac{\partial}{\partial t} f(\bar{R}, t) = -\frac{1}{2} \sum_i \nabla_i f(\bar{R}, t) + \nabla \cdot [\bar{v}_D(\bar{R}) f(\bar{R}, t)]$$

⁸It has been shown recently, that under fairly general conditions there are just two pockets divided by a single plane [98].

$$+ [E_L(\bar{R}) - E_T] f(\bar{R}, t) \quad (191)$$

where

$$\bar{v}_D(\bar{R}) \equiv \nabla \ln |\Psi_T(\bar{R})| = \Psi_T(\bar{R})^{-1} \nabla \Psi_T(\bar{R}) \quad (192)$$

is the $3N_e$ -dimensional drift velocity and E_L is the local energy defined in Eq. (173). The analogy of Eq. (178) now reads

$$f(\bar{R}, t + \tau) = \int \tilde{G}(\bar{R} \leftarrow \bar{R}', \tau) f(\bar{R}') d\bar{R}', \quad (193)$$

where the modified Green's function $\tilde{G}(\bar{R} \leftarrow \bar{R}', \tau)$ is by definition equal to

$$\Psi_T(\bar{R}) G(\bar{R} \leftarrow \bar{R}', \tau) \Psi_T(\bar{R})^{-1}$$

and the analogy of the short-time approximation to $\tilde{G}(\bar{R} \leftarrow \bar{R}', \tau)$ reads

$$\begin{aligned} \tilde{G}(\bar{R} \leftarrow \bar{R}', \tau) &\approx (2\pi\tau)^{-3N_e/2} \exp \left\{ -\frac{[\bar{R} - \bar{R}' - \tau\bar{v}_D(\bar{R}')]^2}{2\tau} \right\} \\ &\times \exp \left\{ -\tau [E_L(\bar{R}) + E_L(\bar{R}') - 2E_T] / 2 \right\} \\ &= G_{diff}(\bar{R} \leftarrow \bar{R}', \tau) G_{branch}(\bar{R} \leftarrow \bar{R}', \tau). \end{aligned} \quad (194)$$

The importance sampling has several consequences. If $\Psi_T(\bar{R})$ were equal to $\Psi(\bar{R})$ there would be no sign problem, and of course, the DMC procedure would be obsolete. This is never the case and it is easy to see that the effect of the drift velocity, $\bar{v}_D(\bar{R})$ (Eq. (192)), is to drive the walkers away from the nodal surface toward regions where the weight of the wave function is large. In addition, there is a modification of the branching term (183), where now instead of the potential energy and the offset E_T difference a difference between the local energy and the offset E_T appears. For a good trial wave function E_L is close to the ground-state energy and roughly constant which means that the fluctuations are largely reduced. Combination of these two factors has a dramatic improvement on the efficiency of the DMC simulations.

In summary the DMC algorithm with importance sampling consists of the following basic steps:

- 1) Generate a set of walkers drawn from $|\Psi_T|^2$ obtained from VMC.
- 2) Evaluate the drift velocity $\bar{v}_D(\bar{R})$ for each walker.
- 3) Propagate each walker for a time step τ , moving it from old \bar{R}' to the new position

$$\bar{R} = \bar{R}' + \chi + \tau\bar{v}_D(\bar{R}'), \quad (195)$$

where χ is a $3N_e$ -dimensional random number vector with normal distribution with variance τ and zero mean.

- 4) Check the sign of Ψ_T and move it back if Ψ_T changed sign.

5) Accept the step with probability⁹

$$p_{accept}(\bar{R} \leftarrow \bar{R}') = \min \left[1, \frac{G_d(\bar{R}' \leftarrow \bar{R}) |\Psi_T(\bar{R})|^2}{G_d(\bar{R} \leftarrow \bar{R}') |\Psi_T(\bar{R}')|^2} \right] \quad (196)$$

6) For each walker calculate the number of copies

$$M_{new} = INT \{ \eta + \exp[-\tau (E_L(\bar{R}) + E_L(\bar{R}') - 2E_T) / 2] \} \quad (197)$$

that will continue in the evolution, where $\eta \in \langle 0, 1 \rangle$ is a random number with uniform distribution.

7) Accumulate the quantities of interest, e.g. E_L .

8) After an initial equilibration, repeat steps 2)–7) until the error bars for averages of interest are sufficiently small. Adjust E_T occasionally during the simulation to keep the average number of walkers roughly constant.

The energy can be estimated either by the offset E_T or by the so-called mixed-estimator

$$\begin{aligned} E_{DMC} &= \lim_{\tau \rightarrow \infty} \frac{\langle e^{-\tau \hat{H}/2} \Psi_T | \hat{H} | e^{-\tau \hat{H}/2} \Psi_T \rangle}{\langle e^{-\tau \hat{H}/2} \Psi_T | e^{-\tau \hat{H}/2} \Psi_T \rangle} = \lim_{\tau \rightarrow \infty} \frac{\langle e^{-\tau \hat{H}} \Psi_T | \hat{H} | \Psi_T \rangle}{\langle e^{-\tau \hat{H}} \Psi_T | \Psi_T \rangle} \\ &= \frac{\langle \Psi_0 | \hat{H} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle} = \lim_{\tau \rightarrow \infty} \frac{\int f(\bar{R}, \tau) E_L(\bar{R}) d\bar{R}}{\int f(\bar{R}, \tau) d\bar{R}} \approx \frac{1}{M} \sum_m E_L(\bar{R}_m). \end{aligned} \quad (198)$$

Examples of the accuracy in total energy calculation achievable using QMC techniques were demonstrated and compared with DFT and HF results in standard approximations above (Fig. 12) for Si_n clusters and small hydrocarbon molecules. The superiority is of the QMC over mean-field methods is evident. Similar superiority is found also for more complex systems, see, e.g. Ref. 88 and chapter 4.7.1 below.

3.1.6.4 Excited states

It is well-known that many-body correlations have a significant effect on excitations and gaps. Although QMC methods are inherently ground-state techniques, they can also provide information on excited states. The lowest band gap, measurable by photoemission experiments can be formulated in terms of a difference between ground-state energies as

$$E_G = (E_{N_e+1} - E_{N_e}) - (E_{N_e} - E_{N_e-1}), \quad (199)$$

⁹The need for this rejection step results from the approximate nature of the Green's function (194). The exact Green's function satisfies the detailed balance condition

$$\tilde{G}_{exact}(\bar{R} \leftarrow \bar{R}', \tau) |\Psi_T(\bar{R}')|^2 = \tilde{G}_{exact}(\bar{R}' \leftarrow \bar{R}, \tau) |\Psi_T(\bar{R})|^2.$$

If the trial wavefunction were equal to the exact ground-state wavefunction, the detailed balance could be imposed by accepting trial moves $\bar{R}' \rightarrow \bar{R}$ with probability p_{accept} (196) [87].

Tab. V. Excitation energies, in eV, for nitrogen atomic solid (I2₁3 structure) [102] and diamond [103] compared with experiment.

	HF	LDA	DMC	Exp.
Nitrogen $\Gamma \rightarrow H$	18.0	6.1	8.5(4)	N/A
Diamond $\Gamma \rightarrow X$	13.2	4.6	5.9(4)	6.1

i.e. in terms of differences between the energies to add and subtract an electron from the N_e -electron system. One complication with this approach is that by this procedure we create a charged system, which, in the case of an extended system, causes spurious long-range image interactions. The standard procedure is to use a charged background to maintain charge neutrality. This background interacts with all electrons, which creates an artificial contribution to the energy. Fortunately, this contribution is exactly known. This procedure was for the first time used by Ceperley and Alder, who calculated the minimum energy zone center gap of the molecular hydrogen crystal as a function of pressure [100].

Another straightforward way of obtaining excitation energies is to devise a many-electron wave function by promotion of the valence electron to conduction band. This creates an exciton, i.e. an electron-hole pair. If we suppose that the exciton is weakly localized (so-called Mott-Wannier type), then its energy is given by E_G minus the exciton binding energy. If the symmetries of the excitonic and ground state are different, the gap can be easily calculated by QMC methods. If this is not the case, more complicated procedures need to be used [101]. Unless the simulation cell is very large, the exciton is artificially localized and the energy dependent on the simulation cell size. The Mott-Wannier exciton energy can be expressed as

$$E_{exc} = \frac{1}{2\epsilon r_0} \quad (200)$$

where ϵ is the static dielectric constant and r_0 average distance between the hole and electron, which is roughly equal to the edge length of the simulation cell. We note that the exciton binding energy in large gap insulators is relatively small, of the order of ≈ 0.1 eV or smaller. As an example of the accuracy achievable we give in Table V examples for the gaps in nitrogen and diamond.

As can be seen, the accuracy of QMC techniques for both ground state as well as excited states is comparable to the best quantum chemistry methods. The question is how the QMC methods compare with the best quantum chemistry methods in terms of numerical cost. It turns out that the scaling with the number of electrons is of the form $O(N^3 + \epsilon N^4)$ with $\epsilon \approx 10^{-3} - 10^{-4}$, hence it has a dominantly cubic scaling, similarly to the mean-field techniques. This attractive scaling, compared with accuracy similar to the best quantum chemistry methods, make the QMC methods a very strong candidate for extreme accuracy, large-scale calculations.

3.1.7 Basis sets, pseudopotentials, extended systems and ionic forces

This section will deal with several important ingredients linked to the electronic subsystem. These are of paramount importance if we want to solve the coupled Eqs. (11), (12) or (15),

(16) in any of the approximations mentioned above.

3.1.7.1 Basis sets

The single particle orbitals $\{\varphi_i\}$ are usually represented in terms of simple analytic functions $\{\chi_\nu\}$ with well-known properties. In general a linear combination of such basis functions

$$\varphi_i = \sum_{\nu=1}^M c_{i\nu} \chi_\nu(\vec{r}) \quad (201)$$

is used, which represents exactly any reasonable function in the limit of complete set of basis functions. The basis functions $\{\chi_\nu\}$ may or may not be dependent on positions of atoms. This ansatz, which is quite general, was for the first time used by Hall [15] and Roothahn [16] in connection with the solutions to the HF equation.

3.1.7.1.1 Gaussians and Slater functions

Typically, basis functions are constructed to mimic true atomic orbitals. The hydrogen atom can be described rigorously, and the eigenfunctions found. The $1s$ orbital looks something like $N \exp[-\zeta r]$. It satisfies all the appropriate boundary conditions, having a cusp at the nucleus and exponentially decaying to zero at infinity. Higher angular momentum functions, like $2p$'s and $3p$'s, can be built from these basic functions through adding the angular nodes by multiplying in factors of x , y , and z . This type of basis functions, called *Slater-type basis* (STOs) is popular in quantum chemistry

$$\chi_{\nu m}^S(\vec{r}) = N_m^S x^{m_x} y^{m_y} z^{m_z} \exp[-\zeta_m |\vec{r}|]. \quad (202)$$

If instead of the exponential $N \exp[-\zeta r]$ we use a Gaussian function $N \exp[-\alpha r^2]$, we loose the boundary conditions but generate a more tractable problem when it comes to calculating integrals (see, e.g. chapter 3.1.1.). The *Gaussian basis functions* (GTOs) of the form

$$\chi_{\nu m}^G(\vec{r}) = N_m^G x^{m_x} y^{m_y} z^{m_z} \exp[-\alpha_m r^2] \quad (203)$$

have received widespread use. Here N_m , ζ_m , α_m are constants that are typically kept fixed during electronic structure calculation, so that only the orbital expansion coefficients $\{c_{i\nu}\}$ need to be optimized. In addition, fixed linear combinations of the above “primitive” basis functions can be used for a given angular momentum channel m , which defines the “contracted” basis sets

$$\chi_\mu^{CG} = \sum_{\nu} c_\nu \chi_\nu^G(\vec{r}). \quad (204)$$

The Slater and Gaussian basis functions are in general centered at the positions of the nuclei, i.e. $\vec{r} \rightarrow \vec{r} - \vec{R}_I$ in Eqs. (202) and (203), leading to linear combination of atomic orbitals (LCAO) ansatz. This ansatz, which is intuitively very appealing, leads to several complications. Firstly, as we will see below in the chapter 3.1.7.4 on ionic forces, localized basis sets coupled to mobile ions generate additional force contribution, so-called *Pullay force*, which needs to be

evaluated. Secondly, use of this *basis set* leads to *superposition error* (BSSE) [104]. This error arises if binding/dissociation energies are computed, because the basis functions localized on one dissociation product affect the energies on the other one. For that reason, the energies of both dissociation products are often computed with all basis functions in position before dissociation.

3.1.7.1.2 Plane waves

A very different approach has its roots in solid-state theory. The periodicity of the underlying lattice produces a periodic potential and imposes the same periodicity on the density. Hence, in this case the natural basis set is *plane waves* as the generic basis set to expand the periodic part of the orbitals. The plane wave basis set is defined as

$$\chi_{\vec{G}}^{PW}(\vec{r}) = N \exp \left[i\vec{G} \cdot \vec{r} \right], \quad (205)$$

where the normalization $N = 1/\sqrt{\Omega}$ with Ω the volume of the periodic supercell and \vec{G} the reciprocal lattice vector. Since plane waves form a complete and orthogonal basis set, they can be used in expansion (201) to expand the orbitals in terms of plane waves. The convergence in terms of plane waves is very simply controlled by a single number, the plane wave cut-off by including the G-vectors up to G_{\max} satisfying: $E_{cut} = G_{\max}^2/2$ (a.u.). Moreover, planewaves are originless, i.e. they don't depend on the positions of the nuclei $\{\vec{R}_I\}$. This, as we show in chapter 3.1.7.4 below, facilitates calculation of ionic forces, and other properties. This last feature is also important in the sense, that they do not favor any part of the space and they are capable of forming any basis function that is needed for system description, provided that the plane wave cutoff E_{cut} is chosen large enough.

As always, there are also weak points coupled with the use of plane waves. The flexibility of the plane waves to form any basis function means that the plane waves are merely a Fourier transform decomposition of the orbitals. If the orbitals exhibit rapid changes on short distances, very large G_{\max} are required, making the calculation expensive. Large G_{\max} may cause instabilities, as G often enters in the denominator in the energy expression (such as in the Hartree term) making the calculation prone to divergent behavior. This is, for example the case, if we perform an "all electron" calculation. Each single-particle wavefunction has to orthogonalize to all preceding ones, which causes their rapid oscillations. Moreover, the inner electrons are more localized than the valence electrons. Similar situation appears also in the case, when we have a large unit cell in all or just in one dimension. The plane waves are useful mainly in connection with pseudopotentials, see chapter 3.1.7.2, which screen out the inner electrons and do away with the wavefunction oscillations due to the orthogonality constraints to the inner electrons. An additional complication arises when finite systems are treated, such as clusters or molecules, or systems with broken 3-dimensional periodicity, such as surfaces. In these situations plane waves may still be used in combination with an artificial unit cell sufficiently large to preclude image interactions. For example, a cluster may be placed in a large supercell with sufficiently large vacuum region surrounded the molecule and the convergence of the results monitored as a function of the supercell edge. In such a case, effectively an array of clusters is calculated. Examples of use of the supercell geometry for treating a variety of systems lacking 3-d periodicity are shown

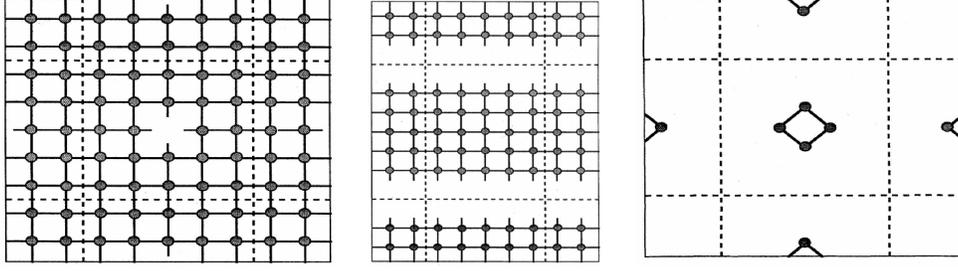


Fig. 17. Schematic illustration of a supercell geometry for a vacancy in a bulk crystalline solid (left panel), for a surface (middle panel) and for an isolated molecule (right panel). The edges of the supercell are shown by dashed lines.

in Fig. 17. The down side is the need to fill equally all the space with plane waves, i.e. also the vacuum space, which may significantly increase the associated computational cost.

3.1.7.1.3 Generalized plane waves

A very elegant generalization of the plane wave concept, which eliminates some of the limitations, was suggested by defining them in a curved ξ -space [105, 106],

$$\chi_{\vec{G}}^{GPW}(\vec{r}) = N \det^{1/2} J \exp \left[i \vec{G} \cdot \vec{r}(\xi) \right], \quad (206)$$

$$\det J = \left| \frac{\partial r^i}{\partial \xi^j} \right|,$$

where J is the Jacobian of the transformation from Cartesian to *curvilinear coordinates* $\vec{r} \rightarrow \xi(\vec{r})$ with $\xi = (\xi^1, \xi^2, \xi^3)$ and $N = 1/\sqrt{\Omega}$. These functions are orthonormal and can be manipulated by FFT techniques. The important feature is that the curvilinear coordinate system needn't be guessed from outset but is rather treated as a variational parameter and determined by a fully adaptive-coordinate approach [105, 106]. The attractive feature is that an initially flat Euclidean space ($\xi(\vec{r}) = \vec{r}$) is transformed into a non-uniform curved Riemannian space such that the density of grid points is highest in the spaces where it is needed, e.g. close to the nuclei, and lowest where it is least needed, e.g. in the vacuum region, see Fig. 18.

As can be seen, this technique eliminates one of the unpleasant features of plane waves, namely the need to fill equally the space with plane waves regardless of the real needs. The flip-side of the coin is that the curvilinear coordinates induce occurrence of the Pulay forces. Besides the formal beauty, the method of generalized plane waves has found also real applications [107, 108].

3.1.7.1.4 Mixed Basis Sets

This approach attempts to merge the advantages of two very different basis sets, namely Gaussian basis sets and plane waves. It should be clear from the discussion above what the merits of the combined approaches could be. More details can be found in Refs. [109–111].

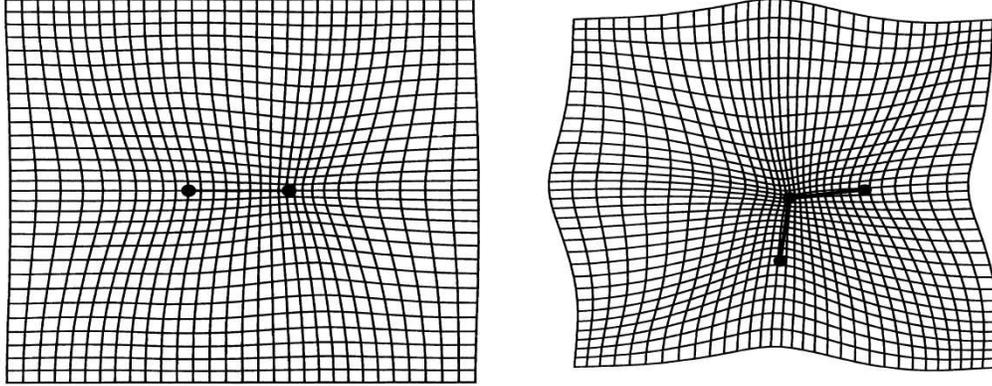


Fig. 18. Image of a regularly spaced rectangular space grid in $\xi(\vec{r})$ space [106]. The left panel shows the grid in the plane of CO bond; the right panel shows the grid in a plane of a H₂O molecule.

3.1.7.1.5 Real Space Grids

A radically different approach consists of giving up the conventional basis set altogether and representing the wavefunction $\psi(\vec{r})$ on a finite grid with a certain *uniform grid* spacing h , $\psi(\vec{r}) \rightarrow \psi(x_i, y_j, z_k)$ [112,113]. This is very attractive, compared to plane waves especially when dealing with finite systems. On the other hand, the kinetic energy operator, trivial in plane wave and Gaussian representation, has to be discretized. High-order central differences are typically used, leading to

$$\begin{aligned}
 & -\frac{1}{2}\nabla^2\psi(\vec{r}) = \\
 & -\frac{1}{2}\left[\sum_{i=-N}^N C_i\psi(x+ih, y, z) + \sum_{j=-N}^N C_j\psi(x, y+jh, z) + \sum_{k=-N}^N C_k\psi(x, y, z+kh)\right] \\
 & + O(h^{2N+2}). \tag{207}
 \end{aligned}$$

Here $\{C_i\}$ are known expansion coefficients that depend on the selected expansion order. The combination of the grid spacing h and expansion order can be tuned for any given application. The discretization points in continuous space can be regarded as a “special” basis set. The “infinite basis set” is reached as $h \rightarrow 0$ for N fixed. A variety of other more sophisticated implementations followed [114, 115].

Note that a special real-space basis is the walkers in the QMC. The $\delta(\vec{R} - \vec{R}')$ can be regarded as a form of δ -function basis set. This representation is probably the most efficient one to represent a wavefunction albeit not the most convenient to work with.

3.1.7.2 Pseudopotentials

There are a number of reasons not to have to deal with all electrons in the system. It has been found that most physical and chemical properties of condensed matter systems depend,

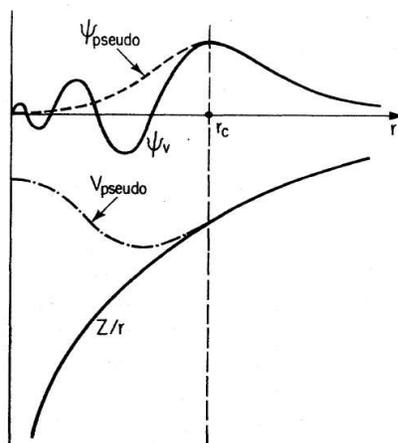


Fig. 19. Schematic illustration of an all-electron radial wave function (solid line) and the corresponding atomic pseudo wavefunction (dashed line) together with the respective external Coulomb potential and pseudopotential.

to a very good approximation, only on valence electrons. The core electrons, strongly localized around the nucleus, neither overlap strongly with core electrons in neighboring atoms, nor do they participate in chemical bond. One can, therefore assume, that the distribution of core electrons does not depend on the chemical environment in which it is placed. This leads to the concept of “frozen” core kept equal to the isolated atom core also in a condensed matter environment.

This has a number of consequences. First, we have to deal with fewer electrons. Moreover, instead of keeping the core electrons frozen, one can screen them out by use of an appropriate *pseudopotential*. This has the added advantage, that the resulting wavefunctions are smoother, as the valence wavefunctions do not have to orthogonalize to the core electrons, a feature which is important especially in connection with plane wave basis sets. Finally, it is easier to deal with fewer electrons also because we deal with smaller numbers. Many properties are related to energy differences, which, in general are very small; often of the order of correlation energy. Hence, reducing the magnitude of the numbers entering the differences pays dividends. The downside being, that absolute energies do not have any meaning, leaving just the relative energies meaningful quantities. Surprisingly, until very recently, the use of pseudopotentials was almost entirely limited to the physics community and largely ignored by the quantum chemistry community.

3.1.7.2.1 Norm-conserving pseudopotentials

The first pseudopotentials were of empirical nature, fitted to some property deemed important [116]. More accurate and reliable scheme was developed later and was based on *norm-conservation* [117]. The norm-conserving pseudopotentials are constructed as follows. (1) First, Schrödinger’s equation is solved for an isolated atom. A typical result for a radial function is shown in Fig. 19. The angular part is the spherical harmonics, due to the spherical symmetry of

the atoms. (2) A cutoff radius R_{cut} is chosen so that the effective all-electron potential within R_{cut} is replaced by a new weaker potential- pseudopotential- which generates a nodeless ground-state wavefunction to the same eigenvalue as the original all-electron state which matches exactly the all-electron wavefunction at and outside R_{cut} (dashed line in Fig. 19).

This is possible, because the physical solution $\phi_l(r)$ is regular for $r \rightarrow 0$. The logarithmic derivative

$$L_l(\varepsilon) = \frac{d}{dr} \ln \phi_l(r, \varepsilon)|_{R_{cut}} = \frac{\phi_l'(R_{cut}; \varepsilon)}{\phi_l(R_{cut}; \varepsilon)} \quad (208)$$

is therefore a well-defined function of the energy ε . On the other hand, for a given energy ε and logarithmic derivative L_l at R_{cut} the solution of the radial Schrödinger equation is uniquely defined. From this follows that if the potential inside R_{cut} is modified in such a way that L_l is left unchanged, the wavefunction outside R_{cut} remains unchanged. In order this procedure to work correctly one has to maintain the norm conservation, i.e., that the all-electron wavefunction and the pseudo-wavefunction have equal norms inside R_{cut} . This latter condition guarantees that the all-electron and pseudo-wavefunctions generate identical electron densities inside R_{cut} . In fact, there is a fair degree of freedom how to devise the pseudowavefunction following the above principles and over the last decades a number of different procedures have been proposed [118–121].

Since the logarithmic derivative (208) depends on angular momentum l , we have to construct separate pseudopotential $V_l^{PS}(r)$ for each l -channel. This means that the full pseudopotential has to be a non-local operator

$$\hat{V}^{PS} = V_{loc}^{PS}(r) + \sum_l V_{nl,l}^{PS}(r) \hat{P}_l, \quad (209)$$

where $\hat{P}_l = |l\rangle \langle l|$ is the projector on l -th momentum component, which by acting on a wave function projects out the just its l -th component. This guarantees that when a full pseudopotential operator \hat{V}^{PS} is applied to a general wavefunction each angular momentum component of the wavefunction experiences only its corresponding part $V_l^{PS}(r)$. In practical application, the infinite sum over l has to be truncated and the maximum value of l limited by some l_{max} , the choice being largely system dependent. The pseudopotential of some specific angular momentum, typically the one of l_{max} , is taken to be the local part $V_{loc}^{PS}(r)$. The nonlocal components $V_{nl,l}^{PS}(r)$ are defined as $V_{nl,l}^{PS}(r) = V_l^{PS}(r) - V_{loc}^{PS}(r)$. Since all $V_l^{PS}(r)$ are identical outside R_{cut} , the nonlocal components $V_{nl,l}^{PS}(r)$ are short-ranged and strictly confined within R_{cut} . Note, that the projection operators \hat{P}_l in Eq. (209) act only on the angular variables of \vec{r} and the pseudopotential $V^{PS}(r)$ is a local operator with respect to the radius r . Hence, the pseudopotential in the form of (209) is called a semilocal pseudopotential. The semilocal form (209) of the pseudopotential is computationally very inefficient. One way to efficiently compute the matrix elements over the semilocal pseudopotentials was proposed in Ref. [122]. A more radical solution was proposed by Kleinman and Bylander [123], who managed to cast the pseudopotential in a fully separable form

$$V^{PS}(r) = V_{loc}^{PS} + \sum_{mn} |p_m(\vec{r})\rangle C_{mn} \langle p_n(\vec{r})|, \quad (210)$$

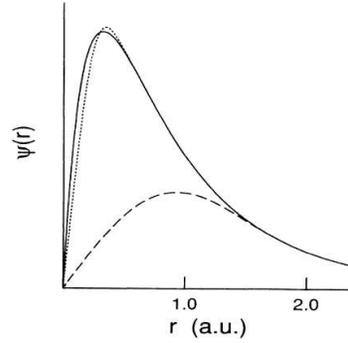


Fig. 20. Illustration of Vanderbilt ultrasoft pseudopotential. Full line: Oxygen all-electron $2p$ radial wave function, dotted line: corresponding “ordinary” valence pseudowave function, dashed line: modified wavefunction in the Vanderbilt pseudopotential scheme.

where $p_m(\vec{r})$ are suitably chosen projection functions strictly localized in R_{cut} . As can be seen, the bits depending on \vec{r}, \vec{r}' are totally decoupled and can be computed independently.

3.1.7.2.2 Vanderbilt ultrasoft pseudopotentials

It turns out that the pseudopotential scheme is problematic for elements with nodeless valence states, such as, e.g. $1s$, $2p$ or $3d$ valence electrons. It may seem useless to use pseudopotential for hydrogen (the $1s$ case), as there are no core electrons to screen out. Yet, if the purpose is to produce a pseudopotential generating a smoother wavefunction, as is the case with plane wave basis sets, the pseudopotential concept is an indispensable tool also in this case. Unfortunately, in those cases the pseudo- and all-electron wavefunctions are almost identical. This can be rationalized by invoking the so-called pseudopotential cancellation theorem [124], which states that in order to generate a “soft” pseudopotential for a given state, there must be a state with the same angular momentum number in the atomic core.

A way out of this problem was found by Vanderbilt, who introduced a new class of pseudopotentials, the so-called *ultrasoft pseudopotentials* [125, 126]. The ultrasoft pseudopotentials give up the normconservation and instead of representing the full valence wavefunction by plane waves, only a small part of the wavefunction is treated by plane waves; see, Fig. 20. This allows a substantial reduction of the size of the plane wave basis set. On the other hand, there are additional complications in dealing with the part of the wavefunction left out, the fact that the eigenstates $\{\varphi_i\}$ are not orthonormal any more, etc. Nevertheless, the Vanderbilt pseudopotentials have been widely used as they allow treatment of a wide range of systems with fairly modest computational cost.

3.1.7.3 Extended systems

If the system has a natural 3-dimensional periodicity, as is the case with the crystals or is treated as 3-dimensionally periodic, as is the case with defects, surfaces, or even disordered systems (see also chapters 3.1.7.1.2 and 4.1), one has to deal with the crystal or supercell symmetry,

and the concept of periodicity in the reciprocal space [127]. In particular, the crystal potential of a periodic system exhibits the same periodicity as the lattice

$$V_{cryst}(\vec{r}) = V_{cryst}(\vec{r} + \vec{L}), \quad (211)$$

with an accordingly periodic wavefunction

$$\psi(\vec{r}) = \psi(\vec{r} + \vec{L}), \quad (212)$$

where the direct lattice vectors \vec{L} connect equivalent points in different cells. The wavefunctions are of the Bloch form¹⁰ [127]

$$\psi(\vec{r}) = \psi_{i,\vec{k}}(\vec{r}) = u_{i,\vec{k}}(\vec{r}) \exp[i\vec{k} \cdot \vec{r}], \quad (213)$$

where \vec{k} is a vector in the first Brillouin zone, The functions $u_{i,\vec{k}}(\vec{r})$ have the periodicity of the direct lattice

$$u_{i,\vec{k}}(\vec{r}) = u_{i,\vec{k}}(\vec{r} + \vec{L}). \quad (214)$$

The index i runs over all (single-particle) states. The states have an occupation number $f_{i,\vec{k}}(\beta)$ (see chapter 3.1.5.9. and Eq. (141)). Due to dispersion of the single-particle states into bands in a periodic system, in all energy expressions above (e.g., Eqs. (32), (101)) the summation over the orbitals of the type \sum_i has to be substituted by a summation over orbitals and integration over the Brillouin zone: $\sum_i \int_{BZ} d\vec{k}$. As an example we give below a reformulation for extended system of the KS expression for energy functional (Eqs. 100–101)

$$E^{KS} = \sum_i \int_{BZ} d\vec{k} f_{i,\vec{k}} \varepsilon_{i,\vec{k}} - \frac{1}{2} \int d\vec{r} V_H(\vec{r}) n(\vec{r}) + E_{xc}[n] - \int d\vec{r} V_{xc}(\vec{r}) n(\vec{r}), \quad (215)$$

$$n(\vec{r}) = \sum_i \int_{BZ} d\vec{k} |\varphi_{i,\vec{k}}|^2. \quad (216)$$

We have to approximate the integral over the Brillouin zone by a finite sum over so-called “special” \vec{k} -points [128–130]

$$\int_{BZ} d\vec{k} \approx \sum_{\vec{k}} w_{\vec{k}}, \quad (217)$$

where $w_{\vec{k}}$ are weights of the integration points. The idea of approximating the integral by a sum of integrand values computed over a set of carefully chosen points in the integration domain is akin to what is done in the usual Gaussian quadrature.

¹⁰Bloch’s theorem is one of the cornerstones of the independent-particle electronic structure theory of solids. However, the reduction of the problem to one single primitive unit cell imposed by the theorem, is possible only in the independent-particle theories; in many-electron theories, such as QMC, it cannot directly be applied. For more details, see for instance Ref. [88].

There is a technical problem with computation of electrostatic energy of an infinite system. Electrostatic energy consists of three parts: Hartree energy of the electrons with density $n(\vec{r})$, interaction of electrons with the nuclei and the internuclear interaction

$$E_{ES} = \frac{1}{2} \int \int d\vec{r} d\vec{r}' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} + \sum_I \int d\vec{r} V_I^{core} n(\vec{r}) + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|}. \quad (218)$$

The singularities in individual terms in (218) for an infinite system can be avoided by the so-called Ewald summation method [1]. This is done by smearing each ionic nuclear charge, Z_I , into a Gaussian distribution centered at \vec{R}_I

$$n_c^I(\vec{r}) = -\frac{Z_I}{(R_I^c)^3} \pi^{-3/2} \exp \left[-\left(\frac{\vec{r} - \vec{R}_I}{R_I^c} \right)^2 \right]. \quad (219)$$

It is convenient to make use of the arbitrariness in the definition of the core potential, and define it to be the potential of the Gaussian charge distribution (219),

$$V_I^{core}(\vec{r}) = \int d\vec{r}' \frac{n_c^I(\vec{r}')}{|\vec{r} - \vec{r}'|} = -\frac{Z_I}{|\vec{r} - \vec{R}_I|} \operatorname{erf} \left[\frac{|\vec{r} - \vec{R}_I|}{R_I^c} \right], \quad (220)$$

where “erf” is the error function. The energy expression (218) can be rewritten by adding and subtracting the interaction energy of the Gaussian charge distribution as follows

$$\begin{aligned} E_{ES} &= \frac{1}{2} \int \int d\vec{r} d\vec{r}' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{1}{2} \int \int d\vec{r} d\vec{r}' \frac{n_c(\vec{r})n_c(\vec{r}')}{|\vec{r} - \vec{r}'|} + \int \int d\vec{r} d\vec{r}' \frac{n_c(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} \\ &+ \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} - \frac{1}{2} \int \int d\vec{r} d\vec{r}' \frac{n_c(\vec{r})n_c(\vec{r}')}{|\vec{r} - \vec{r}'|}, \end{aligned} \quad (221)$$

where $n_c(\vec{r}) = \sum_I n_c^I(\vec{r})$. The first three terms can be combined to electrostatic energy of the total charge distribution $n_{tot}(\vec{r}) = n(\vec{r}) + n_c(\vec{r})$. The remaining terms can be rewritten as a double sum over nuclei and a sum over self-energy terms of the Gaussian charge distributions

$$\begin{aligned} E_{ES} &= \frac{1}{2} \int \int d\vec{r} d\vec{r}' \frac{n_{tot}(\vec{r})n_{tot}(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J|} \operatorname{erfc} \left[\frac{|\vec{R}_I - \vec{R}_J|}{\sqrt{R_I^c{}^2 + R_J^c{}^2}} \right] \\ &- \sum_I \frac{1}{\sqrt{2\pi}} \frac{Z_I^2}{R_I^c}, \end{aligned} \quad (222)$$

where erfc denotes the complementary error function.

Now, for a periodic system, the total electrostatic energy is computed from the above expression. The first term is evaluated in reciprocal (Fourier) space,

$$\begin{aligned} n_{tot}(\vec{G}) &= n(\vec{G}) + \sum_I n_c^I(\vec{G}) S_I(\vec{G}) \\ &= n(\vec{G}) - \frac{1}{\Omega} \sum_I \frac{Z_I}{\sqrt{4\pi}} \exp \left[-\frac{1}{2} G^2 R_I^c{}^2 \right] S_I(\vec{G}), \end{aligned} \quad (223)$$

where $S_I(\vec{G}) = \exp[-i\vec{G} \cdot \vec{R}_I]$ is the structure factor. This yields the electrostatic energy of a periodic system in the form

$$E_{ES} = 2\pi\Omega \sum_{\vec{G} \neq 0} \frac{|n_{tot}(\vec{G})|^2}{G^2} + E_{ovrl} - E_{self}, \quad (224)$$

$$E_{ovrl} = \sum_{I,J}' \sum_{\vec{L}} \frac{Z_I Z_J}{|\vec{R}_I - \vec{R}_J - \vec{L}|} \operatorname{erfc} \left[\frac{|\vec{R}_I - \vec{R}_J - \vec{L}|}{\sqrt{R_I^{c^2} + R_J^{c^2}}} \right] \quad (225)$$

and

$$E_{self} = \sum_I \frac{1}{\sqrt{2\pi}} \frac{Z_I^2}{R_I^c}. \quad (226)$$

The sums in Eq. (225) run over all atoms in the cell, all direct space lattice vectors \vec{L} , and the prime in the first sum indicates that $I < J$ is imposed for $\vec{L} = 0$. As can be seen, the factorization of the electrostatic energy in (224–226) is done so, that both reciprocal and direct space sums are rapidly convergent.

3.1.7.4 Ionic forces

As explained in chapter 2, *forces exerted on the ions* are indispensable ingredient in any system dynamics. In most cases, we do not solve the wave equation for the ions and, instead, treat ions as classical objects moving on a PES derived by electronic structure methods of different sophistication. In other words, in order to solve Eqs. (9), (12), (17), we need to evaluate the derivatives

$$\vec{F}_I = -\nabla_I \langle \psi_0 | \hat{H}_e | \psi_0 \rangle. \quad (227)$$

In principle, this can be done either by finite-difference approximation of the total energy or analytically. Let us follow first the latter possibility.

Analytic evaluation of the derivative yields

$$\nabla_I \langle \psi_0 | \hat{H}_e | \psi_0 \rangle = \langle \psi_0 | \nabla_I \hat{H}_e | \psi_0 \rangle + \langle \nabla_I \psi_0 | \hat{H}_e | \psi_0 \rangle + \langle \psi_0 | \hat{H}_e | \nabla_I \psi_0 \rangle. \quad (228)$$

The contributions from derivatives of the wavefunction $\langle \nabla_I \psi_0 | \hat{H}_e | \psi_0 \rangle$ vanish if the wavefunction is an exact eigenfunction of the Hamiltonian under consideration, and the force can be computed as

$$\vec{F}_I^{HF} = -\langle \psi_0 | \nabla_I \hat{H}_e | \psi_0 \rangle, \quad (229)$$

which corresponds to the *Hellmann-Feynman theorem* [131, 132]. The theorem is valid for variational wavefunctions, such as DFT, HF, provided a complete basis sets are used. This is not the case if position-dependent basis functions are used, and the single-particle states $\varphi_i = \sum_{\nu} c_{i\nu} \chi_{\nu}(\vec{r}; \{R_I\})$ (see chapter 3.1.7.1 above). The explicit dependence of the basis functions

as well as the implicit dependence of the expansion coefficients on atomic positions generates two additional contributions to the forces

$$\nabla_I \varphi_i = \sum_{\nu} (\nabla_I c_{i\nu}) \chi_{\nu}(\vec{r}; \{R_I\}) + \sum_{\nu} c_{i\nu} (\nabla_I \chi_{\nu}(\vec{r}; \{R_I\})). \quad (230)$$

In other words, the complete (generalized) force acting on a nucleus at \vec{R}_I can be obtained by taking a total derivative, of the energy [133]

$$\vec{F}_I = \vec{F}_I^{HF} + \vec{F}_I^{IBS} + \vec{F}_I^{NSC}, \quad (231)$$

where \vec{F}_I^{IBS} is the correction due to the use of incomplete basis set, or the (in)famous *Pullay force* [134]

$$\vec{F}_I^{IBS} = -2 \sum_i f_i \text{Re} \left\{ \sum_{\nu} c_{i\nu} \langle \nabla_I \chi_{\nu} | \hat{H}_e^{NSC} - \varepsilon_i | \psi_i \rangle \right\} \quad (232)$$

and \vec{F}_I^{NSC} is the correction due to non-selfconsistent charge densities

$$\vec{F}_I^{NSC} = - \int d\vec{r} (\nabla_I n) (V^{SCF} - V^{NSC}). \quad (233)$$

Assuming that self-consistency is achieved, which never happens in a numerical calculation, \vec{F}_I^{NSC} vanishes and \vec{F}_I^{IBS} has to be computed with \hat{H}^{SCF} . Similarly, the Pullay force vanishes in the limit of a complete basis set. The additional contributions to the force, \vec{F}_I^{IBS} and \vec{F}_I^{NSC} , complicate the force evaluation. The Pullay force can be eliminated completely by expanding the wavefunctions in terms of originless basis sets, such as plane waves, Eq. (205). This is valid irrespective of the type of dynamics, i.e. both for Ehrenfest as well as Born-Oppenheimer dynamics as long as the number of plane waves remains constant. If the number of plane waves changes, as is the case in constant pressure dynamics (not to be discussed in this paper), where the volume/shape of the simulation cell changes and the plane wave cutoff is kept fixed the *Pullay stress* appears. If instead of plane waves, localized atom-centered basis sets are used, the Pulay forces need to be computed [109, 135]. \vec{F}_I^{NSC} vanishes if self-consistency within the subspace spanned by a given incomplete basis set is reached. Note that this is a weaker requirement than that required by the Hellman-Feynman theorem, which requires that Ψ_0 be an exact eigenstate of the Hamiltonian and a complete basis set to be used.

In addition to the mean-field methods, outlined above, most correlated electronic structure methods, including MCSCF [136, 137], MP [138-140], CISD [141], CCSD(T) [142,143], etc., allow for analytic force calculations. This represents an enormous savings, as instead of $3N_I - 6$ independent total energy calculations required if finite differences are used, only a single force calculation is sufficient if analytic derivatives are used.

Here we mention briefly how forces can be obtained in the QMC method. There are a number of obstacles, which have to be taken into account: (a) unlike DFT and standard quantum chemistry, use of the Hellman-Feynman theorem is not practical as the wavefunction is not obtained by energy minimization, (b) in fixed-node DMC the Hellman-Feynman forces have errors due to discontinuities in the derivative of the fixed-node wave function at nodes. Hence, the forces have to be computed from energy differences. However, QMC methods have statistical errors, which

make computation of energy differences very expensive. Fortunately, one can *use correlated sampling* [144, 145]. Instead of performing two independent runs with independent statistical errors, one generates configurations for a reference situation only sampled from ψ^2 of the reference configuration. Unbiased expectation values for a slightly different secondary wavefunction ψ_s are obtained by reweighting configurations sampled from ψ^2 . In the VMC this means

$$E_s - E = \frac{\langle \Psi_s | H_s | \Psi_s \rangle}{\langle \Psi_s | \Psi_s \rangle} - \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{1}{M} \sum_{i=1}^M \left[\frac{H_s \Psi_s(\bar{R}_i)}{\Psi_s(\bar{R}_i)} W_i - \frac{H \Psi(\bar{R}_i)}{\Psi(\bar{R}_i)} \right]$$

with

$$W_i = \frac{M |\Psi_s(\bar{R}_i)/\Psi(\bar{R}_i)|^2}{\sum_{i=1}^M |\Psi_s(\bar{R}_i)/\Psi(\bar{R}_i)|^2}.$$

However, the walkers sampled from ψ^2 won't be optimal for computing E_s and they need to be moved along with the ions by some amount, $\Delta \vec{r}_i$

$$\vec{r}_{is} = \vec{r}_i + \Delta \vec{r}_i.$$

The same idea can also be used also in the context of DMC where the walker distribution reweighted by a factor

$$W_i \cong \frac{\Psi_T^2(\bar{R}_i)}{\Psi_{Ts}^2(\bar{R}_i)} \exp[-\Delta\tau(E_{Ls} - E_L)/2].$$

is used and subsequently propagated for a short time τ . Obviously, if this is to work, the structural changes need to be very small. Surprisingly, while extremely simple in essence, the QMC ionic forces world-wide have not been extensively used as yet and only a couple of realistic applications have been published, see chapter 4.7.1.3.

3.2 Ionic dynamics

3.2.1 Classical ions

Let us initially assume that we want to solve the system dynamics for classical ions (massive ions at relatively high temperatures) following either the Ehrenfest or Born-Oppenheimer dynamics. Let us further assume that the electronic structure part of the problem can be solved and the ionic forces are known. Hence, all that is needed is the solution of the Eq. (12) or (17)

$$M_I \ddot{\vec{R}}_I(t) = \vec{F}_I(t). \quad (234)$$

These are *Newton's equations of motion* which can easily be discretized and solved; see, for instance, chapter 3.3.1 for one way of discretization. Provided the system is *ergodic*, the system dynamics corresponds to dynamics in microcanonical (N, V, E) ensemble. In this ensemble, in addition to the particle number, simulation cell, and energy, also the total momentum is conserved [1]. From this follows, that system temperature is not a control parameter, and consequently it cannot be fixed.

From a statistical physics point of view a legitimate question arises why we from the outset limit ourselves to MD sampling of the phase space and do not consider Monte-Carlo techniques as well. There are two reasons. Firstly, most of the computational work comes into electronic structure computation of the points on the PES visited. As mentioned above, those are computed “on fly”. If MD scheme is used, the ionic configurations change only marginally from one step to the other. In such a procedure the electronic structure responds also only weakly and most of the numerical work done on the preceding ionic configurations needn’t be redone. One can see this as a reformulation of the perturbation theory. This, though, is not true any more, if one instead uses MC techniques, where by the very nature of the MC procedure the changes in the atomic structure and hence also in the electronic structure need not necessarily be small. The downside of the MD sampling is, that a correct sampling is only achieved if the system is ergodic. The other reason for favoring MD over MC techniques is that the MD procedure automatically yields also information on system dynamics, which gives access to computation of other properties, such as transport properties, entirely from the equilibrium dynamics.

Formally all other ensembles can be obtained from the microcanonical ensemble. Here we limit ourselves to just one more ensemble, the canonical, or (N, V, T) ensemble. The system temperature is related to the total momentum via *equipartition theorem*. Temperature control can be exerted by adding a friction term, which suitably scales the particle velocities. A deterministic algorithm was proposed by Nosé and Hoover [146–148]. The underlying equations of motion read

$$M_I \ddot{\vec{R}}_I(t) = \vec{F}_I(t) - M_I \dot{\xi}(t) \dot{\vec{R}}_I(t), \quad (235)$$

$$Q \ddot{\xi}(t) = \sum_I M_I \dot{\vec{R}}_I^2(t) - 3N_I k_B T, \quad (236)$$

where ξ is an extra variable, which scales the particle velocities, Q is a fictitious mass of the extra variable, k_B Boltzman’s constant, and T the required system temperature. Nosé has shown [146, 147] that a microcanonical dynamics in the extended $3N_I + \xi$ system leads to *canonical dynamics* in the physical system with $3N_I$ degrees of freedom.

The way the thermostat works can intuitively be inferred by inspection of Eq. (235). $\dot{\xi}$ can be regarded as a dynamical friction coefficient, which can acquire both positive and negative value. This leads to either damping or acceleration of the nuclei and hence, via equipartition theorem to cooling if the instantaneous kinetic energy exceeds $k_B T$ or heating otherwise. The velocity scaling also means, that the system dynamics is perturbed in a way dictated by Eq. (236). While it can be explicitly proven that, provided the system is ergodic, the static averages computed along the MD trajectory computed from Eqs. (235), (236), correspond to averages in canonical ensemble, nothing can be explicitly proven concerning the dynamics. As a matter of fact, the ξ -dynamics is just one of many, leading to canonical averaging. However, practical experience has shown, that the effect of the arbitrary coupling of the extra variable ξ to the system dynamics is in most cases negligible and transport coefficients computed in microcanonical and canonical ensemble are identical.

It is well known that the standard Nosé-Hoover thermostat suffers from *non-ergodicity problems* for certain classes of Hamiltonians. The harmonic oscillator and the Fermi-Ulam-Pasta lattice of weakly coupled harmonic oscillators are probably the best examples of non-ergodic

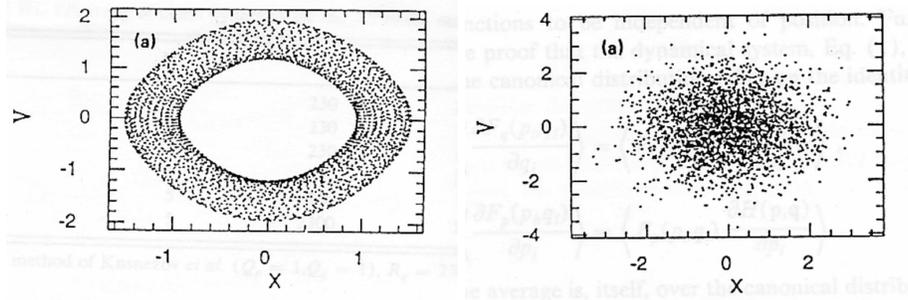


Fig. 21. Coverage of the phase space of a harmonic oscillator [149]. Left: thermostating with a single Nosé-Hoover thermostat; Right: thermostating with a chain of Nosé-Hoover thermostats.

systems [149]. A solution to MD sampling of non-ergodic systems was recently proposed [149]. The solution consists of using a *chain of Nosé-Hoover thermostats*. The equations of motion for the thermostat chain read

$$M_I \ddot{R}_I(t) = \vec{F}_I(t) - M_I \dot{\xi}_1(t) \dot{\vec{R}}_I(t), \quad (237)$$

$$Q_1 \ddot{\xi}_1(t) = \left[\sum_I M_I \dot{R}_I^2(t) - 3N_I k_B T \right] - Q_1 \dot{\xi}_1(t) \dot{\xi}_2(t), \quad (238)$$

$$Q_k \ddot{\xi}_k(t) = \left[Q_{k-1} \dot{\xi}_{k-1}^2(t) - K_B T \right] - Q_k \dot{\xi}_k(t) \dot{\xi}_{k+1}(t) (1 - \delta_{k,K}), \quad k = 2, \dots, K. \quad (239)$$

As can be seen, in the thermostat chain the first thermostat is attached to the physical system, the second thermostat to the first one, and so on. The fictitious mass parameters of the thermostats should be chosen so as to maximize the overlap of their power spectra with the power spectra of the thermostated system,

$$Q_1 \approx \frac{3N_I k_B T}{\omega^2} \quad ; \quad Q_k \approx \frac{k_B T}{\omega^2}, \quad (240)$$

where ω is a typical phonon vibrational frequency of the ions. It has been shown (Fig. 21) that even a fairly short thermostat chain is able to assure ergodic sampling of phase space even for the harmonic oscillator.

3.2.2 Quantum ions

What happens, if the ions cannot be treated as classical objects? Indeed, there are systems where the relation $k_B T \gg \hbar \omega$ does not hold. Imagine dealing with very light ions, such as hydrogen or helium at very low temperatures. In such a case we should go back to the TDSCF equations (5), (6). On practical grounds, this is rarely done. More often, the problem is reformulated so, that the crossdependence of the nuclear and electronic wave functions is eliminated but the quantum nature of the ions retained. The final equations then resemble the BO dynamics Eqs. (16), (17)

but with ions treated quantum-mechanically by solving the time-dependent Schrödinger equation for the ionic system [150]

$$i\hbar \frac{\partial \chi \left(\left\{ \vec{R}_I \right\}; t \right)}{\partial t} = \hat{H}_I \chi \left(\left\{ \vec{R}_I \right\}; t \right) \quad (241)$$

The ionic Hamiltonian $\hat{H}_I = \hat{T}_I + V \left(\left\{ \vec{R}_I \right\} \right)$. There are several problems with this formulation. First, the PES $V \left(\left\{ \vec{R}_I \right\} \right)$ needs to be known. The PES is typically mapped out by direct computation on a grid of points using some of the electronic structure methods outlined in chapter 3.1. For obvious reasons, the dimensionality of the PES which can be explicitly constructed is extremely limited. Similarly is limited also the dimensionality of the problem tractable by direct solution of Eq. (241). Typical upper limit is ~ 5 . Given these limitations, this approach has found applications mainly in molecular scattering and dissociation on surfaces. Here, provided the surface is kept frozen, the dimensionality for a diatomic molecule can be as low as 5–6 (see the example in the Introduction, chapter 2) [151].

3.2.2.1 Path integral approach

A more general approach is provided by a *finite-temperature variant of quantum Monte-Carlo*. The basic quantities allowing computation of statistical averages in quantum statistics is the partition function Z . Let us suppose the system is described by a Hamiltonian \hat{H}_I (temporarily abbreviated to \hat{H}) with wave functions $\{\chi_i\}$ and eigenenergies $\{E_i\}$. The quantum mechanical average values in canonical ensemble (N, V, T) at an inverse temperature $\beta = 1/k_B T$ of any operator \hat{O} can be computed as

$$\langle \hat{O} \rangle = Z^{-1} \sum_i \langle \chi_i | \hat{O} | \chi_i \rangle \exp[-\beta E_i] = Z^{-1} \text{Tr} \left(\hat{O} \exp[-\beta \hat{H}] \right), \quad (242)$$

$$Z = \sum_i \exp[-\beta E_i] = \text{Tr} \left(\exp[-\beta \hat{H}] \right), \quad (243)$$

where in each line the first expression is written in the diagonal energy representation and the second in invariant representation.

Alternatively, one can reformulate Eqs. (242), (243) in terms of *finite-temperature density matrix* ρ . For that purpose we will switch from energy representation to position representation (i.e. eigenfunctions of the position operator \hat{R}_I : $\delta \left(\vec{R} - \vec{R}_I \right)$). We define a composite vector of ionic positions $\mathfrak{R} = \left(\vec{R}_1, \vec{R}_2 \dots, \vec{R}_{N_I} \right)$. The density matrix ρ is defined as

$$\rho \left(\mathfrak{R}, \mathfrak{R}'; \beta \right) = \sum_i \chi_i^* \left(\mathfrak{R} \right) \chi_i \left(\mathfrak{R}' \right) \exp[-\beta E_i] \equiv \langle \mathfrak{R} | \exp[-\beta \hat{H}] | \mathfrak{R}' \rangle. \quad (244)$$

The $\rho \left(\mathfrak{R}, \mathfrak{R}'; \beta \right)$ is a $6N_I + 1$ -dimensional object; the “+1” dimension stands for the temperature dependence. Equations (242), (243) expressed in terms of the density matrix ρ read

$$\langle \hat{O} \rangle = Z^{-1} \int d\mathfrak{R} \int d\mathfrak{R}' \rho \left(\mathfrak{R}, \mathfrak{R}'; \beta \right) \langle \mathfrak{R} | \hat{O} | \mathfrak{R}' \rangle, \quad (245)$$

$$Z = \int d\mathfrak{R} \rho(\mathfrak{R}, \mathfrak{R}'; \beta). \quad (246)$$

The question is, can we compute the quantum-mechanical averages $\langle \hat{O} \rangle$ without having to diagonalize the Hamiltonian \hat{H} ? The answer is positive, and has been given by Feynman in the fifties in *his path-integral (PI) formulation* of quantum mechanics [152, 153]. Here we follow a slightly different line of thoughts. Ceperley published recently an excellent review on the path integral technique [154]. The essence of the path-integral technique is the fact that the product of density matrices (DM) is a density matrix: $\text{DM} \times \text{DM} = \text{DM}$

$$\exp[-\beta_1 \hat{H}] \exp[-\beta_2 \hat{H}] = \exp[-(\beta_1 + \beta_2) \hat{H}], \quad (247)$$

or in position representation

$$\int d\mathfrak{R}_2 \rho(\mathfrak{R}_1, \mathfrak{R}_2; \beta_1) \rho(\mathfrak{R}_2, \mathfrak{R}_3; \beta_2) = \rho(\mathfrak{R}_1, \mathfrak{R}_3; \beta_1 + \beta_2), \quad (248)$$

which shows that a density matrix at a certain temperature, can be written as a convolution of density matrices at different temperatures. Note that in writing Eq. (247) in position representation (248), we have used the identity operator $\int d\mathfrak{R}_2 |\mathfrak{R}_2\rangle \langle \mathfrak{R}_2|$.

3.2.2.2 Discrete formulation of PI: high-temperature expansion

In order to move forward, it is useful to realize one important fact. The density matrix at a very high temperature ($\beta \rightarrow 0$ limit) must behave classically. In other words, if we succeed to reformulate the low-temperature density matrix in terms of a high-temperature matrix (matrices), the high-temperature matrix (matrices) will behave classically and hence, classical statistical mechanics will suffice to study quantum statistics. Let us use the following transformation: $T \Rightarrow MT$, i.e. instead of considering temperature T we consider temperature MT . The density matrix after transformation becomes

$$\exp[-\beta \hat{H}] = \left(\exp\left[-\frac{\beta}{M} \hat{H}\right] \right)^M = \left(\exp[-\tau \hat{H}] \right)^M. \quad (249)$$

The above equation is an identity with the “time” step

$$\tau = \frac{\beta}{M} = \frac{1}{k_B(TM)}, \quad (250)$$

or in position representation

$$\begin{aligned} \rho(\mathfrak{R}_0, \mathfrak{R}_M; \beta) &= \int \int \cdots \int d\mathfrak{R}_1 d\mathfrak{R}_2 \cdots d\mathfrak{R}_{M-1} \rho(\mathfrak{R}_0, \mathfrak{R}_1; \tau) \\ &\quad \times \rho(\mathfrak{R}_1, \mathfrak{R}_2; \tau) \cdots \rho(\mathfrak{R}_{M-1}, \mathfrak{R}_M; \tau). \end{aligned} \quad (251)$$

Equation (251) is a *PI discretization of the DM* $\rho(\mathfrak{R}_0, \mathfrak{R}_M; \beta)$. Note that the discretization coordinate is the temperature. If $M \gg 1$ is finite, points $\mathfrak{R}_1, \mathfrak{R}_2 \cdots \mathfrak{R}_M$ form a discrete “path” over the temperature, if $M \rightarrow \infty$, the points form a continuum path $\{\mathfrak{R}_t\}$; $0 \leq t \leq \beta$. This type

of discretization is also called *imaginary time discretization* [154]. The point is that for $M \gg 1$, i.e. at high temperatures the density matrices $\rho(\mathfrak{R}_{L-1}, \mathfrak{R}_L; \tau)$ behave classically and we will see that a good approximation can be constructed.

3.2.2.3 Primitive approximation

In order to construct an approximation in the $\beta \rightarrow 0$ limit, we need to take into account the fact that in the Hamiltonian $\hat{H} = \hat{T} + \hat{V}$ only the potential energy term \hat{V} is diagonal in the \vec{R} -representation. The kinetic energy operator is instead diagonal in the reciprocal space. Therefore it is reasonable to separate the kinetic and potential energy operators in the density matrix ρ . The problem is that the two operators do not commute and hence the density matrix cannot be straightforwardly factorized

$$\exp\left[-\tau\left(\hat{T} + \hat{V}\right)\right] = \exp\left[-\tau\hat{T}\right] \exp\left[-\tau\hat{V}\right] - \exp\left[\frac{\tau^2}{2}\left[\hat{T}, \hat{V}\right]\right], \quad (252)$$

where $[\hat{X}, \hat{Y}]$ is the commutator between \hat{X} and \hat{Y} operators. The primitive approximation¹¹ consists in neglecting the commutator terms in Eq. (252)

$$\exp\left[-\tau\left(\hat{T} + \hat{V}\right)\right] \approx \exp\left[-\tau\hat{T}\right] \exp\left[-\tau\hat{V}\right] \quad (253)$$

for $\tau \rightarrow 0$. Equation (253) shows that the exact density matrix can be approximated by the product of the density matrices for \hat{T} and \hat{V} alone. One might worry that this will lead to a cumulative error in the $M \rightarrow \infty$ limit due to build up of small errors to a finite error. Fortunately, Trotter has proven that under fairly general assumptions the above formula (253) is actually exact in the $M \rightarrow \infty$ limit [155]

$$\exp\left[-\beta\left(\hat{T} + \hat{V}\right)\right] = \lim_{M \rightarrow \infty} \left[\exp\left(-\tau\hat{T}\right) \exp\left(-\tau\hat{V}\right) \right]^M. \quad (254)$$

Let us now rewrite the primitive approximation in position representation

$$\rho(\mathfrak{R}_0, \mathfrak{R}_1; \tau) \approx \int d\mathfrak{R}_2 \langle \mathfrak{R}_0 | \exp\left[-\tau\hat{T}\right] | \mathfrak{R}_2 \rangle \langle \mathfrak{R}_2 | \exp\left[-\tau\hat{V}\right] | \mathfrak{R}_1 \rangle. \quad (255)$$

The \hat{V} operator is diagonal in the position representation, and hence

$$\langle \mathfrak{R}_2 | \exp\left[-\tau\hat{V}\right] | \mathfrak{R}_1 \rangle = \exp\left[-\tau V(\mathfrak{R}_2)\right] \delta(\mathfrak{R}_1 - \mathfrak{R}_2). \quad (256)$$

The kinetic energy matrix can be evaluated using eigenfunction expansion of \hat{T} [154],

$$\langle \mathfrak{R}_0 | \exp\left[-\tau\hat{T}\right] | \mathfrak{R}_2 \rangle = \left(\frac{M_I M}{2\pi\beta}\right)^{\frac{3}{2}N_I} \exp\left[-\beta\left\{\frac{1}{2}M_I\omega_M^2(\mathfrak{R}_0 - \mathfrak{R}_2)^2\right\}\right], \quad (257)$$

where the squared frequency

$$\omega_M^2 = \frac{M}{\beta^2}. \quad (258)$$

¹¹The more sophisticated methods [154] than the primitive approximation exist but we will not pursue them here.

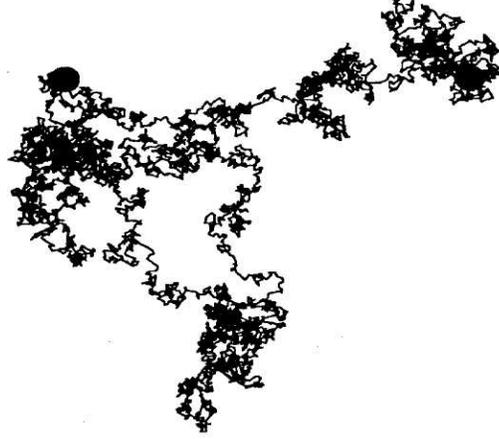


Fig. 22. Trace of a free-particle path using the mapping in Eq. (259) onto polymers.

Combining Eqs. (251), (255), (256), and (257) yields the discrete *PI expression for the density matrix in primitive approximation*¹²

$$\rho(\mathfrak{R}_0, \mathfrak{R}_M; \beta) = \int \int \cdots \int d\mathfrak{R}_1 d\mathfrak{R}_2 \cdots d\mathfrak{R}_{M-1} \left(\frac{M_I M}{2\pi\beta} \right)^{\frac{3}{2} N_I M} \times \exp \left[-\beta \sum_{m=1}^M \underbrace{\left\{ \overbrace{\frac{1}{2} M_I \omega_M^2 (\mathfrak{R}_{m-1} - \mathfrak{R}_m)^2}_{V_{harm}} + \overbrace{\frac{1}{M} V(\mathfrak{R}_m)}_{V_{scaled}} \right\}}_{V_{eff}} \right]. \quad (259)$$

The V_{eff} is a completely classical consisting of a harmonic spring term V_{harm} and scaled potential V_{scaled} . The density matrix (259) can be sampled using exclusively methods of classical statistical mechanics. Note, that V_{scaled} is local, and depends only on \mathfrak{R}_m . The only nonlocal term, which connects different points along the temperature path is the harmonic spring term. The number of discretization points necessarily depends on the temperature β . In the low-temperature limit, $\beta \rightarrow \infty$, also $M \rightarrow \infty$. In this limit V_{eff} is dominated by the harmonic spring term and hence, reduces to a potential of a system of weakly coupled harmonic oscillators. Eq. (259) is the famous mapping from a quantum system (left-hand side) to a classical system (right-hand side). The classical system is a system of interacting “polymers”, see Fig. 22. Obviously, sampling of a classical system is much simpler, than sampling of the original quantum system. The price to be paid is that the dimensionality of the original system has increased from $3N_I$ to $3MN_I$. The Feynman-Kacs formula [153] is obtained by taking the $M \rightarrow \infty$ limit, i.e. by making a continuous path. Note, that in this approximation, Eq. (259) the potential energy surface can

¹²In addition to computing static quantum-mechanical averages, Eqs. (242)-(246), the PI techniques can also be used to approximately compute quantum dynamics. Recently, the so-called centroid path dynamics technique was introduced [156, 157]. It was found, that the time evolution of the centroids $\bar{R}_I^c(t) = \frac{1}{M} \sum_{m=1}^M R_I^m(t)$ of closed Feynman paths of the nuclei contains information on the quantum dynamics.

be rather straightforwardly modeled in the same fashion as for a classical system. For some systems, such as, for instance He, excellent empirical interatomic or intermolecular potentials are available [154]. For many other systems, there are no reliable empirical potentials, and hence the interaction potentials have to be computed from PES constructed by total-energy methods. The PI method in connection with total-energy methods was for the first time used in Refs. [157–159]. One advantage of using the PI formalism is the possibility of an easy quantification of the quantum effects. One can simply perform two simulations one, where the particles are described by polymers, and, for comparison, a simulation, where the particles are completely classical.

Formula (243) for partition function Z contains the trace, i.e. the discretization points have to satisfy the periodic boundary condition $\mathfrak{R}(0) = \mathfrak{R}(\beta)$ and hence also $\mathfrak{R}_{M+1} = \mathfrak{R}_1$. Let us now rewrite Z directly in terms of individual ionic positions

$$Z = \lim_{M \rightarrow \infty} \prod_{m=1}^M \prod_{I=1}^{N_I} \left[\left(\frac{M_I M}{2\pi\beta} \right)^{\frac{3}{2}} \int d\vec{R}_I^m \right] \times \exp \left[-\beta \sum_{m=1}^M \left\{ \underbrace{\sum_{I=1}^{N_I} \frac{1}{2} M_I \omega_M^2 (R_I^m - R_I^{m+1})^2}_{V_{eff}} + \frac{1}{M} V(R_I^m) \right\} \right], \quad (260)$$

where $\vec{R}_I^{M+1} = \vec{R}_I^1$ and $V(\{R_I^m\}) = E_0(\{R_I^m\})$, the ground-state energy in the BO approximation.

The mapping onto a classical system makes it possible to estimate the degree of quantum delocalization simply by measuring the special extent of the particle by the radius of gyration

$$R_{quant}^g = \left\langle \frac{1}{N_I M} \sum_{I=1}^{N_I} \sum_{m=1}^M (\vec{R}_I^m - \vec{R}^c)^2 \right\rangle^{\frac{1}{2}}, \quad (261)$$

$$\vec{R}^c = \frac{1}{N_I M} \sum_{I=1}^{N_I} \sum_{m=1}^M \vec{R}_I^m. \quad (262)$$

The degree to which the different particles behave like quantum-mechanical particles can be measured also in terms of imaginary time correlation functions [160]. Especially useful is the rms position displacement correlation function [161]

$$\tilde{R}_I(\Delta\tau) = \left\langle \left| \vec{R}_I(\tau) - \vec{R}_I(\tau') \right|^2 \right\rangle^{\frac{1}{2}}. \quad (263)$$

where $0 \leq \Delta\tau = \tau - \tau' \leq \beta\hbar$ and $\vec{R}_I(\tau)$ denotes the coordinate of particle I at imaginary time τ . The midpoint value $\tilde{R}_I(\beta\hbar/2)$ of this function is a measure of the particle's spatial "size", whereas its variation in imaginary time determines the degree of ground-state dominance and thus localization [161]. Localized states with a large energy gap ΔE between ground and excited states are characterized by a constant $\tilde{R}_I(\Delta\tau)$ in the range of roughly $\hbar/\Delta E < \Delta\tau < \beta\hbar - \hbar/\Delta E$.

3.2.2.4 Symmetry of the wave function

In the discussion above we have ignored the symmetry of the wave function under particle exchange. Hence, all formulae above are only valid for *distinguishable particles*, the so-called *Boltzmannons*. For indistinguishable particles, we have to incorporate the appropriate symmetry of the wave function. It has been shown that this is essential for systems like liquid helium at low temperature, where exchange plays a crucial role [162]

Let us for simplicity consider system of two particles, which we assume to be Bosons, and denote the symmetric eigenfunctions $\{\chi_i^S(\vec{R}_1, \vec{R}_2)\}$ and the antisymmetric ones by $\{\chi_i^A(\vec{R}_1, \vec{R}_2)\}$. The density matrix for distinguishable particles,

$$\begin{aligned} \rho_D(\vec{R}_1, \vec{R}_2; \vec{R}'_1, \vec{R}'_2; \beta) &= \sum_i \chi_i^S(\vec{R}_1, \vec{R}_2) \exp[-\beta E_i^S] \chi_i^{S*}(\vec{R}'_1, \vec{R}'_2) \\ &+ \sum_i \chi_i^A(\vec{R}_1, \vec{R}_2) \exp[-\beta E_i^A] \chi_i^{A*}(\vec{R}'_1, \vec{R}'_2). \end{aligned} \quad (264)$$

The only difference when we are dealing with Bosons is that we now sum only over the symmetric states

$$\rho_B(\vec{R}_1, \vec{R}_2; \vec{R}'_1, \vec{R}'_2; \beta) = \sum_i \chi_i^S(\vec{R}_1, \vec{R}_2) \exp[-\beta E_i^S] \chi_i^{S*}(\vec{R}'_1, \vec{R}'_2). \quad (265)$$

From this it follows that ρ_B can be expressed in terms of ρ_D

$$\begin{aligned} \rho_B(\vec{R}_1, \vec{R}_2; \vec{R}'_1, \vec{R}'_2; \beta) &= \frac{1}{2} \left[\rho_D(\vec{R}_1, \vec{R}_2; \vec{R}'_1, \vec{R}'_2; \beta) \right. \\ &\left. + \rho_D(\vec{R}_1, \vec{R}_2; \vec{R}'_2, \vec{R}'_1; \beta) \right]. \end{aligned} \quad (266)$$

If the particles are Fermions, we would obtain

$$\begin{aligned} \rho_F(\vec{R}_1, \vec{R}_2; \vec{R}'_1, \vec{R}'_2; \beta) &= \frac{1}{2} \left[\rho_D(\vec{R}_1, \vec{R}_2; \vec{R}'_1, \vec{R}'_2; \beta) \right. \\ &\left. - \rho_D(\vec{R}_1, \vec{R}_2; \vec{R}'_2, \vec{R}'_1; \beta) \right]. \end{aligned} \quad (267)$$

The corresponding formulae for the partition functions follow immediately. For Bosons, we have

$$\begin{aligned} Z_B &= \int d\vec{R}_1 d\vec{R}_2 \rho_B(\vec{R}_1, \vec{R}_2; \vec{R}_1, \vec{R}_2; \beta) \\ &= \frac{1}{2} \left[\int d\vec{R}_1 d\vec{R}_2 \rho_D(\vec{R}_1, \vec{R}_2; \vec{R}_1, \vec{R}_2; \beta) \right. \\ &\left. + \int d\vec{R}_1 d\vec{R}_2 \rho_D(\vec{R}_1, \vec{R}_2; \vec{R}_2, \vec{R}_1; \beta) \right]. \end{aligned} \quad (268)$$

When we express ρ_D in the usual way in terms of discrete chains, we therefore have Z_B as the sum of two terms, one of which involves the integral over configurations of two separate chains each of M beads, and the other integral over configurations of one larger chain of $2M$ beads, the whole thing multiplied by the factor $\frac{1}{2}$. Schematically, where the white circles indicate the

$$Z_B = \frac{1}{2} \left[\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right] + \frac{1}{2} \left[\begin{array}{c} \text{Diagram 3} \end{array} \right] \quad (269)$$

positions integrated over in Eq. (268) and the arrows show the sense of increasing bead index (“time”) m .

If we go through the same argument for three particles, we find

$$Z_B = \frac{1}{3!} \left[\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \\ \text{Diagram 3} \end{array} \right] + \frac{1}{2} \left[\begin{array}{c} \text{Diagram 4} \\ \text{Diagram 5} \end{array} \right] + \frac{1}{3} \left[\begin{array}{c} \text{Diagram 6} \end{array} \right] \quad (270)$$

In general ρ_B is related to ρ_D by

$$\rho_B(\mathfrak{R}_0, \mathfrak{R}_1; \beta) = \frac{1}{N_I!} \sum_P \rho_D(\mathfrak{R}_0, P\mathfrak{R}_1; \beta), \quad (271)$$

where P is a permutation of particle labels. Examples of the effect of the symmetry of the wave function are illustrated in Fig. 23. As can be seen, the symmetry has profound consequences at low temperature. Many of the unusual low-temperature properties of ^4He , such as superfluidity, have to do with the symmetry of the wave function and with the permutations of the density matrix. On the other hand, at more moderate temperatures both ρ_B and ρ_D are almost identical (Fig. 23). For most systems the threshold temperature is a few Kelvin.

A legitimate question is why the PI technique is used only for the ionic subsystem and not also for description of the electrons. Electrons are much lighter than protons and hence, unlike for protons, the symmetry of the wave function is always of paramount importance. This is the reason why the simplest reasonable form of the electronic wave function is the HF wave function, Eq. (21) and not the Hartree form, Eq. (18). In general, while the bosonic density matrix ρ_B can be relatively easily be constructed, the Fermionic density matrix ρ_F represents much harder problem and a systematic way of constructing ρ_F is still lacking. The reason is related to the so-called fermionic sign-problem, which is still an open problem; see chapter 3.1.6.2 for approximations to the fermionic sign-problem in zero temperature QMC. Nevertheless, an all-PI technique based on variants of the above equations, where both electronic (in an approximate way) and ionic degrees of freedom are described by PI mapping has been proposed [163].

3.2.2.5 Technical issues

As mentioned above, in the strongly quantum regime (light particles, $\beta \rightarrow \infty$), the PI mapping leads to a system of weakly perturbed stiff harmonic oscillators. Such a system is non-ergodic [149]. At the first glance, this means that the PI formulation is much better suited to MC, rather than MD sampling. However, MC is introducing pronounced discontinuities in motion trough configuration space, which are counterproductive (see, chapter 3.2.1 on Nosé-Hoover

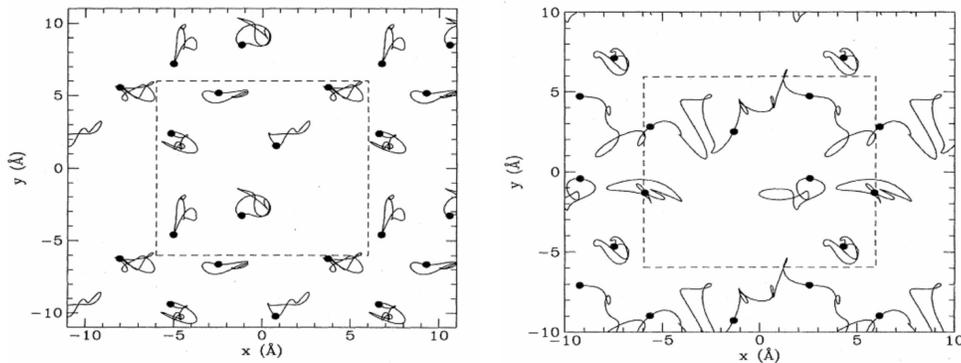


Fig. 23. Left: The trace of the phase of six ^4He atoms at a temperature of 2K. The filled circles are markers for the (arbitrary) beginning of the path. The dashed line indicates the box boundaries. Right: The same as on the right, except for the temperature, which is 0.75 K.

thermostats). Fortunately, the Nosé-Hoover chains of thermostats can restore ergodicity, and hence the MD sampling of the configuration space.

Furthermore, the primitive approximation (259), (260) is computationally rather inefficient. The problem is that the slow paths motion through the phase-space is very difficult to sample, as the MD time step must be adjusted to capture the high-frequency spring oscillations. A more efficient approach consists of transformation from the primitive path variables $\{\tilde{R}_I^m\}$ to either *normal mode variables* $\{\tilde{u}_I^m\}$ [164] or *staging variables* [165]. The normal modes $\{\tilde{u}_I^m\}$ diagonalize the nearest-neighbor harmonic term, which is typically the dominant term. While the technical details of the staging transformation are different, the spirit of the transformation is very similar to that of the normal mode transformation. These algorithms effectively separate the fast and slow variables and so speed up the phase-space sampling efficiency. With these ingredients, the MD sampling can be made comparable to the best MC sampling [154].

If one has to include the exchange, a straightforward evaluation for large N_I as in (268)-(271) is impossible, since there will be $N!$ terms to evaluate. For Bosons, each term in the sum is positive, which opens the possibility to sample the contributions, rather than to explicitly evaluate each term. A Bosonic simulation consists of a combination of random walks through path and permutation space [154]. Sampling of the permutation space for Fermions is not possible due to the cancellation between the contributions of both odd and even permutations (the (in)famous Fermionic sign-problem). Note also, that sampling of the permutation space is a discontinuous process, which rules out MD sampling. Hence, only Boltzmannonic ρ_D can be MD sampled. This is a very serious limitation and the prime reason why up to date the PI techniques in combination with *ab-initio* computed PESs have been applied only to distinguishable protons at temperatures ≥ 5 K.

3.3 Solving the basic equations

In the previous paragraphs we have considered the different theories of electronic and atomic structure. Let us now consider the possibility of really solving the coupled equations of the

combined system consisting of both electronic and ionic degrees-of-freedom. Basically, we have the choice between the Ehrenfest molecular dynamics (EMD), Eqs. (110, (12), and Born-Oppenheimer molecular dynamics (BOMD), Eqs. (16), (17). Both approaches have one key feature in common, namely that *the high-dimensional PES* for the ions is not explicitly constructed, but rather *computed “on fly”* only for the points really visited.

3.3.1 Ehrenfest dynamics

The indisputable advantage of the EMD over the BOMD is the fact that EMD provides *real electronic dynamics* as the wave function propagation follows from solution of time-dependent Schrödinger equation which conforms to unitary propagation [166, 167]

$$\begin{aligned} \Psi(t_0 + m\Delta t) &= \exp\left[i\hat{H}_e(t_0 + (m-1)\Delta t)\frac{\Delta t}{\hbar}\right] \\ &\quad \vdots \\ &\quad \times \exp\left[i\hat{H}_e(t_0 + 2\Delta t)\frac{\Delta t}{\hbar}\right] \\ &\quad \times \exp\left[i\hat{H}_e(t_0 + \Delta t)\frac{\Delta t}{\hbar}\right] \\ &\quad \times \exp\left[i\hat{H}_e(t_0)\frac{\Delta t}{\hbar}\right]\Psi(t_0), \end{aligned} \quad (272)$$

where the time step $\Delta t = t_{\max}/m$ and $H_e(t)$ ¹³ is the electronic Hamiltonian (Eq. (1)) implicitly time-dependent via $\{\vec{R}_I(t)\}$. The *velocity Verlet algorithm*¹⁴ [169] can be used to integrate the equations-of-motion for the ionic degrees-of-freedom¹⁵ which consists of the following steps

- (1) estimate approximate velocities

$$\dot{\vec{R}}_I(t_0 + m\Delta t) = \dot{\vec{R}}_I(t_0 + (m-1)\Delta t) + \frac{\Delta t}{2M_I}\vec{F}_I(t_0 + (m-1)\Delta t), \quad (273)$$

- (2) update atomic positions

$$\vec{R}_I(t_0 + m\Delta t) = \vec{R}_I(t_0 + (m-1)\Delta t) + \Delta t\dot{\vec{R}}_I(t_0 + m\Delta t), \quad (274)$$

- (3) calculate $\vec{F}_I(t_0 + m\Delta t)$,

- (4) correct velocities

$$\dot{\vec{R}}_I(t_0 + m\Delta t) = \dot{\vec{R}}_I(t_0 + m\Delta t) + \frac{\Delta t}{2M_I}\vec{F}_I(t_0 + m\Delta t). \quad (275)$$

¹³ $H_e(t)$ at time t can be evaluated using split-operator techniques [168].

¹⁴The simpler Verlet algorithm is perfectly adequate for most applications [170].

¹⁵We consider here the microcanonical ensemble. If canonical ensemble is considered, the thermostat equation-of-motion needs to be discretized in a similar way except that the corresponding coupling to the ions is added.

There are two very different time-scales involved. The timescale for the electronic dynamics τ_e , and the corresponding time-scale for dynamics of the ions, τ_I . The slower time-scale τ_I is dictated by the fastest vibrational frequencies, typically an optic phonon of the lightest particle in the system. This for hydrogen is $\sim 4000 \text{ cm}^{-1}$ and $\tau_I \sim 10^{-14}$. About 10–20 samples of the highest frequency have to be taken in order to generate a correct, conservative, dynamics, which gives $\Delta t_{\text{max}} \sim \frac{\tau_I}{10}$. A realistic estimate of for the electronic time-scale is $\tau_e \sim 10^{-16}$ s. This gives $\tau_e \approx \frac{\tau_I}{100}$ and hence also $\Delta t_{\text{max}} \sim \frac{\tau_e}{10}$. As can be seen, the price to be paid for following the real system dynamics is rather high, as the discretization time step is about two orders of magnitude slower than what would be the case if the ions dictated the clock speed. Several attempts have been made to tip the scale toward a more acceptable integration time-scale, such as use of different integration steps for electrons and ions [171], multiple time-step integration theory [172], or even use of reduced ionic masses rendering the ionic dynamics fictitious as opposed to the real electronic dynamics [173].

3.3.2 Born-Oppenheimer dynamics

In BOMD, Eqs. (16), (17), the ions set the clock speed. In practice it means that Δt_{max} is about two orders of magnitude larger compared to the EMP. In other words, the ions, which are now the primary species, are moving on the PES generated by electronic structure method of preselected accuracy, in most cases a DFT method. The rationale being, that the DFT methods strike an optimum balance between accuracy and feasibility. The disadvantage is that the electronic dynamics cannot be followed and that the electronic Hamiltonian \hat{H}_e has to be diagonalized for each new position of the ions $\{\vec{R}_I(t_m)\}$. Given the above estimate $\Delta t_{\text{max}} \sim \frac{\tau_I}{10} \sim 1$ fs, this is computationally a very demanding procedure as the diagonalization has to be made thousands of times, depending on the system/process under consideration. Let us now briefly outline the most important diagonalization procedures.

3.3.2.1 Diagonalization of \hat{H}_e

The most straightforward procedure is to transform the solution of the partial differential equation (16) to an algebraic equation. If we consider a single-particle variant of Eq. (16)¹⁶, and expand the solution in a basis, $\{\chi_\nu\}$, Eq. (201), we have after multiplication by χ_λ^* and integration over \vec{r} the following set of *homogeneous algebraic equations*

$$\sum_{\nu} (H_{\lambda\nu} - \varepsilon_{\lambda} \delta_{\lambda\nu}) c_{\nu} = 0. \quad (276)$$

The eigenvalues are determined by the condition

$$|H_{\lambda\nu} - \varepsilon_{\lambda} \delta_{\lambda\nu}| = 0. \quad (277)$$

For each $\{\varepsilon_{\lambda}\}$ a different set of coefficients c are obtained and hence, also a set of the solution $\{\varphi_{\lambda}\}$. The above procedure is a highly non-linear problem as the Hamiltonian itself depends on the set of occupied wave functions $\{\varphi_{\lambda}^{\text{occ}}\}$. Hence, the wave functions have to be found by an iteration procedure, where an initial guess for the wave functions $\{\varphi_{\lambda}^{m=0}\}$ is used to construct

¹⁶If the same procedure is used in a many-body approach, the determinants play the role of the many-body basis.

the initial guess for the Hamiltonian matrix $H_{\lambda\nu}^{m=0}$. Diagonalization of $H_{\lambda\nu}^{m=0}$ yields the first iteration on the wave functions $\{\varphi_{\lambda}^{m=1}\}$ from which a new guess $H_{\lambda\nu}^{m=1}$ for the Hamiltonian matrix is constructed, and so on. However, such a procedure is usually divergent and in order to stabilize the procedure, the so-called mixing scheme has to be adopted. In the *mixing scheme*, the new charge density is computed as a weighted combination of the charge densities from the previous steps: $\rho^{m+1}(\vec{r}) = \alpha\rho^{m+1}(\vec{r}) + (1 - \alpha)\rho^{m-1}(\vec{r})$, ($\alpha \ll 1$). The computational cost of direct diagonalization is $O(M^3)$. Typically $M \gg N_e$, especially with basis sets such as plane waves, where M may easily as large as $O(10^5)$. Most of the computational work is wasted to generate unoccupied states.

There are several ways to overcome the above $O(M^3)$ scaling deadlock. The full matrix diagonalization, yielding all M eigenstates, can be replaced by a *partial diagonalization* [174, 175], yielding only the lowest N_e eigenstates. In this way the numerical cost of the diagonalization can be reduced to $O(N^2M)$.

A more radical approach abandons diagonalization schemes altogether and replaces them by direct minimization schemes [122]. Obviously, such an approach is only valid if the total energy is a single minimum function with respect to the expansion coefficients (Eq. (201)). In most cases this is indeed the case. This assumption can be tested by repeated minimizations with different initial conditions.

Perhaps the simplest way to find a minimum of a function with many variables¹⁷ is provided by the *steepest descent (SD) method*, which can be formulated as

$$\begin{aligned} \dot{\varphi}_i(\vec{r}, t) &= - \underbrace{\frac{\bar{\delta}E}{\bar{\delta}\varphi_i^*(\vec{r}, t)}}_{\text{constrained gradient}} = - \underbrace{\frac{\delta E}{\delta\varphi_i^*(\vec{r}, t)}}_{\text{unconstrained gradient}} + \underbrace{\text{constraints}}_{\text{orthonormality}} \\ &= -H_e\varphi_i(\vec{r}, t) + \text{constraints}, \end{aligned} \quad (278)$$

where the dot indicates derivative with respect to a fictitious time variable t labeling the successive approach steps $\{\varphi_i^{(m)}\}$ toward the minimum and $\bar{\delta}/\bar{\delta}\varphi_i^*(\vec{r}, t)$ indicates the constrained functional derivative that preserves the orthonormality of the wave functions. The constrained derivative can be replaced by an unconstrained derivative $\delta/\delta\varphi_i^*(\vec{r}, t)$ followed by reorthonormalization of the new set of wave functions using for instance the well-known *Gram-Schmidt orthogonalization* scheme. Note, that if one neglects the constraints, Eq. (278) is equivalent to time-dependent Schrödinger equation in imaginary time. Hence, the SD procedure can be thought of as a way of projecting the trial wave function onto ground-state wave function. In discretized form Eq. (278) becomes

$$\varphi(\vec{r}, t + \Delta t) = \varphi(\vec{r}, t) - \Delta t H_e \varphi_i(\vec{r}, t) + \text{constraints}, \quad (279)$$

where the time step Δt fixes the time-scale and the convergence rate of the SD scheme. It plays the role of the ordinary mixing parameter. The procedure described in Eq. (279) must be started from an initial guess $\{\varphi_i(\vec{r}, t = 0)\}$ that is nonorthogonal to the ground-state. In such a case, applying the Gram-Schmidt orthonormalization, iteration of Eq. (279) will provide in the in sufficiently large limit ($t \rightarrow \infty$) the ground-state wave functions $\{\varphi_i(\vec{r}, t \rightarrow \infty)\}$.

¹⁷In large-scale computations the number of optimization parameters may be as large as $O(10^5 - 10^7)$.

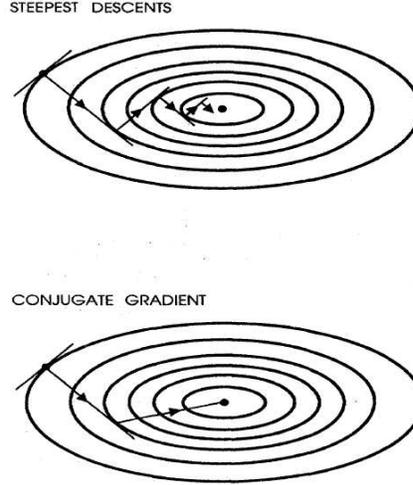


Fig. 24. Schematic illustration of nearly orthogonal steps in minimizing a quadratic function using SD algorithm compared to efficient CG algorithm.

The procedure defined in Eqs. (278), (279) is usually far from efficient. The function near the minimum has often quadratic shape and the SD procedure generates steps which are often orthogonal or nearly orthogonal to one another which results in very many steps required to reach the minimum; Fig. 24.

Such a drawback can be eliminated using methods taking account of the matrix of second derivatives, i.e. the *Hessian* A . The problem is that the size of A is $N_e M \times N_e M$ and hence, its computation is impractical. Fortunately, there are methods, such as *conjugate gradients* (CG) [176], which allow to take advantage of the information contained in A without explicitly constructing it.

Let us outline the basics of the CG method and suppose that the function to be minimized can be approximated by a multidimensional quadratic form around some point P taken as the origin of the coordinates ,

$$f(X) \approx c - \langle b|X \rangle + \frac{1}{2} \langle X|A|X \rangle, \quad (280)$$

where

$$X \equiv (x_1, x_2, \dots, x_L), \quad c = f(P), \quad b = -\nabla f|_P, \quad A_{ij} = \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_P, \quad (281)$$

with a symmetric positive-definite $L \times L$ Hessian matrix. An iterative minimization procedure is then defined by the sequence

$$P^{(m+1)} = P^{(m)} + \lambda^{(m)} h^{(m)}; \quad m = 0, 1, 2, \dots, \quad (282)$$

where $\lambda^{(m)}$ is a scalar obtained by a one-dimensional minimization along the direction $h^{(m)}$ in the multidimensional space of the expansion coefficients defined by

$$h^{(m)} = \begin{cases} g^{(m)}, m = 0 \\ g^{(m)} + \gamma^{(m-1)}h^{(m-1)}, m = 1, 2, 3, \dots \end{cases} \quad (283)$$

and

$$\begin{aligned} g^{(m)} &= -\nabla f(P^{(m)}), \\ \gamma^{(m)} &= \frac{\langle g^{(m+1)} | g^{(m+1)} \rangle}{\langle g^{(m)} | g^{(m)} \rangle}. \end{aligned} \quad (284)$$

The h^m are conjugate and for a quadratic function, Eq. (280), satisfy the *conjugacy property* [176]

$$\langle h^{(n)} | A | h^{(m)} \rangle = 0 \quad \forall n \neq m. \quad (285)$$

The property (285) guarantees that each step is a definite improvement over all the preceding ones and eliminates the difficulties with canyon-like quadratic functions. The important ingredient of the CG algorithm is that the Hessian matrix A is never explicitly required. The CG method is widely used in optimization problems and the cost of a single step is comparable¹⁸ to that of the much less efficient SD algorithm. It is therefore natural to apply it to the electronic structure problem [122]. A difficulty arises in the electronic structure problem due to the orthonormality constraints between the single-particle orbitals. The constraints originate forces of constraint that must be taken into account in the line minimizations. The orthonormal orbitals $\{\varphi_i\}$ may be related to independent but not orthonormal orbitals $\{\phi_i\}$ via

$$\varphi_i = \sum_j S_{ij}^{-1/2} \phi_j, \quad (286)$$

where $S_{ij} = \langle \phi_j | \phi_i \rangle$ is the overlap matrix. The functional derivative with respect to the non-orthonormal orbitals $\{\phi_i\}$ becomes [122]

$$-\frac{\delta E}{\delta \phi_i^*(\vec{r}, t)} = -\hat{H}_e \phi_i + \underbrace{\sum_m \langle \phi_m | \hat{H}_e | \phi_i \rangle \phi_m}_{\text{constraint force}}. \quad (287)$$

This is all we need to set up the CG procedure defined in Eqs. (282)-(284). In order to evaluate the line minimizations along the conjugate directions $\{h^m\}$, we may adopt a parabolic approximation [177] or can treat the orbitals in a non-selfconsistent manner along the conjugate directions [122]. Initially orthonormal orbital will become non-orthonormal after a CG step and hence it is numerically convenient to reorthonormalize the $\{\phi_i\}$ orbitals after each step.

An important issue is related to existence of different length scales, which often hampers the rate of convergence. The problem can be solved by *preconditioning*, i.e. by introduction of a metric that brings all degrees-of-freedom onto the same length scale. The preconditioning is

¹⁸A count of the dominant operations yields an overhead factor of ~ 2 .

particularly simple for the plane wave basis set. For large \vec{G} vectors, the total energy is dominated by the kinetic energy, which is diagonal in the plane wave representation. Based on this observation, several preconditioners have been proposed [177–179]. We take the preconditioner described in Ref. [177] as a representative example,

$$K_{\vec{G}\vec{G}'} = \delta_{\vec{G}\vec{G}'} \frac{27 + 18x + 12x^2 + 8x^3}{27 + 18x + 12x^2 + 8x^3 + 16x^4}; \quad x = \frac{E_{kin}(\vec{G})}{E_{kin}^{\max}}, \quad (288)$$

where E_{kin}^{\max} is an energy cutoff below which the components of the gradient are to be left unchanged. The K has the required properties: approaches $x = 0$ with a value of 1, has derivatives up to third which guarantees that the low-wave-number components are left unchanged. Above $x = 1$, K approaches $1/[2(x-1)]$. At small $|\vec{G}|$ the algorithm is equal to the $|\vec{G}|$ the preconditioning avoids spending of most of the optimization making small changes to the high-wave-number components.

The efficiency of the CG method is compared with other methods in Fig. 25. As can be seen, the CG approach to minimum is significantly faster than the alternative methods, such as CG or the Davidson [174] method, especially in difficult to converge situations, such as for disordered systems. The effect of preconditioning is equally pronounced. As mentioned in chapter 3.1.5.9, problems are bound to arise if dealing with the electronic free energy functionals. The optimization can be stabilized by techniques of density mixing akin to what is done in straightforward matrix diagonalization.

Above we have introduced the *all-band formulation* of the CG method. An alternative formulation uses the CG iteration individually for each orbital. The advantage of the *band-by-band method* [177] is that if the rate of convergence of individual bands exhibits pronounced variations, the slowly converged orbitals are not hampering the convergence of the faster converging orbitals. The example of a molecule impinging on a surface from chapter 2, is a prototypical case. Only a few molecular and surface states around the Fermi level are affected as the molecule approaches the surface the rest showing only little variation. Hence, the optimization can be tailored individually for each orbital. The downside is that many operations (for example all terms depending on the charge density), which were performed only once in the all band-method have now to be recomputed over and over again for each orbital. Moreover, the fast-converging orbitals will benefit little from the conjugacy property (285), as only few steps or perhaps only a single step will suffice to bring them to convergence.

A different approach to diagonalization of the electronic Hamiltonian is based based on extrapolation. A well-known representative is the so-called *Direct Inversion of the Iterative Subspace (DIIS) method* [181, 182]. DIIS uses the information on n previous optimization steps of both the wave function coefficients $\{c_i^k\}$ and error vector $\{e_i^k\}$. The best approximation to the final solution within the subspace spanned by the n stored vectors obtained in a least square sense

$$c_i^{(n+1)} = \sum_{k=1}^n d_k c_i^{(k)} \quad (289)$$

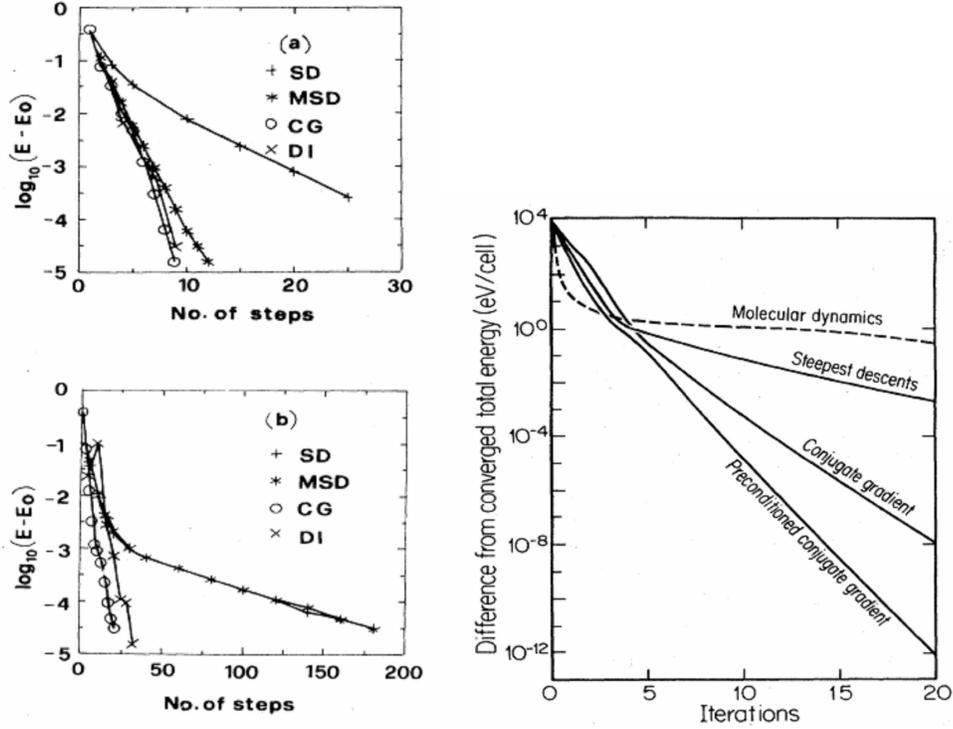


Fig. 25. Left panel: The rate of convergence of the total energy E in SD, modified SD (MSD) [180] and Davidson iterative minimization [174] for (a) Si crystal and (b) a strongly disordered system of Si atoms. Right panel: Effect of the preconditioning on the rate of convergence of the CG method for the Si crystal system. Original method refers to the MSD algorithm.

subject to the restriction

$$\sum_{k=1}^n d_k = 1. \quad (290)$$

The expansion coefficients $\{d_k\}$ are calculated from the following system of linear equations

$$\begin{pmatrix} b_{11} & \cdots & b_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ b_{n1} & \cdots & b_{nn} & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ \vdots \\ d_n \\ -\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \quad (291)$$

where $\{b_{kl}\}$ are given by

$$b_{kl} = \sum_i \langle e_i^{(k)} | e_i^{(l)} \rangle. \quad (292)$$

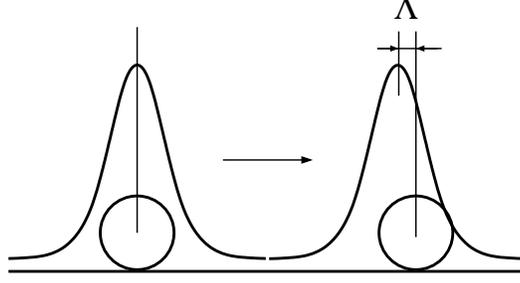


Fig. 26. Illustration of the drag-effect on the ions due to systematic errors induced by use of first-order optimization of the electronic structure. The ions are schematized by an open circle and the electronic structure by a hat-shaped structure over the ion. Left: initial configuration of the combined system of electrons plus ions; Right: situation after the ions start moving.

The error vectors are unknown but can be approximated. Within the quadratic model we have

$$e_i^{(k)} = -K^{-1}\delta_i^{(k)}.$$

In the same approximation, assuming constant K , the new trial vectors become

$$c_i = c_i^{(n+1)} - K^{-1}\delta_i^{(n+1)} \quad (293)$$

where the estimate of the derivative of the energy is

$$\delta_i^{(n+1)} = \sum_{k=1}^n d_k \delta_i^{(k)}. \quad (294)$$

The new trial vectors are not orthonormal and hence a reorthonormalization is required. The efficiency of the DIIS procedure is comparable to that of CG.

3.3.2.2 Accuracy issues

BO dynamics that combines ionic dynamics, either classical (chapter 3.2.1) or quantum via path integral description (chapter 3.2.2), in combination with the optimization techniques for the electronic degrees-of-freedom described above has to be applied with care. One problem that occurs is that this procedure may generate a systematic error in computation of the forces acting on the ions. The problem can be understood by inspection of Fig. 26. Imagine that the electronic structure has initially been perfectly optimized and accurate ionic forces obtained. Using these forces, ionic degrees-of-freedom can be propagated using, for instance, Eqs. (273) – (275). At that point the electronic structure is “left behind” the ions by some distance. What the (“first-order”) electronic optimization is doing is to push the wave functions and the associated charge densities “on top” of the new atomic positions. This may be done with arbitrarily high, but always finite accuracy. As a consequence, the electrons even after reoptimization will remain lagging behind the ions by some amount Δ . As the ions keep moving, the situation will repeat itself and as a result, the finite accuracy of the electronic optimization will lead to generation

of a systematic artificial “braking” force. As a result, the ionic dynamics will not be conservative. In microcanonical ensemble, there will be a spontaneous decrease of the temperature. In canonical ensemble the thermostats will keep the temperature constant, but the thermostat will be constantly compensating the energy drained from the ions. In either case this procedure will lead to violation of the constants-of-motion.

A way to prevent this situation from happening is to provide at each step some degree of randomness in the initial set of wave functions. This can be achieved by wave function extrapolation schemes, where the new set of guess functions are generated by the following algorithm [173, 184]

$$\begin{aligned} \tilde{\varphi}_i(\vec{r}, t + \Delta t) = \varphi_i(\vec{r}, t) &+ \alpha [\varphi_i(\vec{r}, t) - \varphi_i(\vec{r}, t - \Delta t)] \\ &+ \beta [\varphi_i(\vec{r}, t - \Delta t) - \varphi_i(\vec{r}, t - 2\Delta t)] + \underbrace{\text{constraints}}_{\text{orthonormality}}. \end{aligned} \quad (295)$$

Hence, the initial guess is constructed by extrapolation, using three previous electronic configurations. The extrapolation coefficients $\{\alpha, \beta\}$ are fixed by minimizing the norm $\|\vec{R}(t + \Delta t) - \tilde{\vec{R}}(t + \Delta t)\|$ with $\{\tilde{\vec{R}}(t + \Delta t)\}$ being the ionic coordinates extrapolated in the same way as the wave functions. In most applications setting $\beta \stackrel{!}{=} 0$ and using just a first-order extrapolation is perfectly sufficient. The extrapolation scheme (295) leads to a more efficient electronic optimization but does not solve the problem of the drag force on the ions. Fortunately, one can setup a procedure which can extrapolate the wave functions so that they can both leg behind as well as overtake the ions depending, for instance, on the conservation of the constants-of-motion [185]. In this way a perfectly correct, conservative and rather efficient BO dynamics can be generated.

3.3.3 Car-Parrinello dynamics

A completely *new approach* to solving the dynamics for the combined system of electrons and ions was proposed by *Car and Parrinello* (CP) [186]. The aim is to set up an automatic process where time-scales of both electrons and ions are comparable. In the EMD the time-scale is completely determined by the electrons, which, in turn are propagated in time without the need to diagonalize the electronic Hamiltonian \hat{H}_e . Just the opposite is true of the BOMD. From this point-of-view the CP method in a clever way combines the advantages of both previous approaches. In simple terms, the time-scale is set somewhere inbetween the electronic and ionic time-scale and the wave functions are optimized automatically as the ions evolve on the PES.

Let us now outline the CP method. Each point on the PES is considered as being a functional of the electronic degrees-of-freedom, i.e. of the single-particle wave functions $\{\varphi_i\}$ used to build the many-body wave function Ψ_0 and of the ionic positions $\{\vec{R}_I\}$. Hence we have $E[\{\varphi_i\}, \{\vec{R}_I\}] = \langle \Psi_0(\vec{R}_I) | \hat{H}_e(\vec{R}_I) | \Psi_0(\vec{R}_I) \rangle$ (where $\Psi_0(\vec{R}_I) = \det\{\varphi_i\}$ or a combination of more determinants, Eqs. (20), (21)). We have seen above (Eqs. (278), (279)), how local optimization procedures can be used to find the minimum of $E[\{\varphi_i\}, \{\vec{R}_I\}]$ for fixed ions. Let us now assume that we want to optimize both degrees-of-freedom to the global minimum in the spirit of simulated annealing [187]. This can be done using MD techniques, which are usually formulated in terms of Lagrangean dynamics with a Lagrangean $L = K - U$. Clearly, $U = E[\{\varphi_i\}, \{\vec{R}_I\}]$, and $K_I = \frac{1}{2} \sum_{I=1}^{N_I} M_I \dot{\vec{R}}_I^2$. What remains to be fixed is the classical “kinetic energy” K_e of the

single-particle orbitals. By analogy with the first-order Eqs. (278), (279), where the “velocity” of the wave functions was fixed by the time-step Δt , let us assume that the velocity is fixed by a fictitious “kinetic energy” $K_e = \frac{1}{2} \sum_{i=1}^{N_e} \mu_i \langle \dot{\varphi}_i | \dot{\varphi}_i \rangle$, where μ_i is a fictitious mass of the orbital φ_i . With this in mind, following Car and Parrinello [186], we can write the following Lagrangean

$$L_{CP} \left[\{\varphi_i\}, \{\dot{\varphi}_i\}, \{\vec{R}_I\}, \{\dot{\vec{R}}_I\} \right] = \underbrace{\sum_{I=1}^{N_I} \frac{1}{2} M_I \dot{\vec{R}}_I^2 + \sum_{i=1}^{N_e} \frac{1}{2} \mu_i \langle \dot{\varphi} | \dot{\varphi} \rangle}_{\text{kinetic energy}} - \underbrace{E \left[\{\varphi_i\}, \{\vec{R}_I\} \right]}_{\text{potential energy}} + \underbrace{\text{constraints}}_{\text{orthonormality}}. \quad (296)$$

The corresponding Newton’s equations-of-motion are obtained from the Lagrange equations

$$\frac{d}{dt} \frac{\delta L_{CP}}{\delta \dot{\varphi}_i^*} = \frac{\delta L_{CP}}{\delta \varphi_i^*} \quad (297)$$

$$\frac{d}{dt} \frac{\delta L_{CP}}{\delta \dot{\vec{R}}_I} = \frac{\delta L_{CP}}{\delta \vec{R}_I}. \quad (298)$$

The CP equations-of-motion take the following form

$$\mu_i \ddot{\varphi}_i(\vec{r}, t) = - \frac{\delta}{\delta \varphi_i^*(\vec{r}, t)} E \left[\{\varphi_i\}, \{\vec{R}_I\} \right] + \frac{\delta}{\delta \varphi_i^*(\vec{r}, t)} \text{constraints} \quad (299)$$

$$M_I \ddot{\vec{R}}_I(t) = - \frac{\partial}{\partial \vec{R}_I} E \left[\{\varphi_i\}, \{\vec{R}_I\} \right] + \frac{\partial}{\partial \vec{R}_I} \text{constraints}. \quad (300)$$

The constraints are very important as they induce forces of constraint. We have seen an example of the orthonormality constraint, $\int d\vec{r} \varphi_i^*(\vec{r}) \varphi_j(\vec{r}) = \delta_{ij}$, above (Eq. (287)), when discussing the CG algorithm. In the present case we have

$$\text{constraint} = \sum_{ij} \varepsilon_{ij} \left(\int d\vec{r} \varphi_i^*(\vec{r}) \varphi_j(\vec{r}) - \delta_{ij} \right), \quad (301)$$

where ε_{ij} is the Hermitian matrix of Lagrange multipliers. Assuming there is no constraint on ions, equations (299), (300) thus take the form¹⁹

$$\mu_i \ddot{\varphi}_i(\vec{r}, t) = -\hat{H}_e \left[\{\varphi_i\}, \{\vec{R}_I\} \right] \varphi_i(\vec{r}, t) + \sum_j \varepsilon_{ij} \varphi_j, \quad (302)$$

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I E \left[\{\varphi_i\}, \{\vec{R}_I\} \right]. \quad (303)$$

¹⁹In an expansion into a basis set (201), the derivative implies the derivative of the expansion coefficients. As an example of Eq. (302) we show the explicit form in a plane wave basis set $\mu_i \ddot{c}_i(\vec{G}) = -H_{\vec{G}\vec{G}, c_i}(\vec{G}) + \sum_j \varepsilon_{ij} c_j(\vec{G})$.

In principle one could start solving these equations and simultaneously control the temperature of the ions $T_I \propto \sum_I \frac{1}{2} M_I \dot{R}_I^2$ and electrons $T_e \propto \sum_I \frac{1}{2} \mu_i \langle \dot{\varphi}_i | \dot{\varphi}_i \rangle$. However, the equations are almost always used so, that $-\nabla_I E[\{\varphi_i\}, \{\vec{R}_I\}] \sim \vec{F}_I$, which means that $E[\{\varphi_i\}, \{\vec{R}_I\}]$ is close to $\min_{\{\varphi_i\}} E[\{\varphi_i\}, \{\vec{R}_I\}]$. Notice, that if the condition $\min_{\{\varphi_i\}} E[\{\varphi_i\}, \{\vec{R}_I\}]$ is satisfied, $\dot{\varphi}_i(\vec{r}, t) \sim 0$ and

$$\hat{H}_e [\{\varphi_i\}, \{\vec{R}_I\}] \varphi_i(\vec{r}, t) = \sum_j \varepsilon_{ij} \varphi_j. \quad (304)$$

Equation (304) looks like an ordinary single-particle Schrödinger equation with solutions $\{\varphi_i\}$ which converge to linear combinations of true solutions. In order for the algorithm to work correctly, the initial guess for $\{\varphi_i(\vec{r}, t = 0)\}$ must be such that they are nonorthogonal to the ground-state. This may be achieved for example by starting from a randomly generated orthonormal set, subsequently converged to the ground-state by some of the first order methods, such as CG or DIIS. The solution may converge either to the true single-particle states or to their linear combinations, the result depending on the orthonormalization scheme used. If the symmetry-breaking Gram-Schmidt is used, $\{\varphi_i\}$ will converge to true eigenstates of the Hamiltonian \hat{H}_e [188]. However, if the objective is to solve Eqs. (302), (303), the constraints have to be implemented in a way, that does not perturb the dynamics. For example, if we integrated Eq. (302) with Gram-Schmidt orthonormalization, the resulting dynamics would be nonconservative. The constraints for all degrees-of-freedom have to be implemented via Lagrange multiplier techniques, such as RATTLE [189] or SHAKE [190]. The fact that the solutions converge to linear combination of eigenstates would not matter if the Schrödinger equation was linear. Due to the self-consistency, this is not the case. It still does not matter if all occupied states have equal occupation numbers $\{f_i\}$, see Eqs. (139) – (141), as is the case for systems with even number of electrons at $T = 0$. In such a case, a rotation of equation (302) which diagonalizes ε_{ij} matrix is a unitary transformation. This is the regime in which Eqs. (302), (303) have to be used.

The question remains how the CP method works. The system is out of equilibrium with the electronic-degrees-of-freedom at a “fictitious” low temperature $T_e \sim 0$ and ions at a finite physical temperature T_I . Clearly, the two subsystems will equilibrate in the $t \rightarrow \infty$ limit. Nevertheless, for many systems the CP technique provides a perfectly stable simulation condition over relevant observation time-scales. The reason for this was analyzed in Ref. [191]. The authors studied a model system of 8 silicon atoms arranged in a perfect diamond lattice, described by DFT, with orbitals expanded in plane waves, time step $\Delta t = 0.3$ fs, $\mu_i = \mu = 300$ a.u., with atoms initially displaced from their equilibrium positions in the direction corresponding to the optic phonon, in total 20000 time steps (6.3 ps) in microcanonical ensemble [191]. The computed classical power spectrum of the electronic degrees-of-freedom obtained by Fourier transform of the velocity autocorrelation function

$$\gamma(\omega) = \int_0^\infty dt \cos(\omega t) \overline{\sum_i \langle \dot{\varphi}_i(t) | \dot{\varphi}_i(0) \rangle} \quad (305)$$

(the horizontal bar indicates statistical, MD, averaging) are compared with the highest frequency phonon mode of the nuclear subsystem in Fig. 27. It is evident, that for the chosen parameters

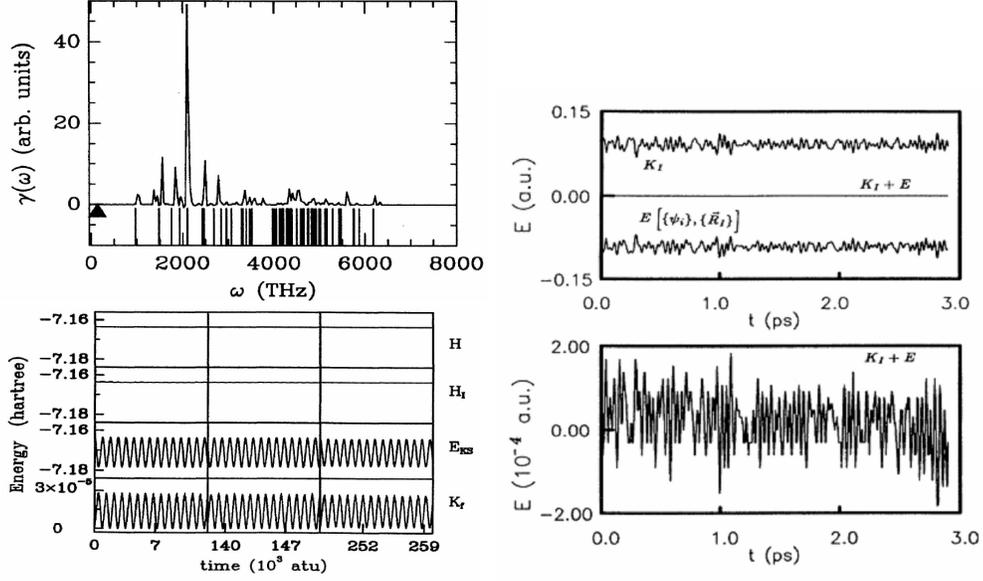


Fig. 27. Left panel: Model system (see text). Upper panel: Power spectrum of the electronic subsystem and harmonic approximation (Eq. (15)) thereof (line spectrum) [191]. The triangle indicates the highest optic phonon frequency of the ionic subsystem. Lower panel: Various energies and constants-of-motion, Eqs. (306) – (309) [191]. Right panel: Results for a real system, model of amorphous silicon described by 64 atoms arranged in a cubic cell ($E_{\text{cut}} = 12$ Ryd, $\mu_i = \mu = 612$ a.u., time step $\Delta t = 0.24$ fs in microcanonical ensemble) [192]. Various energies and constants-of-motion.

the electronic and ionic subsystems are dynamically decoupled, their power spectra do not have any overlap and hence the equilibration and the associated energy transfer from the “hot” ionic subsystem to the “cold” electronic subsystem does not exist, or is taking place on time-scale irrelevant compared to simulation time scales. Further details corroborating this argument are shown in the same figure in terms of the following quantities

$$E_{\text{cons}} = H = \sum_i \frac{1}{2} \mu_i \langle \dot{\varphi}_i | \dot{\varphi}_i \rangle + \sum_I \frac{1}{2} M_I \vec{R}_I^2 + E \left[\{ \varphi_i \}, \{ \vec{R}_I \} \right], \quad (306)$$

$$E_{\text{phys}} = H_I = \sum_I \frac{1}{2} M_I \vec{R}_I^2 + E \left[\{ \varphi_i \}, \{ \vec{R}_I \} \right] = E_{\text{cons}} - K_e, \quad (307)$$

$$V_e = K_e = E \left[\{ \varphi_i \}, \{ \vec{R}_I \} \right] \quad (308)$$

$$K_e = \sum_i \frac{1}{2} \mu_i \langle \dot{\varphi}_i | \dot{\varphi}_i \rangle. \quad (309)$$

The quantity E_{cons} is a classical constant of motion which should remain constant as all constraints are holonomic. As shown in Fig. 27, this is indeed the case. Note also that the kinetic

energy of the electrons K_e exhibits bound small amplitude oscillations with no signs of the electronic degrees-of-freedom heating up. This is very important, as K_e is proportional to the deviation of the electronic structure from the ground-state, which has to be very small.

Another important feature is that the electronic degrees-of-freedom execute high-frequency oscillations around the minimum. This is because Eqs. (299) and (302) are second order differential equations. In BOMD first order equations ($\dot{\varphi}_i(\vec{r}, t) = -\hat{H}_e \varphi_i(\vec{r}, t) + ortho$) are used and, unless precautions are taken (see chapter 3.3.2.2 above), there is a systematic error generating a drag force on the ions (see, Fig. 26).

Under which conditions can the adiabatic separation between the electronic and the ionic degrees-of-freedom, we have seen in the model example above, be achieved and sustained? It can be shown by simple harmonic analysis that the frequency spectrum of the classical electronic degrees-of-freedom close to the minimum is [191]

$$\omega_{ij} = \left(\frac{2(\varepsilon_i - \varepsilon_j)}{\mu} \right)^{\frac{1}{2}} \quad (310)$$

where ε_i and ε_j are eigenvalues of occupied and unoccupied orbitals, respectively. Comparison of the power spectrum and the harmonic approximation of Eq. (15) in Fig. 27 shows, that the harmonic approximation works well. In the same spirit we can estimate the lowest electronic frequency to be related to the electronic gap, E_{gap} , by

$$\omega_e^{\min} \propto \left(\frac{E_{gap}}{\mu} \right)^{\frac{1}{2}} \quad (311)$$

and, in plane wave expansion of the orbitals, by the same argument the highest frequency is related to the energy cut-off, E_{cut} , by

$$\omega_e^{\max} \propto \left(\frac{E_{cut}}{\mu} \right)^{\frac{1}{2}}. \quad (312)$$

By the same token, the maximum time step in the discretization of the equations-of-motion (302), (303) is inversely proportional to the highest frequency in the system, ω_e^{\max} , and hence

$$\Delta t^{\max} \propto \left(\frac{\mu}{E_{cut}} \right)^{\frac{1}{2}}. \quad (313)$$

Typical values for system with a band gap are $\mu = 500 \div 1500$ a.u. and $\Delta t = 5 \div 10$ a.u.

So far we have only discussed results of one model system [191]. In Fig. 27 we show results from a realistic simulation, amorphous silicon at room temperature modeled by 64 atoms with periodic boundary conditions applied [192]. The results are of similar quality and identical conclusions to those from the model system can be drawn also for the realistic system.

More generally the following theorem can be proven [193]. There are constants $C > 0$ and $\mu^* > 0$ such that

$$\Delta_\mu = \left| \vec{R}^\mu(t) - \vec{R}^0(t) \right| + \left| |\varphi^\mu(t)\rangle - |\varphi^0(t)\rangle \right| \leq C\mu^{\frac{1}{2}}, \quad 0 \leq t \leq T \quad (314)$$

and the fictitious kinetic energy satisfies

$$K_e = \frac{1}{2}\mu \langle \dot{\varphi}^\mu(t) | \dot{\varphi}^\mu(t) \rangle \leq C\mu, \quad 0 \leq t \leq T \quad (315)$$

for all values of the parameter μ satisfying $0 \leq \mu \leq \mu^*$, where up to time $T > 0$ there exists a unique nuclear trajectory on the Born-Oppenheimer surface with $\omega_e^{\min} > 0$ for $0 \leq t \leq T$, i.e. there is always a finite electronic excitation gap. In Eqs. (314), (315), μ and 0 indicate trajectory obtained by CP dynamics and BO dynamics, respectively. Note that the theorem above guarantees that not only the nuclear trajectory but also the wave function will stay close to the correct one.

Do the above arguments mean that the CP dynamics is only applicable to systems with gaps? What happens if we want to treat gapless systems, or systems with very small gaps, $E_{gap} \rightarrow 0$? An obvious solution is to control the deviation of the electronic degrees-of-freedom from the BO PES. This can be done by reoptimization of the electronic degrees-of-freedom periodically, as was done in the first applications of the CP method to metallic systems [194, 195]. A more elegant solution was proposed in Ref. [196]. The electronic and ionic subsystems were coupled to separate Nosé-Hoover thermostats [146–148], Eqs. (235), (236). In the most general case (see chapter 3.2.1 on application of thermostating above) a chain of L thermostats can be attached to electron subsystem

$$\mu_i \ddot{\varphi}_i(\vec{r}, t) = -\hat{H}_e \varphi(\vec{r}, t) + \sum_{ij} \varepsilon_{ij} \varphi_j(\vec{r}, t) - \mu_i \dot{\eta}_1(t) \dot{\varphi}_i(\vec{r}, t), \quad (316)$$

$$Q_1^e \ddot{\eta}_1(t) = 2 \left[\sum_i^{occ} \mu_i \langle \dot{\varphi}_i(\vec{r}, t) | \dot{\varphi}_i(\vec{r}, t) \rangle - T_e^0 \right] - Q_1^e \dot{\eta}_1(t) \dot{\eta}_2(t), \quad (317)$$

$$\vdots \quad \quad \quad \vdots$$

$$Q_l^e \ddot{\eta}_l(t) = [Q_{l-1}^e \dot{\eta}_{l-1}^2(t) - k_B T_e^0] - Q_l^e \dot{\eta}_l(t) \dot{\eta}_{l+1}(t) (1 - \delta_{lL}), \quad l = 2, 3, \dots, L, \quad (318)$$

and a chain of K thermostats to the ions

$$M_I \ddot{\vec{R}}(t) = -\nabla E - M_I \dot{\xi}_1(t) \dot{\vec{R}}(t), \quad (319)$$

$$Q_1^n \ddot{\xi}_1(t) = \left[\sum_I M_I \ddot{\vec{R}}_I - 3N_I k_B T \right] - Q_1^n \dot{\xi}_1(t) \dot{\xi}_2(t), \quad (320)$$

$$\vdots \quad \quad \quad \vdots$$

$$Q_k^n \dot{\xi}_k(t) = [Q_{k-1}^n \dot{\eta}_{k-1}^2(t) - k_B T] - Q_k^n \dot{\eta}_k(t) \dot{\eta}_{k+1}(t) (1 - \delta_{kK}), \quad k = 2, 3, \dots, K. \quad (321)$$

The thermostats counterbalance the energy transfer from the “hot” ions to the “cold” electrons and thus significantly contribute to maintaining adiabaticity [197]. The CP method has to be used with extreme caution when applied to gapless systems. The additional complication is that the orbitals cannot have variable occupation numbers $\{f_i(\beta)\}$. In most cases the simulation temperatures are low enough to be approximated by $T \approx 0$ and the variable occupation numbers are used to allow charge flow between different k -points used to sample the Brillouin zone [128–130]. Hence, the use of fixed occupation numbers also means that the system should be large enough to consider the Brillouin zone to be point-like. There are also accuracy issues.

Consider the system from chapter 2, of a molecule dissociating over a surface. The dissociation is normally electronically driven by populating more antibonding molecular state(s). The repopulation is induced by the molecule-surface interaction. Such a process may be hard to correctly capture by the CPMD, while relatively easy to include in the explicit diagonalization of the electronic Hamiltonian in the BOMD. These arguments suggest, that for metallic systems or systems where the electronic structure exhibits pronounced qualitative variations, the BOMD may be the preferred approach [26].

How does the numerical efficiency of the CP method compare to the EMD and the BOMD? Within plane wave basis sets, fixed by the plane wave cutoff energy E_{cut} , the highest electronic frequency in the EMD is $\omega_e^{EMD} \sim E_{cut}$, whereas in CPMD $\omega_e^{CPMD} \sim (E_{cut}/\mu)^{\frac{1}{2}}$ (Eq. (312)). Hence, the electronic frequency increases much more slowly with E_{cut} in the CPMD than in the EMD. Taking realistic values for μ yields $\tau_e^{CPMD} \approx 10^{-15}$, which is about one order of magnitude larger than $\tau_e^{EMD} \approx 10^{-16}$. Comparison with the BOMD is harder to draw, as the electronic time-scale is irrelevant. For very slow ionic processes and/or metallic systems the BOMD, which allows much larger ionic time steps to be taken, may be the method of choice. On the other hand, for systems with a gap and a reasonably fast ionic dynamics, the CPMD is almost invariably the faster and also more convenient method.

3.3.4 From dynamics to optimization

Often we are not interested in full system dynamics. Imagine, for example, that the task was to determine the surface structure of a reconstructed surface. We could either use our intuition and come up with a series of competing surface structures or could use the experimental information in terms of, e.g., scanning tunneling microscopy (STM) images. In the first case we would have to find the equilibrium ground-state structures for the surface structure models under consideration. In the second case, we would have to guess the surface structure from the experimental image²⁰ and compute the simulated STM image for the proposed structure and compare the simulated and experimental image. In either case a *relaxed surface structure* is a necessary prerequisite. Perhaps the simplest way of stating this problem is by starting from the dynamics equations (302), (303), and reformulating them as *steepest descent (SD) equations*

$$\mu_i \dot{\varphi}_i(\vec{r}, t) = -\hat{H}_e \left[\{\varphi_i\}, \{\vec{R}_I\} \right] \varphi_i(\vec{r}, t) + \underbrace{\text{constraints}}_{\text{orthonormality}}, \quad (322)$$

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I E \left[\{\varphi_i\}, \{\vec{R}_I\} \right]. \quad (323)$$

The equations should be used so, as to first allow the electronic degrees-of-freedom to converge to the ground-state for the initial configuration of atoms $\{\vec{R}_I^0\}$. In such a case $-\nabla_I E[\{\varphi_i\}, \{\vec{R}_I\}] = \vec{F}_I^0$. Using these forces in solving Eq. (323) will lead to a series of ionic configurations $\{\vec{R}_I^m\}$ which in the limit $m \rightarrow \infty$ will yield the (local) atomic structure. In a more elaborate search one could use the ideas of simulated annealing [187]. In many cases an educated initial guess for the starting structure $\{\vec{R}_I^0\}$ will help to speed up the search for the desired relaxed atomic structure. Clearly, similarly as in the case of electronic structure optimization (see above) the

²⁰The STM images provide information on electronic rather than atomic structure.

approach outlined in Eqs. (322), (323) is very inefficient. An alternative and more efficient way is based on use of CG [176] or DIIS [181, 182] optimization methods for both electronic and ionic degrees-of-freedom. A very significant body of applications is in the spirit of optimization.

3.3.5 Parallelization and $O(N)$ methods

The computational load associated with the combined electron-ion dynamics is often prohibitive. This load can be to a certain degree reduced by choosing the optimal technique (EMD, BOMD, CPMD), but even with the best technique many simulations are requiring production runs of length which is impractical. The length- and time-scale are fixed by the nature of the problem studied, and hence the computational speed up must be achieved by other means than reduction of the system size or time. Two techniques are widely used: (a) spreading the load over more processors, i.e. use of *parallel computing*, and (b) use of methods that scale linearly with the system size or so-called *$O(N)$ methods*. Both are dealing with the length-scales. To make the time-scales more manageable still represents a major challenge²¹.

We briefly introduce the techniques which help to deal with the length-scales. With the development of parallel supercomputers, it has become clear, that it will be possible to reduce the workload by distributing it over more processors. At variance with classical molecular dynamics, due to huge imbalance between the workload needed for integrating the equations-of-motion for the electronic and ionic degrees-of-freedom, it makes sense to parallelize only over the electronic degrees-of-freedom. In detail, the parallelization strategies depend on the method to be parallelized. Nowadays, most of the electronic structure methods, including the many-body quantum chemistry methods and QMC methods, have been parallelized. One of the first attempts at parallelization in this field was parallelization of the DFT plane wave pseudopotential codes [184, 200, 201]. We take that method as an example. Using plane wave basis representation we have the following mapping of the wave function $\varphi_i(\vec{r})$ into an array C ,

$$\varphi_i(\vec{r}) = \varphi_{n,\vec{k}}(\vec{r}) = \frac{1}{\Omega} \sum_{\vec{G}} c_{\vec{G}}^{n,\vec{k}} \exp \left[i \left(\vec{k} + \vec{G} \right) \cdot \vec{r} \right] \Rightarrow C(M, N_b, N_k), \quad (324)$$

where n is the band index, \vec{k} indicates the k -points used to sample the Brillouine zone, and Ω is the unit cell volume. In array C , M represents the number of plane waves used to expand the wave functions, N_b represents the number of electronic bands, and N_k denotes the number of k -points. These quantities typically scale with the number of atoms N_a in the following way

$$\begin{aligned} M &\sim (100 - 1000) N_a, \\ N_b &\sim (1 - 10) N_a, \\ N_k &\sim 1/N_a. \end{aligned}$$

Observing the structure of the wave function representation, Eq. (324) it is easily seen that one can consider a number of data driven approaches to parallelizing the calculations.

The simplest, and technically least demanding solution is parallelization over the k -points. However, as we want to use parallel computers to solve the large systems, which large unit cells

²¹Recently techniques have been introduced, such as metadynamics [198, 199], which in a controlled way renormalize the underlying potential energy surface (PES) and hence make simulations of even slow processes feasible.

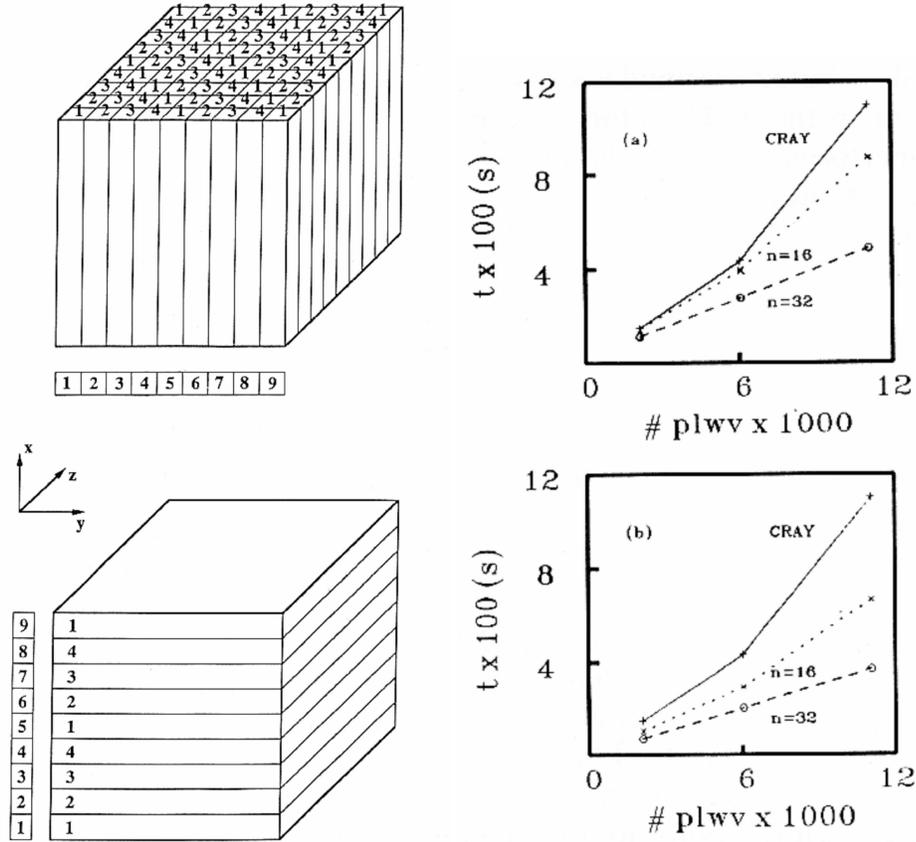


Fig. 28. Left panel: Example of division of parts of reciprocal (upper panel) and real space (lower panel) over nodes. A cubic cell, a uniform $9 \times 9 \times 9$ FFT grid and four computing nodes are assumed. The x -column index in reciprocal space and the yz -plane index in real space is indicated by indices (1-9) in square boxes, the node index (1-4) is also shown. Right panel: Variation of CPU time with basis set size for two different parallel platforms (upper and lower panel) compared with a single processor computer (CRAY X-MP) is shown as a function of the load (# of plane waves) for otherwise identical system. The number of processors ($n = 16$ and 32) are shown as parameters.

and correspondingly small Brillouin zones, this way of distributing the load over the processors is the least efficient one. More appealing is the divide by band, which represents a viable alternative. However, the most benefit from parallel execution can be taken from a *division of reciprocal and real space over the processors*. An example of division of a $9 \times 9 \times 9$ grid over 4 processors is shown in Fig. 28. All quantities are computed by parts on different nodes and the results obtained by summations over the nodes. The summations over nodes require very simple communications (simple real numbers) and hence do not represent a serious overhead. Each part of energy is computed in the space in which it is diagonal. For example the kinetic energy is computed in the

reciprocal space, while the energy due to the external potential is computed in real space. The spaces are connected by Fast Fourier Transforms (FFT). The FFTs can also be parallelized [184]. However, parallelization of the FFTs require long-range communications.

The performance tests are shown in Fig. 28. As can be seen a parallel computer can always outperform even a powerful single processor computer. The computational time for a well-loaded parallel computer always scales with the number of processors, Fig. 28. For example, by doubling the number of processors from $n = 16$ to 32, results in roughly halving the execution time around the highest computational load considered. The scaling is slightly sublinear, as there are communications overheads as well as some residual operations which are or cannot be parallelized. In the opposite limit, Fig. 28, using too many processors for too small systems, results in execution times almost independent of the number of processors. This happens, because in this case the communications overhead is larger than the time spent in computing. Examples of parallel computing for real-world applications will be shown below (chapter 4).

The parallelization strategy and efficiency is determined both by the method and also by the basis set used. Obviously, for example parallelization of QMC techniques will use completely different strategies than the above mentioned DFT plane wave pseudopotential technique. Currently most of the electronic structure methods are parallelized. As for the basis sets, the highest parallel execution is typically achieved by using plane waves or real-space grids. Use of localized basis sets results normally to somewhat lower parallelization efficiency.

Another way of approaching large length-scales in electronic structure calculation is use of methods that scale linearly with the system size, say the number of electrons. These methods are called order N , or $O(N)$ methods. $O(N)$ methods evoke the “shortsightedness” of the electronic structure. This means, that, for example, the charge density at point \vec{r} , $\rho(\vec{r})$, depends only weakly on the charge density at point \vec{r}' for $|\vec{r} - \vec{r}'| > R_c$, R_c being a cutoff beyond which the fragments are considered to be non-interacting.

If such a scheme could be established without compromising the accuracy, the consequences would be far reaching both in the single-particle as well as in the many-body description. In the single-particle case, one could split the “large” system into smaller fragments which would be much cheaper to deal with. In the many-body approach the multireference wave function description this would lead to substantial reduction in the number of integrals that need to be evaluated. As a byproduct, solution of the $O(N)$ problem would automatically solve also the embedding problem²².

Many variants of $O(N)$ single-particle methods exist [202–208]. $O(N)$ variants even of many-body methods, such as CI or QMC [209], have been proposed. We shall not go into details of these techniques. In general, the above mentioned division into non-interacting/weakly interacting subunits is never perfect. In other words, if the “perfect” results are to be reproduced, the numerical overhead is equal to the one for the “perfect” calculation. In addition, use of the $O(N)$ methods often requires an “expert” control over a number of parameters which control the efficiency and the accuracy of the calculation. Only an optimum choice of these control parameters leads to the desired simultaneous reduction of the computational cost and a tolerable error in the obtained results. On the other hand, many of the applications in areas where system size

²²In the embedding problem the system under consideration is divided in different areas, which are treated at different levels of accuracy. For example, the most important part of the system may be treated using the most accurate quantum method, followed by a part treated by less accurate method, semiempirical method, empirical method, and finally by a continuum model.

is the major issue, such as materials science or bioscience [210] would not have been achieved without the $O(N)$ methods.

4 Applications: From the silicon crystal to technological processes

In this part we present a selection of applications deemed to demonstrate both the enormous power of the techniques in dealing with description of matter at atomic/molecular scale as well as the huge leap forward the techniques have made over the last two decades. A very important milestone in this trail was the publication of the seminal paper by Car and Parrinello [186]. On one hand, it may be seen merely as a technical contribution to solving the BOMD. However, a critical look back reveals that that contribution was far deeper. One can safely say that a new window on the atomic/molecular/nano-scale has been opened. What has changed was the ability to open a computer-based “*virtual laboratory*”. For the first time it became possible to “shuffle” atoms on the computer and watch the outcome of this virtual nanoscience or nanomanipulation. Obviously, real laboratory atomic/molecular-scale nanoscience has witnessed a similarly dramatic development. This is a very fortunate harmony, as the experiment without the simulation backup is typically perceived as incomplete, and obviously vice-versa, simulations without an initial experimental stimulation are futile. In what follows is a very personal view of the applications of the *ab-initio* modeling methods, mainly drawn from the author’s own work, with no aspiration to cover all the facets of the atomic/molecular-scale modeling. Rather than hopping through many different topics, applications which fall into natural groups, such as surfaces, clusters or surface dissociation of molecules have been given priority.

4.1 Structurally disordered solids and liquids

4.1.1 Liquid and amorphous silicon

One of the earliest applications of the *ab initio* molecular dynamics techniques was to the disordered Si phases: *liquid Si* (l-Si) and *amorphous Si* (a-sSi). This system was the testing ground for the then new Car-Parrinello technique [186]. While for a-Si several structural models existed and the properties were fairly well known both experimentally and theoretically, the structure and properties of the l-Si were in the nineties a mystery. In addition, l-Si is still of key technological importance in production of Si wafers for electronic application. Understanding the melt structure plays a crucial role in efforts to limit the defect concentration in the grown wafers.

Upon melting l-Si goes from a fourfold semiconducting structure to a metallic low coordinated liquid with a density $\sim 10\%$ higher than in the solid and structure factor dissimilar to that of most liquid metals. The average coordination number is between 6 and 7, while most liquid metals are closely packed with coordination number $\sim 12 - 14$. The low coordination number indicates a persistence of covalent bonding in the liquid. The problem point was also how to reconcile the persistence of covalent bonding with metallic behavior and fast diffusion.

Such a complex wealth of contradictory physical properties was an ideal ground for techniques of *ab initio* statistical mechanics. Not surprisingly, approaches based on empirical interatomic potentials [211–214] have had only limited success.

The results [194, 195] of one of the first plane wave pseudopotential CPMD simulation (Eqs. (302), (303)) to real materials (64 atoms arranged in a simple cubic cell, $T_I = 1800K$,

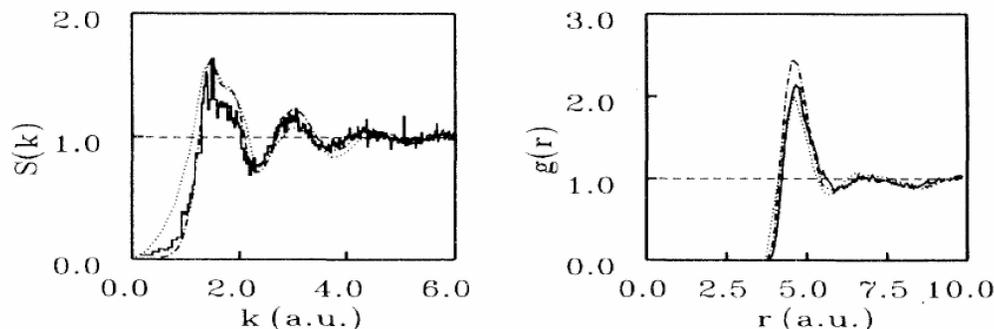


Fig. 29. (a) Static structure factor and (b) pair correlation function. Full line: simulation; dotted line: neutron diffraction experiment [215]; dash-dotted line, x-ray diffraction experiment [216].

$T_e = 0K$, DFT in LDA approximation for exchange and correlation energies, Brillouin zone sampled just at a single k -point, $E_{cut} = 12$ a.u., $\Delta t = 5.5$ a.u., $\mu = 300$ a.u., $t_{sim} = 1.2$ ps after equilibration) are shown in Figs. 29 – 32. As l-Si is metallic, an approximate adiabaticity of the CPMD simulation was maintained by reoptimization of the electronic degrees-of-freedom every 500 time steps. The electronic reoptimization was used in combination with a Nosé-Hoover thermostat (Eqs. (235), (236)); see chapter 3.3.3 for alternative solutions.

The results obtained for the structure of l-Si, shown in Fig. 29, are in impressive agreement with both neutron [215] and x-ray [216] diffraction experiments. The coordination number obtained by integration of $g(r)$ to the first minimum is ~ 6.5 in perfect agreement with the experimental value of ~ 6.4 [215, 216]. Estimation of persistence of covalent bonding may be assessed by triplet correlations, measured, for instance by bond-angle distribution function $g_3 \cdot g_3$ (not shown) exhibits a broad distribution peaked around angles close to the tetrahedral value.

The simulation generates simultaneously the ionic trajectories, Eq. (303) and the corresponding ground-state electronic structure, Eq. (302). In this way the concept of persistence of covalent bonding effects in l-Si can be put on a quantitative basis. In Fig. 30 the evolution of electronic pseudo-charge-density $\rho(\vec{r})$ is shown in a plane defined by three neighboring atoms and compared (Fig. 30(a)) with $\rho(\vec{r})$ is the (110) plane of crystalline silicon (c-Si). In Figs. 30(b) and (c) the two distances between the triplet of atoms are quite close to the bond length in c-Si. Correspondingly, the electronic densities have several common characteristics, such as a strongly nonuniform $\rho(\vec{r})$ and charge accumulation between pairs of adjacent atoms. This provides evidence of persistence of covalent bonding in the liquid. Its subsequent evolution is shown in Figs. 30(c)–(i). We can see how the charge densities respond to the atomic motion. When the interatomic distances increase, the electronic pile-up between the atoms disappears. Also evident is the strong correlation between the covalent bonding (Fig. 30(b)–(d)) and the triplet correlations (bond-angles close to the tetrahedral angle of $\sim 109^\circ$).

The atomic trajectories allow study of time-dependent phenomena and transport properties²³.

²³Note that transport coefficients can be obtained (within linear response, or Kubo theory) from equilibrium simulation.

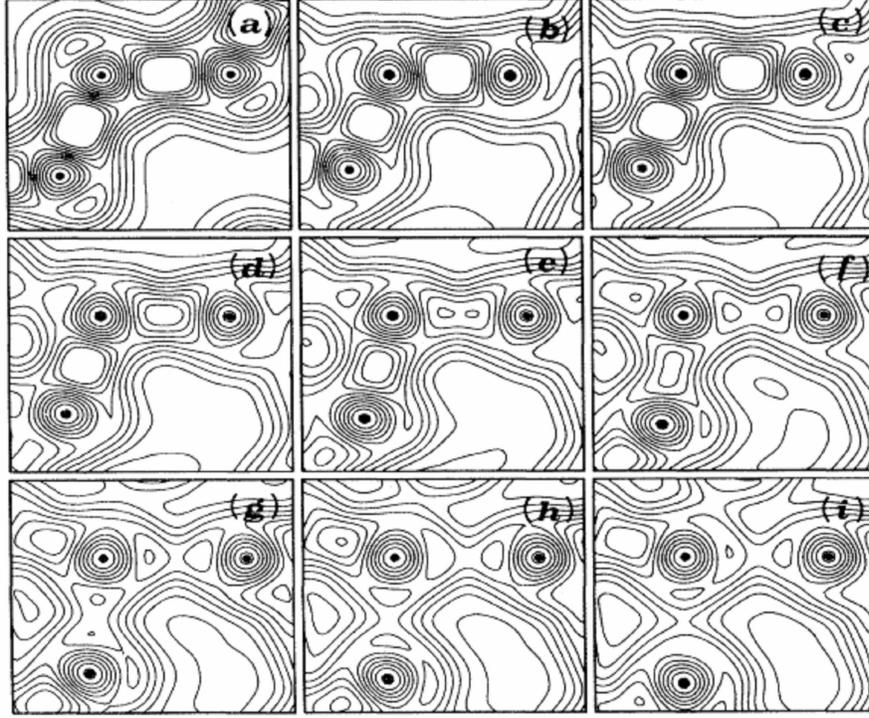


Fig. 30. Contour plots of the valence electronic density $\rho(\vec{r})$. (a) c-Si in the (110) plane. (b)–(i) evolution of $\rho(\vec{r})$ in l-Si at time intervals of $\sim 5.5 \times 10^{-3}$ ps. The dots indicate the positions of the ions.

The diffusion coefficient can be obtained from the mean-square displacement [194, 195, 217]

$$R^2(t) = \frac{1}{N_I} \overline{\sum_{I=1}^{N_I} \left[\vec{R}_I(t) - \vec{R}_I(0) \right]^2} \sim 6Dt + c, \quad \text{as } t \rightarrow \infty, \quad (325)$$

where D is the self-diffusion coefficient and c a constant. The horizontal bar indicates statistical averaging. The long-time behavior of $R^2(t)$ exhibits a quasilinear behavior from which we extract $D \sim 2.3 \times 10^{-4} \text{ cm}^2 \text{ s}^{-1}$. This extremely high value is at first glance at variance with the persistence of covalent bonding. However, as can be inferred from Fig. 30, the dominant process determining the properties of l-Si is a very rapid process of continuous bond breaking and forming processes. The equilibrium between the two processes gives rise to simultaneous existence of seemingly contradictory processes of fast atomic diffusion and covalent bonding.

The persistence of covalent bonding can be traced also in the dynamical properties. The velocity autocorrelation function

$$Z(t) = \frac{\overline{v(t) \cdot v(0)}}{\overline{v(0) \cdot v(0)}} \quad (326)$$

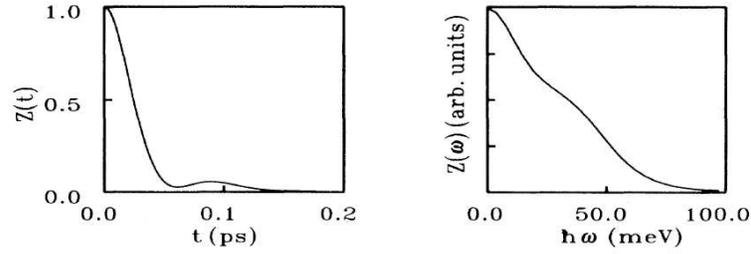


Fig. 31. (a) Velocity autocorrelation function and (b) the corresponding power spectrum.

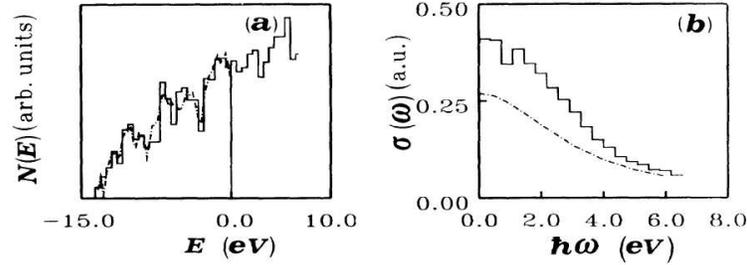


Fig. 32. Electronic properties of l-Si. (a) densities of Kohn-Sham eigenvalues from the simulation. The vertical line indicates the Fermi level. (b) Electrical conductivity from the Kubo-Greewood formula. Drude fit [219] is shown by dashed-dotted line.

along with the spectral density

$$Z(\omega) = \frac{2}{\pi} \int_0^{\infty} dt Z(t) \cos(\omega t) \quad (327)$$

is shown in Fig. 31. The horizontal bars in formula (326) indicate statistical averaging. $Z(t)$ is always positive, leading to a high value of the diffusion coefficient D . Note, that $Z(t)$ has an oscillatory decay to zero after ~ 1.15 ps. This is unusual, because most simple liquids exhibits have a negative oscillation in $Z(t)$. This is due to the so-called caging effect of the shell of neighboring atoms [218]. The caging affect can be thought of as ion bouncing against the cage formed by the first coordination shell of atoms and, as effect of the collision, reversing the velocity. l-Si has much more open structure and exhibits no caging, but only effects due to the occasional formation of covalent bonds. The power spectrum $Z(\omega)$ exhibits the low frequency diffusive modes. In addition, vibrational modes reflecting covalent bonding are visible at around 40 meV, close to the optical vibrational frequency of c-Si below the melting point.

Another manifestation of the insight into the electronic properties of the system is shown in Fig. 32, where the single-particle electronic density of states $N(E)$ is reported. The $N(E)$ exhibits metallic behavior as evidenced by the absence of a gap at the Fermi energy. Given the persistence of covalent bonding, the nearly free-electron behavior of $N(E)$ may come as a surprise. However, the free-electron like behavior is in agreement with x-ray-emission experiments [220].

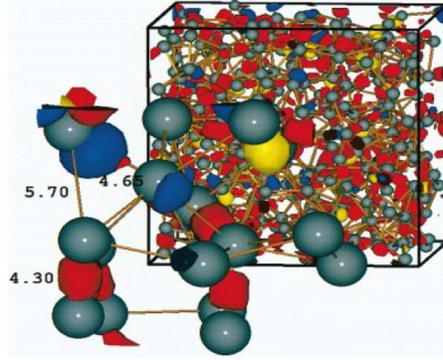


Fig. 33. Snapshot of a low-spin configuration of l-Si. Superimposed on the ball-stick model are isosurfaces corresponding to the total charge of covalent bond (red) spin \uparrow (yellow), spin \downarrow density (blue). The superimposed zoom shows the lower left part of the unit cell. The numbers correspond to the lengths (in a.u.) of the bonds.

In Fig. 32 we also report the electronic conductivity calculated from the Kubo-Greenwood formula [221]

$$\sigma(\omega) = \frac{2\pi}{\Omega} \sum_{i,j} \frac{|M_{ij}|^2}{\omega_{ji}} \delta(\varepsilon_j - \varepsilon_i - \hbar\omega), \quad (328)$$

where $M_{ij} \propto \langle \varphi_i | -i\partial/\partial\mathbf{r} | \varphi_j \rangle$. Despite the short electronic mean free path τ , the calculated $\sigma(\omega)$ exhibits Drude-like falloff $\sigma(\omega) = \sigma(0)/(1 + \omega^2\tau^2)$ in agreement with experiments [219]. The extrapolated value of $\sigma(0)$ gives $\sigma_{dc} = 0.38$ a.u. in good agreement with the experimental value of 0.27 a.u.

One might wonder why the electronic properties calculated from DFT theory, where the single particle energies formally play merely the role of the Lagrange multipliers (see chapter 3.1.5.3), exhibit overall good agreement with the experimentally measured values. The experience shows, that the single-particle levels, except for the discontinuity of E_{xc} and the associated error in the band gap (see chapter 3.1.5.10), which are irrelevant here, may often be a reasonable approximation to the excitation energies.

One might also worry about the system size. Can 64 atoms arranged in a $4 \times 4 \times 4$ cube with applied periodic boundary conditions really reliably capture the liquid disorder? Finally, the first set of simulations was done employing the spin unpolarized version of DFT [194, 195, 217]. A legitimate question arises about spin polarization effects as the covalent bonds are broken. l-Si exhibits also interesting temperature dependence of density, viscosity, surface tension, and magnetic properties [222]. The latter raise the question of what the magnetic field couples to. In order to provide the answers to some of those nagging questions, another simulation was performed [223] with much larger simulation cell ($8 \times 8 \times 8$) employing BOMD and the spin-polarized version of the GGA functional [58]. We have found that the original small simulation cell was entirely adequate. The spin polarization does indeed play a role, but the effect is quantitatively relatively modest. The spin polarization does develop on the bonds of intermediate lengths which are about to be broken, see Fig. 33. This spin fluctuations result in modifica-

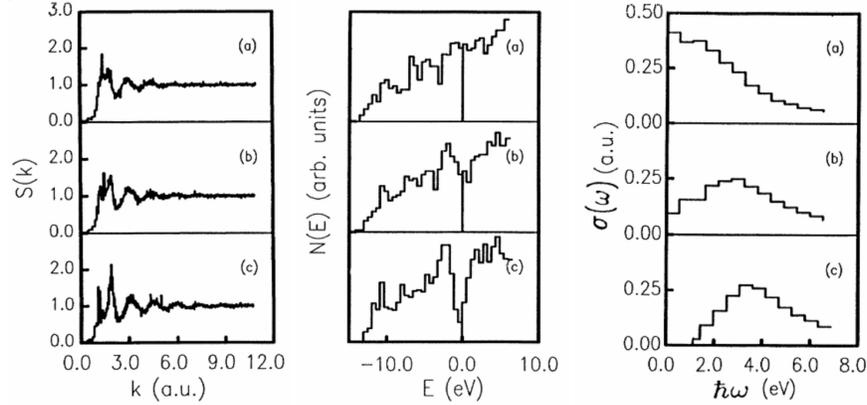


Fig. 34. Temperature dependence of the properties of the Si melt. Left panel: Static structure factor $S(k)$, Middle panel: Density of the Kohn-Sham states $N(E)$, and Right panel: a.c. electrical conductivity $\sigma(\omega)$. The temperatures in (a), (b), and (c) are $T \sim 1550$ K, $T \sim 1250$ K, and $T \sim 950$ K, respectively.

tion of the attractive part of the interatomic potential leading to a slightly higher value of the self-diffusion coefficient $D \sim 3.1 \times 10^{-4} \text{ cm}^2 \text{ s}^{-1}$ to be compared to the original value of $D \sim 2.3 \times 10^{-4} \text{ cm}^2 \text{ s}^{-1}$ and a slightly improved agreement of computed static structure factor with experiment [224].

All in all, the original simulations of l-Si, which were among the very first applications of the *ab-initio* CPMD techniques to real systems, confirmed the enormous power and wealth of properties hitherto unheard of.

The next objective in this study was the amorphous phase of Si (a-Si) [192]. Despite the fact that the structure and properties of a-Si were far better known than those of the liquid counterpart, there were still a number of issues requiring the “first-principles” insight, such as the nature of the dominant defects in disordered tetrahedral networks and their link with their physical properties, or the dependence of the amorphous structures on the preparation conditions. There is a huge difference between the liquid phase, which is a well defined thermodynamic phase and an amorphous structure, which is not. The amorphous structure corresponds to a (metastable) local minimum structure, and hence is strongly dependent on the preparation conditions. A special problem was that most of the knowledge up to date was obtained with empirical interatomic potentials which were heavily biasing the obtained results. With regard to the *ab-initio* techniques the big question was how well these techniques will be able to describe the change from the liquid phase, characterized by higher coordination (~ 6.5), metallic and diffusive behavior, to the predominantly 4-fold coordinated semiconducting amorphous phase.

The liquid system was first cooled from 1800 K to room temperature at a rate $\sim 10^{14}$ K/s, about two orders of magnitude faster than the estimated laboratory cooling rates. During cooling the lattice parameter was gradually adjusted so that at the end it corresponded to the smaller lattice parameter of a-Si²⁴. As the system was still metallic, the same type of dynamics and parameters was used as in the case of l-Si. The changes which occur in the properties of the melt upon cooling are shown in Fig. 34 [192]. As can be seen, the crossover from liquid-like

²⁴This could, of course, be done by running the simulation at a constant pressure, but a higher plane-wave cutoff would be necessary.

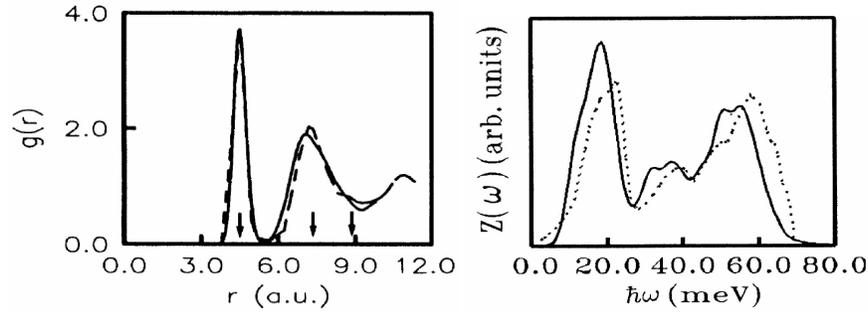


Fig. 35. Characteristics of the annealed a-Si structure. Left panel: Pair-correlation function $g(r)$. Solid line: simulation; dashed line: neutron diffraction experiment [225]. The arrows indicate the positions of the first three crystalline coordination shells. Right panel: Vibrational density of states $Z(\omega)$. Solid line: simulation; dotted line: experiment [226].

to amorphous-like character is evident from the splitting of the broad asymmetric first peak of $S(k)$ in l-Si. At $T \sim 950$ K $S(k)$ is practically indistinguishable from that of a-Si at room temperature (see below). The densities of the Kohn-Sham levels evidently start to develop a pseudogap, which is accompanied by formation of a tetrahedral network. This is reflected in the a.c. conductivities which lose the metallic, Drude-like behavior and exhibit zero conductivity in the d.c. limit. This is a remarkable demonstration of how well the CPMD dynamics was able to describe the liquid (metal) to amorphous (semiconductor) transition without any need to interfere with the interatomic potential²⁵.

The system so prepared was not stable and exhibited structural relaxations. In order to stabilize the structure, the system was led through an annealing cycle during which the temperature was raised to $T \sim 900$ K let evolve and finally again reduced to 300 K [192]. The annealed system was perfectly stable with the tetrahedral order significantly improved over the as-quenched sample [192]. As a representative example we show in Fig. 35 the pair-correlation function $g(r)$ and the vibrational density of states $Z(\omega)$ computed for the annealed sample using formulas (326), (327).

The comparison of the calculated $g(r)$ [192] with experiment [225] is impressive. The agreement in the first peak is almost perfect, while some slight differences between theory and experiment are visible in the second peak, pointing to a larger bond-angle disorder in the simulated structure. Similarly, the comparison of the simulated $Z(\omega)$ [192] with experiment [226] is very good. The main differences are a small and almost rigid (~ 3 meV) shift toward lower frequencies and an overestimation of the weight of the transverse acoustic (TA) peak with respect to transverse optic (TO). The shift of the weight from the TO to the TA peak is one of the distinctive features of the phonon spectrum of a-Si compared to the crystal [226]. The fact that this is overestimated in the simulated vibrational spectrum is another indication that the structure is more disordered than real a-Si.

²⁵Such an ability is absent from the customarily used empirical interatomic potentials [213]. For example the prototypical Stillinger-Weber potential is able to describe correctly (after reparametrization) two of the three condensed phases (crystalline, amorphous, liquid), but not all three.

The slight excess of disorder, which is to be expected given the rapid cooling rate and the small (64-atom) simulation cell which add strain to the simulated sample, does not preclude study of the structural defects, which will bear some similarity with those present in real a-Si. For a long time, the dominant structural defect in a-Si, giving rise to the ESR (electron spin resonance) signal was believed to be a threefold coordinated atom, known as a dangling bond [227]. Later, instead, a fivefold coordinated atom or floating bond [228] was suggested as a viable candidate. Analysis of the annealed a-Si sample gave a substantial support to the fivefold coordinated site as the dominant structure defect [192]. Many of these defects are of weak bond type, such as a fourfold coordinated site with a nearby threefold coordinated site which, due to slight network distortions convert to a fivefold plus fourfold coordinated sites: $T_4 + T_3 \rightarrow T_5 + T_4$. In addition to the coordination defects, a-Si contains also topological defects. Unlike c-Si, which consists only of sixfold rings in the structure, there is a large variety of ring sizes, both odd and even in a-Si. It has been suggested that topological defects may have profound consequences on physical properties of amorphous solids [229], such as, e.g. the shape of the electronic density of states $N(E)$. Analysis of the topological defects has shown that no atom in the simulated structure lies in a local diamond-structure environment.

4.2 Solid surfaces

Solid surfaces represent a very important class of systems. They embody a massive defect where the atomic structure deviates from that found in the bulk. In the simplest case the *atoms relax* from their ideal bulk positions but qualitatively the bulk structure persists also on the surface. In more complicated cases, such as for example on the semiconductor surfaces, the *atoms undergo reconstruction*, whereby the atoms form surface structures significantly and qualitatively deviate from the structure in the bulk. Moreover, surfaces are very important systems for stabilizing nanostructures in nanotechnology, many industrial catalytic reactions take place on surfaces, etc. Not surprisingly, *surface science* is one of the most active fields of solid-state physics and chemistry and materials science. The necessary prerequisite to all those applications is the knowledge of the atomic and electronic structure, energetics, dynamics, chemical activity, etc. of the surface. Some of these aspects will be described for selected examples below.

4.2.1 Takayanagi reconstruction of the Si(111) surface

The Takayanagi reconstruction [230] is observed on the (111) surface of silicon, silicon-germanium alloys [231] and strained (compressed) germanium [232]. It is the one of the most complex reconstruction so far observed on any semiconductor surface. The structure of the Si(111) surface has been subject of continued interest since the first observation of the observed 7×7 reconstruction in 1959 [233]. A large number of models have been proposed for the structure but only the model proposed by Takayanagi [230] is consistent with all the available experimental measurements. The *Takayanagi model* incorporates the following features: (i) dimerization of second-layer atoms, (ii) adatoms, and (iii) a stacking fault between the first and second layers of atoms over one-half of the unit cell, see Fig. 36. Hence, the Takayanagi structure is commonly referred to as the *DAS model*.

The extremely large area and vertical extent of the observed reconstruction presented a formidable challenge to a fully *ab initio* calculation of the total energy and structural proper-

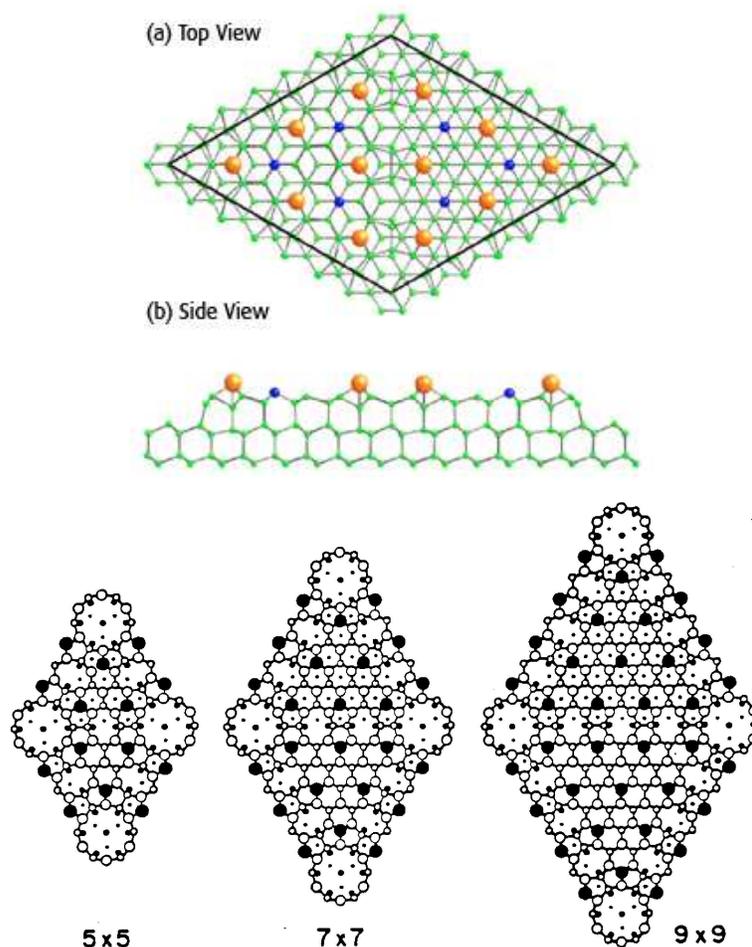


Fig. 36. Upper panel: Top and side views of the 7×7 structure. Yellow color: adatoms; Blue color: restatoms. Lower panel: Top view of a generalized series of 5×5 , 7×7 , and 9×9 Takayanagi reconstructions.

ties of this surface. As there are very subtle charge-transfer processes between dangling bonds on different surface atoms (adatoms and three-fold coordinated atoms in the first layer, the so-called restatoms), an accurate treatment of the electronic structure is required in order to capture these effects. For a long time this structure was considered to be too large for total energy methods. *Ab initio* calculations for this structure became feasible only after introduction of *parallel computing* [184] in the spirit described in chapter 3.3.5 above. These technical advances made it possible to perform a fully *ab initio* (plane wave pseudopotential) investigation of a generalized series of 3×3 , 5×5 , and 7×7 structures, see Fig. 36 and Tab. VI for details, and to compare

Tab. VI. Number of characteristic features in the Takayanagi DAS structures. The symbol $1/2$ means stacking fault over one half of the unit cell.

	dimer	adatom	stacking fault	restatom	cornerhole	atoms in supercell
3×3	3	2	$1/2$	0	1	68
5×5	6	6	$1/2$	2	1	200
7×7	9	12	$1/2$	6	1	400

their energetics [200]. Experimentally the 7×7 structure is the ground state structure for unstrained Si(111) surface. Patches of 5×5 and 9×9 are observed on strained Si and Ge surfaces. Simultaneously another group performed relaxation and total energy calculation for the 7×7 structure and compared it to the energy of the Si(100) surface [201].

Technical details of this first, truly *large-scale calculation* were as follows. The electron-electron interactions were included in the local-density approximation of density-functional theory in Perdew-Zunger's parametrization [38]. For the 5×5 and 7×7 reconstruction the unit cell is large enough for the Brillouine zone to be sampled by a single k point, the Γ point. The calculation for the 3×3 structure was performed using four k -points to provide a comparable accuracy of Brillouine zone sampling. The electronic wave functions were expanded in plane waves with an energy cutoff of 7 Ry. The calculations were performed using slabs eight layers thick with the central two layers kept fixed to simulate the bulk crystal termination of the surface. A slab with two surfaces was considered with center of symmetry through the slab center.

The results of the calculations for the surface energy are shown in Tab. VII. The 7×7 DAS structure has the lowest energy and the 5×5 and 3×3 are decreasingly stable. However, it can be seen that the changes in the surface energies are saturating towards the 7×7 structure, which indicates that the 9×9 and larger structures may not be energetically favorable. These conclusions are corroborated also by analysis of the trends in the structural parameters [200]. The dominant process in the DAS structure stabilization is an increased stabilization of the adatoms which move inward towards the ideal tetrahedral positions across the series from 3×3 to 7×7 . A similar trend is observed also for the restatoms in the 5×5 and 7×7 structures. This trend can, in turn be explained by charge transfer processes which can be assessed from the computed charge densities $n(\vec{r})$ shown in Fig. 37. The main trend we observe is an increasing transfer of charge from dangling bonds on the adatoms to the dangling bonds on the restatoms across the series from 3×3 to 7×7 , which is to be expected from the increasing ratio of restatoms to

Tab. VII. Results for the surface energy calculations.

	3×3	5×5	7×7
Energy per unit cell [eV]	10.765	29.205	56.509
Energy per surface atom [eV]	1.196	1.168	1.153

adatoms. The decrease in the height of the adatoms can be explained by the reduction in the charge in the dangling bonds on these adatoms and the consequent tendency towards sp^2 rather than sp^3 hybridization. The charge density in the dangling bonds on the restatoms is smaller in the 7×7 reconstruction than in the 5×5 because of the higher ratio of rest atoms to adatoms. This explains the reduction in the height of the restatoms on moving from the 5×5 to the 7×7 structure. Interesting variations between the faulted and unfaulted halves of the unit cell was observed. In the 3×3 and 5×5 reconstructions the charge densities on the adatoms in the two halves of the unit cell are broadly similar. In the 7×7 reconstruction the charge densities in the two halves of the unit cell are very different. In the unfaulted half of the unit cell the charge transfer from the adatoms is significantly larger than in the 5×5 structure and it is the same for all the adatoms. In the faulted half of the unit cell the adatoms adjacent to the corner hole have more charge than the adatoms in the middle of the unit cell and both adatoms have more charge than the adatoms in the unfaulted half of the unit cell. These results are consistent with scanning tunneling microscope images of the surface; see Fig. 37.

These simulations [200, 201] have had a very large impact on the field of the atomic/molecular-scale computer physics and chemistry as they paved the way for the truly large-scale applications which followed.

4.2.2 Dynamics of the Takayanagi reconstructed Si(111)- 7×7 surface

After having established the energetics and the mechanism of stabilization of the Si(111)- 7×7 surface we turn to the *dynamics* of this surface. We show below, how the dynamics can be determined in a nonstandard and computationally very appealing way. The dynamics of this surface has been subject of interest since electron energy-loss spectroscopy (EELS) [235] found a split-off 570 cm^{-1} vibrational mode, 10 % higher than the highest bulk mode of Si, and a broader feature near 240 cm^{-1} , see Fig. 38. The presence of the high-frequency mode is difficult to reconcile with the usual picture of a reconstructed semiconductor surface, where bonds typically undergo stretching.

While the EELS experiment identifies the unusual vibrational modes, it provides no answer to the origin of these modes. This is a typical situation where a theoretical backup is required. Several attempts have been made to shed light on the experimental results. Cluster model and *ab initio* DFT calculation [235] assigned both the ~ 570 and $\sim 240 \text{ cm}^{-1}$ modes to a z-polarized vibration localized on the adatom and second-layer atom underneath the adatom moving out of and in phase, respectively, see Fig. 38(a). Empirical interatomic potentials [236] found a more complicated picture, with both modes localized on six atoms, see Fig. 38(b). However, both models are based on methods which have well-known limitations. The cluster model is typically a very poor representation of a surface which suffers from finite-size effects. The charge transfer processes discussed above are in this model ignored altogether. The empirical interatomic potentials are known not to be capable of a realistic modeling of semiconductor surfaces, and hence it is not clear why the dynamical properties should not be plagued by the same problem. In particular, the charge transfer processes so important for this surface are certainly not included in this model.

Not surprisingly, a fully *ab initio* model based on accurate DFT treatment with a complete model of the Takayanagi reconstruction (DAS surface model) [237, 238] provides not only quantitatively more accurate picture but, more importantly, the results, schematically shown in Fig. 38

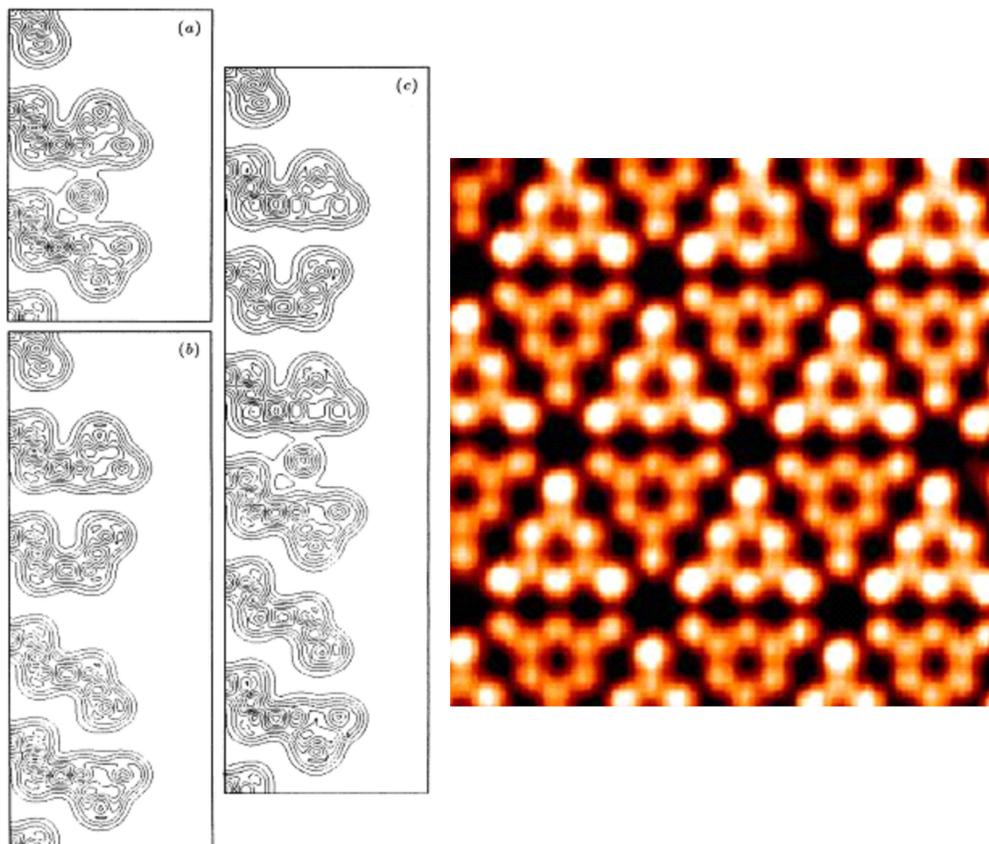


Fig. 37. Left panel: Contour plots of the charge density along the long diagonal in the unit cell of (a) 3×3 , (b) 5×5 , and (c) 7×7 Takayanagi DAS structures. The faulted half of the unit cell is in the upper part of the unit cell. Right panel: Experimental STM image of the Si(111) — 7×7 structure [234]. The faulted part of the unit cell is in the upper part of the image. Notice the correspondence between the computed charge densities and the experimental image.

(c), exhibit also important qualitative differences. The vibrational properties are computed from a short-time (~ 1.2 ps) low-temperature ($T \sim 200$ K) MD trajectory. How can we determine the dynamics of a system as large as this? Given the system size it is preferable to use alternative methods based on *signal processing techniques* rather than methods based on direct diagonalization of the dynamical matrix. We show below how system dynamics can be extracted from a very short and poorly equilibrated MD trajectory. Moreover, the simulation, at least in principle is not limited by the harmonic approximation, which offers the possibility to estimate this effect. The simulation indeed provides evidence for enhanced surface anharmonicity with respect to the bulk which should play a role in the $7 \times 7 \rightarrow 1 \times 1$ phase transition at $T \sim 1100$ K. The dynamical MD simulation provides a wealth of information on a number of properties of

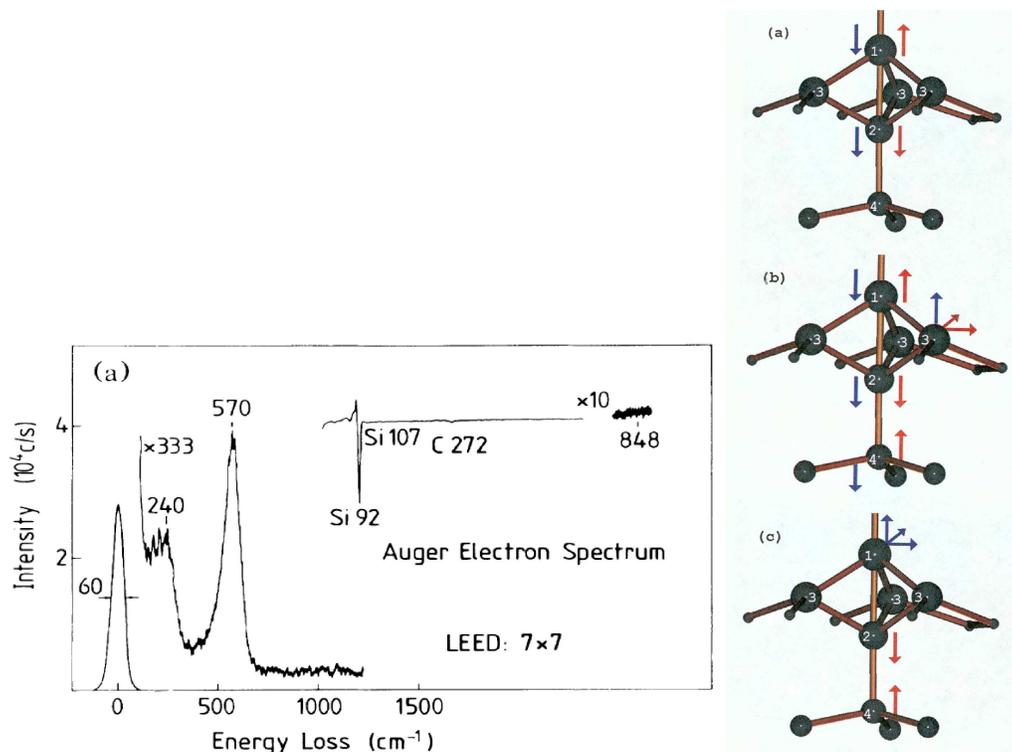


Fig. 38. Left panel: experimental EELS spectra [235]. Right panel: Labeling atoms accounting for the surface modes, and a schematic sketch of the modes determined by EELS in various models. (a) *ab initio* cluster model [235], (b) model based on empirical interatomic potential [236], (c) fully *ab initio* MD model [237, 238]. Red arrows depict the high-frequency mode, and blue arrows the 240 cm^{-1} mode.

the system. For example, at least the upper two layers are “chemically soft”, by which we mean that the longer bonds associated with those layers exhibit large amplitude, lower frequency oscillations with vastly varying bond charge associated with them. On the other hand, unusually strong, compressed bonds have been found between atoms in the second- and third-layer atoms. These findings give intuitive hints as to the unusual vibration properties of the surface.

The simulation was based on the 400-atom model described above. The simulation was run in (E, N, V) ensemble using the BO dynamics with CG electronic optimization and wave function extrapolation technique at each MD step, see chapter 3.3.2, 3.3.2.1 and 3.3.2.2, at a fairly low temperature of $T \sim 200\text{ K}$ for 1.2 ps after initial equilibration.

In addition to the simple estimator Fourier transform (FT) estimator (326), (327) a more sensitive class of spectral estimators based on the *MUltiple Signal Classification* (MUSIC) algorithm [239] was also used. The reason is that the estimator (326), (327) has a very limited resolution, namely $\Delta\omega > 2\pi/\tau$ for equally spaced modes, where τ is the total simulation time. For very short simulation times the spectral resolution becomes very limited; in the present case

$\Delta\omega \approx 35\text{cm}^{-1}$. The resolution is even lower when two modes are excited with different amplitudes. Thermal equilibration proceeds via mode-mode coupling, which is usually weak at low temperatures. Hence, it may be difficult to extract information about weakly excited modes. In particular, some modes may be undetectable because the limited resolution due to insufficient MD observation time hides them inside the main or first aliasing lobe of a close, strongly excited mode. Nowadays, somewhat longer simulation times would certainly be feasible. However, this would still not solve the problem of a proper equilibration, unless one led the system through a set of annealing/cooling cycles.

A more efficient way of dealing with the problem consists of use of a better spectral estimator which can detect even very weakly excited modes. In the MUSIC algorithm [239] the space spanned by the collected data is first separated into *signal and noise space*. In a quasiharmonic regime the MD trajectory $\{|\hat{\mathfrak{R}}(t_k)\rangle\}$ can be modeled by a linear combination of sinusoids,

$$|\hat{\mathfrak{R}}(t_k)\rangle = \sum_{l=1}^P |\mathfrak{N}_l\rangle \{A_l \sin(\omega_l t_k) + B_l \cos(\omega_l t_k)\} + C(t_k), \quad k = 1, 2, \dots, M, \quad (329)$$

where $|\dots\rangle$ denotes a $3N_I$ -dimensional column vector, with N_I being the number of atoms, $|\mathfrak{N}_l\rangle$ the vibrational eigenvectors, $|A_l\rangle$ and $|B_l\rangle$ the amplitudes, $|\omega_l\rangle$ vibrational frequencies, and C additive noise. The MUSIC algorithm exploits the *orthogonality between signal and noise subspace* to construct the estimator

$$P_M(\omega) = \left[e_M^H(\omega) \left(\sum_{m=2P+1}^{M+1} q_m q_m^H \right) e_M(\omega) \right]^{-1}, \quad (330)$$

where $e_M^T(\omega) = (1, \exp[i\omega(\tau/M)], \dots, \exp[i\omega M(\tau/M)])$, $\{q_1, q_2, \dots, q_{2P}\}$ span the signal subspace, and $\{q_{2P+1}, \dots, q_{M+1}\}$ span the noise subspace. $P_M(\omega)$ has sharp peaks at $\omega = \omega_l$, and is negligible elsewhere. Compared to the FT estimator, $P_M(\omega)$ typically achieves comparable resolution with $\sim 30\%$ collected data. An even better signal-to-noise separation can be achieved by fully exploiting the information contained in the MD trajectory, using the self-consistent variant of the MUSIC algorithm, wherein $\{|\mathfrak{N}_l\rangle\}$, $\{|A_l\rangle\}$ and $\{|B_l\rangle\}$ are all determined by minimizing the cost function,

$$\begin{aligned} \Xi(\{|\mathfrak{N}_l\rangle\}, \{A_l\}, \{B_l\}, \{\omega_l\}) &= \sum_{k=0}^M \left\langle \mathfrak{R}(t_k) - \hat{\mathfrak{R}}(t_k) \mid \mathfrak{R}(t_k) - \hat{\mathfrak{R}}(t_k) \right\rangle \\ &- \sum_{i,j} \Lambda_{i,j} (\langle \mathfrak{N}_i \mid \mathfrak{N}_j \rangle - \delta_{i,j}), \end{aligned} \quad (331)$$

where $\Lambda_{i,j}$ is the matrix of Lagrange multipliers imposing the orthogonality constraints. A minimum of Ξ can be obtained by an iterative procedure using a constrained steepest descent techniques [239]. The frequencies $|\omega_l\rangle$, on the other hand, are always estimated using the high-resolution MUSIC estimator, but now applied to the partial trajectories obtained by projecting the original MD trajectory $|\mathfrak{R}(t)\rangle$ onto the fitted eigenvectors $\{|\mathfrak{N}_l\rangle\}$

$$|\mathfrak{R}_l(t)\rangle = |\mathfrak{N}_l\rangle \langle \mathfrak{N}_l \mid \mathfrak{R}(t)\rangle \quad (332)$$

in a self-consistent manner. The new frequencies are used to determine a better estimate of the eigenvectors, so improving the result until $P_M(\omega)$ produces only a single well-resolve peak for each projected trajectory. Compared to the FT estimator, $P_{SCF}(\omega)$ typically achieves comparable resolution with $\sim 10\%$ of collected data. In the present case the resolution is $\Delta\omega \approx 5 \text{ cm}^{-1}$. A further improvement can be achieved by applying symmetry consideration and decomposing the trajectory into a set of trajectories transforming according to a given irreducible transformation.

Within our model we are left with 151 independent atoms and 453 frequencies. It is not possible or even meaningful to determine all of them. The search is limited to determining the *dominant dynamical features*. The 7×7 cell is sufficiently large for translational symmetry to have little effect on the results, and hence, we consider only the $Q = 0$ point in the Brillouin zone. Decomposition of the trajectory according to the point-group symmetry C_{3v} yields three partial trajectories two of them transforming according to one-dimensional irreducible representations (A_1, A_2) and one according to a two-dimensional representation (E). According to group theory [240] A_1 and E are IR and Raman active and A_1 is also EELS active; 41 modes transform according to A_1 , 110 according to A_2 , and 151 according to E .

The most important results from the dynamics analysis are shown in Fig. 39. A detailed analysis of all modes found can be done [238]. To make a connection with the EELS experiment [235], we limit ourselves to just the two experimentally determined modes. The $\sim 240 \text{ cm}^{-1}$ modes can be associated with peaks of all three symmetries (A_1, A_2 , and E), while the high-frequency $\sim 528 \text{ cm}^{-1}$ split-off peak is composed of three peaks of A_1 and E symmetries. The calculated A_1 modes (left panel of Fig. 39 (d)) should correspond to the measured EELS spectra (235). Both peaks can be regarded as in excellent agreement with experiments. The different peaks can be associated with the motion of individual atoms. This assessment can easily be made by inspection of the eigenvectors determined by the SCF MUSIC method. Instead, we use a simpler procedure, namely the lower-resolution FT estimator projected onto individual atoms (right panel of Fig. 39). We can clearly see that the signal derives from atoms 2 and 4, i.e. from the atoms underneath the adatoms. Those pairs of atoms create pairs of strongly compressed bonds with bond lengths $\sim 2.25 \text{ \AA}$ as opposed to the bulk Si-Si distance of 2.35 \AA , the 2.46 \AA for the 1–3 atoms or 2.40 \AA for the 1–2 atoms [237, 238]. This provides the answer to the striking fact, how unusually high frequencies nonexistent in the bulk may exist on reconstructed surfaces. The analysis of these results yields the assignment as shown in Fig. 38.

The aforementioned strong anharmonicity due to the “softness” of the surface can be assessed from Fig. 39. The most striking feature is the width of the highest frequency mode peak of E symmetry at $\sim 528 \text{ cm}^{-1}$ with a satellite at $\sim 400 \text{ cm}^{-1}$. This results from a strong mode mixing due to the anharmonicity of the potential.

This simulation nicely demonstrates that often alternative techniques are required if one wants to make most of the huge amount of data from limited, imperfect or incomplete simulations, which is often all one can realistically hope for.

4.3 Interaction of molecules with surfaces

After having dealt with energetics and dynamics of solid surfaces it is of interest to know how other objects, such as atoms and molecules interact with them. The interaction of particles with solid surfaces is of interest from both technological and scientific point of view [241]. This is

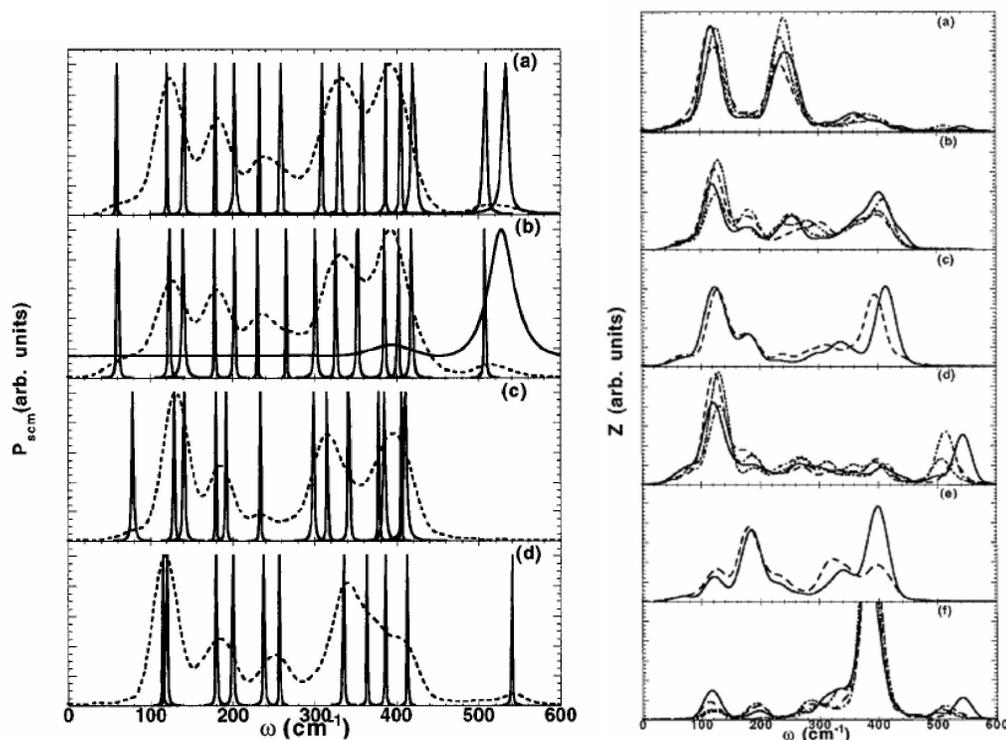


Fig. 39. Left panel: The self-consistent MUSIC spectra (full line) and the corresponding FT spectra (dashed line). (a) FT spectra and MUSIC fit for the total trajectory; (b) The same but for the trajectory transforming according to E ; (c) The same but for the trajectory transforming according to A_2 ; (d) The same but for the trajectory transforming according to A_1 . All spectra are normalized to 1. Right panel: Projected FT spectra onto individual atoms. (a) Adatom spectra corresponding to the four different adatom types. (Spectra of atom 3 (for labeling see Fig. 38)). (c) Spectra for restatoms. (d) Spectra of atom 2 (for labeling see Fig. 38). (e) spectra for dimers. (f) Spectra of atom 4 (for labeling see Fig. 38). The different lines in each figure correspond to different type of adatoms, such as center or corner and/or faulted or unfaulted part of the unit cell.

a vast field which ranges from helium atom scattering to etching, heterogeneous catalysis, formation of self-assembled monolayers, etc. There is a number of associated questions, such as, are the species chemisorbed or are they just physisorbed? Do the molecules adsorb in molecular state or do they dissociate as they adsorb? What is the energy transfer between the molecule and the substrate? Is the energy dissipated by phonons or do other dissipation channels, such as electron-hole pair creation enter? What is the sticking probability for a molecule prepared in a given state? How do molecules form at and desorb from surfaces? Many of these questions can be studied experimentally using techniques such as molecular beam scattering techniques [242]. Theory has proven to be an extremely useful tool in understanding the experiments. Simulations have been used to integrate the time-dependent Schrödinger equation (see the discussion of

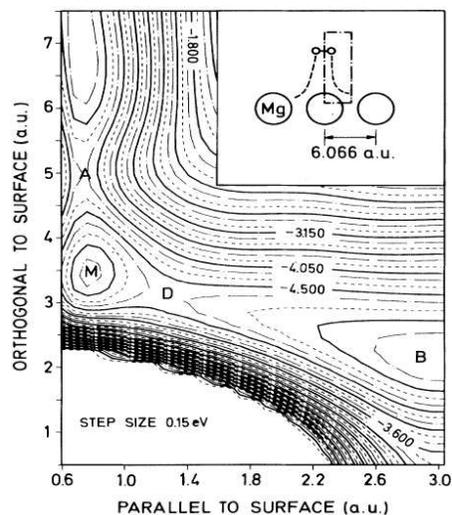


Fig. 40. PES for H_2 dissociating over the atop site into bridge sites (see inset) of $\text{Mg}(0001)$ calculated using jellium pseudopotential perturbation technique [243].

Eq. (241)) [150] or simpler classical dynamics. A major problem with dynamical simulations is that they require an adiabatic potential energy surface for the molecule-surface interaction, whose calculation was made possible only in the last decade. In the following we will show a few examples.

4.3.1 Explicit determination of PES: Dissociation of H_2 on $\text{Mg}(0001)$ surface

This system is important as it was considered a textbook example of a successful modeling of molecule-surface interaction. As discussed in chapter 2 (see, Fig. 2), the PESs for molecule-surface interaction are at least 6-dimensional. Hence, it is not surprising that until recently it has not been possible to calculate realistic PESs for even the simplest systems. Even the simplified surfaces concentrating on two-dimensional sections of the PES, where the molecular axis is parallel to the surface and the energy is calculated as a function of height above the surface and molecular bond length, were not feasible before introduction of parallel computing. Note, that both the electronic structure of the system as well the individual points calculated on the PES can be spread over the processors. In this situation the benchmark calculation for this system was based on jellium approximation for the surface with pseudopotential corrections [243]. The result is shown in Fig. 40. The most prominent feature is the existence of a *molecular precursor state*, labeled M in Fig. 40. This state is a weakly chemisorbed molecular species (adsorption energy is only ~ 0.4 eV). From this precursor state the molecule may dissociate into bridge sites by surmounting a barrier of ~ 0.5 eV. This result was regarded as a strong support to the existence of molecularly chemisorbed species on surfaces. Experimentally, there is strong evidence that molecularly chemisorbed species exist in a number of O_2 dissociation processes, for example $\text{O}_2/\text{Pt}(111)$ [244] and $\text{O}_2/\text{Ag}(111)$ [245]. However, there appears to be little evidence for such

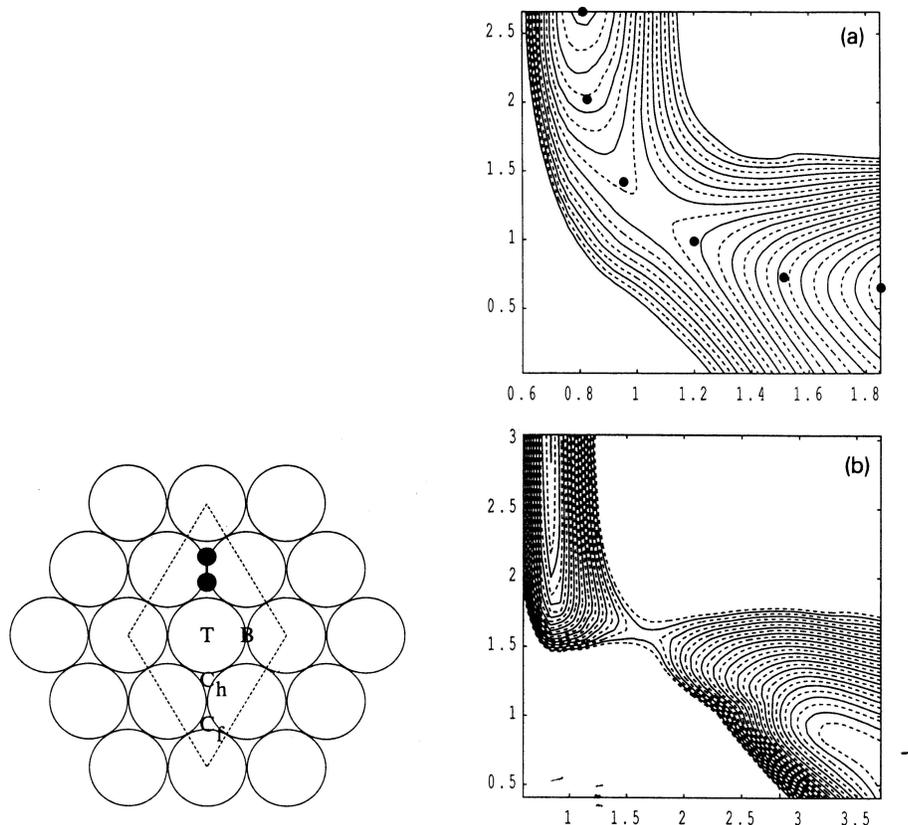


Fig. 41. Left panel: Plan view of the Mg(0001) surface, showing Mg atoms (large circles), a dissociating H_2 molecule (filled circles) and the 2×2 supercell (dashed lines). T stands for top, C_f for fcc center and C_h for hcp center. Right panel: PESs for dissociation about (a) bridge and (b) top sites into neighboring sites. Vertical axis is height of H_2 above surface, horizontal axis is H_2 bond length. Contour spacing is 0.005 eV per molecule, all distances in Å.

precursor states in H_2 dissociation [246].

In this situation a “benchmark” calculation was required. Such a benchmark was provided by a first realistic PES calculation [247]. The geometry of the $H_2/Mg(0001)$ system is shown in Fig. 41. The slab is two unit cells thick with a large vacuum space separating the adjacent slabs. A 2×2 supercell is used in the basal plane and inversion symmetry is applied throughout. The calculations are based on pseudopotential LDA methods. The bare Coulomb potential is used for H. Three Monkhorst-Pack [130] special k-points were used with a large (1 eV) selectronic smearing with an analytic correction to zero smearing [248]. Plane waves with energy up to 300 eV are included. As in other H_2 /metal calculations, any surface relaxation is neglected on the grounds that in a dynamical dissociation process the large mass difference between hydrogen and magnesium will ensure that recoil effects are small.

The results shown in Fig. 41 suggest that the natural dissociation path is where the molecule is incident over a bridge site with the hydrogen atoms oriented towards the center sites (Fig. 41). In order to compare more directly with the results of Ref. [243], the calculations have been also performed for dissociation about the top site into adjacent center sites. The most striking feature of the PESs is that they show *no sign of any precursor state*. Comparison of Figs. 40(a) and (b) shows that dissociation about the bridge site is strongly favored, with the barrier being about 0.6 eV lower per molecule than at the top site. Hence both the preferred dissociation site as well as the qualitative form of the PES strongly differ in the accurate treatment, compared to PESs calculated with more approximate methods [243]. These early calculations paved the way for many more calculations of similar spirit [249–251]. With advance of computer power, calculation of more dimensional PESs became feasible [252, 253].

4.3.2 Statistical sampling of PES: Dissociation of Cl_2 on $\text{Si}(111)\text{-}2 \times 1$ surface

The above approach, while totally adequate for the case of light molecules on relatively heavy surfaces, there is in the limit $M_{mol}/M_{surf} \ll 1$, suffers from a total neglect of any surface recoil effect. Such an effect is bound to arise if $M_{mol}/M_{surf} \sim 1$ [254, 255]. Hence, the 6-dimensional PES $V(x, y, z, d, \phi, \theta)$ for a full treatment of molecule-surface interaction over static surfaces has to be extended to allow for interaction with surface degrees of freedom $\{u_I\}, V(x, y, z, d, \phi, \theta, \{u_I\})$. This significantly increases the dimensionality of the problem to at least a few hundred. The high dimensionality of the PES has several implications. Because of the large number of degrees of freedom this approach is limited to the case where classical treatment of the atomic dynamics applies (large M_{mol}), which is in line with the assumption $M_{mol}/M_{surf} \sim 1$. Returning to the example of Fig. 2 for the 6-dimensional problem, where the number of points required was $O(10^6)$ the number of required points for the $V(x, y, z, d, \phi, \theta, \{u_I\})$ PES would increase to $O(10^{n \times 100})$. An explicit calculation of this high-dimensional PES in the spirit of Figs. 40 and 41 is intractable. What can be done to explore this type of PES? The simplest and most direct way is to apply sampling techniques. The same trick is done also when a high dimensional integral is to be evaluated; for more details see the discussion of formula (171), chapter 3.1.6.1. At modest kinetic energies most parts of the PES are inaccessible, and hence we need to sample only the relevant parts of the PES. One way of doing that is to use MD sampling. MD has the advantage over MC sampling in that it provides access to real molecule-surface dynamics. A series of dynamical simulations started from different initial conditions will determine which regions of the PES are accessible and what degrees-of-freedom are relevant in the molecule-surface interaction at a given incident energy E_i . The information gathered during the sampling can be used as the basis of further importance sampling in order to increase the accuracy achieved in later sampling steps. In absence of such information the PES may be sampled at random starting from a large number of initial conditions compatible with the chosen external conditions such as average translational energy, internal beam temperature, etc.

The incident energy of the impinging molecule E_i is of paramount importance in the molecule-surface interaction. Experimentally incident energies ranging from thermal to keV are used. In collisions at low incident energies the kinetic energy of the molecules is insufficient to break the molecular bond, i.e. $E_{diss} > E_i$, and the dissociation is driven by electronic or chemical process. At higher translational energies the collision-induced dissociation becomes increasingly important [256]. The chemical nature of the breakup process places extreme demands on the

interatomic potentials since it must be capable of describing significant changes in bond lengths and atomic environments, breaking of chemical bonds within the molecule, and creation of new bonds with the substrate. This is very difficult to achieve with any empirical interatomic model and suggests the necessity for an *ab initio* approach.

The approach outlined above is extremely powerful and in principle should be able to provide valuable qualitative insights, or even quantitative characterization of many processes of technological importance, such as molecular beam epitaxy, implantation, or experimental importance, such as molecular beam scattering technique [242] and the associated sticking coefficients s . As shown below, the main current limitation of this approach is in the coarse sampling of the PES. Such a limitation is being relaxed with advance of the computer power which is significantly improving the achievable statistics in the PES sampling.

The approach is demonstrated [257, 258] on the study of Cl_2 dissociation on the π -bonded $\text{Si}(111)\text{-}2 \times 1$ surface [259]. In this system clearly $M_{mol}/M_{surf} \sim 1$. The PES is sampled at incident energy of $E_i = 1$ eV. This incident energy remains in the region of chemically driven breakup, as $E_{diss} \sim 2.5$ eV for a free Cl_2 . This choice of E_i is motivated by experimental results indicating a high value of the initial sticking coefficient s_0 at incident energies around 1 eV [260]. The study is based on five trajectories, see Fig. 42. Slab geometry with four double layers and an almost square $2\sqrt{3} \times 3$ surface unit cell was used with inversion symmetry imposed on the slab. Electronic structure of the system was treated in the DFT theory using LDA approximation for the exchange-correlation energy. Pseudopotentials were used to describe all atoms [119]. The electronic states were expanded at the Γ -point with an energy cutoff of 10Ry. The classical equations of motion were integrated in microcanonical ensemble for between 200 and 400 fs which is enough elapsed time to permit collision with the surface to occur and to classify the outcome.

The results are shown in Fig. 42. The first two trajectories represent two extreme choices of initial conditions. In the former the molecule strikes the surface in a region of high electronic density generated by the unsatisfied Si bonds on the π -bonded chains. In the latter the molecule is directed towards the region of low electronic density in the valley. However, in both cases the Cl_2 molecule spontaneously dissociates upon collision with the surface, the outcome being dichlorination of the substrate. In response the π -bonded chains deform appreciably and locally rehybridize towards sp^3 bonding, lifting the silicon atoms above the π -bonded chain. It is evident that the π -bonded chains play a critical role in both both dissociation processes. Interestingly, dissociation occurs also in trajectory No. 3, where the molecular axis is parallel to the π -bonded chain and the molecule-chain interaction is much less direct. The dynamical scenario in the cases where the molecular axis is oriented along the surface normal (trajectories No. 4 and No. 5) initially involves monochlorination. The Cl_2 molecule remains in a precursor state with the other Cl atom weakly bonded to the chemisorbed Cl and the bond length stretched by $\sim 50\%$. The initial conditions for trajectory No. 4 were chosen to determine whether Cl_2 molecule could penetrate into the subsurface region since the hole underneath the sixfold ring is intuitively the most natural entrance channel. Interestingly, the Cl_2 molecule does not show any tendency to penetrate the subsurface region and is instead displaced laterally towards the π -bonded chain. This demonstrates that even at rather high incident energies parts of the surface unit cell are inaccessible to the molecule. In this geometry the energy gained from formation of the Si-Cl bond is largely deposited in the substrate. At longer times the molecule tilts parallel to the π -bonded chains in trajectory No. 5 and across the valley towards the neighboring π -bonded chain

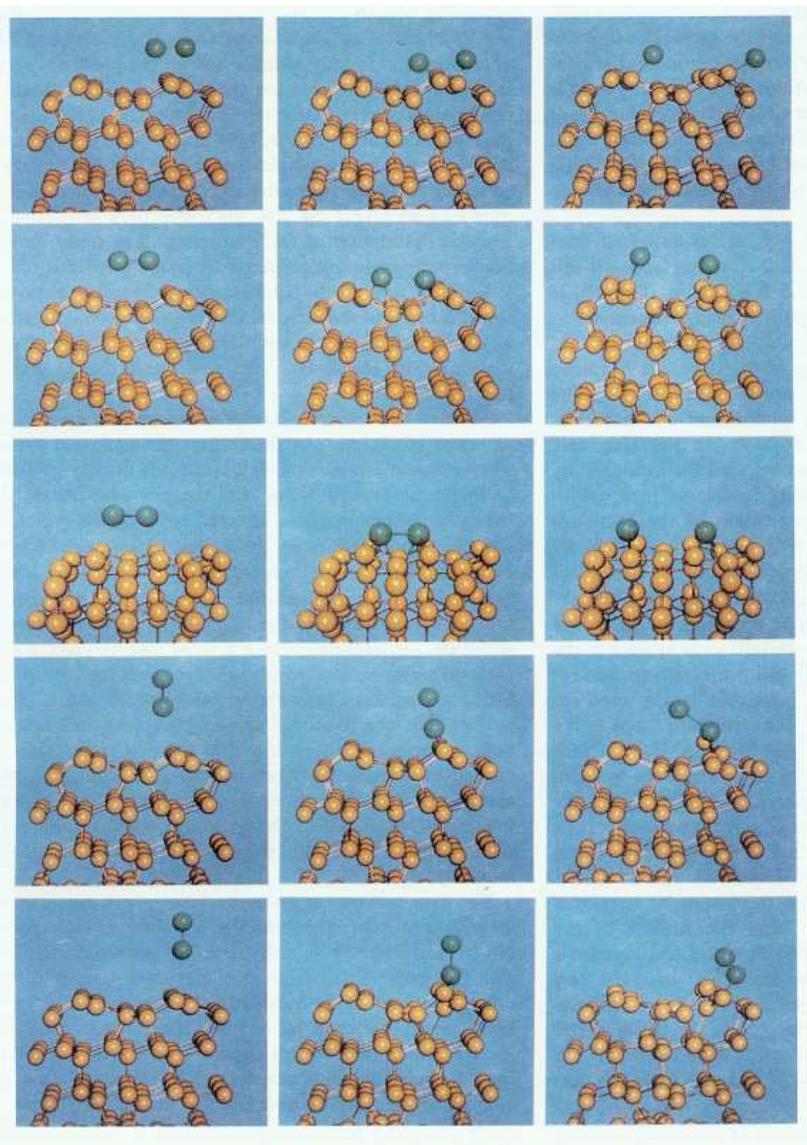


Fig. 42. Ball-and-stick model of chlorination of the Si(111)- 2×1 surface for five trajectories. Left, middle, and right panels show three different dissociation stages.

in trajectory No. 4. In both cases the length of the molecular bond increases as the molecule tilts. In trajectory No. 5 this clearly results in dissociation, with the final reaction product again being dichlorination of the surface. An interesting feature in this simulation is that a local defect, which changes the buckling of the π -bonded chain and the topology of the bonding in the silicon

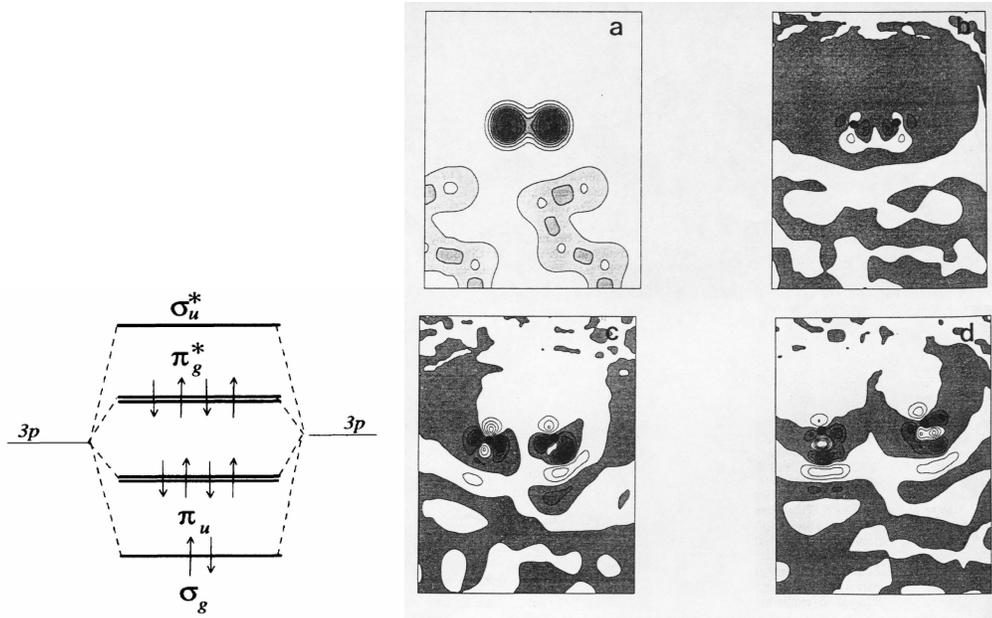


Fig. 43. Left panel: Molecular orbital diagram of Cl_2 . Right panel: Analysis of the dissociation process in terms of charge densities for trajectory 2 showing the repopulation of $3p\pi_g^*$ and $3p\sigma_g^*$ molecular orbitals. (a) ρ_{SCF} at the beginning of the trajectory, (b) ρ_{SCF}^{diff} at the beginning of the trajectory, (c) after collision with the surface, and (d) at the end of the trajectory. The position of Cl atoms is indicated by dots. Positive densities are on greyscale. Note the pronounced charge depletion from the $3p\pi_g^*$ (white regions) and charge accumulation in the $3p\sigma_g^*$ (black regions) around the molecule in (c).

substrate, is created by the collision.

The selfconsistent electronic structure generated along the reaction coordinate can be used to analyze the driving force to dissociation. Intuitively very appealing is the molecular orbital (MO) picture. The MO diagram of the free Cl_2 molecule is shown in Fig. 43. A Cl_2 molecule has bond order 1 with the $\dots 3p\sigma_g$, $3p\pi_u$, and $3p\pi_g^*$ MO's filled and only the $3p\sigma_u^*$ empty. Hence, a chemically driven breakup will be expected to be driven by the filling of the strongly antibonding $3p\sigma_u^*$ MO by transferring charge from the surface and/or another Cl_2 MO. The applicability of this picture to $\text{Cl}_2/\text{Si}(111)\text{-}2 \times 1$ can directly be assessed by analyzing the induced charge upon molecule-surface interaction

$$\rho_{SCF}^{diff}(\vec{r}) = \rho_{SCF}[Si_{subst} + Cl_2](\vec{r}) - \rho_{SCF}[Si_{subst}](\vec{r}) - \rho_{SCF}[Cl_2](\vec{r}). \quad (333)$$

Here $\rho_{SCF}[Si_{subst} + Cl_2](\vec{r})$ denotes the charge density of the interacting molecule-substrate system, $\rho_{SCF}[Si_{subst}](\vec{r})$ ($\rho_{SCF}[Cl_2](\vec{r})$) is the charge density for the substrate alone (adsorbate alone) with atomic positions corresponding to the interacting adsorbate-substrate system.

As an example we show in Fig. 43 the analysis of trajectory 2. The charge density difference $\rho_{SCF}^{diff}(\vec{r})$ in Fig. 43(b) – (d) represent three very different stages during the dissociation process: the first shows intact molecule, the second the onset of dissociation in which the molecule is

highly stretched, the third shows a fully dissociated molecule with two atoms bonded to the surface. In the configuration corresponding to Fig. 43(b) the molecule is far from the surface and the molecule-surface interaction is relatively weak. Accordingly, $\rho_{SCF}^{diff}(\vec{r})$ is rather featureless. By contrast, due to the significantly increased molecule-surface interaction, $\rho_{SCF}^{diff}(\vec{r})$ in Fig. 43(c) exhibits two pronounced features: charge is transferred into regions along the molecular axis, and regions perpendicular to the molecular axis are depleted of charge. Finally, $\rho_{SCF}^{diff}(\vec{r})$ in Fig. 43(d) shows creation of polar bonds between Cl atoms and the substrate. The most important information on the molecular dissociation process is contained in Fig. 43(c). The regions of charge accumulation around the molecule can be interpreted within the MO picture in terms of increasing occupancy of the $3p\sigma_u^*$ MO. We recall that in terms of charge density a characteristic feature of a σ_u^* MO are four lobes distributed along the molecular axis. Similarly, the four lobes of charge depletion off the molecular axis in Fig. 43(c) can be identified as $3p\pi_g^*$. Thus the molecule-surface interaction splits the degeneracy of the $3p\pi_g^*$ doublet by affecting the MO that couples strongly to the substrate while leaving the other one relatively unchanged. It can be seen from Fig. 43(c) that more charge is transferred to the more antibonding $3p\sigma_u^*$ MO's than is depleted from the $3p\pi_g^*$. The effect of charge transfer into the molecule as well as the repopulation of the MO's both destabilize the chemical bond. This analysis indicates the insight provided by applied method and the power of this approach.

Because of the enormous associated numerical overhead this approach has been relatively rarely applied up to date. Few years ago this limited such an approach to just qualitative insights. However, with the steady increase of computer power, already nowadays this approach can yield more quantitative results, such as e.g. the information on sticking coefficient s . All that is needed is a better statistical sampling of the underlying PES, i.e. more generated trajectories.

4.4 Simulation of surface probe microscopy (SPM)

So far we dealt with different aspects of surfaces such as their atomic and electronic structures, their interaction with molecules and possible implications for surface catalytic properties. The quality of computer modeling typically has to be validated against some set(s) of experimental data. The most direct experimental surface probes are surface probe microscopy (SPM). The two most important SPM methods are scanning tunneling microscopy (STM) [261] and *atomic force microscopy (AFM)* [262]. Both STM and AFM have the ability not only to “see” surfaces, adsorbed atoms and molecules on surfaces with *atomic resolution* but have also the *capability to nanomanipulate* the surfaces and adsorbates on them. The problem with both techniques is that while they exhibit an impressive arsenal of capabilities, the understanding of how these capabilities come about remains very limited. Moreover, the interpretation of what is actually being measured remains unclear until and unless the results are back-upped by a simulation. Let us take the more widely used STM as an example. STM measures tunneling currents from the scanning tip into the surface. Applied voltage, typically ± 1 eV around the Fermi level, serves as a tool to switch on and off different surface orbitals to/from which to tunnel. This, in turn means, that the STM is primarily an electronic structure rather than atomic structure tool. In other words, in order to relate the measured STM image to surface atomic structure, typically a simulated STM image for a set of anticipated atomic structures is required, before the STM identification of a surface structure is possible. This shows the enormous importance of computer simulation even in experimental context. We now turn to the other important SPM technique, the AFM. One

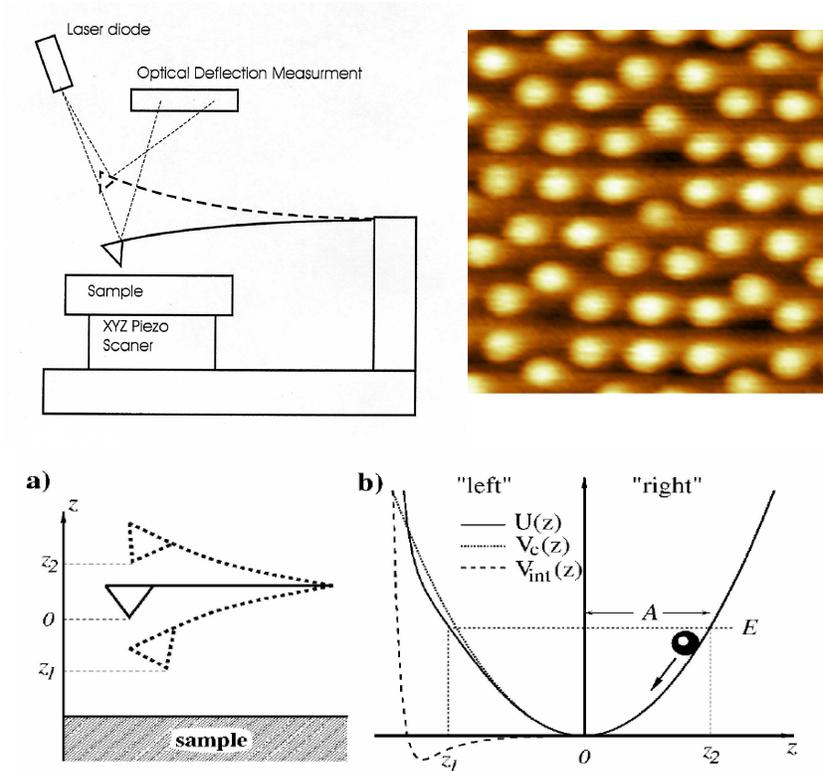


Fig. 44. Upper left: Schematic illustration of the NC AFM apparatus. Upper right: High quality image of the Si(111)- 7×7 surface. Lower panel: The basis of the NC AFM imaging: harmonic oscillation $V_c(z)$ is perturbed in close proximity to the surface by the tip-sample interaction potential $V_{int}(z)$ to produce the total potential $U(z)$.

advantage of AFM is that by its nature AFM, probing the tip-surface interaction, is generally believed to be more directly related to the surface atomic structure than the STM.

4.4.1 Simulation of imaging with dynamic AFM apparatus $\{v_1\}$

The AFM [262] has developed to one of the most powerful probes of atomic structure of both conducting and insulating surfaces. Several different operation modes exist [263] but only the AFM operating in *non contact (attractive) regime (NC AFM)* has demonstrated the ability to achieve a true atomic resolution. The apparatus typically works in attractive regime. For that reason, in order to prevent the tip from snapping into the surface, the tip is oscillated at kilohertz frequencies which minimizes the time the tip spends in close proximity of the surface; see Fig. 44. This is why it is also called Dynamic Surface Force Microscopy (DSFM). The NC AFM apparatus can be operated in several different modes [263] but the simplest and most direct is the technique of *Frequency Modulated Atomic Force Microscopy (FM AFM)* [264] in which, in

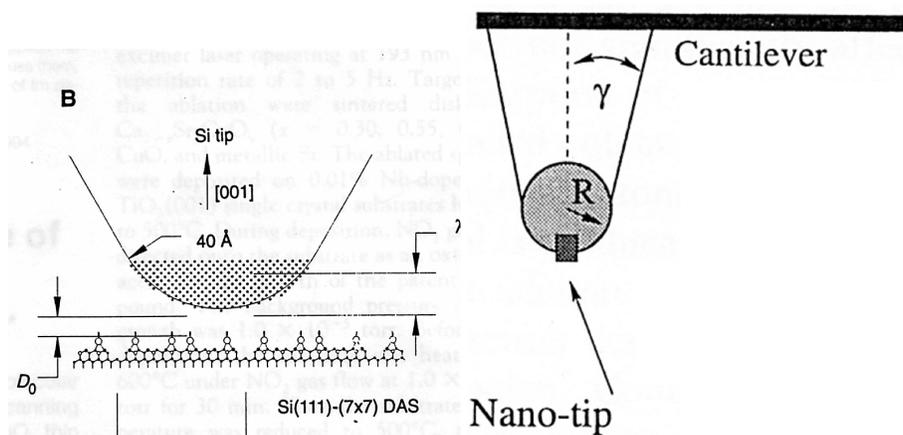


Fig. 45. Left panel: Imaging by a typical tip. Right panel: Schematization of the tip by a sphere and formation of the nanotip.

close proximity to the surface sample, a harmonic motion of the tip is perturbed by tip-surface interaction; Fig. 44. In the simplest case the (negative) frequency shift $\Delta\omega$ is kept constant and the corresponding surface topograph collected. The technique was used to image a host of different reactive surfaces, such as Si(111)- 7×7 [264], Si(001)- 2×1 [265], the polar III-V (110) surfaces: InP(110) [266], GaAs(110) and InAs(110) [267], as well as defects on these surfaces [265, 266]. Theory has provided very strong indications that the short-range chemical type of tip-surface interaction significantly enhances the atomic resolution of the FM AFM [268, 269]. Experiments have corroborated these conclusions [270]. Furthermore, the fact that the atomic resolution is primarily mediated by a dangling bond type of interaction between the atoms on a sample surface and a tip is in accordance with the ability to resolve a number of different atoms (dangling bonds) on the Si(111)- 7×7 surface [270], whereas on polar surfaces, such as InP(110) and GaAs(110) only the anion sublattice could be resolved under usual experimental conditions [266, 267]. These are manifestations of the tip having different “local reactivity” with the surface. Hence, the experimental image is not a genuine property of the surface but also of the tip scanning it. The local reactivity is especially pronounced on polar surfaces where the dangling bonds on anions are more charged than those on cations. The concept of local reactivity, specifically the local modifications of the atomic structure or morphology of the tip apex is instrumental in understanding the imaging mechanism of the NC AFM. It has been observed that true atomic resolution was only possible after an “on purpose” contact of the tip with the sample. This creates an “atomically sharp” tip apex able to image with atomic resolution. The process of formation of a nanoasperity at the tip apex is shown in Fig. 45. Hence, it is not the mesoscopic tip which is important for the atomic-scale imaging, but rather a termination with nanodimensions which determines the collected image. The process of nanotip formation/modification may very easily occur also due to accidental contact of the tip with the surface which may lead to a pick-up process of a surface atom and ultimately to modification of the apex, which is also observed experimentally.

Let us now show how the surface imaging by NC AFM apparatus can be modeled. As an example we describe the application to the InP(110)- 1×1 [271] surface. The same approach has been used also to other systems, such as Si(111)- 7×7 [268, 269] or GaAs(110)- 1×1 [272]. The model for the FM AFM image consists of two key ingredients: (1) the model for the frequency shift $\Delta\omega$, the quantity measured in the experiments, and (2) the model for the potential energy surface $U_{tip-sample}(d, \alpha, x, y)$ for the tip-surface interaction as a function of d , the tip-surface distance, α , tip orientation with respect to the substrate, and (x, y) , the position of the tip in the surface unit cell.

To compute the frequency shift $\Delta\omega$ the *classical perturbation theory* [273] is used, where the tip-surface interaction $V_p(q)$ is the perturbation on the periodic Hamiltonian describing the oscillating tip. The result being [271, 274, 275],

$$\Delta\omega \approx \frac{\partial \overline{V_p(q)}}{\partial J_0} = \frac{\partial}{\partial J_0} \frac{1}{T} \int_0^T V_p(q(t)) dt, \quad (334)$$

where J_0 is one of the action-angle variables and $\overline{V_p(q)}$ the tip-surface potential averaged over one oscillation period T [273]. The $\Delta\omega$ is computed for each tip position (x, y) and orientation α . The $V_p(q)$ is then $U_{tip-surface}$ expressed in the q -coordinate system [276]. In our study $V_p(q)$ is a chemical bond between the tip and the sample surface. The question arises whether breaking/reforming a chemical bond as a result of oscillatory tip motion may be considered a small perturbation. Typical energy scale associated with such a process is of the order of a few eV. This is to be compared with the energy of an oscillating tip. Considering typical experimental values for the amplitude of the tip motion (≈ 200 Å) and spring constant (≈ 30 N/m) yields a value of ≈ 40000 eV, which exceeds the energy of a chemical bond by many orders of magnitude. Another important observation from Eq. (334) is that the experimentally measured $\Delta\omega$ is a complicated non-local quantity which requires knowledge of the tip-surface interaction over the entire oscillation period and not only in one point. More details can be found elsewhere [275].

The model for the tip-surface interaction $V_p(q)$ in the near contact short-range regime dominated by quantum chemical interaction is based on the Density Functional Theory in its plane-wave pseudopotential formulation [26]. The InP (110)- 1×1 surface was modeled in a slab geometry using 4×3 primitive surface unit cells, see Fig. 46. Experiments used Si tips [266] which we modeled by a 10 Si atom tip saturated at the tip base by 15 H atoms [268, 269]. Note, that in this model only the interaction of the very tip apex (nanoasperity) with the surface is considered. The modifications of the tip morphology are described below. The main characteristic of the pure Si tip is the presence of one singly occupied dangling bond sticking out of the tip apex. This dangling bond is instrumental for the atomic resolution of the FM AFM on a reactive surface. The apex and its three nearest neighbor Si atoms of the tip are allowed to relax. The sum of the forces exerted on the fixed atoms of the tip is taken to be the total tip-surface force. The properties of this tip have been extensively studied in Refs. [268, 269] and it was shown that the results are relatively insensitive to the tip size. On the other hand, the results very sensitively depend on the twist angle α of the tip with respect to the substrate [271]. The reason is that the dangling bonds on the InP(110) surface are located close by and multibond tip-surface interactions may occur. However, here we consider just one tip orientation, see Fig. 46.

The total energy and the tip-surface forces are calculated within the DFT in the Generalized Gradient Approximation (GGA) [59]. The energy cut-off was set to $E_{cut} = 8$ Ryd. Optimized

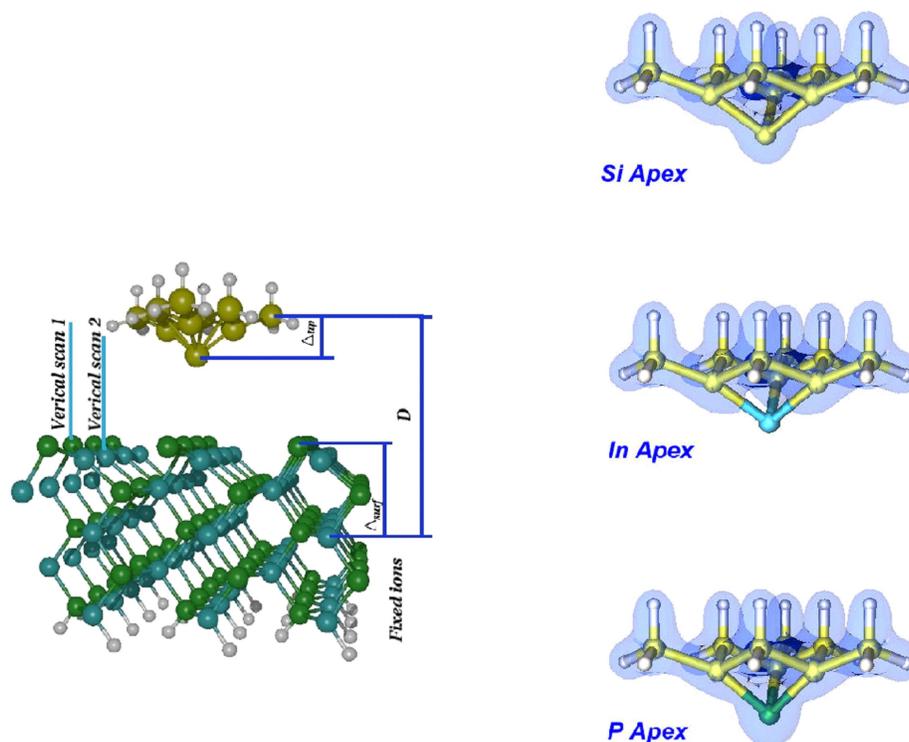


Fig. 46. Left panel: Geometry of the simulation. Right panel: Tips used in the simulation shown by a ball and stick model with a superimposed isosurface of valence electronic charge density. Upper panel: Si/Si tip; middle panel: In/Si tip; bottom panel: P/Si tip.

nonlocal pseudopotentials were used for all elements except for H which was described by bare Coulomb potential. Brillouin zone of the surface unit cell was sampled at the Γ -point. The computationally most expensive part was the DFT calculation of the vertical scans to compute the perturbing tip-surface potential $V_p(q)$ to compute the NCAFM image. Typically 10–15 points were computed for each vertical scan. The scan of the 1×1 surface unit was performed with a lateral step of $\approx 0.5 \text{ \AA}$. This amounts to ≈ 60 points in the surface unit cell.

The effect of modification of tip morphology of the Si tip by exchanging the apex atom can clearly be appreciated in Fig. 46. Henceforth, these tips will be denoted by Si/Si, In/Si, and P/Si tip. The four-valent Si apex forms a singly occupied dangling bond, the three-valent In has an unoccupied dangling bond, whereas the five-valent P has a doubly occupied dangling bond. These tips will have vastly different local reactivity with the InP(110)- 1×1 surface. The prominent feature of this surface is the stabilization by buckling which causes the P (In) atoms to relax upwards (downwards) and the dangling bond states on P (In) atoms to populate (depopulate). As a result, most of the surface reactive charge will float on the P atoms.

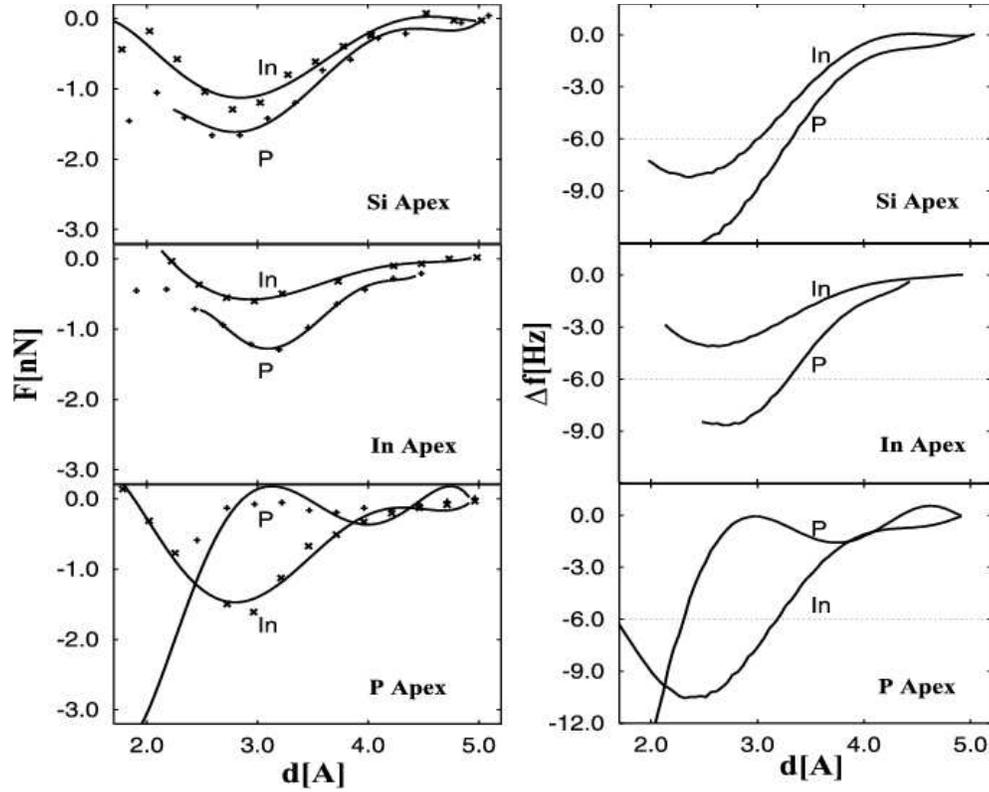


Fig. 47. Results for short-range normal tip-surface forces (left panel) and frequency shifts (right panel) in vertical scans performed with the Si/Si tip (upper panel), In/Si tip (middle panel), and P/Si tip (bottom panel) over In and P sites. Experimental parameters were taken from Ref. [266]. The horizontal line at -6 Hz corresponds to the experimental $\Delta\omega$. The wiggles in the curves through the computed points in the bottom panel are artifacts due to the use of a polynomial fit.

The results for the normal tip-surface force and frequency shifts over the In and P sites for the three different tips are shown in Fig. 47. As expected, the Si apex interacts more strongly with the P site than with the In site $\{v_1^{26}\}$. Interaction of the In/Si tip with the surface is reduced by a factor of ≈ 2 , compared to the Si/Si tip. Nevertheless, qualitatively both apices yield similar results. Situation changes dramatically in the case of a P apex. Due to the doubly occupied P dangling bond state there is very little interaction with the surface P site up to distances of ≈ 3 Å. At that distance a strong rebonding at the tip apex starts.

The tip-surface bonding translates into the behavior of the frequency shifts. The experimental value $\Delta\omega = -6$ Hz [266] yields for the Si/Si tip a computed AFM corrugation of ≈ 0.3 Å, in excellent agreement with the experimental value [266]. In the case of an In apex, our model

²⁶Simulation of NC AFM imaging of InP(110) surface by a model Si tip over P an In sites. A representative value of (pseudo)valence charge density is shown by a superimposed blue isosurface.

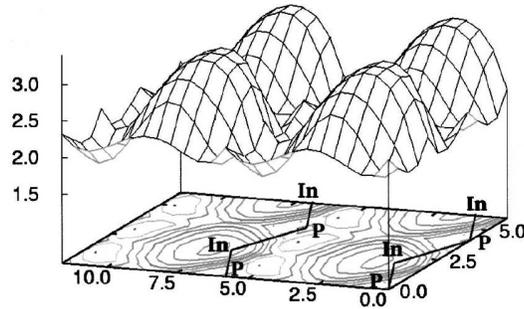


Fig. 48. Simulated AFM topograph of the surface unit cell of the InP(110)- 1×1 surface with the P/Si tip at $\Delta\omega = -6$ Hz. All length units are in Å.

predicts that a measurable In-P corrugation should start appearing for $\Delta\omega \leq -4$ Hz. The P apex represents a qualitatively different scenario where the In (cation) sublattice may have a higher apparent AFM height. In fact, these results suggest that the FM AFM with a P apex should yield the largest value of surface corrugation (≈ 1 Å). These results are qualitatively similar to those obtained for the GaAs(110) surface [272].

The results for the P/Si tip, shown above, represent a very interesting scenario which could explain how the experiments can image the cation (In) sublattice observed on the InAs(110) surface [267]. In order to verify further the findings a simulation of the *FM AFM image of the entire surface unit cell* was performed. This proves that simulation of a FM AFM image of the entire surface unit cell from first principles is feasible.

The results are shown in Fig. 48. The InP(110)- 1×1 surface exhibits a large (≈ 1 Å) AFM corrugation. As expected, with the P apex the protrusions are associated with the In sublattice. Hence, in the case of a P apex, the AFM corrugation is inverted with respect to the corrugation obtained with a Si (In) apex as well as with respect to the “true” geometric corrugation. Moreover, the protrusions do not perfectly match the In sublattice and exhibit a characteristic elliptical shape. This is due to the direction of the In dangling bonds.

4.4.2 Simultion of nanomanipulation with dynamic AFM apparatus $\{v_2, v_3\}$

One of the hottest current topics in DSFM is the nanomanipulation capability of the DSFM. We now show simulation of atomic nanomanipulation using DSFM microscope. Concerning atomic-scale manipulation, although the scanning tunneling microscopy (STM) has been the method of choice for vertical and lateral manipulation of individual atoms, molecules, and bonds, DSFM can overcome its fundamental limitation to conductive samples and pave the way to atomic/molecular nanomanipulation for the widest range of materials.

However, DSFM was only very recently, for the first time, used for vertical [277] and also lateral [278] manipulation of single atoms on semiconductor surfaces. In the vertical manipulation an adatom on the Si(111)- 7×7 surface was first removed by action of the tip. The defect so created was subsequently healed and a Si atom restituted on the adatom site by a “soft nanoindentation” by the tip of a NC AFM apparatus [277], see Fig. 49. These remarkable experiments go beyond the lateral displacement of physisorbed or weakly adsorbed atoms on top of a surface

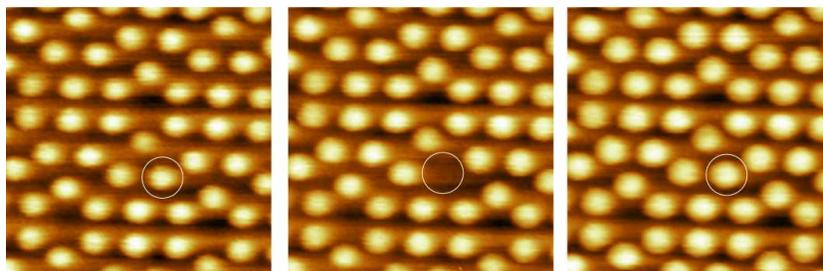


Fig. 49. Experimental vertical DSFM manipulation of an adatom on a Si(111)- 7×7 surface [277]. Left panel: perfect surface prior to manipulation; middle panel: surface after removal of an adatom; right panel: surface with a healed defect after a “soft nanoindentation”.

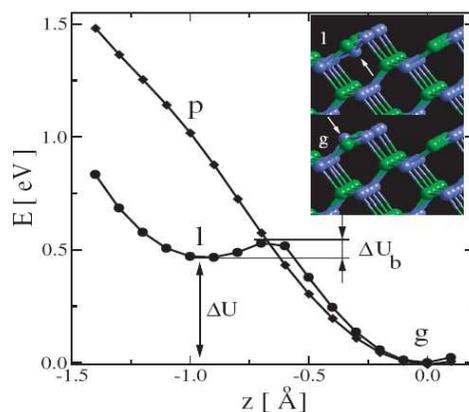


Fig. 50. Potential energy surface for the P antisite defect (dots) and P atom in a perfect In-P chain (diamonds) as the atom moves along surface normal. The insets show ball-and-stick models of the g - and l -structures.

and show that manipulation of atoms that are strongly chemisorbed or even part of the surface (including removal and deposition of single atoms without further surface damage) can be reproducibly performed using only the mechanical energy stored in the oscillating cantilever in DSFM.

To this end a very simple but challenging model system, an anionic antisite defect on a III-V-(110) surface is used [279]. One can demonstrate that the defect, with two stable configurations that differ by $\sim 1 \text{ \AA}$ in the normal direction, can be manipulated with the DSFM tip to switch its position from one minimum to the other.

The model system is built by substituting one of the In surface atoms by P to form an anionic antisite defect. This defect is commonly present on the III-Vs. Two (meta) stable geometries for the defect were found. One with the defect positioned roughly in the plane of the other phosphorus atoms (ground-state geometry), and the other where the defect atom is positioned well below (local-minimum geometry; Fig. 50). These geometries are labeled g and l , respectively.

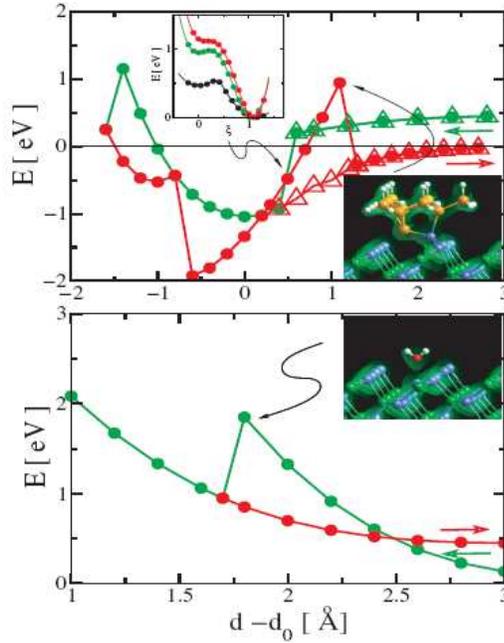


Fig. 51. The PES for nanomanipulation of the P defect atom. Upper panel: $l \rightarrow g$ process using model Si tip. Lower panel: reverse manipulation using model H₂O tip. Green: tip approach; red: tip retraction. Green/red arrows indicate the direction of the tip motion. Note that two tip approach/retraction cycles are shown in the upper panel. Flipping in attractive regime: empty triangles; flipping in repulsive regime: full dots. $d - d_0$ is the tip-surface distance. The insets show ball-and-stick models with superimposed isosurfaces of constant valence electronic charge density at the point of PES indicated by black arrows. PESs for flipping the defect with the tip positioned before (green), after (red) the discontinuity and in infinity (black) are shown in the inset of the upper panel. The reaction coordinate ξ was determined by linear interpolation between structures before and after the discontinuity.

Constrained total energy calculation along a path where the defect atom is moved from the g - to the l -position with all the other free atoms allowed to relax reveals that the defect atom moves vertically in a double well potential with $\Delta U \sim 0.5$ eV, and $\Delta U_b \sim 0.085$ eV (Fig. 50). Hence, the defect could be thermally stabilized at low temperatures in the local minimum. Moreover, as shown in Fig. 50, the existence of the local minimum is an exclusive property of the defect site absent from a perfect phosphorus site (p -site).

It can be shown that the defect can be manipulated from one minimum to the other and determine the influence of the character of the interaction (attractive or repulsive) and the tip reactivity in the process. In order to flip the defect from the local into the global minimum ($l \rightarrow g$ process) we first consider the standard model Si tip (Fig. 46).

Within the limited search the desired nanomanipulation was observed only with the model tip positioned over the trench between the In-P chains close to the defect. The potential energy surface (PES) for that process is shown in Fig. 51. Flipping the defect into the ground state geometry

may occur both in the attractive and repulsive mode, as shown by the two approach/retraction cycles displayed in the upper panel of Fig. 51 $\{v_2^{27}\}$. In the attractive mode, a small reduction of the tip surface-distance ($\sim 0.5 \text{ \AA}$) from the stable imaging range leads to the manipulation. Each discontinuity observed on the total energy curves corresponds to the plastic transition to a new stable configuration with the formation/breaking of either a tip-sample bond, a bond in the nanotip, or in the surface. These discontinuities will reflect in the force and thus on the measured frequency shift.

It was not possible to perform a reverse manipulation ($g \rightarrow l$) with the same model tip. Other tip configurations devised to simulate tip contamination, where the Si apex atom was replaced by either In or P, although less reactive, provide the same results. For that reason a purely repulsive tip was used. One very simple-minded tip functionalization is to consider an oxidized Si tip. Unfortunately, to model the oxide layer on a Si substrate is a problem in its own right. However, all models proposed share the following features: (a) twofold- O and fourfold-Si coordination, (b) O protruding out of the SiO_2 layer. Therefore the simplest model was considered, namely a H_2O molecule [280]. The oxygen atom points towards the surface, as would be the case in an oxidized tip. The H-termination mimics the local twofold coordination and chemistry of the O apex in the oxide layer. Hence, this simple model is a relatively close mimic of an apex formed by an oxide layer. The result is shown in Fig. 51 $\{v_3^{28}\}$. This tip indeed represents a pure repulsive tip-sample interaction and a gentle nanoindentation of the surface leads to flipping the defect atom into the local minimum geometry ($g \rightarrow l$ process).

The three different ways the P atom is bonded on the surface also provide also a critical test of a key issue in DSFM: the capability to achieve *chemical resolution*. The computed approach scans over the three different P sites (not shown) indicate that, at larger tip-sample distances, the bonding strength decreases roughly with the height of the P atom (p , g , and l). The comparison with the approach curve on the In atom shows that the antisite in its stable configuration cannot be distinguished from the nearby In atoms unless we come really close to the surface and enter the short-range repulsion regime. This theoretical result can explain why vacancies are the only point defects observed in the DSFM images of III-V (110) surfaces [281] in the topographic mode.

4.5 Simulation of chemical reactions: the importance of entropy

One important subject of computer simulation of matter is the issue of chemical reactions. This is a field customarily dominated by quantum chemistry and quantum chemistry methods. However, traditional quantum chemistry rarely deals with finite temperature effects²⁹. It is clear that at *finite temperatures* the relevant *thermodynamical potential is free*, rather than *total energy*. This

²⁷Simulation of NC AFM manipulation of an antisite defect (P atom substituting an In atom) on the InP(110) surface by a model Si tip. The animation shows tip-induced modification of the defect atom from a local minimum structure to a global minimum structure. A representative value of (pseudo)valence charge density is shown by a superimposed green isosurface.

²⁸Simulation of a reverse process of manipulation of the antisite defect on the InP(110) surface by a model H_2O tip from a global minimum structure to a local minimum structure. A representative value of (pseudo)valence charge density is shown by a superimposed green isosurface.

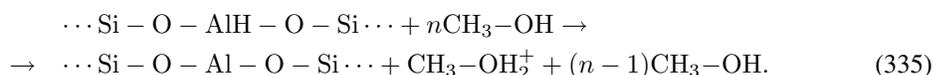
²⁹If finite temperature effects are included within the quantum chemistry approach, they are typically handled within harmonic approximation. Such an approach is extremely cumbersome for large systems and not accurate enough at high temperatures.

amounts to sampling of the vibrational entropy contribution. How can the *entropy* be “measured” on the computer? The entropy contribution is naturally sampled by any statistical method, such as molecular dynamics, but their quantification requires use of additional methods, which will briefly be described below. In addition, many chemical reactions do not take place in vacuum, but rather require a *catalyst*, say a surface, a supported nanoparticle, etc. This environment, in addition to finite temperature, is instrumental for understanding of “*real world*” chemistry. We demonstrate importance of both ingredients in an *industrial reaction*, the *methanol-to-gasoline process*. This reaction was at the center of interest of many quantum chemists for the last few decades and yet, no clear understanding of this complicated reaction has ever emerged from those studies. As shown below, the reason was that due to the computational limits imposed by the quantum-chemistry approach. The quantum chemistry model ignored both the proper environment in which the reaction proceeds as well as the effect of the entropy.

4.5.1 Simulation of the methanol-to-gasoline process

One of the most studied industrial applications of zeolites in current commercial production is the methanol-to-gasoline (MTG) process [282] for catalytic conversion of methanol to hydrocarbons. There is a large volume of experimental evidence that methanol, when catalyzed by an acidic zeolite, is first dehydrated to dimethyl ether (DME) which is then, in combination with methanol, converted to hydrocarbons up to C₁₀. The industrial process is catalyzed by ZSM-5; see Fig. 52, and proceeds at high methanol loadings of ~5–6 methanol molecules per acidic hydroxyl group at a temperature of 700K. The whole process involves a number of steps of increasing complexity: (i) the initial methanol adsorption; (ii) activation of the adsorbed species; (iii) dehydration to DME; (iv) formation of the C-C bond. It has, however, proved difficult to understand the role of the zeolite catalyst in the MTG process, because experiments, typically IR spectroscopy, do not provide a sufficiently complete and detailed atomistic picture of the process since they rely on interpretation in terms of an existing model. Recently, *ab initio* theoretical studies have started to shed light on these processes [283-288]. Here we only briefly mention the first two issues and focus instead on formation of DME in the zeolite under reaction conditions [288].

A consensus has emerged that as soon as the methanol loading reaches two molecules per acid site (... Si-O-AlH-O-Si ...) methanol is chemisorbed as a methoxonium cation (CH₃-OH₂⁺) [286, 287]



A more complete statistical sampling of the underlying potential energy surface was required to elucidate the activation of the adsorbed species [286, 287]. Solvation of the methoxonium ion in the methanol solvent has been found to further soften the methoxonium C-O bond [287] and lead to *activation*. Such strongly activated methoxonium complexes are expected to form DME easily. Experiments suggest a dramatic increase in DME formation in a narrow temperature interval around 500K [289]. Different mechanisms have been proposed for the formation of DME. Bandiera and Naccache have proposed the existence of two surface species [CH₃-OH₂]⁺ and [CH₃O]⁻ forming at the Brønsted acid and its adjacent Lewis basic sites [290] which can condense to produce DME and water. In the “*indirect*” pathway [291] the DME formation

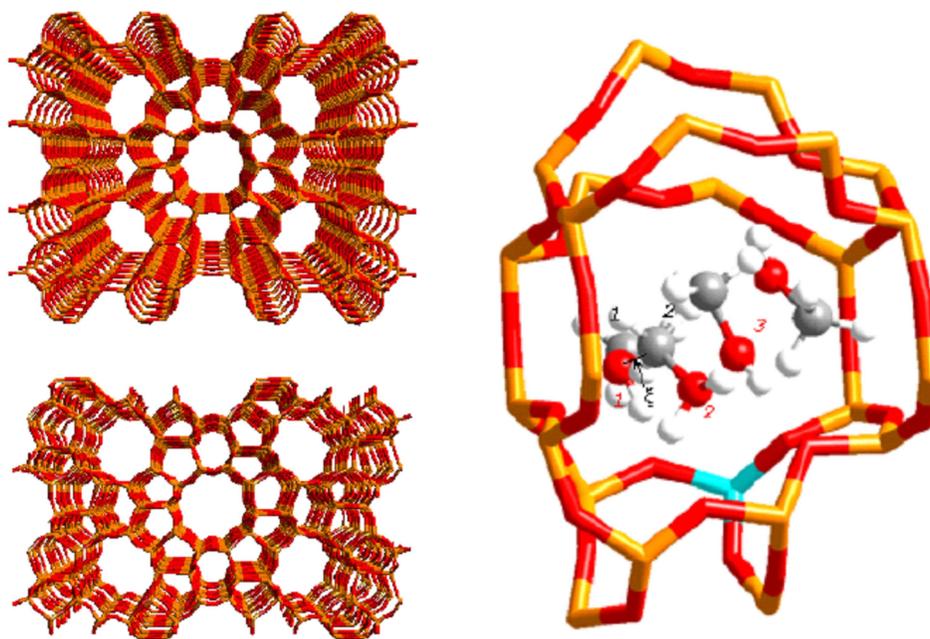
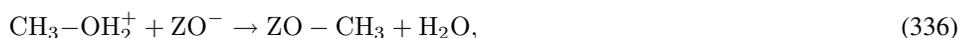
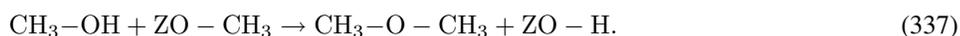


Fig. 52. Left panel: Examples of zeolite structures. Perspective view of ferrierite (upper panel) and ZSM-5 (lower panel) along the (straight) 10-ring channel. The ZSM-5 zeolite is the industrially used catalyst. Right panel: The model for DME formation. Four methanol molecules are loaded in the 8-ring. The black numbers label the carbon atoms, the red numbers the oxygen atoms in the methanol molecules. Molecule #2 underwent spontaneous protonation and forms a methoxonium cation. The holonomic constraint ξ is applied to oxygen #1 and carbon #2. The aluminum defect which is, before the methanol adsorption, compensated by hydrogen is shown in blue.

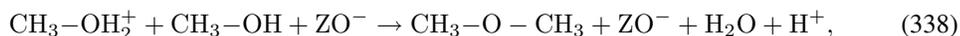
proceeds via a surface methoxy intermediate



which subsequently reacts with another methanol molecule to form DME



Here, Z stands for the zeolite framework. We assume here that the methanol is chemisorbed at an acid site as the protonated complex, since this is expected to be more susceptible to nucleophilic attack. Alternatively, in the “*direct*” pathway [283, 284]



both methanol molecules react with each other inside the zeolite environment which acts merely as a solvent. In reaction (338) the final destiny of the zeolitic proton was not specified, as the simulations performed under conditions which emulate those of the true reaction, suggest that

this proton may be mobile and need not necessarily permanently bind to the Brønsted site. Theoretical studies based on total energy conclude that the direct mechanism is preferred over the indirect one [283, 284]. These previous studies were based on transition state theory³⁰ [292] with the free energy profile approximated either by the internal energy or with the estimation of the entropy from the internal energy within the harmonic approximation. However, as shown below, the maxima in the free and total energy may not coincide which makes such an approach invalid for the present reaction. Nevertheless, only the direct process will be considered below. The study of formation of the first intermediate, the DME, in the MTG process is one of the most complex and comprehensive simulation of a chemical reaction performed. The model considered has been carefully chosen. The commercial catalyst ZSM-5 has a unit cell with 300 atoms, which is too large for the present simulations to be practical. Simulations for ZSM-5 on a picosecond time-scale have already been performed [286, 287]. Simulations for the time-scales relevant for the present simulation are gradually becoming feasible. The simulations were performed in *ferrierite*, see Fig. 52, which has a much smaller unit cell, with only 54 atoms, but a structure very similar to that of ZSM-5 (Fig. 52). The ferrierite structure is the closest mimic to the ZSM-5 structure one is able to find. In particular, it has two channels, a 10-ring channel similar to the straight channel in ZSM-5 and another straight channel with an 8-ring aperture, perpendicular to the 10-ring channel. Only one Brønsted acid site is considered, corresponding to a Si/Al ratio of 18. Given the dearth of experimental data and the computational cost of the present simulations no other Si/Al ratio was attempted. The reaction conditions have been simulated by loading four methanol molecules into the 8-ring channel and associated intersection regions of ferrierite (Fig. 52). The system was prepared so that two methanol molecules #1 and #2 in Fig. 52 can react without the need of reorientation. The temperature in the simulation was taken to be 700 K. This system was shown to form strongly activated methoxonium species [286, 287]. The activated species are expected to be susceptible to a nucleophilic attack by another methanol molecule to give rise to reaction (338). The ability of this system to exhibit activation makes this system a strong contender for the direct reaction pathway.

Ab initio MD simulations [288] have been performed for the formation of DME. It suffices to say that simulations were run in the (N,V,T) ensemble using DFT in its plane-wave pseudopotential formulation. Hence, periodic boundary conditions allow us to consider the full zeolite topology. Gradient corrected functionals are required for an accurate description of the DME formation and the PW'91 [58] variant of the GGA approximation to the DFT was used. Norm-conserving pseudopotentials were used to represent the core electrons and the wave functions of the valence electrons are expanded in plane waves at the Γ -point of the supercell with a cut-off of 40 Ryd. The huge nonuniform entropic corrections computed beyond the harmonic approximation are intuitively expected to be the major ingredient missing in all previous DFT calculations using identical or similar GGA energy functionals.

Free energies can be calculated using the “Blue Moon” ensemble [293, 294] variant of thermodynamic integration to overcome the large reaction barrier and to evaluate the *entropic contribution* along the reaction coordinate. This is a well known approach which, however, has not been applied very often to complex chemical systems. The constrained MD was performed by

³⁰The Transition State Theory (TST) locates the saddle point separating the reactant and product well of either total or free energy PES. In most cases TST considers only one-dimensional PESs.

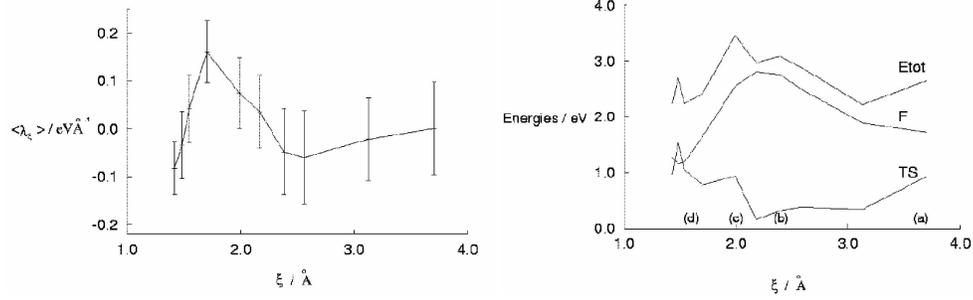


Fig. 53. Left panel: Variation (in $\text{eV}\text{\AA}$) along the reaction coordinate ξ of the Lagrange multipliers λ_ξ . Right panel: Variation along the reaction coordinate ξ of the free energy profile $\Delta F(T = 700\text{K})$; total energy profile $\Delta E(T = 700\text{K})$, and the entropy contribution TS . The zero of the vertical scale is arbitrary. (a)–(d) label the configurations shown in Fig. 54.

adding a holonomic constraint to the Lagrangean generating the MD,

$$L = \sum_{I=1}^N \frac{1}{2} M_I \dot{\vec{R}}_I^2 - U(\{\vec{R}_I\}) + \lambda_\xi [\xi \{\vec{R}_I\} - \xi_0], \quad (339)$$

where the first term on the r.h.s of Eq. (339) is the kinetic energy of the N_I ions, $U(\{\vec{R}_I\})$ is the many-body potential energy, equal to the Kohn-Sham energy, and λ_ξ is the Lagrange multiplier for the reaction coordinate $\xi\{\vec{R}_I\}$. In the present case of DME formation via the direct path, the reaction coordinate is the distance $|\vec{R}_{-CH_3} - \vec{R}_{-OH}|$ between the methyl group of the methoxonium ion and the hydroxyl group of one of the other methanols (c.f. Fig. 52) and ξ_0 is the externally fixed value of that distance. The constrained dynamics run was started from a well equilibrated unconstrained configuration. The constrained dynamics was run for ten values of the constraining distance ξ_0 ; the length of each run was ~ 2 ps after equilibration. The Helmholtz free energy profile was computed by

$$\Delta F(\xi, T) = \int_R^P \langle \lambda_\xi \rangle_{\xi_0, T} d\xi_0, \quad (340)$$

where R refers to reactants and P to products and the thermodynamic averaging $\langle \dots \rangle_{\xi_0, T}$ is performed by averaging over the MD trajectories. We note that formula (340) is correct only for a constraint consisting only of a distance $\xi = |\vec{R}_I - \vec{R}_J|$. The formulas for a more general case can be found in Ref. [294].

The Lagrange multipliers along the reaction coordinate, that ensure that the distance between the C atom on the methoxonium cation and the O atom on the other methanol molecule (Fig. 52) remains fixed, are shown in Fig. 53. These Lagrange multipliers are required to compute the free energy profile from formula (340). The computed free energy profile, the total energy

$$E_{tot} = \sum_{I=1}^N \frac{1}{2} M_I \dot{\vec{R}}_I^2 + U(\{\vec{R}_I\}),$$

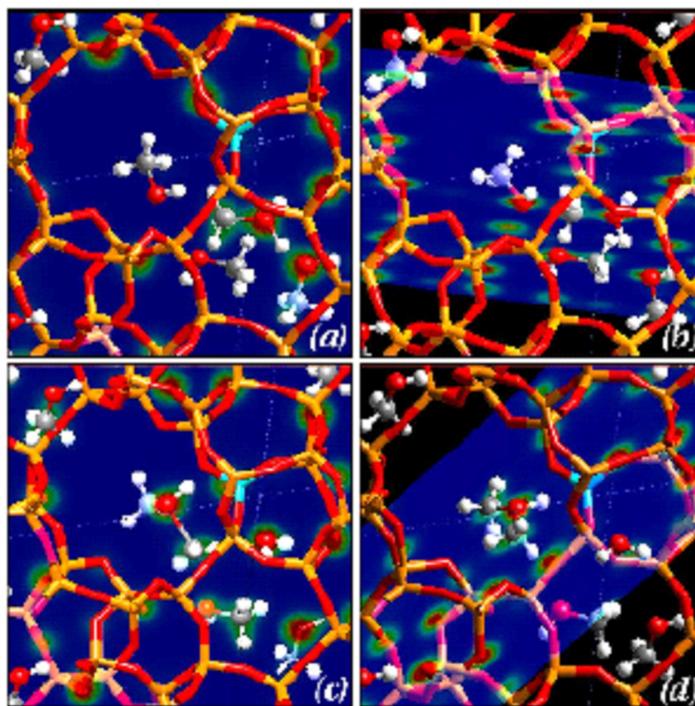


Fig. 54. Ball and stick models with superimposed valence electronic charge densities for points (a)–(d) along reaction coordinate defined in Fig. 53. The electronic charge density is shown on a plane defined by the oxygen #1, carbon #2, and the Al defect (Fig. 52).

and the entropy profile

$$TS(\xi) = E_{tot}(\xi) - F(\xi)$$

are shown in Fig. 53. Note that, unlike $E_{tot}(\xi)$, the zeroes of the $F(\xi)$ and $S(\xi)$ profiles are arbitrarily defined, and we only determine the profiles relative to the thermodynamic state of the reactants.

To give a better insight into the reaction process, we show in Fig. 54 characteristic configurations sampled from the MD trajectories of the reacting molecules in (a) the reactant well, (b) and (c) near the transition state, and (d) in the product well. Fig. 54(a) shows a configuration with the constraint ξ_0 corresponding to the value of the Lagrange multiplier $\langle \lambda_\xi \rangle = 0$. In this configuration the zeolitic proton, which originally formed part of one methoxonium cation $\text{CH}_3\text{-OH}_2^+$, is shared by two methanol molecules and executes a motion in a double well type of potential surface by hopping between the two methanol molecules. The configuration in Fig. 54(b) corresponds to the transition state in the free energy profile, where the water starts to dissociate from the methoxonium cation. In Fig. 54(c) the water molecule is completely dissociated and a new

chemical bond is forming between the methyl group of the methoxonium cation and one of the methanol molecules. This configuration corresponds to the transition state from the total energy curve. Figure 54(d) shows the configuration corresponding to the minimum of the free energy profile, where protonated DME was formed. Only a further compression of the bond by the applied constraint ($\xi_0 \sim 1.42 \text{ \AA}$) led to deprotonation of DME. This proton then became rather mobile on the MD scale. It moved away from the DME to form a hydroxonium cation. As a result the total energy E_{tot} in Fig. 53 started to decrease again, which suggests that this configuration is energetically more stable than the configuration with protonated DME. Given the fact that a very simple form of the reaction coordinate ξ was assumed and a single constraint applied to control the reaction it is important to assess the correctness of this choice. Two processes take place as the system climbs the reaction barrier. The applied constraint forces the formation of a chemical bond between the oxygen in the methanol #1 and the carbon atom on the methoxonium ion #2 (Fig. 52). During this process methoxonium is dehydrated which breaks the $\text{H}_3\text{C}-\text{OH}_2^+$ bond. However, these two processes do not take place simultaneously. It is found that the water molecule from the $\text{CH}_3-\text{OH}_2^+$ complex dissociates near $\xi \sim 2.38 \text{ \AA}$, before the other (C-O) bond is formed around $\xi \sim 2.0 \text{ \AA}$. The global maximum/saddle point corresponds to the transition state from the E_{tot} profile. Hence, there is no competition between breaking and forming chemical bonds. In particular, for $\xi \sim 2.38 \text{ \AA}$ the dissociated water does not take any active part in the DME formation and comes to equilibrium by optimizing the alignment of its dipole moment. In order to check the reversibility of the process tests have been made around the maximum/saddle point. In further checks another (metastable) configuration for $\xi \sim 2.16 \text{ \AA}$ was located with water still bonded in the $\text{CH}_3-\text{OH}_2^+$ complex. However, E_{tot} in this arrangement was $\sim 0.4 \text{ eV}$ higher than for the configuration with the water molecule already dissociated. On the other hand, configurations with dissociated water for $\xi_0 > 2.38 \text{ \AA}$ were unstable and spontaneously relaxed to the arrangement corresponding to the $\text{CH}_3-\text{OH}_2^+$ complex. Hence, the thermodynamically stable reaction path is the one given in Fig. 52 and the choice of the reaction coordinate ξ is meaningful.

From Fig. 53 we see that the total energy and free energy profiles along the reaction coordinate differ appreciably even at a qualitative level. In particular the entropic contribution to the reaction barrier is of the same order as the internal energy contribution and hence, any conclusion reached without explicitly including the entropy will be incorrect. This finding may not appear surprising at $T = 700 \text{ K}$. However, to the best of our knowledge, the huge nonuniform entropic corrections have never been properly treated before in theoretical modeling of the MTG.

We now discuss these features more in detail. The internal energy curve $E_{tot}(\xi)$ exhibits two activated processes; dissociation of water from the methoxonium cation around $\xi \sim 2.38 \text{ \AA}$ and reaction of the methyl group with the other methanol around $\xi = 2 \text{ \AA}$, separated by a minimum. The latter process corresponds to the transition state from $E_{tot}(\xi)$. On the other hand, the transition state from the $F(\xi)$ profile corresponds roughly to the former process of dissociation of water from the methoxonium cation. This clearly shows that the customary assumption of the dominance of the internal energy is *not* valid and that a more complex MD sampling of the internal energy surface is required.

The sampling of the flat anharmonic multi-minima internal energy surface leads to the huge and nonuniform variation of the entropic profile $S(\xi)$. As the entropy associated with the zeolite catalyst is approximately constant the complicated $S(\xi)$ profile can be understood in terms of elementary molecular processes as follows. Up to the transition state, the entropy decreases as

a function of ξ . This is caused by two different processes which lead to reduced mobility of the reacting methanol molecules. First, the two molecules are “glued” together by sharing the zeolitic proton (Fig. 54(a)). This proton moves between two methanol molecules (#2 and #3 in Fig. 52) which results in a very large rms fluctuation of the O-H distance when ξ is around 3.17 Å. A consequence of this proton transfer is that these two methanol molecules are effectively glued together by this proton. This results in a reduction of their mobility and hence a lowering of the entropy $S(\xi)$. A similar process can be seen to occur in Fig. 54(b), where now the methyl group is shared between the dissociating water and the reacting methanol and, hence gluing these groups together. Beyond the transition state a new molecule, is formed, initially characterized by one $\text{CH}_3\text{OH}^+-\text{CH}_3$ very loose chemical bond. In the region $\xi \sim 2$ Å this loose bond is responsible for the steep increase in the entropy. As can be deduced from the behavior of entropy around the minimum of $F(\xi)$ the protonated DME is characterized by an increase in number of possible configurations. The steep decrease in entropy for the smallest value of ξ is caused by deprotonation of DME which makes the product stiffer. The variation of the $S(\xi)$ profile explains also the other features above.

The analysis of the reaction in the MTG process leading to the first intermediate, the DME formation, demonstrates the wealth of information and insight which may be gained from a simulation of a complex catalytic reaction. Obviously, the model used above has a number of limitations dictated primarily by the limitations in computer power. However, there is no doubt whatsoever, that it will be possible to refine this model in the future and include the commercial catalyst ZSM-5 in the model, and extend the model to all the steps of the reaction including formation of the C-C bonds. On a more general ground, theory and simulation were promising for several decades to provide a reliable modeling tool for a wide range of catalytic reactions. Due to the limited accuracy and computer power such a tool was never provided. Recent advances in this field give hopeful outlooks.

4.6 Simulations in nanotechnology

The fascinating idea of harnessing the molecular building blocks to *assemble nanometer-scale devices* for applications ranging from electronic to biological, has been around for decades. However, only relatively recently advances in atomic-scale experimental imaging and manipulation techniques, most notably scanning probe microscopy (SPM), atomic force microscopy (AFM); see chapter 4.4 above, and dip-pen lithography have paved the way to the first practical realizations of that idea [295]. Promising *molecular-electronics* building-blocks started to appear, such as fullerenes, carbon nanotubes, self-assembled monolayers (SAMs) of thiolates ($\text{CH}_3-(\text{CH}_2)_m\text{-S-}$) on gold etc. [296]. A more traditional, albeit equally important, application of some of these systems is corrosion protection of metals (e.g. n-alkanethiolates on copper).

Synthesis techniques for these materials have been developed. For example, thiolate-capped nanometer-scale gold particles [297], or immobilized particles used as nano-contacts to molecules attached to an electrode surface can readily be prepared [298, 299]. Isolated metal clusters can be attached to a metal substrate via a molecular gate and interrogated individually by SPM techniques. In this way (i) two-terminal molecular junctions displaying *rectifying behavior* [300], (ii) three-terminal T- or Y-shaped junctions in *transistor architectures* showing field-effect modulation [301], and (iii) arrays of units for chemical, electronic or magnetic storage of information, have been synthesized. Self-assembly of complete integrated circuits from materials ranging

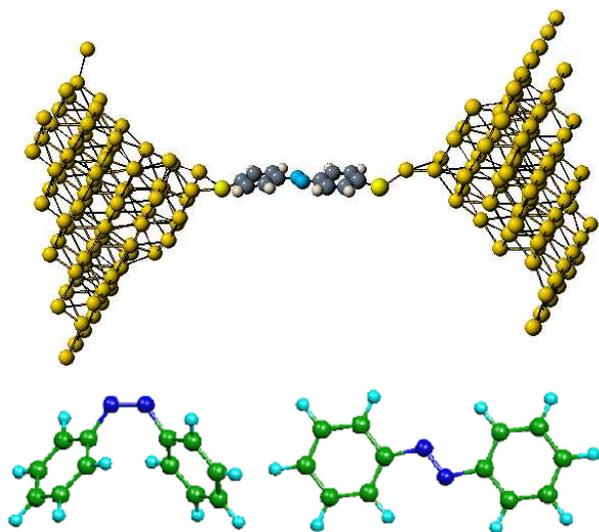


Fig. 55. Upper panel: Computer generated model [305] for a monomolecular nanojunction formed by dithio-trans-azobenzene linked via two S atoms to two realistic models of gold electrodes. Lower panel: cis-azobenzene (left) and trans-azobenzene (right). Note the significantly larger, by ~ 3.5 Å, bond length of the trans- isomer.

from monolayer protected clusters, nanotubes, conducting polymers, or even DNA, start playing an increasing role in nano-fabrication techniques. Research projects are underway to explore the possibilities to self-organize molecular systems into “cross-bar” arrangements for generating molecular storages based on single molecules [302]. These memory devices are based on supramolecular systems such as catenanes representing vectorial molecular systems, such that by making use of redox reaction bistable molecular states can be addressed on a single molecule level. Thus, the ultimate miniaturization limit will eventually lead to true “mono-molecular” devices in which all elements, including wiring, will be part of a single macromolecule. This conceptually appealing avenue will further increase the complexity of intelligent engineering of these extended multifunctional macromolecules.

While progress in the application driven areas of molecular- and nano-electronics proceed at a very rapid pace, the fundamental understanding of the underlying physics and chemistry still remains murky. This is especially true of the theory and simulation, where the complex molecular structures assembled and their measured transport characteristics are still awaiting firmer theoretical understanding. In the same way, even experimentally this rapidly proceeding field lacks so far the vigorous experimental scrutiny customarily done in more traditional areas of materials science. Here we will review work on *single-molecule devices*, such as *organometallic nanojunctions*.

4.6.1 Organometallic nanojunctions { $v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}$ }

As indicated above, thiolates are important molecules serving as convenient links to gold electrodes. Example of use of a thiolate-gold link is shown in Fig. 55. Often thiolates are used

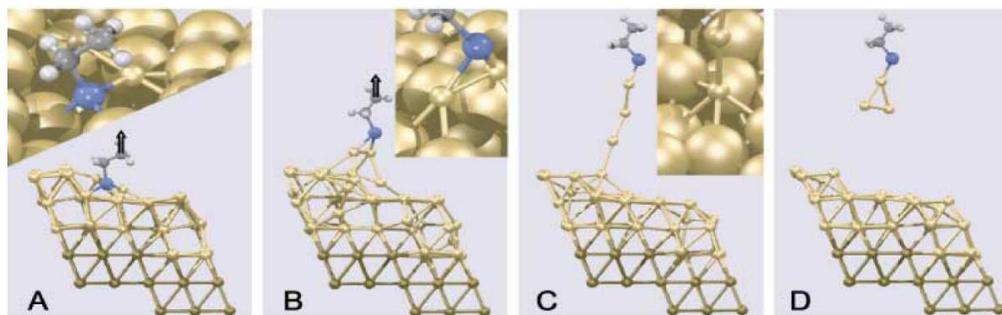


Fig. 56. Pulling an organometallic nanowire formed by ethylthiolate chemisorbed on a stepped gold surface emulated by a high-index Au(221) surface. The simulation was performed at $T = 300$ K by gradually lifting the z-coordinate of the C atom constrained to a plane [313].

to functionalize other, more complex organic molecules, to link them to gold. These technologies open up unprecedented new possibilities to control the *single molecule nanojunctions* and harness their *novel electronic and mechanical properties*. It is now possible to investigate the single-molecule nanojunctions experimentally by a static AFM apparatus which exerts *stress on the junction* (force-distance curves) in combination with *applied voltage* (distance conductance, distance I/V curves) [299, 303, 304] and possibly with other fields, such as e.g. laser light. *Light* can be used to *modify molecular conformation* of *chromophore molecules*, of which azobenzene with the two vastly different cis- and trans- isomers is probably the best known example; see Fig. 55. The electronic properties of such devices are being considered as possible candidates for molecular electronics applications.

In these technologies, the *self-assembled monolayers (SAMs)* of organic molecules on coinage metal surfaces, play important role and therefore are currently among the most studied surface systems [306, 307] with a number of potential applications ranging from surface coating and lubrication [308] to molecular electronics [296, 299]. Of these, the most studied SAM system is *thiolates on gold surfaces*. Relative ease of its preparation and its unusually high degree of order make this system a prototypical SAM. In distinct contrast to gold, comparatively little work has been done on thiolate bonding to other coinage metals. Here copper is especially interesting as a much cheaper potential substitute for gold. However, available experimental evidence suggests that substitution of gold by copper modifies also the underlying chemistry of the metal-molecule bond. Hence, the thiolate-copper bond needs to be analyzed and understood. The metal-sulfur bond for both metals (Cu, Au) is analyzed below. In this context we note that gold is a very special coinage metal as it is characterized by significant *relativistic effects* [309, 310] absent from other such metals. The relativistic effects strongly contract the *s* shell(s) which leads to preference for low coordination/low dimensionality [311, 312]. These phenomena manifest themselves by a tendency to form wires when pulled using mechanically controllable break junctions [311, 312] or a “thiolate hook”, see Fig. 56 { v_4 ³¹} (from Ref. [313]).

³¹Simulation of a process of pulling (mechanochemistry) a methylthiolate ($\text{CH}_3\text{-CH}_2\text{-S}$) molecule chemisorbed on a stepped gold surface at 300 K by static AFM apparatus. The action of the AFM is emulated by constraining the methyl group (CH_3) of the thiolate molecule to a plane and slowly lifting it up.

There are several experimental studies confirming thiolate SAMs formation on low-index Cu surfaces. X-ray standing waves study of adsorption of short-chain alkylthiolates on Cu(111) [314, 315] suggests the existence of a thiolate phase at low temperature, and coexistence of different low- and high-temperature thiolate phases, but also an atomic sulfur phase at room temperature. Similarly, highly ordered thiolate phases, and at higher temperatures, atomic S phases or combinations of the above were observed on Cu(100) and Cu(110) surfaces using thermal desorption and XPS techniques [316, 317]. Experiments also suggest a completely different thermal dissociation of thiolates on gold and copper surfaces. Heating the Cu(111) surface covered by S-CH₃ adlayer desorbs CH₃ radicals [314], whereas molecular S-CH₃ desorbs from the Au(111) counterpart [318].

A novel and totally different way of detaching organic molecules from metals is provided by “mechanochemistry” where mechanical energy, instead of the traditional thermal energy, is used to induce chemical reactions. In such experiments, an external pulling force is applied to a nano-junction consisting of molecule chemisorbed onto a metal surface or cluster all the way to destruction of the junction. Such a proposition was studied in recent simulations, where the pulling-induced rupture process was investigated for thiolate/gold contacts [313, 319], see Fig. 56, mimicking AFM experiments designed to probe individual chemical bonds [320]. These simulations strongly indicated that, instead of detaching a thiolate radical with the associated breaking of the Au-S contact, as one might naively expect based on results of thermal experiments [318], a series of isomerization steps involving gold atoms led to extraction of a monatomic gold nanowire or to isomerization and unfolding of clusters. Hence, for the gold-thiolate system, different bond scission scenarios can be obtained by using mechanochemistry [313, 319], and thermal desorption. The two scenarios (thermal v.s. mechanochemistry) for both systems (Cu, Au) are analyzed below.

We first target systematically the neutral CH₃-(CH₂)_m-S-M_n adducts (where $m = 0, 1$ and $n = 1, 3, 5, 7,$ and 9 , the metal atom M being Cu and Au) as well as their neutral, cationic, and anionic dissociation fragments [321] in order to assess different channels to rupture. For the sake of comparison, relevant results from a related study of the Cu(111) and Au(111) surfaces are also included. We limit the cluster sizes M_n to odd numbers n as to avoid further complications due to unpaired electrons and open shells. In order to facilitate comparison with copper, the gold cluster structures which we employ here are, with the exception of the $n = 3$ adduct, isostructural with those used for copper. The model of the Cu(111) surface consists of a 6-layer slab with periodic boundary conditions hosted in a $c(2 \times 2)$ supercell setup with a methylthiolate coverage corresponding to one molecule per supercell [322]. As a model of the Au(111) surface we also employ a 6-layer Au(111) slab with a SAM in the $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure based on an orthorhombic supercell containing two methylthiolate molecules. These surface results represent the “asymptotic” ($n \rightarrow \infty$) reference values for fragmentation energies.

In experimental studies various thiolate chain lengths are used, from $n = 1$ [314, 315] up to $n = 30$ [323]. A computational study based on these systems would be more expensive employing a fully *ab initio* description of the alkyl chain. However, the previous study of thiolated gold clusters [324] demonstrated that adsorption and fragmentation energies obtained with chains as short as two (one) carbon atoms provide accurate quantitative (qualitative) representatives for results with longer chains. For those reasons in what follows we considered thiolate species with up to two carbon atoms only.

The majority of the results were obtained in the framework of the DFT with the (spin-

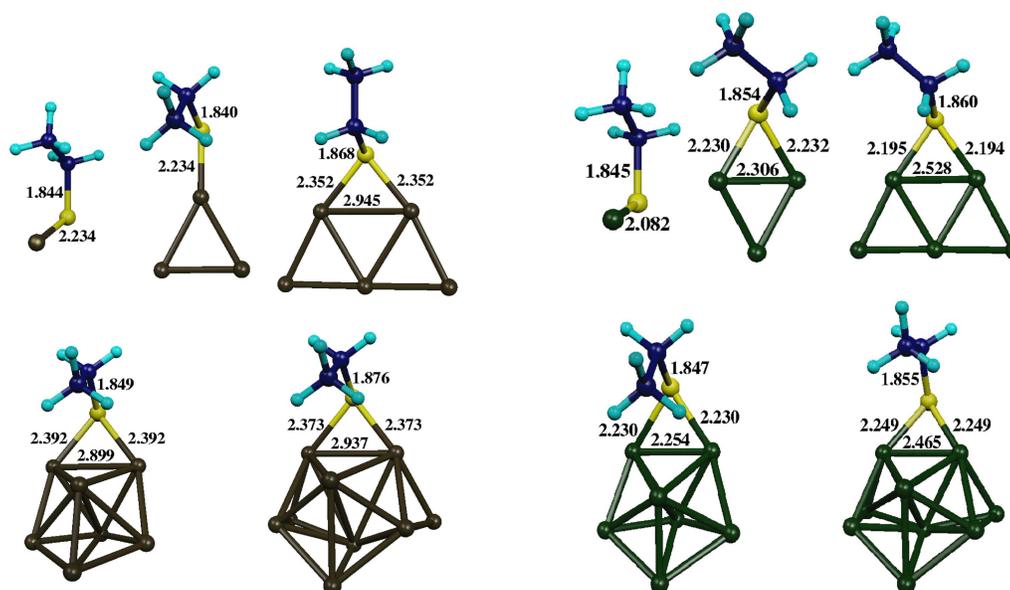


Fig. 57. Relaxed geometries of the $\text{CH}_3\text{-CH}_2\text{-S-Cu}_n$ (left panel) and $\text{CH}_3\text{-CH}_2\text{-S-Au}_n$ (right panel) adducts for $n = 1, 3, 5, 7,$ and 9 . Interatomic distances are given in Å.

polarized) PBE functional [59] and plane-wave basis sets. The calculations involving Cu were performed with ultrasoft pseudopotentials [125, 126] throughout with a plane-wave cutoff of 25 Ry, including the $3d$ and $4s$ valence electrons of scalar-relativistic Cu. Calculations involving Au clusters were performed with a scalar-relativistic normconserving pseudopotential [121] for Au and a cutoff of 60 Ry. All calculations were done in a large supercell with edge length of 13 Å for Cu and 14 Å for Au. For thioliates on Cu(111) a 2×2 Monkhorst-Pack k-point mesh [130] and an electronic smearing of 0.2 eV was used, whereas the Γ -point approximation was used for the larger Au(111) model. The dynamics simulations were done using thermostated Car-Parrinello dynamics.

For the sake of comparison an extensive set of calculations using localized basis sets (with a “small-core” SSD effective core potential and associated basis set for Au whilst for all other atoms all electrons were explicitly treated and the TZVP basis set was used; many of the results were also obtained with the 6-311G** basis set), using the PBE functional and second-order Møller-Plesset perturbation theory (MP2) was performed. This allows to assess the pseudopotential approximation or the importance of van der Waals interactions.

The structures obtained by local optimization methods for M_n -ethylthiolate for $\text{M}=\text{Cu}$ and Au are depicted in Fig. 57. For the Cu(111) surface the thiolate is found to be three-fold coordinated with Cu-S interactions consisting of two bond distances of 2.26–2.29 Å and a third one at 2.62 Å associated with a tilting of the 1.85 Å long S-C bond away from the surface normal. For the case of Au(111), two-fold coordination is indeed obtained but with substantially longer Au-S distances of 2.45 Å and a shorter S-C bond of 1.82 Å. The fact that the cluster structures

do differ from that found on the surface suggests a priori that neither series of clusters will yield fragmentation energies entirely consistent with the bulk metal surface even for the largest values of n considered in this work.

We consider now fragmentations of $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n$ adducts into the following products:

- (1) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow (\text{CH}_3\text{-(CH}_2)_m\text{-S})^- + \text{Cu}_n^+$
- (2) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow \text{CH}_3\text{-(CH}_2)_m\text{-S} + \text{Cu}_n$
- (3) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow (\text{CH}_3\text{-(CH}_2)_m\text{-S})^+ + \text{Cu}_n^-$
- (4) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow (\text{CH}_3\text{-(CH}_2)_m)^- + (\text{Cu}_n\text{-S})^+$
- (5) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow \text{CH}_3\text{-(CH}_2)_m + \text{Cu}_n\text{-S}$
- (6) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow (\text{CH}_3\text{-(CH}_2)_m)^+ + (\text{Cu}_n\text{-S})^-$
- (7) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow \text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_{n-2} + \text{Cu}_2$
- (8) $\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_n \rightarrow \text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_{n-4} + \text{Cu}_4$

as a function of cluster size in the range $n = 1, 3, \dots, 9$ and in this section only for the methyl systems, i.e. $m = 0$. Here, fragmentation pathways (1) – (3) describe scission of the Cu-S bond, pathways (4) – (6) correspond to breaking of the S-C bond, and pathways (7) – (8) describe cleavage of the metal cluster itself. The calculated fragmentation (i.e. dissociation or binding) energies $\Delta E = \sum E_{\text{prod}} - E_{\text{adduct}}$ are given in Tab. VIII. For the fragmentation schemes (1) – (6) we take the two product fragments and optimize their structures separately to the respective nearest local minima. This leads in all cases, with the exception of $(\text{Cu}_5\text{-S})^+$, to isostructural species. Thus, we have found that each fragment (with the above exception) adiabatically relax to a nearby isostructural minimum which is a likely occurrence if the process is thermally activated. On the other hand, it is unlikely to be valid if an external mechanical pulling force is applied to achieve fragmentation [314]. Pulling for instance by an AFM tip, while the cluster itself is fixed by one or several atoms to a surface [314] might break the cluster and/or induce isomerization [325]. Thus, it should be noted that fragmentation under thermal activation and mechanical pulling may lead to very different fragmentation pathways.

Before analyzing the results we validate the different approximations used in calculating the energy differences. To this end in particular the PBE functional and finite basis set size using

Tab. VIII. Reaction energies ΔE (eV) for methylthiolate- Cu_n adducts calculated in the plane-wave basis set and employing the PBE exchange-correlation functional.

#	channel	$n = 1$	$n = 3$	$n = 5$	$n = 7$	$n = 9$
(1)	$(\text{CH}_3\text{-(CH}_2)_m\text{-S})^- + \text{Cu}_n^+$	9.5	7.9	8.6	7.2	6.9
(2)	$\text{CH}_3\text{-(CH}_2)_m\text{-S} + \text{Cu}_n$	2.9	3.6	3.8	2.8	3.1
(3)	$(\text{CH}_3\text{-(CH}_2)_m\text{-S})^+ + \text{Cu}_n^-$	10.5	11.3	10.9	9.9	10.6
(4)	$(\text{CH}_3\text{-(CH}_2)_m)^- + (\text{Cu}_n\text{-S})^+$	11.7	8.3	8.8	7.9	6.9
(5)	$\text{CH}_3\text{-(CH}_2)_m + \text{Cu}_n\text{-S}$	2.9	2.4	2.6	1.6	1.4
(6)	$(\text{CH}_3\text{-(CH}_2)_m)^+ + (\text{Cu}_n\text{-S})^-$	10.8	10.1	9.9	9.3	9.0
(7)	$\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_{n-2} + \text{Cu}_2$	-	2.1	3.0	2.4	2.9
(8)	$\text{CH}_3\text{-(CH}_2)_m\text{-S-Cu}_{n-4} + \text{Cu}_4$	-	-	3.3	3.5	3.4

Tab. IX. Reaction energies ΔE (eV) calculated with: pwPBE, locPBE and MP2 for isostructural Cu_5 and Au_5 ethylthiolate. Pw stands for plane wave basis set and PBE functional, loc for localized basis sets and PBE functional, MP2 for second-order Møller-Plesset perturbation theory.

#	channel	M=Cu	M=Cu	M=Cu	M=Au	M=Cu	M=Cu
		pw	loc	MP2	pw	loc	MP2
(1)	$(\text{CH}_3-(\text{CH}_2)_m\text{-S})^- + \text{Cu}_n^+$	8.5	8.3	7.9	9.6	9.1	8.9
(2)	$\text{CH}_3-(\text{CH}_2)_m\text{-S} + \text{Cu}_n$	3.9	3.9	4.3	3.4	3.3	3.6
(3)	$(\text{CH}_3-(\text{CH}_2)_m\text{-S})^+ + \text{Cu}_n^-$	10.8	11.1	12.2	10.1	9.3	9.8
(4)	$(\text{CH}_3-(\text{CH}_2)_m)^- + (\text{Cu}_n\text{-S})^+$	8.9	10.0	13.4	12.0	12.3	13.3
(5)	$\text{CH}_3-(\text{CH}_2)_m + \text{Cu}_n\text{-S}$	2.6	3.4	3.5	2.6	2.7	3.1
(6)	$(\text{CH}_3-(\text{CH}_2)_m)^+ + (\text{Cu}_n\text{-S})^-$	7.9	9.3	9.5	7.9	10.4	12.7
(7)	$\text{CH}_3-(\text{CH}_2)_m\text{-S-Cu}_{n-2} + \text{Cu}_2$	3.1	2.9	2.6	3.1	3.4	3.3

fragmentation energies for $\text{CH}_3\text{-CH}_2\text{-M}_5$ with $\text{M}=\text{Cu}$ and Au as a test example w.r.t. MP2, see Tab. IX, have been tested. This is an important issue quite in general as in most of the applications considered here plane wave basis sets were used. As discussed above, plane waves have the advantage of being totally unbiased and well controllable by a single parameter, the plane wave cutoff. In general, the qualitative agreement between locPBE and MP2 employing the same basis is quite good, with the notable exception of the highest energy reaction channels such as the fragmentation of $\text{C}_2\text{H}_5\text{S-Cu}_5$ into $(\text{CH}_3\text{-CH}_2)^- + (\text{M}_5\text{-S})^+$, see Tab. IX. For these higher energy channels variation on the order of an eV between the various computational schemes is common; here the anionic metal clusters as reaction fragments might in addition pose problems for the PBE approach relative to the MP2 treatment. However, these errors do not alter the relative ordering of the lowest energy reaction channels. Both methods (locPBE and MP2) find similar differences between S-C and M-S fragmentation energies. Thus, it is concluded that the relative bonding scenarios including various fragmentation scenarios are reasonably well represented within the confines of the PBE functional. Comparison (for Cu-thiolate systems) of plane-wave pseudopotential results (pwPBE) to those of the all-electron localized basis set results (locPBE) are in reasonable (10 %) agreement, with the notable exception that S-C bond cleavage on the copper species is lowered by ~ 0.8 eV. We examine the error resulting from the type of the localized basis by employing a second triple zeta basis set of similar quality as TZVP, which is 6-311G**. In this case the errors were found to be significantly larger than those resulting from the pseudopotential approximation within the plane-wave scheme. In case of the copper systems S-C bond rupture and cluster fragmentation are respectively calculated to be 3.4 eV and 2.9 eV with TZVP but found to be 4.2 eV and 2.1 eV with the 6-311G** basis set with similar sized errors occurring in most other reaction energies. For the gold system the agreement is slightly better. Thus, the relative error induced from basis set size is significant and gives a very strong support to use of the plane wave basis sets, which are from this point of view totally unbiased.

The energy differences ΔE between the thiolated adducts and the dissociation products reported in Tab. VIII show that only two out of the eight pathways are energetically competitive, namely cleavage of the S-C bond, pathway (5), and fragmentation of the copper cluster, path-

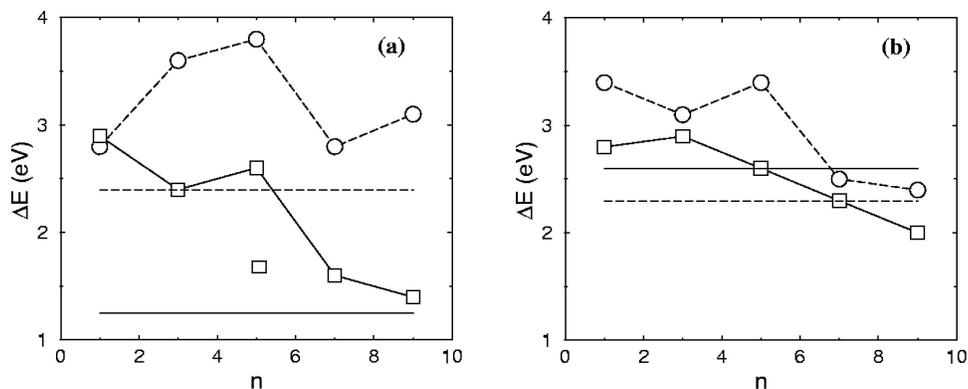


Fig. 58. Fragmentation energies for breaking the S-C (squares and solid lines) and the M-S (circles and dashed lines) bonds for thiolates chemisorbed to M_n clusters. (a): methylthiolate- Cu_n clusters and (b): ethylthiolate- Au_n clusters. The result for the three-dimensional Cu_5 isomer is shown by a filled square (see text) and the surface results for S-C and M-S bonds are marked by horizontal solid and dashed lines, respectively.

way (7). In particular, dissociation into neutral species is always strongly energetically favored over dissociation into charged fragments so that only neutral species will be considered in the following discussion.

The most important fragmentation energies are plotted in Fig. 58 (a) in order to display more clearly trends as a function of cluster size n . The results show systematic and pronounced lowering of the S-C bond fragmentation energy (squares) with increasing cluster size. For the larger clusters, $n = 7$ and 9 , this energy approaches the value of 1.3 eV as obtained for methylthiolate on a $\text{Cu}(111)$ surface [322] using the same electronic structure method. Overall, the binding energy is found to decrease by more than a factor of two with respect to the limiting case $n = 1$ based on both the $n = 9$ cluster and the surface. Correspondingly *the S-C bond in the thiolate adduct is significantly destabilized in larger clusters*. The only exception from the trend seems to be the $n = 5$ case, which needs some clarification. Bare Cu_5^- , Cu_5 , and Cu_5^+ clusters are known to have planar shapes. However, there exists also a three-dimensional (3D) Cu_5 isomer and its associated adduct and thus fragments, which is however ~ 0.8 eV higher in energy compared to the 2D adduct. As mentioned above, the 3D cluster structure is important for the $(\text{Cu}_5\text{-S})^+$ fragment. The relaxation of the higher energy 2D fragment to the 3D non-isostructural isomer is a result of the high electronic gradients observed in the initial structure as well as the relatively close energetics of 2D and 3D Cu_5 clusters. Hence, it is worthwhile to consider both isomers in order to test also the effect of cluster dimensionality upon the results. Calculation of fragmentation path (2) with the 3D Cu_5 constituent leads to a lowering of the S-C bond energy to 1.7 eV, see the filled square in Fig. 58(a). Thus we observe a monotonic decrease of the S-C bond energy provided that Cu_n clusters of maximum dimensionality (for given n) are considered.

A weak S-C bond for large n is consistent with experimental surface studies [314-317] which show desorption of hydrocarbon chains starting at temperatures of few hundred Kelvin and full desorption at around 500 K. On the other hand, no systematic trend can be found for the S-Cu

bond based on studying clusters as a function of n . The surface reference of about 2.4 eV, on the other hand, is consistently lower by about 0.5 to 1.5 eV and reflects the fact that binding to the cluster is different from that on the surface. Likewise, cluster fragmentation energies are also highly sensitive to the cluster structure for these small species as evidenced by the wide variation as a function of n for the energy of reaction (7) which oscillates on the order of 0.5 eV. This makes clear the point that the clusters are not sufficiently large enough to quantitatively reflect the bulk metal surface and hence the detailed energetics is still governed by molecular details. Hence, it is surprising in this light that S-C bond rupture has such a clear dependence on n and argues toward a mechanism that is only grossly dependent upon the cluster geometry but not on the thiolate binding configuration.

For Au it was also found that fragmentation channels into charged species are significantly higher in energy compared to fragmentation into neutrals; see Fig. 58 (b) for the $n = 1, 3, 5, 7$ and 9 Au_n species. In all the obtained species, with the exception of the $n = 3$ cluster, the thiolate prefers a two-fold coordination pattern similar to that reported on the Au(111) surface [326]. The lowest energy channels at all cluster sizes is fragmentation of the cluster core into a smaller Au_k ($k < n$) cluster and a Au_{n-k} /thiolate adduct, the only exception being 2D $n = 5$ species where it is 0.5 eV higher in energy than S-C bond rupture. For the smallest gold clusters studied (≤ 5) cleavage of the S-C bond is found to be about the same as observed on copper clusters. However, unlike the scenario for copper, increasing the cluster size has only a marginal effect on weakening this bond. Specifically, the *S-C bond energy decreases in value by only about 10 %* as the cluster size is increased from $n = 3$ to $n = 9$ as opposed to about 50 % found for ethylthiolate on copper clusters. As a result the S-C bond cleavage energy in the thiolate bound to Au_9 is 1.4 eV higher than that of the copper analogue. Unlike copper, desorption of the thiolate molecule decreases in energy from $n = 3$ to $n = 9$ such that it also becomes competitive with S-C bond rupture for the larger species. Hence, although the smaller n species exhibit fragmentation energetics similar to those found for the copper system, the size evolution of these energies and most notably the pronounced weakening of the S-C bond, is quite distinct.

Why is gold so different from copper? One could naively assume that given the fact that both metals are isoelectronic in their valence shells, their properties would be also very similar. However, just the opposite is true and the two metals couldn't be more different! These differences are behind the nobility of gold. To delve into the reason behind these observations we briefly examine the hypothetical situation where relativistic effects are neglected for gold. *Relativistic effects* have long been recognized to be crucial in understanding the chemistry of gold [309]. In brief, relativistic stabilization and contraction of the Au 6s orbital leads to the large increase in both the ionization potential (IP) and the electron affinity (EA), hence to a large electronegativity of gold, and a shortening of intermolecular bond lengths. Akinaga *et al.* [326] were the first to consider relativistic contributions in the context of SAMs on a gold surface.

To achieve this we have performed a little "alchemy" and constructed a non-relativistic norm-conserving Au pseudopotential [121]. As a test of the reliability of this approach we consider the IP of the 6s electron in the gold atom which is calculated to be 9.1 eV with our scalar relativistic potential whereas it reduces to 6.8 eV in the non-relativistic case. Additionally, the Au_2 diatomic molecule is found to have a bond length of 2.54 Å (using PBE and the 11-electron scalar-relativistic pseudopotential). Removal of the relativistic corrections causes the Au-Au bond to lengthen substantially to 2.98 Å in agreement with the above described qualitative trend.

We now consider fragmentation energies for a non-relativistic thiolate- Au_9 adduct. Our

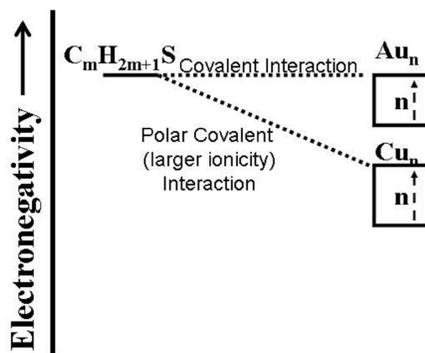


Fig. 59. Schematic representation of trends in electronegativities, EN, between thiolate radicals, Au_n and Cu_n and its effect upon metal-sulfur bonding. The scheme is based upon computed EN values defined as $EN = (IP + EA)/2$, where IP is ionization potential and EA electron affinity.

model finds that the weaker binding of thiolate to Au, relative to Cu, is a result of relativistic effects, resulting into dramatic changes in both structure and energetics of the adduct. The Au-S interaction retains its two-fold bonding structure, however the bond length increases from 2.37 Å to 2.95 Å in the non-relativistic limit, which is accompanied by an expansion of the Au_9 cluster core as expected. Despite this pronounced lengthening of the Au-S bond we observe a significant increase in the thiolate binding energy, from 2.8 eV to 4.0 eV in the non-relativistic case. Even more notable is the observation that dissociation into Au_9S and C_2H_5 is dramatically lowered from 2.5 eV to 0.4 eV. Thus, for the non-relativistic gold cluster, the trend is to strengthen the metal-sulfur bond and, simultaneously, to lower the S-C binding energy signaling a strong preference of S-C bond scission over Au-S rupture as induced by the neglect of relativistic corrections. This trend is remarkable also when comparing copper and relativistic gold, see Fig. 58. In a nutshell, one could say that the neglect of relativistic effects in gold leads to a behavior that mimics that of copper. Hence, it is concluded that the essential difference between copper and gold w.r.t. their interaction with thiolate molecules in particular dissociation pathways is largely due to relativistic electronic structure effects.

We now seek the answer to the question of the nature of the M-S bond in the two systems and to the most notable trend of the S-C bond weakening in the thiolate- Cu_n system with increasing n . In order to provide an understanding of the chemical bonding with a clear focus on the difference between copper and gold it is useful to consider first a simple qualitative descriptor. Given that copper is less electronegative than sulfur one would be tempted to formally assign the sulfur atom with a negative charge and the copper cluster with a positive charge. This conceptual procedure would result in an ionic picture of thiolate-metal bonding along with the associated non-directionality in the thiolate-copper interaction. As already mentioned above however, this is strikingly distinct from the thiolate-gold SAMs where the current view is of a more covalent and directional Au-S interaction.

Some insights can be obtained from the calculated electronegativities (EN) of relevant fragments which can help to distinguish between covalent and ionic bonds, see Fig. 59 for a schematic representation. The calculated EN value, obtained by the Mulliken definition, is found to be 5.4 eV for both methyl- and ethylthiolate. This results from the fact that both HOMO and LUMO orbitals involved in ionization and electron capture are predominantly centered on the S atom. In comparison, with Au_n clusters the EN values are found to range between 4.7 and 5.4 eV with the larger clusters having a propensity toward a small EN value due to a lower ionization potential. This results in a small EN difference between Au_n clusters and thiolate, leading to a rather covalent and thus directional Au-S interaction. In comparison, Cu_n clusters are found to have much smaller EN values ranging between 3.6 and 4.2 eV, again with a propensity toward lower EN value for larger clusters which. Hence, the Cu_n and thiolate radicals have a larger EN difference ranging from 1.2 to 1.8 eV and thus should have a more ionic, or polar covalent bonding interaction than their Au_n counterparts. Note that this interpretation fits qualitatively well with our observations above that non-relativistic Au behaves more like copper in that absence of relativity decreases IP and EA of Au. Hence, non-relativistic gold clusters lower their electronegativity and thus lead to more ionic nature of Au-S bonding akin to copper. In order to understand the pronounced weakening of the S-C bond with the size of the Cu_n , we consider a more detailed electronic structure analysis using the one-electron orbital picture. The main features governing the S-C bond weakening as a function of the metal cluster size can be gleaned from the spatial distribution of just a few eigenstates, see Fig. 60. The other tool used to obtain a more detailed information on electronic structure is the Mulliken population analysis (MPA) by projecting Kohn-Sham molecular orbitals (MOs) expanded in plane-waves onto a minimal basis set composed of numerical (pseudo)valence atomic orbitals of S, C, H and Cu atoms [327]. Taking into account the well-known drawbacks and caveats of MPA itself and the intermediate projection step, this analysis allows to identify the most significant relative trends if studied as a function of cluster size. The spilling parameter, indicating the quality of the projection of eigenstates onto the localized orbitals, is always around 1 %, indicating almost complete projection onto the minimal basis set.

The findings from MPA and wave function analysis in Fig. 60 can be summarized as follows. The two lowest eigenstates at ~ -18.5 eV and ~ -15 eV correspond to bonding and antibonding C(2s)-S(3s) σ -type MOs, respectively. These MOs change only slightly upon variation of the number of Cu atoms in the cluster. The MOs #3 and #4 are mainly S-C π -type bonding orbitals formed largely by C(2p_x), C(2p_y) and by S(3p_x), S(3p_y) AOs. The lower one of these orbitals loses a part of its S-C population with increasing cluster size (about 10 % of the charge moves towards the S-Cu region), which, though, weakens the S-C bond only negligibly. Of all eigenstates there are just two which are significantly populated by C(2p) and/or S(3p) electrons and simultaneously exhibit a strong dependence on the size of the copper cluster. These are MOs #5 and #6 with energies around -9 eV and with their energy separation changing between 2 eV and 0.5 eV, which are formed by C(2p_z), S(3p_z), and Cu(3d) orbitals. These MOs are positioned below the bottom of the Cu(3d) and Cu(4s) range of energies and for large clusters they are similar to “impurity levels” in the Cu(3d) “band” which is in the range ~ -8 eV to ~ -4 eV. For $n = 1$ the lower orbital is strongly populated on the S-C bond (forming a σ -type bonding orbital) and the higher one (which has a more complicated spatial distribution and less bonding character) gives very little population into the S-C bond. Increasing the cluster size causes (apart from the creation of two quasi-symmetrical S-Cu contacts) the S-C bonding contribution in MO

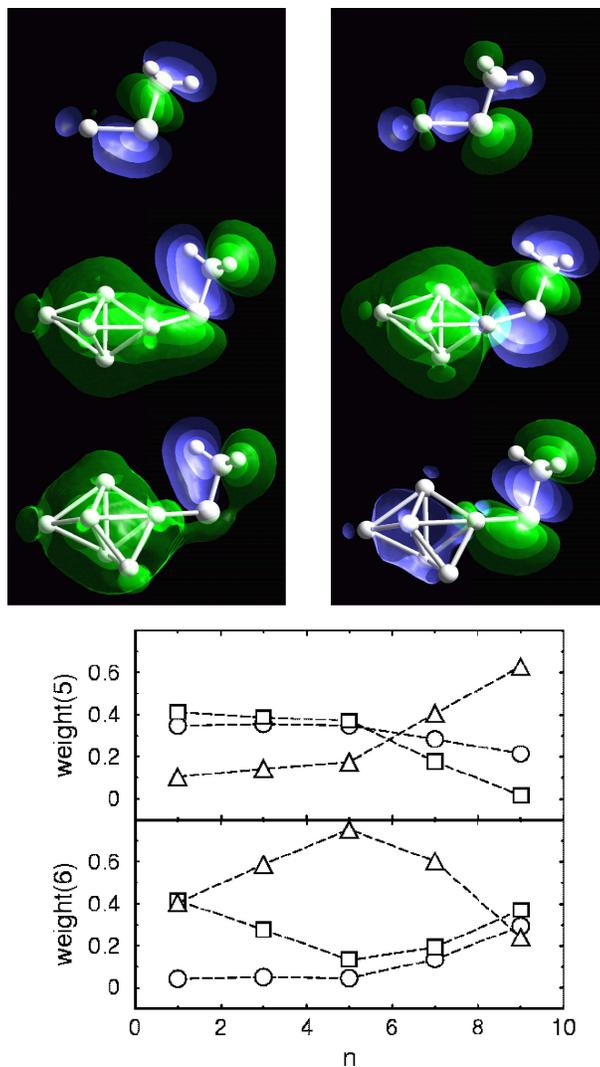


Fig. 60. Analysis of S-C bond weakening in $\text{CH}_3\text{-S-Cu}_n$. Upper panel: Spatial distribution of Kohn-Sham eigenstates #5 (left column) and #6 (right column) for $n = 1, 7$, and 9 . Blue/green color indicates the sign of the wave function. Lower panel: Corresponding Mulliken population analysis for the orbitals #5 (upper subpanel) and #6 (lower subpanel) as a function of cluster size; results for p orbitals on C (circles) and S (squares) are summations over p_x, p_y , and p_z projectors and results for Cu (triangles) are sums over $4s$ and $3d$ projectors.

#5 to decrease. This is counter-balanced by a simultaneous gain of population in MO #6, see Fig. 60. Close examination of Fig. 60 also shows that actually a crossover occurs around $n = 5$ in concurrence with the enhancement of the weakening of the S-C bond upon the 2D-3D cluster

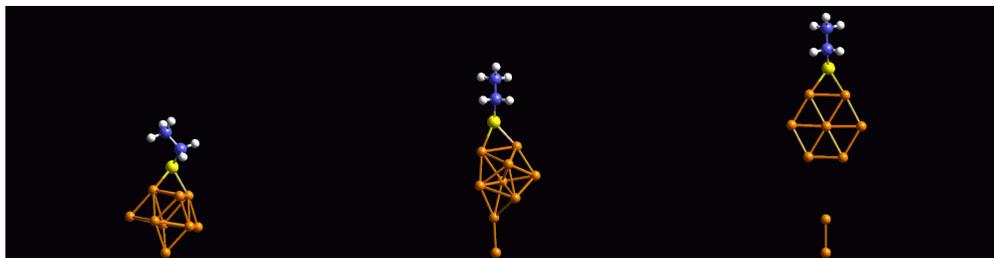
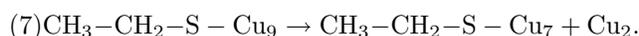


Fig. 61. Static simulation ($T = 0$ K) of mechanochemistry on $\text{CH}_3\text{-CH}_2\text{-S-Cu}_9$. The system was constrained in a way similar to the one used in [314], see Fig. 56, i.e. the methyl group was fixed to a plane and lifted up, while the Cu atom on the bottom of the Cu_9 was kept fixed.

structural transition. In agreement with our simple model based on electrostatics, the increase in the interaction with the larger copper cluster results in a mixing of occupied S-C MOs on the thiolate, such that $\text{S}(3p_z)$ population is transferred from MO #5 to the higher in energy MO #6, thus weakening the S-C bond.

Given the above analysis, one might conclude that the thiolate-copper nanojunction would be unstable against modest heating and that it would fragment along the C-S bond also under mechanical pulling, i.e. if mechanochemistry is applied. As Fig. 61 demonstrates, the last assumption is certainly wrong [325] $\{v_5^{32}, v_6^{33}, v_7^{34}, v_{10}^{35}, v_{11}^{36}\}$. The system instead fragments along the pathway



However, as can be seen, the copper cluster undergoes a substantial reorganization which flattens the copper cluster. Furthermore, instead of weakening the S-C bond by pulling, just the opposite happens: the energy of breaking the C-S bond in $\text{CH}_3\text{-CH}_2\text{-S-C}$ under otherwise equilibrium conditions requires 1.4 eV (see Tab. VIII) but under mechanical pulling [325] this energy increases to ~ 2.5 eV! This makes the C-S absolutely immune against rupture. A more general question may be asked. Imagine a nanoparticle with arbitrary chemical bonds. The weakest bond(s) will start to rupture if the nanoparticle is heated up to sufficiently high temperature. Take the same nanoparticle, but now at a very low temperature and mount it in the (static) AFM apparatus (for example using thiolate bonds) and pull the particle along the weakest bond(s). Will the same bond(s) be broken under the action of mechanochemistry? As the above example indicates, the answer may in general be NO! This line of thought also suggests how the thiolated copper

³²Simulation of mechanochemistry on a thiolated Cu_5 cluster ($\text{CH}_3\text{-CH}_2\text{-S-Cu}_5$). Static simulation at $T = 0$ K by a stepwise lift-up of the methyl group (CH_3). Similar model as above.

³³Simulation of mechanochemistry on a thiolated Cu_9 cluster ($\text{CH}_3\text{-CH}_2\text{-S-Cu}_9$). Static simulation at $T = 0$ K by a stepwise lift-up of the methyl group (CH_3). Similar model as above.

³⁴Simulation of mechanochemistry on a thiolated Cu(111) surface. Static simulation at $T = 0$ K by a stepwise lift-up of one out of four thiolate molecules ($\text{CH}_3\text{-CH}_2\text{-S}$). Similar model as above.

³⁵Simulation of a thermal break-up of a thiolated ($\text{CH}_3\text{-CH}_2\text{-S}$) copper surface at 1800 K. Note that the outcome is different from that shown in the corresponding mechanoprocess.

³⁶Simulation of a thermal break-up of a thiolated Cu_9 cluster ($\text{CH}_3\text{-CH}_2\text{-S-Cu}_9$) at 1800 K. Note that the outcome is different from that shown in the corresponding mechanoprocess.

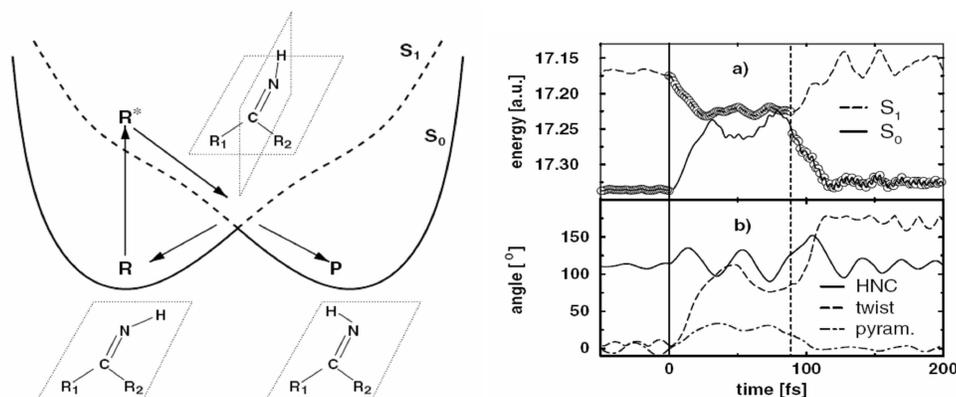


Fig. 62. Left panel: Schematic view of the photoreaction pathways of formaldehyde ($R_1 = R_2 = H$). The S_0 (solid line) and the S_1 (dashed line) energy curves are plotted qualitatively against a hypothetical reaction coordinate whose main contributor is the NH twist angle (the angle between the planes containing the R_1CR_2 and HNC, respectively) [329]. Right panel: (a) Time evolution of S_0 and S_1 energies following photoexcitation in the case of a successful $R \rightarrow P$ reaction. Solid and dashed vertical lines indicate the moment of photoexcitation and nonadiabatic transition to the ground state, respectively. The open circles indicate the PES on which the nuclei are propagated. (b) Corresponding time evolution of the pyramidalization, HNC, and HN twist angles. The HN bond is seen to flip from 0° to 180° resulting in the photoproduct.

nanoparticles could be thermally stabilized: all that is needed is a slight application of strain along the S-C bond! One could try to use the bond selectivity of thermal and mechanochemistry as a novel designing tool in nanotechnology. The bonds could be selectively manipulated either by heating or by application of mechanical stress.

In addition to thermal energy, and mechanical energy, the other important energy source to initiate a reaction is use of light whereby the system is promoted into an electronically excited state. Electronically nonadiabatic processes play a crucial role in many different fields, including femtoscience [328]. A very simple application is shown in Fig. 62 for the cis-trans photoisomerization of the smallest unprotonated Schiff base, formaldehyde $H_2C=NH$ [329]. The system may evolve on the ground-state PES S_0 and the photoexcitation is simulated by promoting the system into a (first) excited state PES S_1 . The construction of the excited state PES was described in chapter 3.1.5.12, 3.1.5.13. The ground- and excited-state PESs intersect (or at least touch) at the so-called *conical intersections*. At such points the system may with a certain probability [329] switch back from the excited- to the ground-state PES whereby a reisomerization may occur. This is illustrated for formaldehyde in Fig. 62.

Returning back to the organometallic nanojunctions we discuss now junctions formed by a chromophore molecule, the azobenzene, linked by sulphur atoms to two gold electrodes, in the way shown in Fig. 55. Shining laser light of appropriate wavelength will lead to photoswitching of the molecule. The ground- and first excited-state PESs are shown in Fig. 63 [330]. The simplest model, which we study here, consists of just the azobenzene molecule with two constraints applied to the end points of the azobenzene molecule and the electrodes completely integrated away. The simulation parameters were as follows: PBE exchange correlation functional [59],

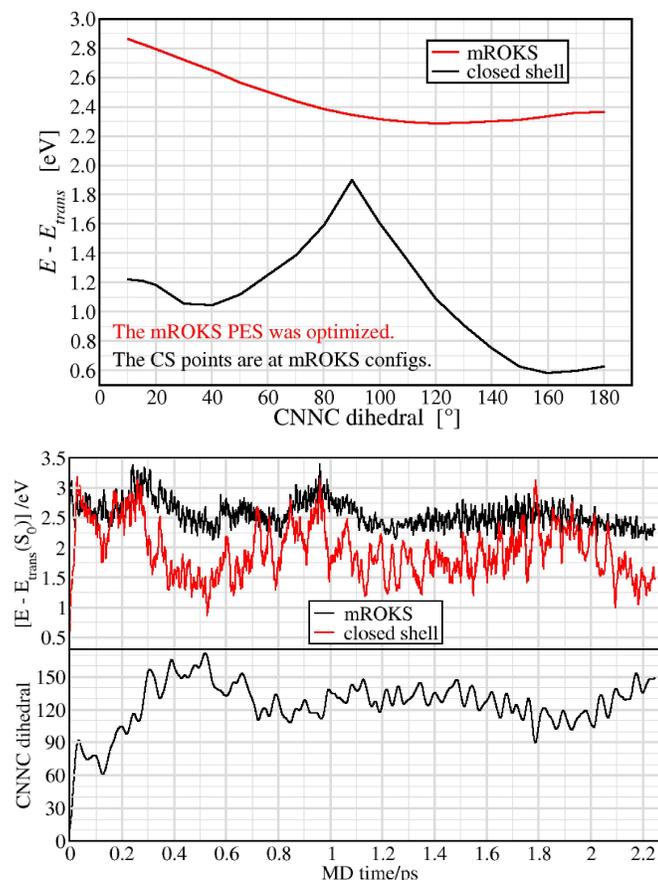


Fig. 63. Upper panel: Ground (black) and first excited (red) states of the azobenzene molecule: the vertical deexcitation spectrum [330]. The reaction coordinate is taken to be the CNNC dihedral angle which corresponds to the rotation reisoerization mechanism. The PESs have been slightly adjusted to fit the experimentally known data for the cis- and trans- isomers. Lower panel: The ground- and excited state PESs S_0 and S_1 for azobenzene molecule and the corresponding CNNC dihedral angle from a MD sampling at 300 K. As in the ground-state PES shown in the upper part, the ground-state PESs have been generated for atomic configuration corresponding to the excited state. The points where the two PESs overlap correspond to the conical intersections.

normconserving pseudopotentials [206], plane wave cutoff of 70 Ry. The dynamical simulations have been performed at room temperature. The first excited state S_1 was calculated using a modification of the ROKS, see chapter 3.1.5.12, the so-called mROKS technique [331].

From the results in the upper panel of Fig. 63, showing the results of a switching by rotation reisoerization (modification of the CNNC dihedral angle) it can be seen that the PESs are qualitatively similar to those schematically shown in Fig. 62. While no conical intersection can

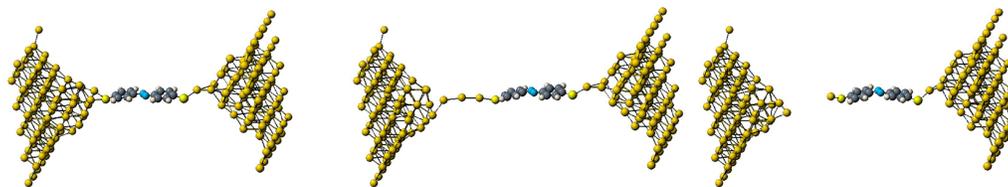


Fig. 64. Mechanochemistry on a *g*o*l*d-*t*r*a*n*s*-azobenzene nanojunction with realistic gold electrodes [325]. Note, that a similar nanowire structure as on a stepped gold surface [314] is formed as a result of mechanical pulling.

be seen in the upper part of Fig. 63, which shows only a section through the PES, the conical intersections are bound to exist somewhere on a more dimensional PES, as experimentally the photoswitching does occur [332]. A typical result from a MD run set up to sample the PESs in order to reveal the possibility of a photoswitching [330] is shown in the lower part of Fig. 63. The results show that an azobenzene molecule initially prepared in a *cis*- isomer geometry quickly changes its conformation after switching to the excited S_1 state, so that the mean value of the CNNC dihedral angle is around $\sim 120^\circ$. As can be seen from the upper part of Fig. 63, that value of the dihedral angle is close to the conical point. As seen from the lower part of Fig. 63, the conical points are indeed encountered. The important point is, that stress exerted on the azobenzene molecule, e.g. by AFM apparatus, see Fig. 55, will control the relative position of the S_0 , S_1 PESs, and hence the switching probability [330].

Can mechanochemistry be used to achieve reversion? Such a possibility has not been much studied up to date. Yet, in principle it could be possible to switch the conformation from say, the *cis*- to *trans*- using exclusively the mechanical energy of the SPM tip. We have been doing simulations of the mechanochemical switching [325]. However, in such a simulation we now need a complete geometry of both tips, see Fig. 64, where, as an example, we show mechanochemistry of a gold-*t*r*a*n*s*-dithioazobenzene-gold nanojunction up to the point of mechanical destruction [325] $\{v_8^{37}, v_9^{38}\}$. Given that both tips are explicitly included in the model, the calculations have been done with localized basis sets (double zeta plus polarization)³⁹. The usual DFT model with PBE exchange-correlation functional [59] and pseudopotentials was applied. The results of one such simulation shown in Fig. 64 give support to the more simplified method used in most of the other mechanochemistry simulations, where one of the two electrodes was integrated away from the simulation and substituted by a simple constraint. The fact that Au nanowires, similar to those shown in Fig. 56, are formed also with a complete set of realistic electrodes gives a strong support to the simplified model with constraints. These simulations

³⁷Simulation of reversion of a dithioazobenzene molecule from *cis* to *trans* isomer using only mechanical energy of the tips. The azobenzene molecule is anchored to two realistic gold model tips by a sulphur bond. Static simulation at $T = 0$ K by a stepwise motion of one of the two electrodes.

³⁸Simulation of mechanochemistry on *trans*-dithioazobenzene using only mechanical energy of the tips. The azobenzene molecule is anchored to two realistic gold model tips by a sulphur bond. Static simulation at $T = 0$ K by a stepwise motion of one of the two electrodes.

³⁹We note that a very careful optimization of the localized basis sets was required, in order to achieve agreement with the more accurate plane wave basis set. With less carefully selected basis sets the fragmentation of the nanojunction resulted in scission of the S-Au or S-C bond, depending on the particular basis set used. This once again demonstrates the advantage of use of plane wave basis sets, wherever feasible.

also indicate that the organic molecule is extremely stiff and the gold electrodes rather ductile. Can under these circumstances mechanichemistry be used to affect the geometry and properties of the molecule, rather than just mechanically destruct the gold electrode? Preliminary results indicate that at least a cis- to trans- mechanoreisomerization should be possible [325].

These few *ad hoc* selected examples, most of which were stimulated by experimental groups, show how modeling in a “virtual nanolaboratory” can be used to both guide the real laboratory experiments and assist interpretation of the obtained results.

4.7 Quantum Monte-Carlo

From the general discussion, chapter 3.1.6 and 3.2.2 above, it follows that there are more variants of quantum Monte-Carlo (QMC). The zero-temperature QMC [88] seeks a solution of the Schrödinger equation, while the finite-temperature version, the Path-Integral Monte-Carlo (PIMD), or molecular dynamics [154], samples the density matrix. Due to the problems to include the proper statistics (symmetry of the wave function) the PIMC is mainly used for ions and bosons or Boltzmannions; see chapter 3.2.2.4 above. The zero-temperature QMC can be used for bosons and, in fixed-node approximation, also to fermions. Below we give a few examples of application of both techniques.

4.7.1 Zero-temperature QMC

4.7.1.1 Homogeneous electron gas

The homogeneous electron gas (HEG) is the simplest yet still relevant manyfermion model [333]. Just for completeness we recall that it consists of electrons of various densities⁴⁰ moving on a homogeneous positive background of a density, corresponding to a neutral system. This means that the discrete nature of the positively charged ions is smeared out in the positively charged homogeneous background. Despite its simplicity, the HEG is of great importance from at least three points of view. It provides the first approximation to describe the valence electrons in metals [334]. It is the basic ingredient of the Density Functional Theory (DFT) [25, 26] meaning that any of the many DFT calculations are actually based on the results for HEG. Finally, beyond the range of r_s relevant for these applications, it displays a complex and fascinating phase diagram that includes the Wigner crystal [335] a variety of magnetic structures, and possibly more complex and exotic structures. Wigner crystal is fascinating as it is a demonstration of pure quantum phase transition: a HEG is unstable as it is diluted below a certain density and the electrons start “crystallizing” on a regular lattice in the space. Despite the nine decades of intensive research of the static and dynamic properties of HEG, the properties have only been rigorously established in the limit of high densities, where the system approaches a perfect gas [336] and at low density where the electrons crystallize [335].

The breakthrough came with Ceperley and Alder [36] who performed essentially absolutely accurate calculation for this system based on QMC approach, see chapter 3.1.6. Four phases

⁴⁰The densities are usually expressed in terms of r_s , where r_s is a measure of the interelectronic distance ($n = \frac{1}{(4/3)\pi r_s^3} \Rightarrow r_s = \sqrt[3]{\frac{3}{4\pi n}}$).

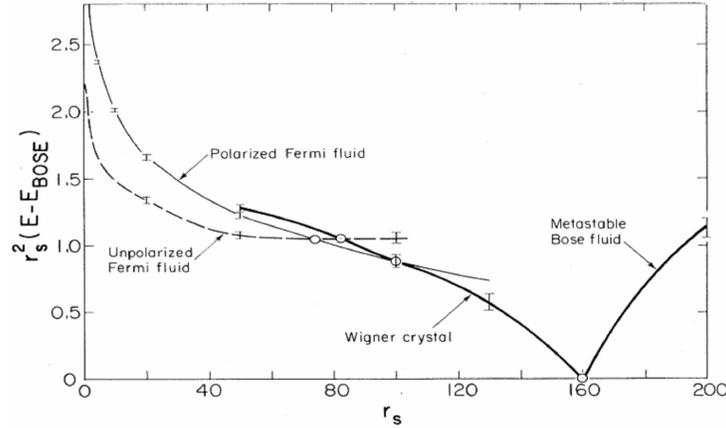


Fig. 65. The energies for the four phases studied relative to that of the lowest Boson state times r_s^2 at different densities r_s .

have been considered: the paramagnetic or unpolarized Fermi fluid, ferromagnetic or the polarized Fermi fluid, the Bose fluid and the Wigner crystal with a bcc lattice. First a standard VMC calculation was performed with up to two-body correlations and proper cusp condition included. For the fluid phase the single-particle orbitals are plane waves with the wave vector lying within the Fermi sea. For the polarized state the Fermi wave vector has been increased to allow for twice as many spatial orbitals. In the crystal phase, the orbitals are Gaussians centered around b.c.c. lattice sites with a width chosen variationally. Next a fixed-node DMC simulation using ~ 100 ensembles taken from the VMC simulation with the nodes fixed by the VMC trial wave function was performed. The third step is *nodal relaxation*. In this procedure if a random walk strays across the node of the trial wave function it is *not* terminated but its contribution to any average is reversed. At any stage of the random walk there is a population of positive walks (those that remained in the same nodal region or crossed an even number of nodes) and population of negative walks (those that crossed an odd number of nodes). It can be shown, that the difference population converges to the antisymmetric eigenfunction. However, both the positive and negative population grows geometrically with a rate equal to the difference between the Fermi and Bose energies. If the relaxation time from the fixed-node distribution times this energy difference is less than unity, the fermion energy *can* be reliably extracted. This condition is met for this system if the nodes of the Hartree-Fock wave function are used. However, this is a very lucky situation, and for most other systems the nodal relaxation cannot be performed. The last step is the *finite-size scaling*. The largest uncertainty is connected with this step. The size dependence was established on simulations for systems ranging from 38 to 248 particles. Extrapolation to infinite-particle results was carried out at each density on the basis of $E(N) = E_0 + E_1/N + E_2\Delta_N$, where the coefficients E_0 , E_1 , and E_2 were empirically determined from the simulations.

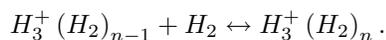
The results of the simulation are shown in Fig. 65. As can be seen the system undergoes Wigner crystallization at $r_s = 160 \pm 10$. The Fermion system has two phase transitions, crystallization at $r_s = 100 \pm 20$ and depolarization at $r_s = 75 \pm 5$. Very small differences in energy

were found between a Boson crystal and a Fermion crystal which made it very difficult to address this question.

These benchmark calculations are of paramount importance for the DFT theory, which uses HEG results as the primary input, see also chapters 3.1.5.3, 3.1.5.4, 3.1.5.7, and 3.1.5.8. For that reason the early QMC results for this system [36] have later been refined in a density range relevant for the DFT calculations, i.e. $0.8 \leq r_s \leq 10$ [37]. The simulations were based on system sizes between 169 and 458 particles which renders the finite-size scaling more accurate. Most importantly, these newer set of simulations covered also the case of partial polarization with the following spin polarizations $\zeta = 0, 0.21, 0.42, 0.56,$ and 1 , see chapter 3.1.5.4 and formula (111) in particular. This makes it possible to validate the different interpolation schemes between the polarized and unpolarized case. The Vosko *et al.* formula [39] was found to be especially accurate.

4.7.1.2 Odd-numbered, positively charged H-clusters

We now proceed from model systems to real physical systems, namely to hydrogen clusters. Several other applications of the method to molecular, cluster, as well as extended systems were briefly introduced in chapters 3.1.5.5, 3.1.5.6, and 3.1.6.4. The clusters of hydrogen molecules represent seemingly the simplest examples of molecular clustering effects. The neutral $(\text{H}_2)_m$ clusters are van der Waals (vdW) bonded systems with intermolecular bonding energies of only a few Kelvin [337]. Although they are difficult to handle experimentally, the simple vdW bonding allows a relatively straightforward theoretical modeling. The previous studies showed that the $(\text{p-H}_2)_m$ clusters with $m \leq 20$ exhibit extreme quantum-liquid-like behavior, absence of any shell structure [338], and superfluid behavior at very low temperatures [339]. On the other hand, the experiments on positive cluster ions have revealed that the odd numbered H_m^+ , $m = 3, 5, 7, \dots$ clusters are much more abundant in the mass spectra [340] than the even numbered H_m^+ , $m = 2, 4, 6, \dots$ ones [341]. In addition, experiments also suggest that the structure of the odd numbered cluster ions consists of a very stable H_3^+ core ion surrounded by more weakly bound H_2 molecules arranged in solvation shells around the core; hence the structure has the form $\text{H}_3^+(\text{H}_2)_n$. The support for this model is supported also by the measured enthalpies (ΔH) of equilibria for the reaction [342–344]



The experimental results [342] for $-\Delta H_{n-1,n}$ yield 6.9 kcal/mol for $n = 1$, ~ 3.3 for $n = 2$, ~ 3.2 for $n = 3$, ~ 1.72 for $n = 4$, etc. The stepwise decrease in the enthalpies with steps at $n = 2, 4, 7, 9, \dots$ which goes in a geometrical series with a quotient $1/2$ has been interpreted in terms of opening of successive, less tightly bound, solvation shells at those n .

The positive odd numbered clusters have been a subject of several theoretical studies based on quantum chemical methods [345–347], quantum Monte Carlo (QMC) [348], and DFT [349]. The first two approaches were mainly concerned with the dissociation energetics, while the DFT techniques have been used in connection with path integral treatment of the protons, see chapter 3.2.2 and 4.7.2.1, in order to explore the quantum and thermal fluctuations. All three groups of methods have confirmed the $\text{H}_3^+(\text{H}_2)_n$ structure emerging from the experimental studies. Given the deficiencies of the DFT techniques in description of the energetics and even structure of

these systems [61, 349], it is not surprising that the delicate structure and tiny energy differences of $-\Delta H_{n-1,n}$ were not very well described by the DFT techniques, see the discussion to Fig. 13 above. It is more surprising that even high-quality quantum chemistry calculations, such as MP2 [347], CCSDT(T) [346] or even QMC [348] failed to describe correctly the dissociation enthalpy $-\Delta H_{n-1,n}$. As shown below, one reason for these large discrepancies (CISD, CCSD(T), and QMC) compared to the experimental values is that total energy differences are compared with experimental dissociation enthalpies and hence the dynamical effects are ignored in the theoretical description. The calculations which include the thermal contribution to enthalpy [347] indeed exhibit a significant improvement up to H_{13}^+ over the theoretical treatment without that contribution [345, 346, 348].

The objective is to resolve the above mentioned shortcomings and produce benchmark calculations for this class of system up to $m=17$ [17]. In particular, two new cluster structures, H_{15}^+ and H_{17}^+ are proposed. The geometry of each cluster was optimized. For smaller clusters this was done by using the CISD method. However, starting from $m \geq 11$, the application of the CISD method to optimize the ionic degrees of freedom becomes very demanding, mainly due to the flatness of the PES, which results in a very slow convergence. For that reason for $m \geq 11$ the geometry optimization was done using the MP2 method. For each optimized geometry the calculation of total energies the CCSD(T) was performed. The total energy was evaluated also within the DMC method within the fixed-node approximation, hence it does not exhibit the basis set errors characteristic of the quantum chemistry approaches. Since the CCSD(T) method includes the correlation effects to a very high level, the differences observed between the DMC and CCSD(T) results are mainly due to the basis set errors and, to a much smaller extent, due to the fixed-node bias in DMC. This is very important especially for present systems, in which very subtle polarization effects are very sensitive to the size of the basis set. The correlation consistent cc-pVTZ basis sets [350] were used. We note that the correlation consistent basis sets are expected to be superior to the standard 6-311G** Gaussian basis sets customarily used. The basis set superposition error was found to be less than 0.1 kcal/mol, and was neglected. The dissociation enthalpies were calculated from the differences between ground-state energies obtained by correlated methods, corrected for zero-point vibrational energy (ZPVE) Δ_0 and thermal contribution $\Delta_{298.15}$, both calculated at $T = 298.15$ K in harmonic approximation at the MP2 level of theory.

Optimized geometries for H_m^+ for $m \leq 17$ are shown in Fig. 66. For $m = 11$, six equivalent minima for the additional H_2 molecule have been found; three minima both above, and below the H_3^+ moiety. The structure of this cluster was used in the search for geometries of the larger clusters. For $m = 13, 15$, and 17 all distinct geometries corresponding to the different placements of the other H_2 molecules have been investigated. The structures shown in Fig. 66 were found to be marginally lower in energy than the other isomers. The shell structure of these clusters is clearly visible: the core ion is formed by H_3^+ , the first, mainly chemically bonded shell, is completed at $m = 9$, and the second, predominantly physically bonded shell, at $m = 15$. We note that at $m = 17$ a new, third shell is opened by the last H_2 dimer which is slightly more distant from the H_3^+ core than the other three ones and does *not* occupy any of the minima identified in the case of H_{11}^+ . This dimer is rather located in a new minimum below the H_3^+ plane at the center of the H_3^+ core (c.f. Fig. 66).

Standard dissociation enthalpies at $T = 298.15$ K, $-\Delta H_{n-1,n}$, evaluated by different methods are compared with experimental results in Fig. 67. The results demonstrate that all methods,

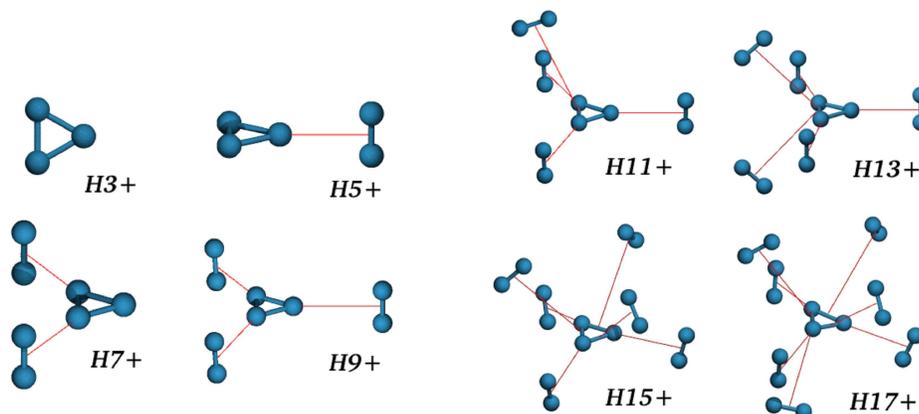


Fig. 66. Optimized geometries of H_n^+ for $n = 3, 5, \dots, 17$ [61]. Note that the geometries of the two largest cluster structures have not been described before.

except for CISD, qualitatively reproduce the step-like dependence of $-\Delta H_{n-1,n}$. The failure of the CISD is attributed to the lack of size-consistency, see chapter 3.1.4. CCSD(T) and MP2 yield results of comparable quality, reproducing qualitative and quantitative features of the experimental $-\Delta H_{n-1,n}$. Best results overall are obtained from QMC energies. The DMC method does not suffer from the basis set errors while the fixed-node approximation is quite small in weakly bonded systems. It is very encouraging that by combining high accuracy description of electronic correlation with a reasonable estimation of ZPVE and thermal effects we are able to reproduce the measured values to within less than 0.5 kcal/mol! The residual errors are almost certainly due to the description of the thermal excitations in the harmonic (MP2) approximation. Remarkably, QMC reproduces even the tiny step at $m = 17$ which appears due to the slight difference in the bonding distance of the last H_2 molecule opening the third coordination shell and has not been identified by any of the previous calculations.

The agreement of the results with experiment for $m = 15$ and 17 indicates that the two isomers are excellent candidates for the ground-state structures. We note that the error for $m = 11$ and 13 is marginally larger than that for the other clusters. This can be qualitatively understood as being due to the more open structure for these two cases, compared to $m = 15$ and 17 in which all three minima in the second solvation shell are occupied (c.f. Fig. 66). This renders them more anharmonic and hence the use of harmonic approximation less accurate. The accuracy of this approximation can be judged from Fig. 67 which shows the comparison between the DMC dissociation energies/enthalpies at $T = 0$ and $T = 298.15$ K. The contribution to the $-\Delta H_{n-1,n}$ of the thermal excitations is a strongly decreasing function of n , and so is the absolute error from this approximation for the larger clusters. Fig. 67 also reveals the importance of the thermal excitations for the smaller clusters (being about 50 % of the step size for $n = 1$) when comparing the theoretical results with experiments and clearly explains the discrepancies between theory and experiments in previous studies [345, 346]. In particular, the DMC results are essentially perfect up to $n = 3$, i.e., in the region dominated by chemical bonding for which the treatment

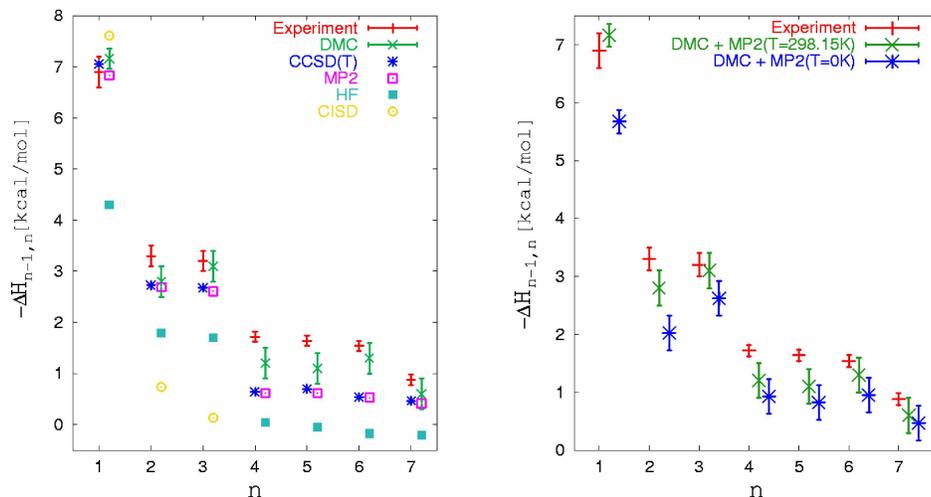


Fig. 67. Left panel: Comparison of the dissociation enthalpies $-\Delta H_{n-1,n}$ in various Quantum Chemistry methods and QMC at room temperature with experimental results [344]. Right panel: Validation of the importance of finite-temperature effects in the QMC approach.

of ZPVE in harmonic approximation is an established assumption. These results show stable clusters for all m . In contrast, the previous calculations [345, 346] gave negative $-\Delta E_{n-1,n}$ for $m = 11$, and 13. In particular, the results compare very favorably with those of Ref. [347] which exhibit a rather poor agreement with experiments for $m \geq 11$. This is most likely the combined result of the differences in the ground-state geometries of the clusters at higher m and basis set errors. Interestingly, for $n = 1$, both DMC and CCSD(T) results are in perfect agreement with the experiments, while perturbative techniques, MP2 [247], are lower by almost 1 kcal/mol.

These clusters exhibit a multitude of different bonds. Clearly, the polarization effect of the H_3^+ core ion will run out of power at some cluster size and for very large clusters very far from the core ion the outer shells will behave quite like the even-numbered clusters. A trace of this behavior can be seen also in the small odd-numbered clusters studied here. It was found that starting from $m = 11$ one can expect that both zero point motion fluctuations and also temperature effects for $T \sim 10$ K and higher will enable the outermost dimer to pass the barriers from one minimum to another with a resulting *fluctional behavior*. Obviously, the H_2 molecules in the successive coordination shells will be even more mobile causing their fluctional behavior.

4.7.1.3 Toward QMC-based *ab-initio* molecular dynamics

So far all correlated techniques for electronic structure calculation were limited to static ions. The reason is the prohibitive numerical cost to obtain the correlated many-body wave function Ψ . Yet, in many occasions, the DFT-based *ab-initio* MD has insufficient accuracy dictated by the well-known deficiencies of the DFT, see chapter 3.1.5.6, 3.1.5.7, 3.1.5.11. Can the accuracy of the *ab-initio* MD be improved? This question is precisely the same as the question asked by Car and Parrinello some 20 years ago [186], when the only MD benchmarks were based

on accuracies achievable with empirical interatomic potentials. The answer was invention of the Car-Parrinello coupled electron-ion dynamics, see chapter 3.3.3. As a matter of fact, in principle the CP dynamics is just a very clever way to adjust the time-scales of the two vastly different subsystems in the simulation: the electrons and the ions. The time-scale is set so, that the electrons naturally “follow” the ions without the need to ever have to explicitly reoptimize them. Can a similar trick be applied also in the domain of correlated calculations, such as QMC?

It turns out that the answer is positive! The indications are provided by the *correlated sampling* [144, 145] which was used to outline the calculation of QMC-based forces on ions, see chapter 3.1.7.4. In the CP dynamics the time-steps are typically very short, of the order of ~ 0.1 fs, which typically induces structural changes of the order of 10^{-3} to 10^{-4} Å. These time-scales make it possible to use the correlated sampling and render the optimization of the many-body wave function Ψ numerically efficient [144, 145].

Recently, a step was made toward a QMC-based MD [351]. Simple hydrogenated Si-clusters (SiH_4 , Si_5H_{12} , and $\text{Si}_{14}\text{H}_{20}$) at the temperature of $T = 1000$ K were simulated. While the forces were still DFT-based and the atomic structures generated by a standard plane wave pseudopotential BO MD simulation ($E_{\text{cut}} = 12$ Ryd, PBE functional [59], large box with periodic boundary conditions, $\Delta t = 0.075$ fs), a QMC electronic structure was generated in parallel for the atomic configurations encountered in the simulation. For the calculations the same formulas as for QMC calculation of forces have been used, see chapter 3.1.7.4. With the above parameters only three imaginary time-steps were required to converge the QMC calculation. Surprisingly, the QMC part of the calculation represented only a 50–100 % overhead over the LDA BO MD! Some results are shown in Fig. 68.

As can be seen from Fig. 68, the agreement between the continuously and discretely sampled data is excellent for all cluster sizes. The importance of QMC accuracy is evident from the variation of the LDA error along the MD trajectory: the smallest error being ~ 0.8 eV and the largest ~ 1.2 eV, i.e., the error is by no means constant.

Continuous QMC calculations are in its infancy. However, it is expected, that it will soon be possible to use the correlated sampling to compute also the forces exerted on ions and treat both electronic and ionic degrees-of-freedom on the same footing in a fully QMC-based simulation. The work along that line is under way.

4.7.2 Finite-temperature QMC

The description of quantum behavior of ions by *path integral treatment* is a well-established method with several applications both in the field of *empirical interatomic potentials* as well as in the *ab-initio* domain. The two directions represent rather disjunct categories: typically the largest quantum effects on the ions are found in systems with very weak bonding; typical examples being 1-He [154] or the systems composed of H_2 molecules [338, 339]. Those systems are predominantly van der Waals bonded and can reliably be described by empirical interatomic potentials. As discussed in chapter 3.1.5.6, 3.1.5.7, this is precisely the type of bond very hard to model by the DFT description. The more accurate electronic structure methods, such as the QMC, chapter 3.1.6, could be applied, but the cost is still prohibitive. In the opposite limit there are systems, bonded in much more complex manner, involving typically chemical, as opposed to physical bonding. Such systems require an explicit electronic structure description of the evolving chemical bonds. Several applications which run on DFT PESs exist [352–254]. There are

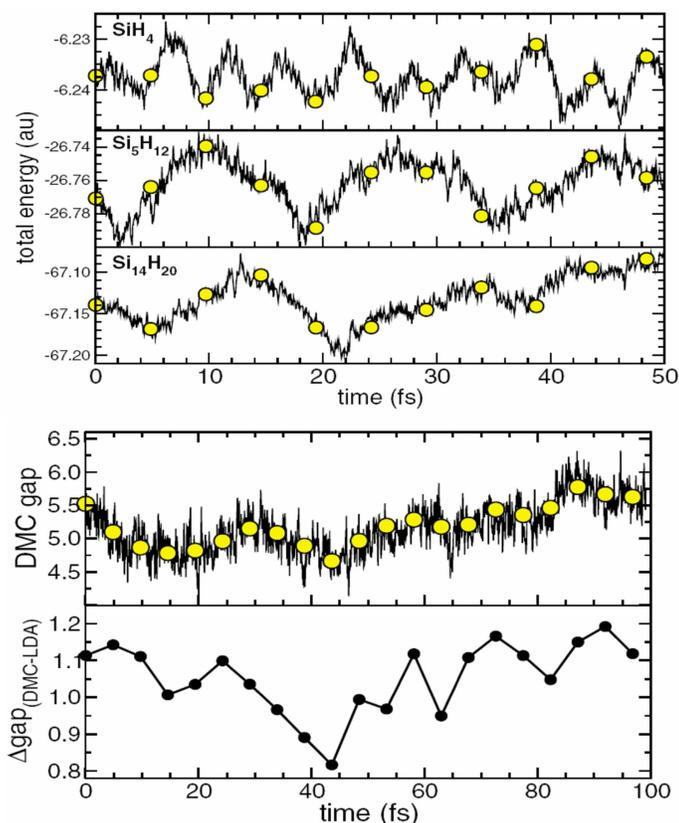


Fig. 68. Upper panel: Time evolution of DMC total ground-state energies [351]. The circles (lines) correspond to discretely sampled (continuous) DMC calculations. The error bars of the discretely sampled data are smaller than the symbol size. Lower panel: Time evolution of the DMC HOMO-LUMO gap and difference between the DMC and LDA gaps [351]. For the DMC gaps the circles (lines) correspond to discretely sampled (continuous) DMC calculations.

also systems characterized by a multitude of bonds, ranging from very strong covalent to very weak van der Waals. One such example is the odd-numbered H-clusters.

4.7.2.1 Odd-numbered, positively charged H-clusters

The system was already introduced in chapter 4.7.1.2, where the many-body electronic correlations in these clusters were addressed by means of zero-temperature QMC. Here we apply the path integral simulation technique and study the quantum delocalization in these clusters at a low temperature of $T = 5$ K [349].

Quantum chemical [345–347] and QMC calculations [61, 348] have confirmed that the positively charged protonated hydrogen clusters lose the quantum-liquid-like behavior of the pure $(\text{H}_2)_n$ clusters [338] and develop instead a pronounced shell structure of the type $\text{H}_3^+(\text{H}_2)_n$ with

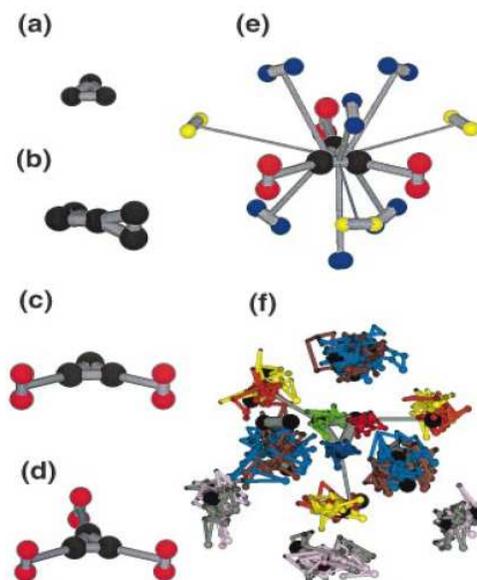


Fig. 69. The protonated hydrogen clusters studied with the respective point group symmetry. (a) H_3^+ (D_{3h}); (b) H_5^+ (C_{2v}); (c) H_7^+ (C_{2v}); (d) H_9^+ (D_{3h}); (e) H_{27}^+ (D_{3h}); (f) snapshot of a full imaginary time path for H_{27}^+ at 5 K. To guide the eye atoms marking the origin of the path are shown by large balls with bonds shown by sticks. Beads in the H_3^+ core ion and in the successive shells are shown in different colors.

“magic numbers”. However, it was also found that the PESs on which the protons move are very flat and anharmonic. Under these circumstances quantum effects are expected to be significant, and it is far from clear whether at low temperatures the structure is dominated by the minimum on the PES, or if the behavior is more quantum-liquid-like as for the unprotonated $(\text{H}_2)_n$ clusters. This point is addressed below.

The clusters studied are shown in Fig. 69. The simulations used a combination [157, 158] of path integral simulation of the nuclei [154] with DFT description of their interactions in the Born-Oppenheimer approximation. The exchange-correlation effects were described at the GGA level [64]. In order to make the protons more amenable to description by plane wave basis sets, the hydrogen cores were described by a norm-conserving pseudopotential. The orbitals were expanded in plane waves with an energy cutoff of 30 Ry in a large cubic cell ($a_{lat} = 30$ a.u.); the positive charge was compensated by a uniform background. As indicated in chapter 4.7.1.2, the accuracy of the DFT description is rather limited. This can be seen for instance from the H_5^+ geometry in Fig. 69. However, as shown below, this accuracy is sufficient to study the quantum delocalization of the protons and the related properties. The canonical path integral was mapped onto $P = 32$ cyclically connected replicas. Nuclear exchange was neglected—the particles were assumed to be Boltzmannions (see chapter 3.2.2.4). The justification for this approximation comes from simulations on the much more quantum-liquid-like $(\text{H}_2)_n$ systems, which show that the effects of the Bose-Einstein statistics is vanishing at 5 K [338]. The sampling efficiency was increased by using the staging action together with Nosé-Hoover thermostat chains of length 4

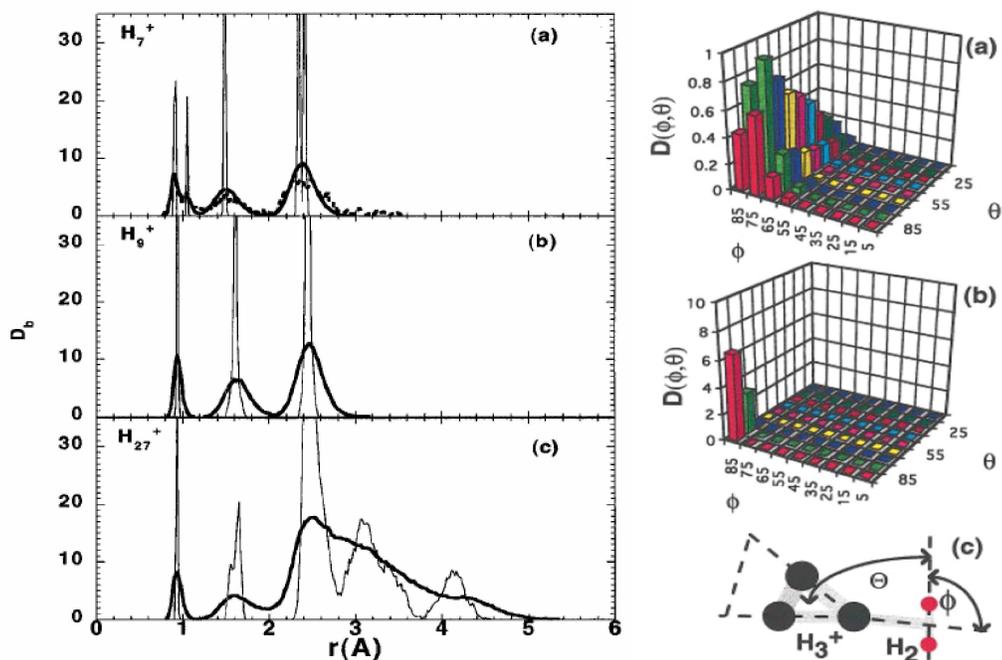


Fig. 70. Left panel: Normalized partial distribution functions of bond distances involving at least one proton from the H_3^+ core. Bold lines: quantum simulations at 5 K; solid lines: classical simulation at 5 K; dashed line in (a): classical simulation at 500 K. Right panel: Angular distribution $D(\Theta, \Phi)$ of the first-shell H_2 molecules relative to the H_3^+ core ion for H_9^+ at 5 K. (a) quantum simulation, (b) classical simulation, (c) definition of angles Θ and Φ .

coupled separately to each nuclear degree-of-freedom, see chapter 3.2.2.5. A time step of 0.6 fs was used. Statistical averages were obtained from $\sim 5 \times 10^3$ configurations after equilibration. In addition, simulations with classical protons of the same length were performed. Comparison of the quantum and a corresponding classical simulation provides a unique opportunity to separate the quantum and thermal fluctuations.

The finite-temperature distribution functions of the distances that involve at least one proton of the H_3^+ core are displayed in Fig. 70. The classical simulations at 5 K reveal a pronounced radial structures corresponding to successive sharp salvation shells that are clearly separated. The general picture that emerges from these classical simulations is that all the investigated clusters are extremely rigid objects. This character changes if quantum fluctuations are properly taken into account at 5 K. All peaks broaden substantially, in particular those of the salvation shells and are not clearly separated as in the classical approximation. The broadening is particularly severe for the largest cluster, H_{27}^+ , that displays only a broad band after the second maximum. The H_7^+ cluster was simulated also classically at 500 K. However, the cluster was unstable and dissociated after ~ 800 steps into H_3^+ and H_2 fragments, which underlines that quantum effects cannot be mimicked simply by thermal fluctuations.

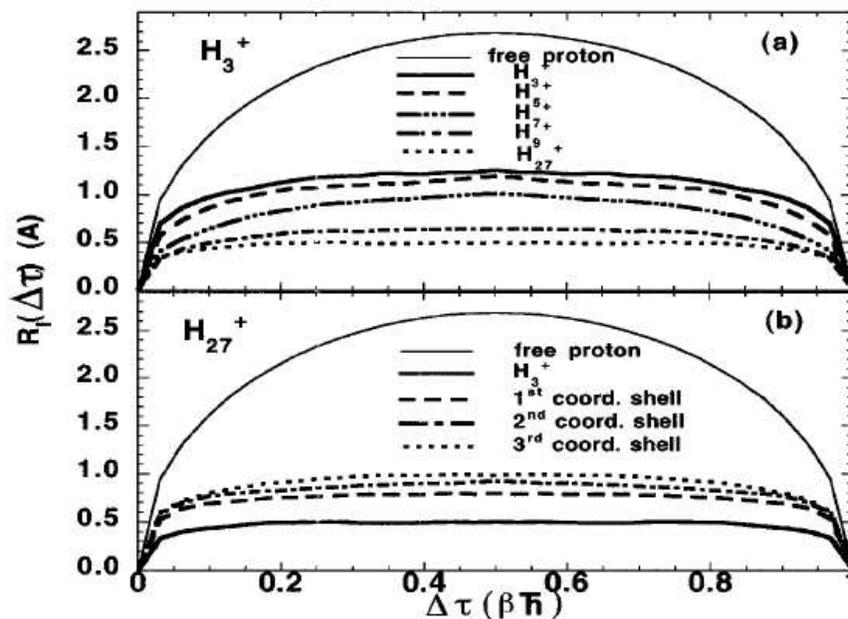


Fig. 71. Root-mean-square position displacement functions $\tilde{R}_I(\Delta\tau)$ at 5 K. (a) H_3^+ core protons in H_n^+ . (b) Protons in the H_3^+ core and in the successive salvation shells in H_{27}^+ . For reference the free proton function is shown in light lines.

The effect brought about by introduction of a proton into a neutral cluster can be quantified by comparing properties of the $(H_2)_{13}$ cluster, which was subject of path integral simulation with a model potential [338] with the largest protonated cluster studied, H_{27}^+ . The floppiness of the clusters may be characterized by the Lindemann's rms bond length fluctuations δ and the cluster "size" measured by the cluster radius of gyration R_g [338], see formulas (261), (262) of chapter 3.2.2.3. For $(H_2)_{13}$ $\delta = (32.0 \pm 2) \%$ and $R_g = 4.5 \pm 02 \text{ \AA}$ based on molecular centroid distances at 5 K. The numbers are in remarkable contrast to $\delta \sim (2 \times 10^{-7}) \%$ and $R_g = 2.7 \text{ \AA}$ for the protonated counterpart at 5 K. The dramatic reduction of δ by many orders of magnitude signals a proton freezing out of the internal translational molecular degrees of freedom. Furthermore, the reduction of the cluster size R_g by a factor of 2 reflects a strong electrorestrictive effect.

What is then the reason for the substantial broadening observed in Fig. 70? The answer can be found in the angular distribution shown also in Fig. 70. As can be seen, the H_2 ligands undergo very large-amplitude rotations. The rotations are much reduced in the classical case. Hence, the rotations are a quantum-induced phenomenon.

The degree to which the different protons behave like quantum particles can be put on a quantitative footing in terms of the rms position displacement correlation function $\tilde{R}_I(\Delta\tau)$, see formula (263) in chapter 3.2.2.3. These functions are depicted in Fig. 71(a) for the three core protons that constitute the H_3^+ . First of all, one observes a clear trend of decreasing size of the

three protons as the cluster grows: the $\tilde{R}_I (\beta\hbar/2)$ extension of the protons shrinks by a factor of more than 2 from the bare H_3^+ ion to the core of H_{27}^+ . It turns out that these protons are tightly localized as proven by the flat correlation function for $\Delta\tau$ larger than $\sim 0.2\beta\hbar$. One can conclude that the closure of the first salvation shell at H_9^+ induces a pronounced ground-state dominance and localization on the H_3^+ core ion. The localization properties in successive shells are analyzed in Fig. 71(b) or the largest cluster. Here it becomes evident that the H_3^+ core protons of H_{27}^+ are clearly ground-state dominated, whereas those of the solvating H_2 ligands are much more delocalized in space, which goes hand-in-hand with the rotational delocalization of the ligands. This behavior is shown pictorially in Fig. 69(f).

5 Outlooks

As should be clear by now, the atomic/molecular-scale simulations develop in several, often contradictory, directions. On one hand, the techniques are applied to ever larger systems and complex processes where in addition to the growing computer power also less accurate approximations are normally required. Typical representative of that type of application is life-sciences [355, 356], biophysics and biochemistry [357-359]. These, almost exclusively DFT applications, have already now a large impact on the development of this field. In this respect, often the methods of embedding [360], and $O(N)$ methods (see chapter 3.3.5) play the key role. Hand-in-hand with this development there is also an increased transfer of these computational technologies from academic institutions into industrial research in pharmacology, chemistry, materials, automotive and other industries.

It should also be clear, that we are still far from the ideal chemical accuracy of ~ 1 kcal/mol. For system sizes of interest here, there is only one method, the QMC, which gives a real promise of achieving that sort of accuracy. However, so far that level of accuracy was only achieved for a few systems. Moreover, these methods still require a lot of human input. For that reason these ultra-accurate technologies still remain almost strictly in the academic domain.

Making a full advantage of all these developments, brings us closer to the ideal of having a cheap, fast, accurate, and efficient “virtual computational nano-scale laboratory” at disposal.

References

- [1] M.P. Allen, D. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987)
- [2] D. Frenkel, B. Smith, *Understanding Molecular Simulation, From Algorithms to Applications* (Academic Press, San Diego, 1996)
- [3] D. Marx, J. Hutter, in *Modern Methods and Algorithms of Quantum Chemistry*, Ed. J. Grotendorst, (NIC Series Vol. 3, 2000) p. 329
- [4] J.C. Tully, in *Modern Methods for Multidimensional Dynamics Computations in Chemistry*, Ed. D.L. Thompson (World Scientific, Singapore, 1998)
- [5] J.C. Tully, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, Eds. B.J. Berne, G. Chiccott, D.F. Coker (World Scientific, Singapore, 1998), Chapter 21, p. 489
- [6] P.A.M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, Oxford, 1947)
- [7] E. Deumens, A. Diz, R. Longo, Y. Öhrn, *Rev. Mod. Phys.* **66** (1994) 917
- [8] A. Messiah, *Quantum Mechanics* (North Holland, Amsterdam, 1964)
- [9] P. Ehrenfest, *Z. Phys.* **45**, 455 (1927)

- [10] J.C. Tully, in *Modern Theoretical Chemistry*, Ed. W.H. Miller, (Plenum Press, New York, 1976), part B, p. 217
- [11] D.R. Hartree, *Proc. Cambridge Philos. Soc.* **24** (1928) 89
- [12] J.C. Slater, *Phys. Rev.* **35** (1930) 210
- [13] V. Fock, *Z. Physik* **61** (1930) 126
- [14] A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989)
- [15] G.G. Hall, *Proc. Roy. Soc. A (London)* **205** (1951) 541
- [16] C.C.J. Roothaan, *Rev. Mod. Phys.* **23** (1951) 69
- [17] T. A. Koopmans, *Physica* **1** (1933) 104
- [18] See, for instance, C. Petrillo, F. Sacchetti, *Phys. Rev. B* **51** (1995) 4755
- [19] T. Kato, *Commun. Pure Appl. Math.* **17** (1975) 87
- [20] R.T. Pack, W. Byers Brown, *J. Chem. Phys.* **45** (1966) 556
- [21] J. Čížek, *Adv. Chem. Phys.* **14** (1969) 35
- [22] L.H. Thomas, *Proc. Cambridge Philos. Soc.* **23** (1927) 452
- [23] E. Fermi, *Z. Physik* **48** (1928) 73
- [24] E.H. Lieb, B. Simon, *Phys. Rev. Lett.* **31** (1973) 681
- [25] R.O. Jones, O. Gunnarsson, *Rev. Mod. Phys.* **61** (1989) 689
- [26] M.C. Payne *et al.*, *Rev. Mod. Phys.* **64** (1992) 1045
- [27] P. Hohenberg, W. Kohn, *Phys. Rev.* **136** (1964) B864
- [28] Levy, *Proc. Nat. Acad. Sci. USA* **76** (1979) 6062
- [29] W. Kohn, L.J. Sham, *Phys. Rev.* **140** (1965) A1133
- [30] K. Capelle, G. Vignale, *Phys. Rev. Lett.* **86** (1983) 243
- [31] H. Eschrich, W.E. Pickett, *Solid State Commun.* **118** (2001) 123
- [32] Kohn, A. Savin, C.A. Ullrich, *Int. J. Quantum Chem.* **100** (2004) 20
- [33] P.A.M. Dirac, *Proc. Cambridge Philos. Soc.* **26** (1930) 376
- [34] J.C. Slater, *Phys. Rev.* **81** (1951) 385; *ibid.* **82** (1951) 538
- [35] R. Gáspár, *Acta Phys. Hung.* **3** (1954) 263
- [36] D.M. Ceperley, B.J. Alder, *Phys. Rev. Lett.* **45** (1980) 566
- [37] G. Ortiz, P. Ballone, *Phys. Rev. B* **50** (1994) 1391
- [38] J.P. Perdew, A. Zunger, *Phys. Rev. B* **29** (1981) 5048
- [39] S. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **58** (1980) 1200
- [40] U. von Barth, L. Hedin, *J. Phys. C: Solid State Phys.* **5** (1972) 1629
- [41] J. Harris, R.O. Jones, *J. Phys. F* **4** (1974) 1170
- [42] O. Gunnarson, B. I. Lundqvist, *Phys. Rev. B* **13** (1976) 4274
- [43] D.C. Langreth, J. P. Perdew, *Solid State Comm.* **17** (1975) 1425
- [44] R.Q. Hood *et al.*, *Phys. Rev. Lett.* **78** (1997) 3350
- [45] R.O. Jones, in *Computational Nanoscience: Do it yourself*, Ed. J. Grotendorst, S. Blügel, and D. Marx, NIC Series, Vol. 31, 2006, p. 45
- [46] L. Mitas, in *Electronic Properties of Solids Using Cluster Methods*, Eds. T.A. Kaplan, S.D. Mahanti, Plenum Press, New York, 1995), p.131
- [47] L. Mitas, J.C. Grossman, in *Recent Advances in Quantum Monte Carlo Methods*, Ed. W.A. Lester Jr. (World Scientific, 1997), p. 133

- [48] J.C. Grossman, L. Mitas, *Phys. Rev. Lett.* **74** (1995) 1323
- [49] J.C. Grossman, L. Mitas, K. Raghavachari, *Phys. Rev. Lett.* **75** (1995) 3870
- [50] F. Herman, J.P. van Dyke, I.B. Ortenburger, *Phys. Rev. Lett.* **22** (1969) 807
- [51] J.P. Perdew, *Phys. Rev. Lett.* **55** (1985) 1665
- [52] E. Engel, S.H. Vosko, *Phys. Rev. B* **50** (1994) 10498
- [53] C. Filippi, C.J. Umrigar, M. Taut, *J. Chem. Phys.* **100** (1994) 1290
- [54] A.D. Becke, *Phys. Rev. A* **33** (1988) 3098
- [55] C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B* **37** (1988) 785
- [56] B.G. Johnson, P.M.W. Gill, J.A. Pople, *J. Chem. Phys.* **97** (1992) 7846; *ibid.* **98** (1993) 5612
- [57] A.D. Boese, N.C. Handy, *J. Chem. Phys.* **114** (2001) 5497
- [58] J.P. Perdew, in *Electronic Structure of Solids '91*, Eds. P. Ziesche, H. Eschrig (Akademie verlag Berlin, 1991), p.11
- [59] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77** (1996) 3865
- [60] M. Ernzerhof, G.E. Scuseria, *J. Chem. Phys.* **110** (1999) 5029
- [61] P. Bokes, I. Štich, L. Mitas, *Int. J. Quant. Chem.* **83** (2001) 86
- [62] See, for instance, M. Benoit, D. Marx, M. Parrinello, *Nature* **392** (1998) 258
- [63] Y. Andersson, D.C. Langreth, B.I. Lundqvist, *Phys. Rev. Lett.* **76** (1996) 102
- [64] See, for instance, J.P. Perdew et al., *Phys. Rev. B* **46** (1992) 6671; Yu-Min Juan, E. Kaxiras, *Phys. Rev. B* **46** (1992) 6671
- [65] J.P. Perdew et al., *Phys. Rev. Lett.* **82** (1999) 2544
- [66] J. Tao et al., *Phys. Rev. Lett.* **91** (2003) 146401
- [67] J. Perdew et al., *J. Chem. Phys.* **120** (2004) 6898
- [68] A. Görling, M. Levy, *J. Chem. Phys.* **106** (1997) 2675
- [69] See, for instance, G. Onida, L. Reining, A. Rubio, *Rev. Mod. Phys.* **74** (2002) 601
- [70] V.I. Anisimov, J. Zaanen, O.K. Andersen, *Phys. Rev. B* **44** (1991) 943
- [71] N.D. Mermin, *Phys. Rev.* **137** (1965) A1441
- [72] A. Alavi et al., *Phys. Rev. Lett.* **73** (1994) 2599
- [73] J.P. Perdew et al., *Phys. Rev. Lett.* **49** (1982) 1691
- [74] L.J. Sham and M. Schlüter, *Phys. Rev. Lett.* **51** (1983) 1888
- [75] M Städele et al., *Phys. Rev. Lett.* **79** (1997) 2089
- [76] C.C.J. Roothaan, *Rev. Mod. Phys.* **32** (1960) 179
- [77] I. Frank et al., *J. Chem. Phys.* **108** (1998) 4060
- [78] T. Ziegler, A. Rauk, E.J. Baerends, *Theor. Chim. Acta* **43** (1977) 261
- [79] M. Head-Gordon et al., *Chem. Phys. Lett.* **219** (1994) 21
- [80] E. Runge, E.K.U. Gross, *Phys. Rev. Lett.* **52** (1984) 997
- [81] E.K.U. Gross, W. Kohn, *Adv. Quant. Chem.* **21** (1990) 2565
- [82] M.E. Casida, in *Recent Advances in Density-Functional Methods*, Ed. D.P. Chong, (Wold Scientific, Singapore, 1995), p. 155
- [83] C. van Caillie, R.D. Amos, *Chem. Phys. Lett.* **308** (1999) 249; *ibid* **317** (2000) 159
- [84] F. Furche, R. Ahlrichs, *J. Chem. Phys.* **117** (2002) 7433
- [85] J. Hutter, *J. Chem. Phys.* **118** (2003) 3928
- [86] N.L. Doltsinis, M. Sprik, *Chem. Phys. Lett.* **330** (2000) 563

- [87] S. Taylor, D.G. Wilden, J. Comer, *Chem. Phys.* **70** (1928) 291
- [88] See, for instance, W.M.C. Faulkes *et al.*, *Rev. Mod. Phys.* **73** (2001) 33
- [89] P. Bokes, I. Štich, L. Mitas, in *Multiscale Computational Methods in Chemistry and Physics*, Eds. A. Brandt, J. Bernholc, K. Binder, (NATO Science Series III, Vol. 177, 2000), p.13
- [90] N. Metropolis *et al.*, *J. Chem. Phys.* **21** (1953) 1087
- [91] P.J. Reynolds *et al.*, *J. Chem. Phys.* **77** (1982) 5593
- [92] D.M. Ceperley, B.J. Alder, *J. Chem. Phys.* **81** (1984) 5833
- [93] Y.S.M. Wu, A. Kuppermann, J.B. Anderson, *Phys. Chem. Chem. Phys.* **1** (1999) 929
- [94] J.B. Anderson, *J. Chem. Phys.* **63** (1975) 1499
- [95] J.B. Anderson, *J. Chem. Phys.* **65** (1976) 4121
- [96] J.W. Moskowitz *et al.*, *J. Chem. Phys.* **77** (1982) 349
- [97] G. Ortiz, D.M. Ceperley, R.M. Martin, *Phys. Rev. Lett.* **71** (1993) 2777
- [98] L. Mitas, *Phys. Rev. Lett.* **96** (2006) 240402
- [99] R.C. Grimm, R.G. Stoner, *J. Comput. Phys.* **7** (1971) 134
- [100] D.M. Ceperley, B.J. Alder, *Phys. Rev. B* **36**(1987) 2092
- [101] W.M.C. Foulks, R.Q. Hood, R.J. Needs, *Phys. Rev. B* **60** (1999) 4558
- [102] L. Mitas, R.M. Martin, *Phys. Rev. Lett.* **72** (1994) 2438
- [103] L. Mitas, *Comput. Phys. Commun.* **96** (1996) 107
- [104] S.F. Boys, F. Bernardi, *Mol. Phys.* **19**(1970) 553
- [105] F. Gygi, *Europhys. Lett.* **19** (1992) 617
- [106] F. Gygi, *Phys. Rev. B* **48** (1993) 11692
- [107] D.R. Hamann, *Phys. Rev. B* **51** (1995) 7337; *ibid.* **51** (1995) 9508; **54** (1996) 1568; **56** (1998) 14979
- [108] D.R. Hamann, *Phys. Rev. Lett.* **81** (1998) 3447
- [109] P.E. Blöchl, *Phys. Rev. B* **50** (1994) 17953
- [110] G. Lippert, J. Hutter, M. Parrinello, *Mol. Phys.* **92** (1997) 477
- [111] G. Lippert, J. Hutter, M. Parrinello, *Chem. Acc.* **103** (1999) 124
- [112] J.R. Chelikowsky, N. Troullier, Y. Saad, *Phys. Rev. Lett.* **72** (1994) 1240
- [113] J.R. Chelikowsky, *et al.*, *Phys. Rev. B* **50** (1994) 11355
- [114] E. Tsuchida, M. Tsukada, *Solid State Commun.* **94** (1995) 5
- [115] E. Tsuchida, M. Tsukada, *Phys. Rev. B* **52** (1995) 5573
- [116] V. Heine, D. Weaire, *Solid State Phys.* **24** (1970) 249
- [117] D.R. Hamann, M. Schlüter, C. Chang, *Phys. Rev. Lett.* **43** (1979) 1494
- [118] G.B. Bachelet, D.R. Hamann, M. Schlüter, *Phys. Rev. B* **26** (1982) 4199
- [119] G. Kerker, *J. Phys. C* **13** (1980) L189
- [120] N. Troullier, J.L. Martins, *Phys. Rev. B* **43** (1991) 1993
- [121] S. Gödecke, M. Teter, J. Hutter, *Phys. Rev. B* **54** (1996) 1703
- [122] I. Štich *et al.*, *Phys. Rev. B* **39** (1989) 4997
- [123] L. Kleinman, D.M. Bylander, *Phys. Rev. Lett.* **48** (1982) 1425
- [124] V. Heine, in *Solid State Physics*, Eds. H. Ehrenreich, F. Seitz, D. Turnbull, Academic, New York, 1980), Vol. 35
- [125] D. Vanderbilt, *Phys. Rev. B* **41** (1990) 7892
- [126] K. Laasonen *et al.*, *Phys. Rev. B* **47** (1993) 10142

- [127] N.W. Ashcroft, N.D. Mermin, *Solid State Physics* (Saunders College Publishing, Philadelphia, 1976)
- [128] A. Baldereschi, *Phys. Rev. B* **7** (1973) 5212
- [129] D.J. Chadi, M.L. Cohen, *Phys. Rev. B* **8** (1973) 5747
- [130] H.J. Monkhorst, J.D. Pack, *Phys. Rev. B* **13** (1976) 5189
- [131] H. Hellmann, *Z. Phys.* **85** (1933) 180
- [132] R.P. Feynman, *Phys. Rev.* **56** (1939) 340
- [133] P. Bendt, A. Zunger, *Phys. Rev. Lett.* **50** (1983) 1684
- [134] P. Pullay, *Mol. Phys.* **17** (1969) 197
- [135] P. Ordejon, E. Artacho, J.M. Soler, *Phys. Rev. B* **53** (1996) R10441
- [136] S. Kato, K. Morokuma, *Chem. Phys. Lett.* **65** (1979) 19
- [137] J.D. Goddard, N.C. Handy, H.F. Schaefer, *J. Chem. Phys.* **71** (1979) 1525
- [138] J.A. Pople *et al.*, *Int. J. Quant. Chem.* **13** (1979) 225
- [139] G.B. Fitzgerald *et al.*, *J. Chem. Phys.* **82** (1985) 4379
- [140] J. Gauss, D. Cremer, *Chem. Phys. Lett.* **153** (1988) 303
- [141] B.R. Brooks *et al.*, *J. Chem. Phys.* **72** (1980) 4652
- [142] G.E. Scuseria, *J. Chem. Phys.* **94** (1991) 442
- [143] T.J. Lee, A.P. Rendell, *J. Chem. Phys.* **94** (1991) 6229
- [144] C. Filippi, C.J. Umrigar, *Phys. Rev. B* **61** (2000) R16291
- [145] R. Assaraf, M. Caffarel, *J. Chem. Phys.* **113** (2000) 4028
- [146] S. Nosé, M.L. Klein, *Mol. Phys.* **50** (1983) 1055
- [147] S. Nosé, *Mol. Phys.* **52** (1984) 255; *J. Chem. Phys.* **81** (1984) 511; *Prog. Theor. Phys. Suppl.* **103** (1991) 1
- [148] W.G. Hoover, *Phys. Rev. A* **31** (1985) 1695
- [149] G.J. Martyna, M.L. Klein, M. Tuckerman, *J. Chem. Phys.* **97** (1992) 2635
- [150] See, for instance, S. Holloway, *J. Phys.: Condensed Matter* **3** (1991) S43
- [151] For a recent reference, see, for instance, A. Dianant, A. Gross, *J. Chem. Phys.* **120** (2004) 5339 (2004)
- [152] R.P. Feynman, *Phys. Rev.* **90** (1953) 1116, *ibid* **91** (1953) 1291; **91** (1953) 1301
- [153] R.P. Feynman, *Statistical Mechanics* (Benjamin, New York, 1973)
- [154] D.M. Ceperley, *Rev. Mod. Phys.* **67** (1995) 279
- [155] F.H. Trotter, *Proc. Am. Math. Soc.* **10** (1959) 545
- [156] J. Cao, G.A. Voth, *J. Chem. Phys.* **99** (1993) 10070
- [157] D. Marx, M.E. Tuckerman, G.J. Martyna, *Comput. Phys. Commun.* **118** (1999) 166
- [158] D. Marx, M. Parrinello, *Z. Phys. B* **95** (1994) 143; *J. Chem. Phys.* **104**, (1996) 4077
- [159] N.P. Cheng, R.N. Barnett, U. Landan, *Chem. Phys. Lett.* **237** (1995) 161
- [160] B.J. Berne, D. Thirumalai, *Ann. Rev. Phys. Chem.* **37** (1986) 401
- [161] D. Chandler, K. Leung, *Ann. Rev. Phys. Chem.* **45** (1994) 557
- [162] D.M. Ceperley, E. Pollock, *Phys. Rev. Lett.* **56** (1986) 351
- [163] F.H. Zhong, D.M. Ceperley, *Phys. Rev. E* **58** (1998) 5123
- [164] M.E. Tuckerman *et al.*, *J. Chem. Phys.* **99** (1996) 2796; *J. Chem. Phys.* **104** (1996) 5579
- [165] M.E. Tuckerman *et al.*, *J. Chem. Phys.* **99** (1993) 2796

- [166] R. Kosloff, *J. Chem. Phys.* **92** (1988) 2087
- [167] R. Kosloff, *Ann. Rev. Phys. Chem.* **45** (1994) 145
- [168] M.D. Feit, J.A. Fleck, Jr., A. Steiger, *J. Comput. Phys.* **47** (1982) 412
- [169] H.C. Andersen, *J. Comput. Phys.* **52** (1983) 24
- [170] L. Verlet, *Phys. Rev.* **159** (1967) 98
- [171] E.S. Fois *et al.*, *J. Phys. Chem.* **92** (1988) 3268
- [172] M. Tuckerman, B.J. Berne, G.J. Martina, *J. Chem. Phys.* **97** (1990) 1990
- [173] J. Teilhaber, *Phys. Rev. B* **46** (1992) 12990
- [174] E.R. Davidson, in *Methods in Computational Molecular Physics*, Eds. G.H.F. Diercksen, S. Wilson, NATO Advanced Study Institute, series C, Vol 113.
- [175] D.M. Wood, A. Zunger, *J. Phys. A* **18** (1985) 1343
- [176] W.M. Press *et al.*, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986)
- [177] M.P. Teter, M.C. Payne, D.C. Allan, *Phys. Rev. B* **40** (1989) 12255
- [178] F. Tassone, F. Mauri, R. Car, *Phys. Rev. B* **50** (1994) 10 561
- [179] G. Kresse, J. Furthmüller, *Phys. Rev. B* **54** (1996) 11169
- [180] M.C. Payne *et al.*, *Phys. Rev. Lett.* **56** (1986) 2656
- [181] P. Pulay, *Chem. Phys. Lett.* **73** (1980) 393
- [182] C. Czászár, P. Pulay, *J. Mol. Struct.* **114** (1984) 31
- [183] T.A. Arias, M.C. Payne, J.D. Joannopoulos, *Phys. Rev. B* **45** (1992) 1538
- [184] L.J. Clarke, I. Štich, M.C. Payne, *Comput. Phys. Commun.* **72** (1992) 14
- [185] I. Štich, *unpublished*
- [186] R. Car, M. Parrinello, *Phys. Rev. Lett.* **55** (1985) 2471
- [187] S. Kirkpatrick, C.D. Gelatt, Jr., M.P. Vecchi, *Science* **220** (1983) 671
- [188] M.C. Payne, M. Needels, J.D. Joannopoulos, *Phys. Rev. B* **37** (1996) 8138
- [189] W.C. Swope *et al.*, *J. Chem. Phys.* **76** (1982) 637
- [190] van Gunsteren, H.J.C. Brendsen, *Mol. Phys.* **34** (1977) 1311
- [191] G. Pastore, E. Smargiassi, F. Buda, *Phys. Rev. A* **44** (1991) 6334
- [192] I. Štich, R. Car, M. Parrinello, *Phys. Rev. B* **44** (1991) 11092
- [193] F.A. Bornemann, C. Schütte, *Numer. Math.* **78** (1998) 359
- [194] I. Štich, R. Car, M. Parrinello, *Phys. Rev. Lett.* **63** (1989) 2240
- [195] I. Štich, R. Car, M. Parrinello, *Phys. Rev. B* **44** (1991) 4262
- [196] M. Sprik, *J. Phys. Chem.* **95** (1991) 2283
- [197] A. Pasquarello *et al.*, *Phys. Rev. Lett.* **69** (1992) 1982
- [198] D. Chandler *et al.*, *J. Chem. Phys.* **108** (1998) 1964
- [199] A. Laio, M. Parrinello, *Proc. Nat. Acad. Sci. USA* **99** (2002) 12562
- [200] I. Štich *et al.*, *Phys. Rev. Lett.* **68** (1992) 1351
- [201] K.D. Brommer *et al.*, *Phys. Rev. Lett.* **68** (1992) 1355
- [202] W. Yang, *Phys. Rev. Lett.* **66** (1991) 1438
- [203] S. Baroni, P. Giannozzi, *Europhys. Lett.* **17** (1992) 547
- [204] X.-P. Li, W. Nunes, D. Vanderbilt, *Phys. Rev. B* **47** (1993) 10891
- [205] D.A. Drabbold, O.F. Sankey, *Phys. Rev. Lett.* **70** (1993) 3631
- [206] S. Goedecker, L. Colombo, *Phys. Rev. Lett.* **73** (1994) 122

- [207] E. Hernández, M.J. Gillan, *Phys. Rev. B* **51** (1995) 10157
- [208] W. Kohn, *Phys. Rev. Lett.* **76** (1996) 3168
- [209] A.J. Williamson, *Phys. Rev. Lett.* **87** (2001) 246406
- [210] See, for instance, J.P. Lewis, P. Ordejón, O.F. Sankey, *Phys. Rev. B* **55** (1997) 6880
- [211] J. Hafner, G. Kahl, *J. Phys. F* **4** (1984) 2259
- [212] F.H. Stillinger, A. Weber, *Phys. Rev. B* **31** (1985) 5262
- [213] J. Broughton, X.P. Li, *Phys. Rev. B* **35** (1987) 9120
- [214] W.D. Luedke, U. Landman, *Phys. Rev. B* **37** (1988) 4656
- [215] J.P. Gabathuller, S. Steeb, *Z. Naturforsch.* **34a** (1979) 1314
- [216] Y. Waseda, K. Suzuki, *Z. Phys. B* **20** (1975) 339
- [217] I. Štich, *Phys. Rev. A* **44** (1991) 1401
- [218] J.P. Hansen, I.R. McDonald, *Theory of simple liquids* (Academic, London, (1976)
- [219] K.M. Shvarev, B.A. Baum, P.V. Geld., *Sov. Phys. Solid. State* **16** (1975) 2111
- [220] C.F. Hague, C. SÉNÉMAUD, H. OSTROVIECKY, *J. Phys.F* **10** (1980) L267
- [221] E.N. Economou, *Green's Functions in Quantum Physics, Springer series in Solid State Physics*, Vol 7, (Springer Verlag, Berlin 1083), p. 152
- [222] H. Sasaki *et al.*, *Jpn. J. Appl. Phys.* **33** (1994) 3803; *ibid* **34** (1995) 432; S.I. Chung *et al.*, *ibid* **34** (1995) L631
- [223] I. Štich, M. Parrinello, J.M. Holender, *Phys. Rev. Lett.* **76** (1996) 2077
- [224] S. Takeda, *Jpn. J. Appl. Phys.* **34** (1995) 4889
- [225] J. Fortner, J.S. Lannin, *Phys. Rev. B* **39** (1989) 5527
- [226] W.A. Kamitakahara *et al.*, *Phys. Rev. B* **36** (1987) 6539
- [227] See, for instance, E.S.R. Elliot, *Physics of Amorphous Materials* (Logman, London, 1984)
- [228] S.T. Pantelides, *Phys. Rev. Lett.* **57** (1986) 2979 (1986); *ibid* **58** (1987) 1344
- [229] See, for instance, N. Rivier, *Adv. Phys.* **34** (1987) 95
- [230] K. Takayanagi *et al.*, *J. Vac. Sci. Technol B* **4** (1985) 1079; *Surf. Sci.* **164** (1985) 367
- [231] R.S. Becker, J.A. Golovchenko, B.S. Swartzentruber, *Phys. Rev. B* **32** (1985) 8455
- [232] H.-J. Gossmann *et al.*, *Phys. Rev. Lett.* **55** (1985) 1106
- [233] R.E. Schlier, H.E. Farnsworth, *J. Chem. Phys.* **30** (1959) 917
- [234] Y.L. Wang *et al.*, *Phys. Rev. B* **70** (2004) 073312
- [235] W. Daum, H. Ibach, J.E. Müller, *Phys. Rev. Lett.* **59** (1987) 1593
- [236] X.-P. Li, P.B. Allen, J.Q. Broughton, *Phys. Rev. Lett.* **61** (1988) 243; X.-P. Li *et al.*, *Phys. Rev. B* **38** (1988) 3331
- [237] I. Štich, K. Terakura, B.E. Larson, *Phys. Rev. Lett.* **74** (1995) 4491
- [238] I. Štich, J. Kohanoff, K. Terakura, *Phys. Rev. B* **54** (1996) 2642
- [239] See, for instance, J. Kohanoff, *Comput. Mater. Sci.* **2** (1994) 221
- [240] See, for instance, M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964)
- [241] See, for instance, A. Zangwill, *Physics at Surfaces* (Cambridge University Press, Cambridge, England, 1988)
- [242] See, for instance, B.E. Hayden, in *Dynamics of Gas-Surface Interactions*, Eds. C.T. Rettner, M.N.R. Ashfold (The Royal Society of Chemistry, London, 1991)
- [243] J.K. Nørskov *et al.*, *Phys. Rev. Lett.* **46** (1981) 257

- [244] A.C. Luntz, J. Grimblot, D.E. Fowler, *Phys. Rev. B* **39** (1989) 12903; W. Wurth *et al.*, *Phys. Rev. Lett.* **65** (1990) 2426
- [245] K.C. Prince, G. Paolucci, A.M. Bradshaw, *Surf. Sci.* **175** (1986) 101; R.J. Gust *et al.*, *Surf. Sci.* **278** (1992) 239
- [246] A.S. Mårtensson, C. Nyberg, S. Andersson, *Phys. Rev. Lett.* **65** (1986) 2426; H.F. Berger, *Surf. Sci. Lett.* **275** (1992) L627
- [247] D.M. Bird, L.J. Clarke, M.C. Payne, I. Štich, *Chem. Phys. Lett.* **212** (1993) 528
- [248] M.J. Gillan, *J. Phys. Condensed Matter* **1** (1989) 689
- [249] B. Hammer, K. W. Jakobsen, J.K. Norskov, *Phys. Rev. Lett.* **69** (1992) 1971
- [250] B. Hammer *et al.*, *Phys. Rev. Lett.* **73** (1994) 1400
- [251] J.A. White, D.M. Bird, M.C. Payne, I. Štich, *Phys. Rev. Lett.* **73** (1994) 1404
- [252] C. Diaz *et al.*, *J. Chem. Phys.* **125** (2006) 114706
- [253] S. Lorentz, A. Gross, M. Scheffler, *Chem. Phys. Lett.* **395** (2004) 210
- [254] M. Hand, J. Harris, *J. Chem. Phys.* **92** (1990) 7610
- [255] J. Harris, A.C. Luntz, *Surf. Sci.* **287/288** (1993) 56
- [256] A.W. Kleyn, *J. Phys. Condens. Matter* **4** (1992) 8375
- [257] A. De Vita, I. Štich, M.J. Gillan, M.C. Payne, L.J. Clarke, *Phys. Rev. Lett.* **71** (1993) 1276
- [258] I. Štich, A. De Vita, M.J. Gillan, M.C. Payne, L.J. Clarke, *Phys. Rev. B* **49** (1994) 8076
- [259] K.C. Pandey, *Phys. Rev. Lett.* **47** (1981) 1913
- [260] M.L. Yu, B.N. Eldridge, *Phys. Rev. Lett.* **58** (1987) 1691
- [261] See, for instance, C.J. Chen, *Introduction to Scanning Tunneling Microscopy* (Oxford University Press, New York, Oxford, 1992)
- [262] G. Binnig, C.F. Quate, Ch. Gerber, *Phys. Rev. Lett.* **56** (1986) 930
- [263] See, for instance, *Noncontact Atomic Force Microscopy*, Eds. S. Morita, R. Wiesendanger, E. Meyer, (Springer Verlag, Berlin 2002)
- [264] F.J. Giessibl, *Science* **267** (1995) 68
- [265] S. Kitamura, M. Iwatsuki, *Jpn. J. Appl. Phys., Part 1* **35** (1996) L668
- [266] Y. Sugawara, M. Ohta, H. Ueyama, S. Morita, *Science* **270** (1995) 1646
- [267] A. Schwarz *et al.*, *Surf. Sci.* **140** (1999) 293
- [268] R. Pérez, M.C. Payne, I. Štich, K. Terakura, *Phys. Rev. Lett.* **78** (1997) 678
- [269] R. Pérez, I. Štich, M.C. Payne, K. Terakura, *Phys. Rev. B* **58** (1998) 10835
- [270] Uchihashi *et al.*, *Phys. Rev. B* **56** (1997) 9834
- [271] J. Tóbič, I. Štich, K. Terakura, *Phys. Rev. B* **63** (2001) 245324
- [272] S.H. Ke, T. Uda, I. Štich, K. Terakura, *Phys. Rev. B* **63** (2001) 245323
- [273] See, for instance, W. Dittrich, M. Reuter, *Classical and Quantum Dynamics* (Springer, New York, 1996), pp. 75–117
- [274] F.J. Giessibl, *Phys. Rev. B* **56** (1997) 16010
- [275] I. Štich, J. Tóbič, R. Pérez, K. Terakura, S.H. Ke, *Prog. Surf. Sci.* **64** (2000) 179
- [276] J. Tóbič, I. Štich, R. Pérez, K. Terakura, *Phys. Rev. B* **60** (1999) 11639
- [277] N. Oyabu *et al.*, *Phys. Rev. Lett.* **90** (2003) 176102
- [278] Y. Sugimoto *et al.*, *Nature Materials* **4** (2005) 156
- [279] P. Dieška, I. Štich, R. Perez, *Phys. Rev. Lett.* **95** (2005) 126103
- [280] P. Dieška, I. Štich, R. Perez, *Phys. Rev. Lett.* **91** (2003) 216401

- [281] M. Kaminska, E.R. Weber, *Imperfections in III/V Materials* (Academic Press, San Diego, 1993)
- [282] S.L. Meisel, *et al.*, *Chem. Technol.* **6** (1976) 86
- [283] S.R. Blaszkowski, R.A. van Santen, *J. Am. Chem. Soc.* **119** (1997) 5020; *ibid.* **118** (1996) 5152
- [284] R. Shah, J.D. Gale, M.C. Payne, *J. Phys. Chem. B* **101** (1997) 478; R. Shah, M.C. Payne, M.-H. Lee, J.D. Gale, *Science* **271** (1996) 1395
- [285] E. Nusterer, P.E. Blöchl, K. Schwarz, *Angew. Chem. Int. Ed.* **35** (1996) 175; E. Nusterer, P.E. Blöchl, K. Schwarz, *Chem. Phys. Lett.* **253** (1996) 448; K. Schwarz, E. Nusterer, P.E. Blöchl, *Int. J. Quant. Chem.* **61** (1997) 369
- [286] I. Štich, J.D. Gale, K. Terakura, M.C. Payne, *Chem. Phys. Lett.* **283** (1998) 402
- [287] I. Štich, J.D. Gale, K. Terakura, M.C. Payne, *J. Am. Chem. Soc.* **121** (1999) 3292
- [288] M. Hytha, I. Štich, J.D. Gale, K. Terakura, M.C. Payne, *Chem. Eur. J.* **7** (2001) 2521
- [289] F. Wakabayashi *et al.*, *Stud. Surf. Sci. Catal.* **105** (1996) 1739
- [290] J. Bandiera, C. Naccache, *Appl. Catal.* **69** (1991) 139
- [291] T.R. Forester, R.F. Howe, *J. Am. Chem. Soc.* **109** (1987) 5076
- [292] See, for instance, C.H. Bennett, in *Algorithms for Chemical Computations*, ACS series No. 46, (Ed. R.E. Christofferson, 1997), p. 63
- [293] E.A. Carter, G. Ciccotti, J.T. Hynes, *Chem. Phys. Lett.* **156** (1989) 472
- [294] M. Sprik, G. Ciccotti, *J. Chem. Phys.* **109** (1998) 7737
- [295] See, for instance, A. Janshoff, M. Neitzert, Y. Oberdörfer, H. Fuchs, *Ang. Chem. Int. Ed.* **39** (2000) 3213
- [296] R.G. Nuzzo, D.L. Allara, *J. Am. Chem. Soc.* **105** (1983) 4481; M.A. Reed *et al.*, *Science* **278** (1997) 252; *Appl. Phys. Lett.* **78** (2001) 3735; K.F. Kelly *et al.*, *Science* **292** (2001) 2303
- [297] M. Brust, *et al.*, *J. Chem. Soc., Chem. Commun.* **801** (1994) 234
- [298] T. Baum, *et al.*, *Science* **272** (1996) 1323
- [299] J. Reichert, *et al.*, *Phys. Rev. Lett.* **88** (2002) 176804; E. Scheer, *Physik Journal* **1** (2002) 20
- [300] C. Zhou *et al.*, *Appl. Phys. Lett.* **71** (1997) 611
- [301] J.H. Schön, H. Meng, Z. Bao, *Nature* **413** (2001) 713
- [302] C. P. Collier *et al.*, *Science* **285** (1999) 391
- [303] B.Q.Q. Xu *et al.*, *Nanoletters* **5** (2005) 1491
- [304] X.L. Li *et al.*, *J. Am. Chem. Soc.* **128** (2006) 2135
- [305] R. Turanský, M. Konôpka, I. Štich, D. Marx, *unpublished*
- [306] A. Ulman, *An Introduction to Ultrathin Organic Films: From Langmuir-Blodgett to Self-assembly*, (Academic, San Diego, CA 1991)
- [307] G. Yang, G.-Y. Liu, *J. Phys. Chem. B* **107** (2003) 8746
- [308] B. Bushan, *Principles and Application of Tribology*, (Wiley, New York, 1999)
- [309] P. Pyykkö, *Chem. Rev.* **88** (1988) 563
- [310] N. Takeuchi, C.T. Chan, K.M. Ho, *Phys. Rev. Lett.* **63** (1989) 1273; *Phys. Rev. B* **43** (1991) 14363
- [311] S.R. Bahn, K.W. Jakobsen, *Phys. Rev. Lett.* **87** (2001) 266101
- [312] R. H. M. Smit, *Phys. Rev. Lett.* **87** (2001) 266102
- [313] D. Krüger, H. Fuchs, R. Rousseau, D. Marx, M. Parrinello, *Phys. Rev. Lett.* **89** (2002) 186402
- [314] G. J. Jackson *et al.*, *Phys. Rev. Lett.* **84** (2000) 119
- [315] S.M. Driver, D.P. Woodruff, *Surf. Sci.* **457** (2000) 11; *ibid* **479** (2001) 1
- [316] S. Vollmer, G. Witte, Ch. Wöll, *Langmuir* **17** (2001) 7660

- [317] Y.-H. Lai *et al.*, *Phys. Chem. B* **106** (2002) 5438
- [318] H. Kondoh *et al.*, *J. Chem. Phys.* **111** (1999) 1175
- [319] D. Krüger, H. Fuchs, R. Rousseau, D. Marx, *Angew. Chem. Int. Ed.* **42** (2003) 2251
- [320] M. Grandbois *et al.*, *Science* **283** (1999) 1727
- [321] M. Konôpka, R. Rousseau, I. Štich, D. Marx, *J. Am. Chem. Soc.* **126** (2004) 12103
- [322] M. Konôpka, R. Rousseau, I. Štich, D. Marx, *Phys. Rev. Lett.* **95** (2005) 096102
- [323] G. Jennings *et al.*, *Langmuir* **14** (1998) 6130
- [324] D. Krüger *et al.*, *J. Chem. Phys.* **115** (2001) 4776
- [325] R. Turanský, M. Konôpka, I. Štich, D. Marx, *to be published*
- [326] Y. Akinaga, T. Nakajima, K. Hirao, *J. Chem. Phys.* **114** (2001) 8555
- [327] D. Sanchez-Portal, E. Artacho, J.M. Soler, *Solid State Commun.* **95** (1995) 685
- [328] A.H. Zewail, *Femtochemistry: Ultrafast dynamics of the chemical bond* (World Scientific, Singapore 1994)
- [329] N.L. Doltsinis, D. Marx, *Phys. Rev. Lett.* **88** (2002) 166402-1
- [330] M. Konôpka, N. Doltsinis, I. Štich, D. Marx, *to be published*
- [331] Ch. Nonneberg *et al.*, *J. Chem. Phys.* **119** (2003) 11585
- [332] Y. Yu, M. Nakano, T. Ikeda, *Nature* **425** (2003) 145
- [333] G.D. Mahan, *Many-Particle Physics* (Plenum, New York, 1991), chap. 5
- [334] N.W. Ashcroft, N.D. Mermin, *Solid State Physics* (Holt-Sounders Editions, New York, 1976).
- [335] E.P. Wigner, *Phys. Rev.* **46** (1934) 1002
- [336] M. Gel-Mann, K.A. Brueckner, *Phys. Rev.* **106** (1957) 364
- [337] K.B. Whaley, *Int. Rev. Phys. Chem.* **13** (1994) 41
- [338] D. Scharf, G.J. Martyna, M.L. Klein, *Chem. Phys. Lett.* **197** (1992) 231; *J. Chem. Phys.* **97** (1992) 3590
- [339] P. Sindzingre, D.M. Ceperley, M.L. Klein, *Phys. Rev. Lett.* **67** (1991) 1871
- [340] R. Clampitt, L. Gowland, *Nature (London)* **223** (1969) 815
- [341] N.J. Kirchner, M.T. Bowers, *J. Chem. Phys.* **86** (1987) 1301
- [342] K. Hiraoka, P. Kebarle, *J. Chem. Phys.* **62** (1975) 2267
- [343] R.J. Beuhler, S. Eherenson, L. Friedman, *J. Chem. Phys.* **79** (1983) 5982
- [344] K. Hiraoka, *J. Chem. Phys.* **87** (1987) 4048
- [345] M. Farizon *et al.*, *Chem. Phys. Lett.* **177** (1991) 451; *J. Chem. Phys.* **96** (1992) 1325
- [346] M. Chermette, H. Razafinjanahary, L. Carrion, *J. Chem. Phys.* **107** (1997) 10643
- [347] E.W. Ignacio, S. Yamabe, *Chem. Phys. Lett.* **287** (1998) 563
- [348] T. Pang, *Chem. Phys. Lett.* **228** (1994) 555
- [349] I. Štich, D. Marx, M. Parrinello, K. Terakura, *Phys. Rev. Lett.* **78** (1997) 3669; *J. Chem. Phys.* **107** (1997) 9482
- [350] T.H. Dunning Jr., *J. Chem. Phys.* **90** (1989) 1007
- [351] J.C. Grossman, L. Mitás, *Phys. Rev. Lett.* **94** (2005) 056403
- [352] D. Marx, M. Parrinello, *Nature* **375** (1995) 216
- [353] M.E. Tuckerman *et al.*, *Science* **275** (1997) 817
- [354] D. Marx *et al.*, *Nature* **397** (1999) 601
- [355] W. Andreoni, *Perspect. Drug Discovery Des.* **9–11** (1998) 161

- [356] P. Carloni, F. Alber, *Perspect. Drug Discovery Des.* **9–11** (1998) 169
- [357] J. Hutter, P. Carloni, M. Parrinello, *J. Am. Chem. Soc.* **118** (1996) 8710
- [358] M. D. Segall *et al.*, *Xenobiotica* **29** (1999) 561
- [359] A. Biffone *et al.*, *Pure Appl. Chem.* **68** (1997) 2105
- [360] U. Rothlisberger *et al.*, *J. Biol. Inorg. Chem.* **5** (2000) 236