

Quantum simulation techniques for electrons and ions

Roman Martoňák

*Department of Experimental Physics
Faculty of Mathematics, Physics and Informatics
Comenius University, Bratislava, Slovakia
martonak@fmph.uniba.sk*

2013

Literature:

Richard M. Martin, *Electronic Structure Basic Theory and Practical Methods*,
Cambridge University Press (2004)

Outline

- I Ab Initio Total Energy Calculations (Introduction to Density Functional Theory)
- II Ab-initio Molecular Dynamics (Car-Parrinello) Method
- III Special topics
- IV Path Integral Quantum Monte Carlo methods

Autor: Roman Martoňák

Názov: Quantum simulation techniques for electrons and ions

Vydavateľ: Knižničné a edičné centrum FMFI UK

Rok vydania: 2013

Miesto vydania: Bratislava

Vydanie: prvé

Počet strán (číslované): 112

Internetová adresa: http://www.fmph.uniba.sk/index.php?id=el_st_m

ISBN: 978-80-8147-014-1

EAN: 9788081470141

Part I

Ab initio total-energy calculations (Introduction to Density Functional Theory)

Need for *better accuracy* of the potential energy

- classical potentials are good for many applications but have clear limits
- good close to equilibrium state for which they are fitted, away from equilibrium the accuracy deteriorates rapidly
- processes of *bond breaking and remaking* are typically described poorly
- *chemistry* is excluded
- because of insufficient accuracy the *predictive power is often limited*
- improving accuracy of the *total energy calculations* is one of the main challenges of computational physics
- we have to *solve the Schrödinger equation for the electrons* (in some approximation)
- hierarchy (in the order of increasing accuracy):
 - semi-empirical methods (e.g. tight-binding schemes)
 - density-functional theory (DFT)
 - Quantum Monte Carlo (QMC)
 - accurate quantum chemistry methods (configurational interaction, coupled cluster), extremely demanding

An accurate description allows us to calculate a *wide spectrum of properties*:

- shapes and size of molecules
- crystal structures of solids
- binding energies
- ionization energies and electron affinities

- heights of energy barriers to various processes
- static response functions
- vibrational and phonon frequencies
- mechanical properties of solids
- phase boundaries
- once we have an accurate potential, we can run MD and directly observe some processes
- provides a *predictive power*

How to solve the electronic problem ?

- we assume the *Born-Oppenheimer approximation*
- we use the *atomic units*, mass unit m_e , charge unit e , length unit Bohr radius $a_0 = \frac{4\pi\epsilon_0\hbar^2}{e^2m_e} = 0.529 \text{ \AA}$, energy unit Hartree $= \frac{\hbar^2}{m_e a_0^2} = 27.21 \text{ eV}$
- electrons in the field of the nuclei

$$\begin{aligned}\hat{T}_e(\{\mathbf{r}\}) &= -\sum_i \frac{1}{2} \nabla_{\mathbf{r}_i}^2 \\ \hat{V}_{ee}(\{\mathbf{r}\}) &= \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ \hat{V}_{eN}(\{\mathbf{r}\}, \{\mathbf{R}\}) &= -\sum_{iJ} \frac{Z_J}{|\mathbf{r}_i - \mathbf{R}_J|} = \sum_{i=1}^N v(\mathbf{r}_i) \\ \hat{H} &= \hat{T}_e(\{\mathbf{r}\}) + \hat{V}_{ee}(\{\mathbf{r}\}) + \hat{V}_{eN}(\{\mathbf{r}\}, \{\mathbf{R}\})\end{aligned}$$

- we want to solve the time-independent Schrödinger equation

$$\hat{H}\Psi(\{\mathbf{r}\}, \{\mathbf{R}\}) = E(\{\mathbf{R}\}) \Psi(\{\mathbf{r}\}, \{\mathbf{R}\})$$

- we are particularly interested in the *ground state energy* $E_0(\{\mathbf{R}\})$

1 The Hartree-Fock method

The variational principle for the ground state

- we define energy $E[\Psi]$ as a functional of the wavefunction Ψ

$$E[\Psi] = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int d\mathbf{r} \Psi^* \hat{H} \Psi}{\int d\mathbf{r} \Psi^* \Psi}$$

- variational principle

$$E[\Psi] \geq E_0$$
$$E_0 = \min_{\Psi} E_0$$

- every eigenstate Ψ is an extremum of $E[\Psi]$
- condition $\delta E[\Psi] = 0$ is equivalent to the Schrödinger equation
- the normalization condition can be included via Lagrange multiplier $\delta[\langle \Psi | \hat{H} | \Psi \rangle - E \langle \Psi | \Psi \rangle] = 0$
- the condition is true also for excited states Ψ_1, Ψ_2, \dots
- the ground-state energy $E[\Psi]$ depends only on the number of electrons N and the nuclear potential $\hat{V}_{eN}(\{\mathbf{r}\})$
- kinetic energy and electron repulsion terms are *universal* (determined only by N)
- the ground-state energy E can be seen as functional of N and $\hat{V}_{eN}(\{\mathbf{r}\})$

The Hartree-Fock approximation (mean-field theory) Hartree (1928), Fock (1930)

- electrons are *fermions*
- we take a simple ansatz for the wavefunction Ψ
- we approximate the wavefunction Ψ_{HF} as an antisymmetrized product (*Slater determinant*) of N orthonormal spin orbitals $\Psi_i(\mathbf{x})$

- each $\Psi_i(\mathbf{r})$ is a product of a spatial orbital $\phi_i(\mathbf{r})$ and spin function $\sigma(s) = \alpha(s)$ (\uparrow) or $\beta(s)$ (\downarrow)

$$\Psi_{HF} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Psi_1(\mathbf{x}_1) & \Psi_2(\mathbf{x}_1) & \dots & \Psi_N(\mathbf{x}_1) \\ \Psi_1(\mathbf{x}_2) & \Psi_2(\mathbf{x}_2) & \dots & \Psi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \Psi_1(\mathbf{x}_N) & \Psi_2(\mathbf{x}_N) & \dots & \Psi_N(\mathbf{x}_N) \end{vmatrix} = \frac{1}{\sqrt{N!}} \det[\Psi_1 \Psi_2 \dots \Psi_N]$$

- the Hartree-Fock wavefunction is normalized $\langle \Psi_{HF} | \Psi_{HF} \rangle = 1$
- the product approximation corresponds to treating each electron in the *mean-field* of the other electrons
- the *optimal orbitals* $\Psi_i(\mathbf{x})$ can be determined by a *variational procedure*

The Hartree-Fock equations

the expectation value of the energy reads

$$E_{HF} = \langle \Psi_{HF} | \hat{H} | \Psi_{HF} \rangle = \sum_{i=1}^N H_i + \frac{1}{2} \sum_{i,j=1}^N (J_{ij} - K_{ij})$$

$$\text{where } H_i = \int \Psi_i^*(\mathbf{x}) \left[-\frac{1}{2} \nabla^2 + v(\mathbf{x}) \right] \Psi_i(\mathbf{x}) d\mathbf{x}$$

$$J_{ij} = \iint \Psi_i(\mathbf{x}_1) \Psi_i^*(\mathbf{x}_1) \frac{1}{r_{12}} \Psi_j(\mathbf{x}_2) \Psi_j^*(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2$$

$$K_{ij} = \iint \Psi_i^*(\mathbf{x}_1) \Psi_j(\mathbf{x}_1) \frac{1}{r_{12}} \Psi_i(\mathbf{x}_2) \Psi_j^*(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2$$

- J_{ij} are *Coulomb integrals*, K_{ij} are *exchange integrals*, $J_{ii} = K_{ii}$
- E_{HF} is minimized with respect to orbitals $\Psi_i(\mathbf{x})$ subject to orthonormalization conditions $\int \Psi_i^*(\mathbf{x}) \Psi_j(\mathbf{x}) = \delta_{ij}$

The Hartree-Fock equations

- set of Hartree-Fock differential equations

$$\hat{F} \Psi_i(\mathbf{x}) = \sum_{j=1}^N \epsilon_{ij} \Psi_j(\mathbf{x})$$

$$\hat{F} = -\frac{1}{2} \nabla^2 + v + \hat{g}$$

- ϵ_{ij} are Lagrange multipliers for the constraints
- \hat{g} is the Coulomb exchange operator, $\hat{g} = \hat{j} - \hat{k}$

$$\hat{j}(\mathbf{x}_1)f(\mathbf{x}_1) = \sum_{k=1}^N \int \Psi_k^*(\mathbf{x}_2)\Psi_k(\mathbf{x}_2)\frac{1}{r_{12}}f(\mathbf{x}_1)d\mathbf{x}_2$$

$$\hat{k}(\mathbf{x}_1)f(\mathbf{x}_1) = \sum_{k=1}^N \int \Psi_k^*(\mathbf{x}_2)f(\mathbf{x}_2)\frac{1}{r_{12}}\Psi_k(\mathbf{x}_1)d\mathbf{x}_2$$

- coupled system of equations can be solved by iteration until *self-consistency* is achieved

The Hartree-Fock equations

- “orbital energies”

$$\epsilon_i = \epsilon_{ii} = \langle \Psi_i | \hat{F} | \Psi_i \rangle = H_i + \sum_{j=1}^N (J_{ij} - K_{ij})$$

- the total energy E_{HF} is *not a sum of orbital energies*

$$E_{HF} = \sum_{j=1}^N \epsilon_j - V_{ee}$$

$$V_{ee} = \int \Psi_{HF}^*(\mathbf{x}^N) \left(\sum_{ij} \frac{1}{r_{ij}} \right) \Psi_{HF}(\mathbf{x}^N) d\mathbf{x}^N = \frac{1}{2} \sum_{i,j=1}^N (J_{ij} - K_{ij})$$

- nonlinear *self-consistent field* method
- E_{HF} is an *exact upper bound* for the energy E (variational principle)

Correlation energy

- the exact wavefunction of an interacting system is *never* a single determinant
- error in calculation by Hartree-Fock is called *correlation energy*

$$E_{corr} = E - E_{HF}, \quad E_{corr} < 0$$

- E_{corr} is the part of energy due to many-body effects *beyond mean-field description*
- calculation of E_{corr} is a *major* problem in many-body theory
- in quantum chemistry, mixing of many determinants is used to calculate accurate energies (*Configurational Interaction CI*)
- many-body perturbation techniques
- only *small systems* can be treated accurately
- E_{corr} is of crucial importance when the *bonds* change

Electron density

$$\rho(\mathbf{r}_1) = N \sum_{s_1 \dots s_n} \int \dots \int |\Psi(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, \dots, \mathbf{r}_N s_N)|^2 d\mathbf{r}_2 \dots d\mathbf{r}_N, \rho(\mathbf{r}) \geq 0$$

- *electron density* $\rho(\mathbf{r})$ is a function of just one position \mathbf{r} and $\int \rho(\mathbf{r}) d\mathbf{r} = N$

$$\begin{aligned} \langle V_{ext} \rangle &= \int \dots \int \Psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N) \sum_{i=1}^N V_{ext}(\mathbf{r}_i) \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) d\mathbf{r}_1 \dots d\mathbf{r}_N \\ &= \sum_{i=1}^N \int \dots \int d\mathbf{r}_1 \dots d\mathbf{r}_N \Psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N) V_{ext}(\mathbf{r}_i) \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \int d\mathbf{r} \delta(\mathbf{r} - \mathbf{r}_i) \\ &= \sum_{i=1}^N \int d\mathbf{r} V_{ext}(\mathbf{r}) \int \dots \int d\mathbf{r}_1 \dots d\mathbf{r}_N \Psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N) \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \delta(\mathbf{r} - \mathbf{r}_i) \\ &= \int d\mathbf{r} V_{ext}(\mathbf{r}) \langle \sum_{i=1}^N \delta(\mathbf{r} - \mathbf{r}_i) \rangle = \int d\mathbf{r} V_{ext}(\mathbf{r}) \rho(\mathbf{r}) \end{aligned}$$

- $\langle V_{ext} \rangle$ depends on the wavefunction Ψ *only through the density* $\rho(\mathbf{r})$

Hellman - Feynman theorem

- Hamiltonian $\hat{H}(\lambda)$ dependent on a parameter λ

- $\Psi(\lambda)$ is a normalized eigenstate with eigenvalue $E(\lambda)$

$$\begin{aligned}\hat{H}(\lambda)\Psi(\lambda) &= E(\lambda)\Psi(\lambda) \\ E(\lambda) &= \langle \Psi(\lambda) | \hat{H}(\lambda) | \Psi(\lambda) \rangle \\ \frac{dE(\lambda)}{d\lambda} &= \frac{d}{d\lambda} \langle \Psi(\lambda') | \hat{H}(\lambda) | \Psi(\lambda') \rangle \Big|_{\lambda'=\lambda} + \langle \Psi(\lambda) | \frac{\partial \hat{H}(\lambda)}{\partial \lambda} | \Psi(\lambda) \rangle\end{aligned}$$

- the first term is zero because of the variational principle
- the *Hellman - Feynman theorem*

$$\frac{dE(\lambda)}{d\lambda} = \langle \Psi(\lambda) | \frac{\partial \hat{H}(\lambda)}{\partial \lambda} | \Psi(\lambda) \rangle$$

- integral form

$$E(\lambda_1) - E(\lambda_2) = \int_{\lambda_1}^{\lambda_2} \langle \Psi(\lambda) | \frac{\partial \hat{H}(\lambda)}{\partial \lambda} | \Psi(\lambda) \rangle d\lambda$$

Electrostatic theorem Feynman (1939)

- assume the electronic Hamiltonian for static nuclei, nuclear positions $\{\mathbf{R}\}$ are parameters

$$\hat{H} + V_{NN} = -\sum_i \frac{1}{2} \nabla_{\mathbf{r}_i}^2 + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{iJ} \frac{Z_J}{|\mathbf{r}_i - \mathbf{R}_J|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

- force acting on the nuclei is

$$\begin{aligned}-\frac{\partial(E + V_{NN})}{\partial \mathbf{R}_I} &= \langle \Psi | -\frac{\partial(\hat{H} + V_{NN})}{\partial \mathbf{R}_I} | \Psi \rangle \\ &= \int d\mathbf{r} \rho(\mathbf{r}) \frac{Z_I(\mathbf{r} - \mathbf{R}_I)}{|\mathbf{r} - \mathbf{R}_I|^3} + \sum_{J \neq I} \frac{Z_I Z_J (\mathbf{R}_I - \mathbf{R}_J)}{|\mathbf{R}_I - \mathbf{R}_J|^3}\end{aligned}$$

- *electrostatic theorem* for the forces
- the quantum expression is *exactly identical to the classical one*
- all we need is the electronic density
- can be used for the geometry optimization of molecular structure as well as for ab initio molecular dynamics

2 Density functional theory (DFT)

Literature:

- Robert G. Parr and Weitao Yang, *Density Functional Theory of Atoms and Molecules*, Oxford University press (1989)
- J. P. Perdew and S. Kurth, *Density Functionals for Non-relativistic Coulomb Systems in the New Century*. Lecture Notes in Physics **620**, pp. 1-55, (2003) Springer Verlag

Density functional theory

- techniques based on wavefunctions are complicated
- for *ground states*, a special theory exists
- instead of the wavefunction $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)$, the density $\rho(\mathbf{r})$ is the basic variable
- conceptually this represents a great simplification
- offers a practical computational scheme
- many free and commercial codes exist
- widespread use in many branches of physics and chemistry

The Thomas-Fermi model Thomas (1927), Fermi (1927)

- how does the kinetic energy of electrons depend on density?
- consider noninteracting electrons at $T = 0$
- we divide space into small cubes of size $\Delta V = l^3$ containing ΔN electrons, $\rho(\mathbf{r}) = \frac{\Delta N}{\Delta V}$ (may be different in different cells)
- density of states $g(\epsilon) = \frac{\pi}{4} \left(\frac{8ml^2}{h^2} \right)^{\frac{3}{2}} \epsilon^{\frac{1}{2}}$
- at $T = 0$ occupation $f(\epsilon) = 1$ for $\epsilon < \epsilon_F$, $f(\epsilon) = 0$ for $\epsilon > \epsilon_F$
- kinetic energy $\Delta E = 2 \int \epsilon f(\epsilon) g(\epsilon) d\epsilon = \frac{8\pi}{5} \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} l^3 \epsilon_F^{\frac{5}{2}}$
- number of particles $\Delta N = 2 \int f(\epsilon) g(\epsilon) d\epsilon = \frac{8\pi}{3} \left(\frac{2m}{h^2} \right)^{\frac{3}{2}} l^3 \epsilon_F^{\frac{3}{2}}$

$$\Delta E = \frac{3h^2}{10m} \left(\frac{3}{8\pi} \right)^{\frac{2}{3}} l^3 \left(\frac{\Delta N}{l^3} \right)^{\frac{5}{3}} = \frac{3h^2}{10m} \left(\frac{3}{8\pi} \right)^{\frac{2}{3}} \rho^{\frac{5}{3}}(\mathbf{r}) \Delta V$$

The Thomas-Fermi model

- integrating over the whole space we get the *Thomas-Fermi kinetic energy functional* (in atomic units)

$$T_{TF}[\rho] = C_F \int \rho^{\frac{5}{3}}(\mathbf{r}) d\mathbf{r}, C_F = \frac{3}{10} (3\pi^2)^{\frac{2}{3}}$$

- the underlying idea is in fact the *local density approximation (LDA)*
- application to an atom

$$E_{TF}[\rho(\mathbf{r})] = C_F \int \rho^{\frac{5}{3}}(\mathbf{r}) d\mathbf{r} - Z \int \frac{\rho(\mathbf{r})}{r} d\mathbf{r} + \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2$$

- one can search for the density $\rho(\mathbf{r})$ which minimizes the functional $E_{TF}[\rho(\mathbf{r})]$ under the constraint $\int \rho(\mathbf{r}) d\mathbf{r} = N$
- fails to predict binding in molecules
- before the advent of DFT (1964) the TF model was considered merely as a simple and inaccurate model
- the Hohenberg-Kohn theorems provide a different interpretation: TF model represents an approximation to the exact *density functional theory*

Hohenberg-Kohn theorems Hohenberg and Kohn 1964

- the external potential $v(\mathbf{r})$ and number of electrons N *determine the Hamiltonian entirely*
- all properties of the ground state are given by $v(\mathbf{r})$ and N
- we assume a *non-degenerate* ground state

Theorem 1 (First Hohenberg-Kohn theorem). *The external potential $v(\mathbf{r})$ is uniquely determined (up to an additive constant) by the electron density $\rho(\mathbf{r})$.*

The theorem allows to replace the wavefunction $\Psi(\vec{r})$ by the density $\rho(\mathbf{r})$, which is much more practical.

First Hohenberg-Kohn theorem

Proof. • we assume a non-degenerate ground state

- $\rho(\mathbf{r})$ determines the number of electrons N
- assume there are two different potentials v, v' defining two Hamiltonians \hat{H}, \hat{H}' such that the respective ground states Ψ, Ψ' are different but give the same density $\rho(\mathbf{r})$
- variational principle

$$E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle = E'_0 + \int \rho(\mathbf{r}) [v(\mathbf{r}) - v'(\mathbf{r})] d\mathbf{r}$$

$$E'_0 < \langle \Psi | \hat{H}' | \Psi \rangle = \langle \Psi | \hat{H} | \Psi \rangle + \langle \Psi | \hat{H}' - \hat{H} | \Psi \rangle = E_0 - \int \rho(\mathbf{r}) [v(\mathbf{r}) - v'(\mathbf{r})] d\mathbf{r}$$

- summing the two equations we get a *contradiction* $E_0 + E'_0 < E'_0 + E_0$
- there cannot be two different potentials v, v' that give for the ground state the same $\rho(\mathbf{r})$

□

First Hohenberg-Kohn theorem *Consequences*

- $\rho(\mathbf{r})$ itself determines N and v and therefore all properties of the ground state
- ground state energy $E[\rho]$ is *functional* of the density

$$E_v[\rho] = T[\rho] + V_{Ne}[\rho] + V_{ee}[\rho] = \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + F_{HK}[\rho]$$

$$\text{where } F_{HK}[\rho] = T[\rho] + V_{ee}[\rho]$$

- $F_{HK}[\rho]$ is independent of $v(\mathbf{r})$ and is a *universal functional* of $\rho(\mathbf{r})$
- we separate from V_{ee} the classical repulsion $J[\rho]$ (*Hartree term*)

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2$$

$$V_{ee}[\rho] = J[\rho] + \text{non-classical term}$$

- the non-classical term will be the main part of the *exchange-correlation energy*

Second Hohenberg-Kohn theorem

represents a *variational principle* for the density

Theorem 2 (Second Hohenberg-Kohn theorem). *For any normalized density $\tilde{\rho}(\mathbf{r}) \geq 0$, $\int \tilde{\rho}(\mathbf{r}) d\mathbf{r} = N$,*

$$E_0 \leq E_v[\tilde{\rho}] ,$$

where $E_v[\rho] = \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} + F_{HK}[\rho]$.

Proof. • $\tilde{\rho}(\mathbf{r})$ determines the $\tilde{v}(\mathbf{r})$, Hamiltonian \tilde{H} and wavefunction $\tilde{\Psi}$

$$\langle \tilde{\Psi} | H | \tilde{\Psi} \rangle = \int \tilde{\rho}(\mathbf{r})v(\mathbf{r})d\mathbf{r} + F_{HK}[\tilde{\rho}] = E_v[\tilde{\rho}] \geq E_v[\rho]$$

□

Second Hohenberg-Kohn theorem *Consequences*

- ground-state energy is stationary with respect to variations of $\rho(\mathbf{r})$

$$\delta \left\{ E_v[\rho] - \mu \left[\int \rho(\mathbf{r})d\mathbf{r} - N \right] \right\} = 0$$

- Euler-Lagrange equation

$$\mu = \frac{\delta E_v[\rho]}{\delta \rho(\mathbf{r})} = v(\mathbf{r}) + \frac{\delta F_{HK}[\rho]}{\delta \rho(\mathbf{r})}$$

- μ is the chemical potential
- in principle the last equation is an exact equation for the ground state $\rho(\mathbf{r})$ but we don't know the exact $F_{HK}[\rho]$
- we can only construct approximations

Levy constrained-search derivation (Levy 1979)

- in the variational principle we split the minimization over all wavefunctions Ψ into *two steps*

$$E = \min_{\Psi} \langle \Psi | \hat{H} | \Psi \rangle = \min_{\rho} \left\{ \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{H} | \Psi \rangle \right\}$$

- first we minimize over all wavefunctions $\Psi \rightarrow \rho(\mathbf{r})$ that yield the given N-particle density $\rho(\mathbf{r})$

$$\min_{\Psi \rightarrow \rho} \langle \Psi | \hat{H} | \Psi \rangle = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle + \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r}$$

- all Ψ that yield the same $\rho(\mathbf{r})$ have the same value of $\int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r}$
- we define the *universal functional*

$$F_{HK}[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle = \langle \Psi_{\rho}^{min} | \hat{T} + \hat{V}_{ee} | \Psi_{\rho}^{min} \rangle$$

- in the second step we minimize over all densities ρ

$$E = \min_{\rho} \left\{ F_{HK}[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} \right\}$$

- the minimizing density is the *ground-state density*
- equivalent and more "constructive" derivation of the DFT (avoids proof by contradiction)
- in principle it provides a path towards the exact functional $F_{HK}[\rho]$, but it is not possible to calculate it explicitly

3 The Kohn-Sham method

The Kohn-Sham method Kohn, Sham (1965)

- DFT provides an enormous simplification due to use of electronic density $\rho(\mathbf{r})$ as basic variable instead of $\Psi(\mathbf{r})$
- problem: the functional

$$F_{HK}[\rho] = T[\rho] + V_{ee}[\rho] = T[\rho] + J[\rho] + \text{non-classical term}$$

is unknown

- Thomas-Fermi theory is an approximation to DFT: $T[\rho]$ is taken from non-interacting electron gas and the non-classical term is neglected completely
- there were attempts to improve the accuracy of the TF approach and construct better approximations for the $F_{HK}[\rho]$, but no much success

- the method of Kohn and Sham employs a different approach, based on re-introducing wavefunctions in the form of one-electron orbitals
- much better accuracy can be achieved
- method provides a practical scheme for numerical calculations and is currently the most widespread use of DFT in physics and chemistry

The Kohn-Sham method

- we introduce a fictitious *non-interacting system* with no electron-electron repulsion V_{ee} and an effective external potential v_s

$$\hat{H}_s = - \sum_{i=1}^N \frac{1}{2} \nabla_{\mathbf{r}_i}^2 + \sum_{i=1}^N v_s(\mathbf{r}_i) = \sum_{i=1}^N \hat{h}_s(\mathbf{r}_i)$$

- by construction the *ground state density of the non-interacting system must be identical to the ground state density of the interacting system*
- the ground-state wavefunction of the non-interacting system is a Slater determinant

$$\Psi_s = \frac{1}{\sqrt{N!}} \det[\Psi_1 \Psi_2 \dots \Psi_N]$$

- Ψ_i are the N lowest eigenstates of the one-electron Hamiltonian \hat{h}_s

$$\hat{h}_s \Psi_i = \left[-\frac{1}{2} \nabla_{\mathbf{r}_i}^2 + v_s(\mathbf{r}_i) \right] \Psi_i = \epsilon_i \Psi_i$$

$$\rho(\mathbf{r}) = \sum_{i=1}^N \sum_s |\Psi_i(\mathbf{r}, s)|^2$$

The Kohn-Sham method

- the exact kinetic energy of the non-interacting system $T_s[\rho]$ is easy to calculate

$$T_s[\rho] = \langle \Psi_s | \sum_{i=1}^N \left(-\frac{1}{2} \nabla_{\mathbf{r}_i}^2 \right) | \Psi_s \rangle = \sum_{i=1}^N \langle \Psi_i | -\frac{1}{2} \nabla_{\mathbf{r}}^2 | \Psi_i \rangle$$

- we express the functional $F[\rho]$ in a different way

$$\begin{aligned} F[\rho] &= T[\rho] + V_{ee}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \\ E_{xc}[\rho] &= T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \end{aligned}$$

- $E_{xc}[\rho]$ is called the *exchange-correlation energy*
- $E_{xc}[\rho]$ contains the difference $T[\rho] - T_s[\rho]$ (hopefully small) and the non-classical part of electron-electron repulsion $V_{ee}[\rho]$
- $T[\rho]$ is typically large while $E_{xc}[\rho]$ is much smaller
- $E_{xc}[\rho]$ is better suited for approximations than $T[\rho]$, many different approximations were constructed
- $E_{xc}[\rho]$ is of *crucial importance for chemical bonding*

Exchange-correlation potential

- the energy functional to be minimized is

$$E[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r}$$

- the Euler variational equation becomes

$$\mu = \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} = v(\mathbf{r}) + \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} = v_{eff}(\mathbf{r}) + \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})}$$

$$\text{where } v_{eff}(\mathbf{r}) = v(\mathbf{r}) + \frac{\delta J[\rho]}{\delta \rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r})$$

- $v_{eff}(\mathbf{r})$ is the *Kohn-Sham effective potential*
- the quantity $v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$ is the *exchange-correlation potential*
- the form of $T_s[\rho]$ is still unknown
- an additional ingredient is necessary to make progress

The Kohn-Sham equations

- we formulate the variational problem in terms of the *Kohn-Sham orbitals* Ψ_i , required to be orthonormal

$$\int \Psi_i^*(\mathbf{r})\Psi_i(\mathbf{r}) d\mathbf{r} = \delta_{ij}$$

- $E[\rho]$ is expressed as functional of the KS orbitals

$$\begin{aligned} E[\rho] &= T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} \\ &= \sum_{i=1}^N \int \Psi_i^*(\mathbf{r})\left(-\frac{1}{2}\nabla_{\mathbf{r}}^2\right)\Psi_i(\mathbf{r}) d\mathbf{r} + J[\rho] + E_{xc}[\rho] + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} \end{aligned}$$

$$\text{where } \rho(\mathbf{r}) = \sum_{i=1}^N \sum_s |\Psi_i(\mathbf{r}, s)|^2$$

- we minimize $\Omega[\{\Psi_i\}]$, ϵ_{ij} are Lagrange multipliers

$$\Omega[\{\Psi_i\}] = E[\rho] - \sum_{i=1}^N \sum_{j=1}^N \epsilon_{ij} \int \Psi_i^*(\mathbf{r})\Psi_j(\mathbf{r}) d\mathbf{r}$$

The Kohn-Sham equations

- imposing the variational condition $\delta\Omega[\{\Psi_i\}] = 0$ provides the *Kohn-Sham equations*

$$\begin{aligned} \left[-\frac{1}{2}\nabla_{\mathbf{r}}^2 + v_{eff}(\mathbf{r}) \right] \Psi_i &= \epsilon_i \Psi_i \\ v_{eff}(\mathbf{r}) &= v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r}) \\ \rho(\mathbf{r}) &= \sum_{i=1}^N \sum_s |\Psi_i(\mathbf{r}, s)|^2 \end{aligned}$$

- $H_{KS} = -\frac{1}{2}\nabla_{\mathbf{r}}^2 + v_{eff}(\mathbf{r})$ is called *the Kohn-Sham Hamiltonian*
- $v_{eff}(\mathbf{r})$ depends on $\rho(\mathbf{r})$
- $\rho(\mathbf{r})$ depends on Ψ_i which depend on $v_{eff}(\mathbf{r})$

- *one-particle Schrödinger equation* is typically solved by diagonalization in some basis, N lowest lying eigenstates are found
- *coupled system of nonlinear equations* that can be solved iteratively to self-consistency
- to start we have to have an initial guess for the density

The Kohn-Sham equations

- total energy E in the KS theory is *not the sum* of the orbital energies

$$\sum_{i=1}^N \epsilon_i = \sum_{i=1}^N \langle \Psi_i | -\frac{1}{2} \nabla_{\mathbf{r}}^2 + v_{eff}(\mathbf{r}) | \Psi_i \rangle = T_s[\rho] + \int \rho(\mathbf{r}) v_{eff}(\mathbf{r}) d\mathbf{r}$$

$$E = \sum_{i=1}^N \epsilon_i - \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[\rho] - \int \rho(\mathbf{r}) v_{xc}(\mathbf{r}) d\mathbf{r}$$

- *many-electron system* is described by *one-electron equations* (similar to the Hartree-Fock theory)
- however, Hartree-Fock theory is approximate by construction while in the KS theory the accuracy can be systematically improved by better approximations to $E_{xc}[\rho]$
- the orbital energies ϵ_i have no direct physical meaning

Local density approximation (LDA)

- the difficult problem was transferred into the terms $E_{xc}[\rho]$ and $v_{xc}(\mathbf{r})$
- we still need some approximation for it
- search for an improved $E_{xc}[\rho]$ is the greatest challenge in the DFT
- the simplest choice is the *local density approximation*

$$E_{xc}^{LDA}[\rho] = \int \rho(\mathbf{r}) \epsilon_{xc}(\rho(\mathbf{r})) d\mathbf{r}$$

- ϵ_{xc} is the exchange and correlation energy per particle of a *uniform electron gas* of density ρ

$$v_{xc}^{LDA}(\mathbf{r}) = \frac{\delta E_{xc}^{LDA}[\rho]}{\delta \rho(\mathbf{r})} = \epsilon_{xc}(\rho(\mathbf{r})) + \rho(\mathbf{r}) \frac{\partial \epsilon_{xc}(\rho)}{\partial \rho}$$

- accurate values for $\epsilon_{xc}(\rho)$ are known from QMC calculations (Ceperley, Alder 1980)
- QMC values have been interpolated to provide an analytic expression
- LDA was a widely used method in the beginnings of DFT, nowadays more elaborated schemes are preferred due to better accuracy

Iterative solution of the Kohn-Sham equations basic iteration loop

- good choice of the *initial density* is important
- for an atom, one can start from the Thomas-Fermi density
- for a molecular or solid state system one can start from a sum of atomic densities

Simple argument about DFT E. Bright Wilson (1965)

- the total density determines the number of electrons

$$N = \int \rho(\mathbf{r}) d\mathbf{r}$$

- at the nucleus the Coulomb potential is singular and the wavefunction is finite but *non-analytic*
- the position of *cusps* in the density determines the position of the nuclei
- the *nuclear cusp condition* (Kato 1957, Steiner 1963) determines the nuclear charges

$$\left. \frac{\partial}{\partial r_A} \bar{\rho}(r_A) \right|_{r_A=0} = -2 Z_A \bar{\rho}(0)$$

where $\bar{\rho}(r_A)$ is angularly averaged density around nucleus A

- the nuclear charges determine *the whole Hamiltonian*, because kinetic energy and electron-electron repulsion depend only on the number of electrons N
- density itself determines *all the properties*

Atoms [1]

- in atoms we assume both the density and the KS potential to be *spherically symmetric* (always true in a closed shell system)
- in open shell atoms, we assume all degenerate orbitals to be equally filled
- Hartree potential

$$v_{Hartree}(r) = \frac{4\pi}{r} \int_0^r dr' r'^2 n(r') + 4\pi \int_r^\infty dr' r' n(r')$$

- KS wave functions have simple separable form with traditional atomic quantum numbers n, m, l

$$\phi_i(\mathbf{r}) = R_{nl}(r)Y_{lm}(\theta, \Phi)$$

- KS equation becomes a simple 1D equation

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} - \frac{1}{r} \frac{d}{dr} + \frac{l(l+1)}{2r^2} + v_{KS}(r) \right] R_{nl}(r) = \epsilon_{nl} R_{nl}(r)$$

Atoms

- 2nd order equation can be transformed in two coupled 1st order equations ($f_{nl}(r) \equiv R_{nl}(r)$)

$$\begin{aligned} \frac{df_{nl}(r)}{dr} &= g_{nl}(r) \\ \frac{dg_{nl}(r)}{dr} + \frac{2}{r}g_{nl}(r) - \frac{l(l+1)}{2r^2}f_{nl}(r) + 2[\epsilon_{nl} - v_{KS}(r)]f_{nl}(r) &= 0 \end{aligned}$$

- asymptotic behaviour for $r \rightarrow \infty$ (we have $v_{KS}(r) \rightarrow 0$)

$$\begin{aligned} \frac{df_{nl}(r)}{dr} &= g_{nl}(r) \\ \frac{dg_{nl}(r)}{dr} + 2\epsilon_{nl}f_{nl}(r) &\simeq 0 \\ f_{nl}(r) &\rightarrow \exp(-\sqrt{-2\epsilon_{nl}}r) \\ g_{nl}(r) &\rightarrow -\sqrt{-2\epsilon_{nl}}f_{nl}(r) \end{aligned}$$

Atoms

- at the origin $r \rightarrow 0$ the solutions have form

$$\begin{aligned} f_{nl}(r) &\rightarrow Ar^\alpha \\ g_{nl}(r) &\rightarrow Br^\beta \end{aligned}$$

where $B = lA$, $\alpha = l$, $\beta = l - 1$

- for fixed ϵ_{nl} and A one can integrate the 1st order equations starting from the above initial condition
- if ϵ_{nl} is not an eigenvalue, the solution will not obey the boundary condition at $r \rightarrow \infty$ (will diverge)
- eigenvalues ϵ_{nl} can be found by integrating the equations *forward* (from $r = 0$) and *backward* (from “practical” ∞) to some intermediate point and matching the two solutions
- *no diagonalization is necessary*
- important to generate *pseudopotentials* (see later)

4 Plane-wave pseudopotential method

Literature

- W. E. Pickett, *Pseudopotential methods in Condensed matter applications*, Comp. Phys. Rep. **9**, 115 (1989)
- M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, J. D. Joannopoulos, *Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients*, Rev. Mod. Phys. **64**, 1045 (1992).
- M. Fuchs and M. Scheffler, *Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory*, Comp. Phys. Comm. **119**, 67-98 (1999)

Condensed matter calculations Choice of a basis set

- to solve the KS equations numerically we need to choose a *basis set* f_ν (simple analytic functions) to expand the orbitals

$$\Psi_i(\mathbf{r}) = \sum_{\nu} c_{i\nu} f_{\nu}(\mathbf{r}, \{\mathbf{R}\})$$

- various choices are possible (accuracy vs. efficiency)
- in quantum chemistry, *localized* basis is often used (Gaussian type orbitals centered on the nuclei)
- if PBC are applied, a completely *delocalized* basis set is provided by *plane waves*

$$f_{\mathbf{G}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \exp(i\mathbf{G} \cdot \mathbf{r}), \quad \Omega \text{ is the volume of the periodic cell}$$

- other possibilities:
 - combinations of localized and originless sets
 - wavelets - localized in both real and reciprocal space
 - real space grids - potential energy operator is diagonal and derivatives are approximated by finite difference formulas

Plane waves

- the plane-wave basis set is complete and orthonormal
- plane waves do not depend on the positions of the nuclei - *originless* basis
- derivatives become just multiplications in the \mathbf{G} space
- operators can be evaluated in the space where there are diagonal and Fast Fourier transform can be efficiently used
- unbiased, all regions within the periodic cell are treated on equal footing (can be an advantage as well as disadvantage)
- problem - a large number of PW is needed to achieve a high spatial resolution in the region close to the nuclei
- solution - *pseudopotentials*
- *plane wave scheme with pseudopotentials* is considered a “standard model” for the total energy calculations

Supercell

- perfect solids have natural periodicity determined by the *unit cell*
- also non-periodic systems (defective solids, molecules, clusters, surfaces, liquids etc.) can be studied by using PBC - *supercell*

Bloch's theorem

- Kohn-Sham potential itself has the periodicity of the cell
- electronic wavefunction in a periodic system

$$\phi_{\mathbf{k},n}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r})u_{\mathbf{k},n}(\mathbf{r})$$

- \mathbf{k} is a vector in the *first Brillouin zone*
- $u_{\mathbf{k},n}$ is a cell-periodic function that can be expanded in plane waves

$$u_{\mathbf{k},n}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k},n}(\mathbf{G}) \exp(i\mathbf{G} \cdot \mathbf{r})$$

- \mathbf{G} are reciprocal lattice vectors, $\mathbf{G} \cdot \mathbf{l} = 2\pi m$, \mathbf{l} is lattice vector of the supercell, m is an integer

$$\phi_{\mathbf{k},n}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k},n}(\mathbf{G}) \exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}]$$

- KS energies are $\epsilon_{\mathbf{k},n}$, n is the *band index*

k-point sampling

- density of \mathbf{k} -points is proportional to the volume of the supercell
- for an infinite system, we would calculate finite number of electronic wave functions at an *infinite* number of \mathbf{k} -points
- wave functions at close \mathbf{k} -points are very similar
- special methods exist to approximate a *filled* electronic band by calculation of states at special set of \mathbf{k} -points in the Brillouin zone (e.g. Monkhorst and Pack 1976)

$$\int d\mathbf{k} \rightarrow \sum_{\mathbf{k}} w_{\mathbf{k}}$$

where $w_{\mathbf{k}}$ are weights of the integration points

- good for insulators and semiconductors, good accuracy can be reached with a very small set of \mathbf{k} -points
- for metals, a dense set of \mathbf{k} -points is needed to define the *Fermi surface*
- the density of \mathbf{k} -points can be increased until convergence of the total energy is reached
- for a large supercell, only the Γ point is often taken

Plane-wave basis sets

- kinetic energy of a plane wave is $\frac{1}{2}|\mathbf{k} + \mathbf{G}|^2$
- we introduce a cutoff energy E_{cutoff} and keep only plane waves with $\frac{1}{2}|\mathbf{k} + \mathbf{G}|^2 < E_{cutoff}$
- we obtain a *finite basis set* with $N_{PW} \sim \frac{\sqrt{2}}{3\pi^2} \Omega E_{cutoff}^{\frac{3}{2}}$ plane waves
- the cutoff introduces an error to the total energy calculation
- cutoff E_{cutoff} can be increased until convergence is reached
- both the Brillouin zone sampling and the cutoff have to be fine tuned in any calculation
- electron density

$$\rho(\mathbf{r}) = \sum_{\mathbf{k},n} \sum_{\mathbf{G},\mathbf{G}'} f(\epsilon_{\mathbf{k},n}) c_{\mathbf{k},n}^*(\mathbf{G}') c_{\mathbf{k},n}(\mathbf{G}) \exp[i(\mathbf{G} - \mathbf{G}') \cdot \mathbf{r}]$$

$$\rho(\mathbf{G}) = \sum_{\mathbf{k},n} \sum_{\mathbf{G}'} f(\epsilon_{\mathbf{k},n}) c_{\mathbf{k},n}^*(\mathbf{G}' - \mathbf{G}) c_{\mathbf{k},n}(\mathbf{G}')$$

- sum over \mathbf{G} in the expansion of the density extends over double the range of the wavefunction expansion

Pseudopotentials (Motivation)

- in the region close to the nuclei, wavefunctions oscillate rapidly over short length scale, because the wavefunctions of different orbitals are orthogonal
- huge number of plane waves would be needed to perform all-electron calculations (to describe the core orbitals)

- most physical and chemical properties of condensed matter systems are mainly dependent on the *valence electrons* and to much lesser extent on the *core electrons*
- *pseudopotential approximation* (originally proposed by Fermi 1934, Philips 1958, Heine and Cohen 1970, etc.) - core electrons are removed
- core electrons and the strong ionic potential V_{AE} are replaced by a *weaker and smooth* pseudopotential V_{PS}
- PP acts on the pseudo wavefunctions $|\Phi_{PS}\rangle$ rather than on true valence wavefunctions $|\Psi_{AE}\rangle$
- Schrödinger equation for all electrons is replaced by equation for valence electrons only

$$\begin{aligned}(T + V_{AE})|\Psi_{AE}\rangle &= \epsilon|\Psi_{AE}\rangle \\ (T + V_{PS})|\Phi_{PS}\rangle &= \epsilon|\Phi_{PS}\rangle\end{aligned}$$

Pseudopotentials

- pseudo wavefunctions approach the true valence wavefunctions in the region outside the core radius r_c
- inside the core radius the nodal structure of the true valence wavefunction is replaced by a smooth function
- this allows a small plane wave cutoff
- pseudopotential should be constructed to be *transferable* - usable in different chemical environments
- $|\psi_c\rangle, |\psi_v\rangle$ are exact solutions of the Schrödinger equation for the core and valence electrons

$$\hat{H}|\psi_c\rangle = E_c|\psi_c\rangle, \hat{H}|\psi_v\rangle = E_v|\psi_v\rangle$$

- valence orbitals are a sum of *smooth pseudo wavefunction* $|\phi_v\rangle$ and an oscillating part resulting from orthogonalization to the core orbitals

$$\begin{aligned}|\psi_v\rangle &= |\phi_v\rangle + \sum_c \alpha_{cv}|\psi_c\rangle \\ \langle\psi_c|\psi_v\rangle &= 0 \Rightarrow \alpha_{cv} = -\langle\psi_c|\phi_v\rangle \\ \hat{H}|\phi_v\rangle &= E_v|\phi_v\rangle + \sum_c (E_c - E_v)|\psi_c\rangle\langle\psi_c|\phi_v\rangle\end{aligned}$$

- $|\phi_v\rangle$ satisfies a Schrödinger-like equation with an *energy-dependent* pseudo Hamiltonian

$$\hat{H}^{PK}(E) = \hat{H} - \sum_c (E_c - E) |\psi_c\rangle \langle \psi_c|$$

Pseudopotentials

$$\hat{w}^{PK}(E) = \hat{v} - \sum_c (E_c - E) |\psi_c\rangle \langle \psi_c|$$

- \hat{v} is the bare potential and $\hat{w}^{PK}(E)$ is effective potential for the valence electrons
- at radius where the core states decay to zero, $\hat{w} \rightarrow v$
- additional advantages of using pseudopotentials:
- fewer electronic states have to be calculated
- the total energy is a fraction of the true total energy, so much lower accuracy is sufficient to get the *energy differences* between different configurations
- however, the total energy has no physical meaning, only differences are meaningful
- originally, pseudopotentials were constructed *empirically* by fitting an analytic form to experimental data (band structure)

Ab-initio pseudopotentials

- much more realistic pseudopotentials can be generated *ab initio*
- in early DFT studies, they were developed by fitting parametrized functional forms to reproduce the atomic eigenvalues and orbital shapes
- general form of the angular momentum dependent pseudopotential

$$\hat{W} = \sum_{lm} |lm\rangle V_l \langle lm|$$

- wavefunction is projected on the angular momentum eigenfunctions and each component is multiplied by the relevant pseudopotential V_l

- the projector is *non-local*

$$\langle \mathbf{r} | lm \rangle \langle lm | \mathbf{r}' \rangle = \langle \Omega | lm \rangle \langle lm | \Omega' \rangle \delta(|\mathbf{r}| - |\mathbf{r}'|)$$

- taking V_l as a local function $V_l = w_l(r)$ we get the PP in the (most common) *semi-local* form (Y_{lm} are spherical harmonics)

$$w(\mathbf{r}, \mathbf{r}') = \sum_l \sum_{m=-l}^l Y_{lm}^*(\hat{\mathbf{r}}) w_l(r, r') Y_{lm}(\hat{\mathbf{r}}')$$

$$w_l(r, r') = w_l(r) \delta(r - r')$$

Norm conserving pseudopotentials Hamann, Schlüter, Chiang (1979) [2]

- the exchange-correlation energy depends on the electron density
- outside the core radius r_c the pseudo wavefunction should produce the *identical density* as the real wavefunction
- the two wavefunctions should be identical in *absolute magnitude*, not just in spatial dependence
- all-electron calculation is performed for an isolated atom
- inversion of the Schrödinger equation is performed to find the PP
- PP found in this way is used for any environment of the atom
- when the PP is used, *the same XC functional* has to be used that was used to generate the PP
- Bachelet, Hamann, Schlüter (1982) [3] published a table of pseudopotentials for all elements from $Z = 1$ to $Z = 94$

Generation of a norm conserving pseudopotential

- integral of the pseudo and real wavefunction inside r_c should be equal so that the Coulomb potential generated outside r_c is identical

$$R_l^{PP}(r) = R_{nl}^{AE}(r) \text{ if } r > r_l$$

$$\int_0^{r_l} dr r^2 |R_l^{PP}(r)|^2 = \int_0^{r_l} dr r^2 |R_{nl}^{AE}(r)|^2 \text{ if } r < r_l$$

$$\epsilon_l^{PP} = \epsilon_{nl}^{AE}$$

- pseudowavefunctions should have no nodal surfaces
- cutoff radii r_l depend on l , cannot be smaller than the outermost nodal surface of the true wavefunction
- r_l are not adjustable parameters
- with smaller r_l the PP becomes more accurate and transferable, but also stronger
- the choice of r_l is a balance between size of the plane-wave basis set and accuracy

Generation of a norm conserving pseudopotential Procedure

1. the free atom KS radial equations are solved

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + v_{KS}^{AE}[n^{AE}](r) \right] r R_{nl}^{AE}(r) = \epsilon_{nl}^{AE} r R_{nl}^{AE}(r)$$

$$v_{KS}^{AE}[n^{AE}](r) = -\frac{Z}{r} + v_{Hartree}[n^{AE}](r) + v_{XC}[n^{AE}](r)$$

2. pseudo wavefunction is defined using norm conservation conditions, the shape of the wavefunction for $r < r_l$ depends on the kind of PP and has to be defined
3. PP is determined by *inverting the radial KS equation*

$$w_{l,scr}(r) = \epsilon_l^{PP} - \frac{l(l+1)}{2r^2} + \frac{1}{2r R_l^{PP}(r)} \frac{d^2}{dr^2} [r R_l^{PP}(r)]$$

4. *unscreening* of the PP - subtraction of the screening effects due to the valence electrons

$$w_l(r) = w_{l,scr}(r) - v_{Hartree}[n^{PP}](r) - v_{XC}[n^{PP}](r)$$

Non-linear core correction Louie, Froyen, Cohen (1982) [4]

- in the unscreening we actually assumed a *linear dependence* of the v_{XC} on the density
- this is only true if the frozen core electrons and the valence states *do not overlap*

- if overlap is present, linearization will lead to poor transferability because v_{XC} is a non-linear function

- solution: the unscreened PP is redefined as

$$w_l(r) = w_{l,scr}(r) - v_{Hartree}[n^{PP}](r) - v_{XC}[\tilde{n}^{core} + n^{PP}](r)$$

and the *partial core density* $\tilde{n}^{core}(r)$ is supplied together with the PP

- in the calculation, we use

$$E_{XC} = E_{XC}[\tilde{n}^{core} + n]$$

$$\text{and } V_{XC} = V_{XC}[\tilde{n}^{core} + n]$$

Non-linear core correction

- $\tilde{n}^{core}(r)$ is equal to the core density of the atomic reference state in the region of overlap with valence states
- in order to eliminate the *rugged core density*, $\tilde{n}^{core}(r)$ joins the true core density at the core cutoff radius r_{nlc} and in the core region equals a smooth polynomial $P(r)$ going to zero at $r = 0$

$$\tilde{n}^{core}(r) = \begin{cases} n^{core}(r) & \text{for } r \geq r_{nlc} \\ P(r) & \text{for } r < r_{nlc} \end{cases}$$

$$\int_0^{r_{nlc}} dr r^2 \tilde{n}^{core}(r) < \int_0^{r_{nlc}} dr r^2 n^{core}(r)$$

- r_{nlc} defines the region of overlap, typically is chosen at the point where the true core density becomes smaller than the atomic valence density
- important for the alkali and transition metals where core orbitals extend into valence density (Zn, Cd)
- the NLC correction substantially increases the *transferability* of the PP

Troullier-Martins PP (Troullier, Martins 1991) [5]

- smooth PP, pseudo wave-functions are defined as

$$R_l^{PP}(r) = \begin{cases} R_{nl}^{AE}(r) & \text{for } r > r_l \\ r^l \exp[p(r)] & \text{for } r < r_l \end{cases}$$

$$p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}$$

- coefficients c_i are imposed by
 - norm conservation
 - continuity of pseudo wavefunctions and their first four derivatives at $r = r_l$
 - zero curvature of the screened PP at the origin

Kleinman-Bylander separable form of PP Kleinman, Bylander (1982) [6]

- a semi-local pseudopotential can be rewritten in a *separable form*

$$w_l(r, r') = v_l(r)v_l(r')$$

- the real-space expression becomes more complicated (truly non-local) but the Fourier space representation is simpler
- separation of short and long-range part (Coulomb), long-range is local

$$\Delta w_l = w_l(r) - w_{local}(r)$$

- $w_{local}(r)$ is chosen so that the rest becomes as small as possible
- typically the most repulsive PP component is chosen as $w_{local}(r)$

$$w(\mathbf{r}, \mathbf{r}') = w_{local}(r) + \sum_l \Delta w_l(r) \sum_{m=-l}^l Y_{lm}^*(\hat{\mathbf{r}}') Y_{lm}(\hat{\mathbf{r}}) \delta(r - r')$$

$$w^{KB}(\mathbf{r}, \mathbf{r}') = w_{local}(r) + \sum_l \sum_{m=-l}^l \frac{\phi_{lm}(\mathbf{r}) \Delta w_l(\mathbf{r}) \Delta w_l(\mathbf{r}') \phi_{lm}(\mathbf{r}')}{\int d\mathbf{r} \Delta w_l(\mathbf{r}) |\phi_{lm}(\mathbf{r})|^2}$$

Kleinman-Bylander separable form of PP

- the action of $w^{KB}(\mathbf{r}, \mathbf{r}')$ on the *valence pseudo wavefunctions* $\phi_{lm}(\mathbf{r})$ is identical to that of the original semi-local one

$$\int d\mathbf{r}' w^{KB}(\mathbf{r}, \mathbf{r}') \phi_{lm}(\mathbf{r}') = w_l(r) \phi_{lm}(\mathbf{r})$$

- nevertheless $w(\mathbf{r}, \mathbf{r}')$ and $w^{KB}(\mathbf{r}, \mathbf{r}')$ are two different potentials
- the construction illustrates another dimension of non-uniqueness of PP

- when applied to molecular or solid state system, the two forms do not produce identical results (both are equally valid)
- if the PP approximation is valid, the results must be very similar
- KB is the most used form in DFT codes
- “ghost” states problem can arise with a bad choice of the $w_{local}(r)$

Quality of the PP approximation - transferability

- the *logarithmic derivative* $\frac{d}{dr} \ln \phi$ (ratio slope-value) plays a role of the boundary condition for the wavefunction at the core cutoff radius r_c
- by construction it is adjusted to have the same value in the all-electron and pseudo wavefunctions *at the atomic eigenvalues*
- in molecular or condensed matter system, the eigenvalues are shifted from the atomic ones by *bonding or banding*
- the *norm conservation is related to the logarithmic derivative* of the wave function

$$-\frac{1}{2}(r\phi_\epsilon)^2 \frac{d}{d\epsilon} \left(\frac{d}{dr} \ln \phi_\epsilon \right) = \int_0^R \phi_\epsilon^2 r^2 dr$$

where ϕ_ϵ is a solution of the Schrödinger equation for energy ϵ (not necessarily an eigenvalue)

- the *first energy derivatives* of the logarithmic derivative in the all-electron and pseudo wavefunctions also agree
- if the norm-conserving PP reproduces the *scattering properties* of the all-electron potential in the neighborhood of the atomic eigenvalues, it does so also in the *entire range of valence bands or molecular orbitals*

Quality of the PP approximation - transferability

- norm-conserving PP provide good transferability
- *quality and computational efficiency are conflicting requirements* but can be reasonably well satisfied by modern norm-conserving PP
- quality of a PP depends on

- correct scattering properties
 - choice of the core cutoff radius r_c , smaller r_c are good for accuracy but bad for efficiency (strong PP)
 - adequate account of the non-linear XC interaction between core and valence electrons
 - quality of the KB form (avoiding the spurious states)
 - validity of the frozen core approximation
- in any case testing in different environments has to be performed
 - other common type of PP is the Vanderbilt *ultrasoft PP* (Vanderbilt 1990) [7] which partially relaxes the requirement of the norm conservation in favour of generating softer PP (useful for oxygen)

Plane-wave representation of the KS equations

- kinetic energy

$$T = \sum_{i=1}^N \langle \phi_i | -\frac{1}{2} \nabla_{\mathbf{r}}^2 | \phi_i \rangle = \frac{1}{2} \sum_{\mathbf{k}, n} \sum_{\mathbf{G}} f(\epsilon_{\mathbf{k}, n}) |c_{\mathbf{k}, n}(\mathbf{G})|^2 |\mathbf{k} + \mathbf{G}|^2$$

- potentials are represented in *the Fourier space*
- Hartree energy

$$E_{Hartree} = \frac{1}{2} \int \rho(\mathbf{r}) v_{Hartree}(\mathbf{r}) d\mathbf{r} = \frac{\Omega}{2} \sum_{\mathbf{G}} v_{Hartree}(\mathbf{G}) \rho(\mathbf{G})$$

$$v_{Hartree}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'$$

- Ω is the volume of the cell
- Poisson's equation

$$v_{Hartree}(\mathbf{G}) = 4\pi \frac{\rho(\mathbf{G})}{\mathbf{G}^2}$$

- $v_{Hartree}(\mathbf{G})$ diverges at $\mathbf{G} = 0$ (positive contribution, interaction between like charges)

Plane-wave representation of the KS equations

- we consider the external potential in a general *non-local* form $w(\mathbf{r}, \mathbf{r}')$ (*pseudopotential*)

$$E_{ei} = \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} \rightarrow \sum_{i=1}^N \int d\mathbf{r} \int d\mathbf{r}' \phi_i(\mathbf{r})w(\mathbf{r}, \mathbf{r}')\phi_i^*(\mathbf{r}')$$

- local potential is a special case $w(\mathbf{r}, \mathbf{r}') = \tilde{w}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')$

$$E_{ei} = \sum_{\mathbf{k}, n} \sum_{\mathbf{G}} \sum_{\mathbf{G}'} f(\epsilon_{\mathbf{k}, n}) c_{\mathbf{k}, n}^*(\mathbf{G}) c_{\mathbf{k}, n}(\mathbf{G}') w(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}')$$

$$w(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}') = \int d\mathbf{r} \int d\mathbf{r}' w(\mathbf{r}, \mathbf{r}') \exp[-i(\mathbf{k} + \mathbf{G})\mathbf{r}] \exp[i(\mathbf{k} + \mathbf{G}')\mathbf{r}']$$

$$w(\mathbf{r}, \mathbf{r}') = \sum_{j, \alpha} w_{\alpha}(\mathbf{r} - \boldsymbol{\rho}_j - \mathbf{R}_{\alpha}, \mathbf{r}' - \boldsymbol{\rho}_j - \mathbf{R}_{\alpha})$$

$$w(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}') = \sum_{\alpha} \exp[-i(\mathbf{G} - \mathbf{G}')\mathbf{R}_{\alpha}] w_{\alpha}(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}')$$

where $\boldsymbol{\rho}_j$ are supercell lattice vectors

- $w_{\alpha}(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}')$ is the Fourier transform of the ionic pseudopotential

Evaluation of $w_{\alpha}(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}')$

- calculation of this matrix element for a general non-local or semilocal PP $w_{\alpha}(\mathbf{r}, \mathbf{r}')$ scales as N_{PW}^2 which is prohibitively expensive

$$w_{\alpha}(\mathbf{k}, \mathbf{k}') = \int d\mathbf{r} \int d\mathbf{r}' w_{\alpha}(\mathbf{r}, \mathbf{r}') \exp[-i\mathbf{k}\mathbf{r}] \exp[i\mathbf{k}'\mathbf{r}']$$

- with PP in separable form (Kleinman-Bylander) it becomes much easier

$$w_{\alpha}(\mathbf{r}, \mathbf{r}') = w_{\alpha}^{local}(\mathbf{r})\delta(\mathbf{r}, \mathbf{r}') + \sum_l \sum_{m=-l}^l \gamma_{lm} \phi_{lm}(\mathbf{r})\Delta w_l(\mathbf{r})\Delta w_l(\mathbf{r}')\phi_{lm}(\mathbf{r}')$$

$$w_{\alpha}(\mathbf{k}, \mathbf{k}') = w_{\alpha}^{local}(\mathbf{k} - \mathbf{k}') + \sum_l \sum_{m=-l}^l \gamma_{lm} \left[\int d\mathbf{r} \exp[-i\mathbf{k}\mathbf{r}] \phi_{lm}(\mathbf{r})\Delta w_l(\mathbf{r}) \right] \left[\int d\mathbf{r}' \exp[i\mathbf{k}'\mathbf{r}'] \phi_{lm}(\mathbf{r}')\Delta w_l(\mathbf{r}') \right]$$

- this procedure scales as N_{PW}

Electrostatics

- the local part of pseudopotential contains the tail of the ionic Coulomb potential $\sim -\frac{1}{r}$
- the electron-ion energy

$$E_{ei}(\mathbf{G}, \mathbf{G}') = \sum_{\mathbf{k}, n} f(\epsilon_{\mathbf{k}, n}) c_{\mathbf{k}, n}^*(\mathbf{G}) c_{\mathbf{k}, n}(\mathbf{G}') w(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}')$$

diverges at $\mathbf{G}, \mathbf{G}' \rightarrow 0$ (negative contribution, interaction between unlike charges)

- ion-ion Coulomb energy

$$E_{nn} = \frac{1}{2} \sum_{\alpha\alpha'jj'}' \frac{Z_\alpha Z_{\alpha'}}{|\boldsymbol{\rho}_j + \mathbf{R}_\alpha - \boldsymbol{\rho}_{j'} - \mathbf{R}_{\alpha'}|}$$

also diverges (positive contribution, interaction between like charges)

- the sum is handled by Ewald summation and the divergencies of $E_{Hartree}$, E_{ei} and E_{nn} at $\mathbf{G} \rightarrow 0$ cancel each other, leaving a *finite contribution* (the total system electrons + ions is *charge neutral*)

Electrostatics

$$\begin{aligned} & \lim_{\mathbf{G}, \mathbf{G}' \rightarrow 0} \left[\sum_{\mathbf{k}, n} f(\epsilon_{\mathbf{k}, n}) c_{\mathbf{k}, n}^*(\mathbf{G}) c_{\mathbf{k}, n}(\mathbf{G}') w(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}') + \frac{\Omega}{2} v_{Hartree}(\mathbf{G}) \rho(\mathbf{G}) \right] \\ & + E_{nn} = E_{rep} + E_{Ewald} \\ & E_{rep} = Z_{total} \frac{1}{\Omega} \sum_{\alpha} \Lambda_{\alpha} \\ & \Lambda_{\alpha} = \frac{1}{\Omega} \int d\mathbf{r} \left[w_{\alpha, local}(r) + \frac{Z_{\alpha}}{r} \right] \\ & E_{Ewald} = \frac{1}{2} \sum_{\alpha, \alpha'} Z_{\alpha} \Gamma_{\alpha, \alpha'} Z_{\alpha'} \\ & \Gamma_{\alpha, \alpha'} = \frac{4\pi}{\Omega} \sum_{\mathbf{G} \neq 0} \frac{\cos[\mathbf{G} \cdot (\mathbf{R}_{\alpha} - \mathbf{R}_{\alpha'})]}{G^2} \exp\left(-\frac{G^2}{4\eta^2}\right) + \sum_j \frac{\text{erfc}(\eta|\boldsymbol{\rho}_j + \mathbf{R}_{\alpha} - \mathbf{R}_{\alpha'}|)}{|\boldsymbol{\rho}_j + \mathbf{R}_{\alpha} - \mathbf{R}_{\alpha'}|} \\ & - \frac{\pi}{\eta^2 \Omega} - \frac{2\eta}{\sqrt{\pi}} \delta_{\alpha, \alpha'} \end{aligned}$$

Total energy

- the Ewald term does not depend on the density and is evaluated only once at the beginning of each iteration cycle
- the total energy is

$$E_{tot} = T + E'_{Hartree} + E'_{ei} + E_{XC} + E_{Ewald} + E_{rep}$$

- the terms with $\mathbf{G}, \mathbf{G}' = 0$ are excluded from the Hartree and pseudopotential contributions
- exchange-correlation energy (LDA)

$$E_{XC} = \int \rho(\mathbf{r}) \epsilon_{xc}(\rho(\mathbf{r})) d\mathbf{r} = \Omega \sum_{\mathbf{G}} \epsilon_{xc}(\rho(\mathbf{G})) \rho(\mathbf{G})$$

- the XC functional $\epsilon_{xc}(\rho(\mathbf{r}))$ is local in real space so it has to be evaluated on a real space grid and $\epsilon_{xc}(\rho(\mathbf{G}))$ is obtained via Fourier transform

KS equations in the plane wave representation

- simple form

$$\sum_{\mathbf{G}'} H_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) c_{\mathbf{k},n}(\mathbf{G}') = \epsilon_{\mathbf{k},n} c_{\mathbf{k},n}(\mathbf{G})$$

$$H_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) = \frac{1}{2} |\mathbf{k} + \mathbf{G}|^2 \delta_{\mathbf{G},\mathbf{G}'} + w(\mathbf{k} + \mathbf{G}, \mathbf{k} + \mathbf{G}') + v_{Hartree}(\mathbf{G} - \mathbf{G}') + v_{XC}(\mathbf{G} - \mathbf{G}')$$

- Hamiltonian matrix for each \mathbf{k} point is constructed
- in principle, it can be solved by direct diagonalization and iteration until self-consistency
- the total energy should be converged both in cutoff and in \mathbf{k} -points, but in practice it is more important to *converge the energy differences* rather than the absolute energy
- diagonalization scales as P^3 where P is the dimension of the matrix
- typically we have ~ 100 plane waves per atom
- severe limitation to the size of the system
- *alternative techniques for minimizing the KS functional* have been developed

Iteration to self-consistency

- iteration to self-consistency is performed until a convergence criterion is reached
- convergence in energy $|E^{(i)} - E^{(i-1)}| < \eta_E$ or in density $\int d\mathbf{r} |\rho^{(i)} - \rho^{(i-1)}| < \eta_\rho$
- to start a the new iteration cycle, the new density is mixed with the previous density to avoid oscillations
- simplest version - linear mixing

$$\rho^{(i+1)} = \beta \rho' + (1 - \beta) \rho^i$$
$$\rho'(\mathbf{r}) = \sum_i^{\text{occ}} |\phi_i(\mathbf{r})|^2$$

- $\beta \approx 0.3$
- more sophisticated versions of mixing exist, using densities from several previous iterations (Broyden 1965) [8]

The band gap problem

- KS orbitals and energies *do not have a direct physical meaning*
- the plane waves KS method provides a *band structure*
- valence and conduction bands are typically reproduced reasonably well
- the difference between the bottom of the conduction band and the top of the valence band (*KS gap*) is in semiconductors typically about *two times smaller* than the experimental *band gap*

The band gap problem

- e.g. in Si the experimental indirect band gap is 1.17 eV and the LDA one is 0.48 eV (from Γ to 0.85X)
- this is not only due to inaccuracy of LDA, even the exact XC functional will not guarantee that the band gap has the correct value (Sham and Schlüter 1983) [9]

- *DFT is a ground state theory*
- experimentally, the electronic excitation spectrum of a solid is measured by photoemission experiments
- the excitation spectrum and the band gap are related to *quasiparticles*
- when self-energy correction (many-body perturbation theory) is taken into account, substantial improvement can be obtained (GW approximation, Hedin 1965) [10]

Example 3. DFT study of solid silicon phases (Yin and Cohen 1982) [11]

- Si has ground state electronic configuration $1s^2 2s^2 2p^6 3s^2 3p^2$
- reference configuration $3s^2 3p^{0.5} 3d^{0.5}$, PP of the Hamann-Schlüter-Chiang type
- for isolated atom the PP reproduces eigenvalues within few mRy and wavefunctions outside the core radius within 1 %
- *plane-wave basis is not biased towards any particular crystal structure*
- $E_{cutoff} = 11.5$ Ry

Example 4. DFT study of solid silicon phases (Yin and Cohen 1982)

- total energy was calculated for 7 crystal structures: fcc, bcc, hcp, cubic diamond (CD), hexagonal diamond (HD), β -tin, simple cubic (sc)
- for semiconducting CD (HD) phases, 10 (6) special k points in the irreducible BZ were sufficient to reach convergence
- other phases were found metallic, number of sampling k-points was 24, 35, 70, 36, 60 in the β -tin, sc, bcc, hcp and fcc, respectively
- with increasing E_{cutoff} the total energy decreases (more variational freedom) and *cohesive energy* $E_c = E_{crystal}/N_{atom} - E_{atom}$ increases

Example 5. DFT study of solid silicon phases (Yin and Cohen 1982)

- upon formation of a crystal E_{kin} grows because of localization of electrons when covalent bonds are formed
- at the same time E_{XC} and E_{pot} decreases
- the CD crystal structure is *stabilized due to both E_{XC} and E_{pot}*

- zero-point (quantum) phonon energy has to be taken into account at $T = 0$

Example 6. DFT study of solid silicon phases (Yin and Cohen 1982)

- equations of state calculated for various phases are fitted on the Murnaghan's equation

$$E_{tot}(V) = \frac{B_0 V}{B'_0} \left[\frac{(V_0/V)^{B'_0}}{B'_0 - 1} \right] + \text{const}$$

where B_0 and B'_0 are bulk modulus and its pressure derivative at the equilibrium volume V_0

- CD phase is slightly more stable than the HD phase
- at $p = 9.9$ GPa a transition from CD to β -tin structure is predicted (expt. at 12.5 GPa at room temperature), driven by the Ewald (ion-ion) energy

Example 7. DFT study of solid silicon phases (Yin and Cohen 1982)

- because of norm-conserving properties of the PP, the pseudovalence charge density reproduces well the true electronic density
- in order to compare with experiment (X-ray diffraction data) the core structure factors have to be added to the valence structure factors
- the agreement is very good
- phases in the order of *decreasing covalent bond character* : CD, HD, β -tin, sc, bcc, fcc, hcp

Small computational project - ab initio calculation of Si crystal

- for *Si crystal in the diamond phase*, calculate the equilibrium lattice constant a and bulk modulus B_0
- use the CPMD code
- use the LDA
- cell of 2x2x2 unit cells (64 atoms)
- Γ point only for BZ sampling
- Troullier-Martins pseudopotential
- $E_{cutoff} = 12$ Ry (check for convergence at one volume)
- calculate few points on the $T = 0$ equation of state $E(V)$
- compare results with other calculations and with exp. data

5 Exchange-correlation functionals (beyond LDA)

Exchange-correlation functionals and their accuracy

- E_{XC} is a relatively small fraction of the total energy of the system
- E_{XC} is about 100 % of the *chemical bonding energy*
- the exchange-correlation functional is the *crucial ingredient* of DFT scheme
- *empirical* and *non-empirical* functionals
- empirical ones include parameters fitted to some database of systems (molecules, solids, ...) and can be seen as *interpolation* schemes
- non-empirical ones try to incorporate as many *exact constraints* (known from theory) as possible - can be seen as *extrapolation* schemes

The pair correlation function

- diagonal element of the second order density matrix

$$\begin{aligned}\rho_2(\mathbf{r}_1, \mathbf{r}_2) &= \rho_2(\mathbf{r}_1\mathbf{r}_2, \mathbf{r}_1\mathbf{r}_2) \\ &= \frac{1}{2}N(N-1) \sum_{s_1 \dots s_n} \int \dots \int d\mathbf{r}_3 \dots d\mathbf{r}_N |\Psi(\mathbf{r}_1s_1, \mathbf{r}_2s_2, \mathbf{r}_3s_3, \dots, \mathbf{r}_Ns_N)|^2 \\ \rho(\mathbf{r}) &= \frac{2}{N-1} \int \rho_2(\mathbf{r}, \mathbf{r}') d\mathbf{r}'\end{aligned}$$

- electron-electron repulsion $\langle V_{ee} \rangle$

$$\langle V_{ee} \rangle = \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho_2(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$$

- we define the *pair-correlation function* $h(\mathbf{r}, \mathbf{r}')$ which includes *all non-classical effects*

$$\rho_2(\mathbf{r}, \mathbf{r}') = \frac{1}{2}\rho(\mathbf{r})\rho(\mathbf{r}') [1 + h(\mathbf{r}, \mathbf{r}')]$$

The exchange-correlation hole

- for every \mathbf{r} , the function $h(\mathbf{r}, \mathbf{r}')$ satisfies the sum rule

$$\int d\mathbf{r}' \rho(\mathbf{r}') h(\mathbf{r}, \mathbf{r}') = -1$$

- we define the *exchange-correlation hole* $\rho_{xc}(\mathbf{r}, \mathbf{r}')$ of an electron at \mathbf{r}

$$\rho_{xc}(\mathbf{r}, \mathbf{r}') = \rho(\mathbf{r}') h(\mathbf{r}, \mathbf{r}')$$

which satisfies the *sum rule*

$$\int d\mathbf{r}' \rho_{xc}(\mathbf{r}, \mathbf{r}') = -1$$

- exchange-correlation hole contains a unit charge with sign opposite to that of the electron
- the electron-electron repulsion $\langle V_{ee} \rangle$ becomes

$$\langle V_{ee} \rangle = J[\rho] + \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r}) \rho_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$$

The XC functional via the *exchange-correlation hole* Gunnarsson and Lundquist (1976) [12]

- XC energy functional

$$\begin{aligned} E_{xc}[\rho] &= T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \\ F[\rho] &= \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \hat{V}_{ee} | \Psi \rangle = \langle \Psi_{\rho}^{min} | \hat{T} + \hat{V}_{ee} | \Psi_{\rho}^{min} \rangle \end{aligned}$$

- in the Kohn-Sham non-interacting system there is no \hat{V}_{ee}

$$T_s[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} | \Psi \rangle = \langle \Phi_{\rho}^{min} | \hat{T} | \Phi_{\rho}^{min} \rangle$$

where $|\Phi_{\rho}^{min}\rangle$ is a Slater determinant

- we consider λ *coupling* between the *non-interacting* ($\lambda = 0$) and *normal (interacting, $\lambda = 1$)* system

$$F_{\lambda}[\rho] = \min_{\Psi \rightarrow \rho} \langle \Psi | \hat{T} + \lambda \hat{V}_{ee} | \Psi \rangle = \langle \Psi_{\rho}^{\lambda} | \hat{T} + \lambda \hat{V}_{ee} | \Psi_{\rho}^{\lambda} \rangle$$

where $|\Psi_{\rho}^{\lambda}\rangle$ minimizes $\langle \hat{T} + \lambda \hat{V}_{ee} \rangle$ and provides density $\rho(\mathbf{r})$

- $\rho(\mathbf{r})$ remains the same as λ changes

The XC functional via the *exchange-correlation hole*

$$\begin{aligned} F_1[\rho] &= F[\rho] = T[\rho] + V_{ee}[\rho] \\ F_0[\rho] &= T_s[\rho] \end{aligned}$$

- we assume a smooth *adiabatic connection* between $\lambda = 0$ and $\lambda = 1$

$$E_{xc}[\rho] = F_1[\rho] - F_0[\rho] - J[\rho] = \int_0^1 d\lambda \frac{\partial F_\lambda[\rho]}{\partial \lambda} - J[\rho]$$

- we apply the Hellman - Feynman theorem and get

$$E_{xc}[\rho] = \int_0^1 d\lambda \langle \Psi_\rho^\lambda | \hat{V}_{ee} | \Psi_\rho^\lambda \rangle - J[\rho]$$

- only $\lambda = 1$ is physical, all other values are fictitious

$$E_{xc}[\rho] = \int_0^1 d\lambda \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho_2^\lambda(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - J[\rho]$$

The XC functional via the *exchange-correlation hole*

$$E_{xc}[\rho] = \int d\mathbf{r} \int d\mathbf{r}' \frac{\bar{\rho}_2(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - J[\rho] \text{ where}$$

$$\bar{\rho}_2(\mathbf{r}, \mathbf{r}') = \int_0^1 d\lambda \rho_2^\lambda(\mathbf{r}, \mathbf{r}') = \frac{1}{2} \rho(\mathbf{r}) \rho(\mathbf{r}') [1 + \bar{h}(\mathbf{r}, \mathbf{r}')]]$$

$$E_{xc}[\rho] = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r}) \rho(\mathbf{r}') \bar{h}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r}) \bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$$

where $\bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}') = \rho(\mathbf{r}') \bar{h}(\mathbf{r}, \mathbf{r}')$ is the *average exchange-correlation hole*

- $E_{xc}[\rho]$ is Coulomb interaction between each electron and the average exchange-correlation hole around it
- three contributions:
 1. *self-interaction correction* (electron does not interact with itself)
 2. the *Pauli exclusion principle* which keeps 2 electrons with parallel spin apart in space
 3. the *Coulomb repulsion* which which keeps any 2 electrons apart in space
- effects (1) and (2) yield *exchange energy* (present also at $\lambda = 0$) while effect (3) yields the *correlation energy* (arises only at $\lambda \neq 0$)

The XC functional via the *exchange-correlation hole*

Two important consequences:

- the sum rule

$$\int d\mathbf{r}' \bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}' \rho(\mathbf{r}') \bar{h}(\mathbf{r}, \mathbf{r}') = -1$$

is a local condition providing stringent test on DFT functionals where an approximate $\bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}')$ is used

- $E_{xc}[\rho]$ depends only on the *spherically averaged behaviour* of $\bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}')$

$$E_{xc}[\rho] = \frac{1}{2} \int d\mathbf{r} \rho(\mathbf{r}) \int_0^\infty 4\pi s ds \rho_{xc}^{SA}(\mathbf{r}, s)$$

where $\rho_{xc}^{SA}(\mathbf{r}, s) = \frac{1}{4\pi} \int_{\Omega:|\mathbf{r}-\mathbf{r}'|=s} \bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}') d\mathbf{r}'$

is spherically averaged exchange-correlation hole

- $\bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}')$ can be further decomposed into exchange and correlation contributions

Why LDA works

- $E_{xc}[\rho]$ depends only on the *system, spherically and coupling-constant averaged* XC hole

$$E_{xc}[\rho] = \frac{N}{2} \int_0^\infty ds 4\pi s^2 \frac{\langle \rho_{xc}^{SA}(\mathbf{r}, s) \rangle}{s}$$

where $\langle \rho_{xc}^{SA}(\mathbf{r}, s) \rangle = \frac{1}{N} \int d\mathbf{r} \rho(\mathbf{r}) \frac{1}{4\pi} \int_{\Omega:|\mathbf{r}-\mathbf{r}'|=s} \bar{\rho}_{xc}(\mathbf{r}, \mathbf{r}') d\mathbf{r}'$

- the XC hole in LDA is that of a *possible physical system* (homogeneous electron gas) and satisfies several important sum rules
- the LDA model for $\langle \rho_{xc}^{SA}(\mathbf{r}, s) \rangle$

$$\langle \rho_{xc}^{SA}(\mathbf{r}, s) \rangle = \frac{1}{N} \int d\mathbf{r} \rho(\mathbf{r}) \rho_{xc}^{unif}(\rho(\mathbf{r}), s)$$

can be quite realistic

- the system average *unweights* the regions where LDA is less reliable
- *cancellation of errors* between the exchange and correlation energy

Properties of LDA

- LSDA is exact for the homogeneous electron gas and good approximation in the long-wavelength limit
- LSDA is reliable moderate-accuracy approximation (reproduces chemical trends)
- LSDA is sufficient for many purposes in solid-state
- *cohesive energies* in solids are systematically *overestimated*
- *lattice constants* are systematically *underestimated* (by few %)
- LDA fails to describe hydrogen bonds

Local spin-density approximation (LSDA)

- local density approximation

$$E_{xc}^{LDA}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}(\rho(\mathbf{r}))$$

where $\epsilon_{xc}(\rho(\mathbf{r}))$ corresponds to homogeneous *spin-compensated* electron gas ($\rho(\mathbf{r}) = \rho_{\uparrow}(\mathbf{r}) + \rho_{\downarrow}(\mathbf{r})$)

- local spin-density approximation LSDA is a generalization

$$E_{xc}^{LSDA}[\rho_{\uparrow}, \rho_{\downarrow}] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}[\rho_{\uparrow}(\mathbf{r}), \rho_{\downarrow}(\mathbf{r})]$$

where $\epsilon_{xc}(\rho_{\uparrow}, \rho_{\downarrow})$ corresponds to homogeneous *spin-polarized* electron gas

- necessary when external magnetic field is present
- also in absence of magnetic field it is better, e.g. in systems with *unpaired electrons* (open-shell molecules)
- various parametrizations exist

Beyond LDA - generalized gradient approximation (GGA) functionals

- ideally, all good properties of LSDA should be preserved and bad ones should be improved

$$E_{xc}^{GGA}[\rho_{\uparrow}, \rho_{\downarrow}] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}^{GGA}[\rho_{\uparrow}(\mathbf{r}), \rho_{\downarrow}(\mathbf{r}), \nabla\rho_{\uparrow}(\mathbf{r}), \nabla\rho_{\downarrow}(\mathbf{r})]$$

- GGA includes a gradient correction to the XC hole near the electron
- while the construction of LSDA is in principle unique, it is not the case with GGA
- BLYP (empirical) - the combination of exchange functional (Becke 1988) [13] with the correlation functional (Lee, Yang, and Parr 1988) [14]
- PBE non-empirical functional (Perdew, Burke, and Ernzerhof 1996) [15]
- GGA corrects LDA and sometimes *overcorrects*

Further developments

- *meta GGA* functionals include additional ingredients

$$E_{xc}^{MGGA}[\rho_{\uparrow}, \rho_{\downarrow}] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}^{MGGA}[\rho_{\uparrow}(\mathbf{r}), \rho_{\downarrow}(\mathbf{r}), \nabla\rho_{\uparrow}(\mathbf{r}), \nabla\rho_{\downarrow}(\mathbf{r}), \tau_{\uparrow}, \tau_{\downarrow}]$$

where $\tau_{\sigma}(\mathbf{r}) = \sum_i^{occ} \frac{1}{2} |\nabla\psi_{i\sigma}(\mathbf{r})|^2$ is the *kinetic energy density* for the occupied KS orbitals

- $\tau_{\sigma}(\mathbf{r})$ is non-local functional of the density $\rho_{\sigma}(\mathbf{r})$
- PKZB (Perdew, Kurth, Zupan, Blaha 1999) [16] - one empirical parameter
- TPSS - (Tao, Perdew, Staroverov, Scuseria 2003) [17] - nonempirical functional
- adding fraction of *exact exchange* - hybrid functionals
- exact exchange techniques - third generation of density functionals
- exact exchange is expensive to calculate

Problem with van der Waals interactions in DFT

- dispersion interactions are *weak but important* (physisorption, soft matter, protein folding, ...)
- van der Waals interactions between distant closed-shell atoms originate from *correlated dipolar fluctuations* $V(R) = -C_6R^{-6}$
- in DFT the interaction should arise from a *long-ranged correlation hole*, very different from that in uniform electron gas
- conventional LDA and GGA miss the van der Waals interactions completely because they are local
- fully *nonlocal* functionals are required to get the long-range forces from distant (nonoverlapping) atoms, molecules, and solids
- various solutions (special functionals) have been proposed
- in practice, classical van der Waals interaction terms are often added to ab-initio calculations (Grimme 2006) [18]

Example 8. The influence of density functional models in ab initio MD simulation of liquid water (VandeVondele, Mohamed, Krack, Hutter, Sprik, Parrinello 2005) [19]

- strong effect of functionals on the structure and dynamics (diffusivity) of liquid water
- BLYP, PBE, and TPSS are too sluggish and overstructured
- some functionals exhibit less structure compared to experiment
- *qualitative differences* observed

Part II

Ab-initio Molecular Dynamics (Car-Parrinello) Method

Literature:

Giulia Galli, Michele Parrinello, *Ab-Initio Molecular Dynamics: Principles and*

Practical Implementation, Computer Simulation in Materials Science, NATO ASI Series Volume 205, 1991, pp 283-304

Dominik Marx, Jürg Hutter, *Ab initio molecular dynamics: Basic Theory and Advanced Methods*, Cambridge University Press (2009)

6 Car-Parrinello Molecular Dynamics

Ab-initio Molecular Dynamics Motivation

- static total energy calculations are good but *MD is better*
- allows to study *chemical reactions* and provides information about time evolution of *electronic properties*
- we assume that the *Born-Oppenheimer* approximation is valid
- true for insulators and semiconductors
- in case of metals, it is more subtle but it is believed to be true as well
- in principle, total energy DFT calculations allow to find an interatomic potential $V(\{\mathbf{R}_I\})$ from first principles
- the Hellman - Feynman theorem allows to find the forces $\mathbf{F}(\{\mathbf{R}_I\})$
- we have to find the interaction potential for each ionic configuration $\{\mathbf{R}_I\}$

$$V(\{\mathbf{R}_I\}) = \langle \Psi_0 | H | \Psi_0 \rangle$$
$$\mathbf{F}(\{\mathbf{R}_I\}) = -\nabla_{\{\mathbf{R}_I\}} V(\{\mathbf{R}_I\})$$

where $|\Psi_0\rangle$ is the *instantaneous* electronic ground state

- in order to do MD, we need to perform at least $\sim 10^4$ *evaluations* of the interaction potential and forces

Dynamical approach to energy functional minimization Car, Parrinello (1985) [20]

- different approach to DFT calculations *without explicit use of the KS equations*

- minimization of the KS functional in the space of electronic wavefunctions Ψ_i is a *non-linear optimization problem*

$$E[\{\Psi_i\}, \{\mathbf{R}_I\}] = \sum_{i=1}^N \langle \Psi_i | -\frac{1}{2} \nabla_{\mathbf{r}}^2 | \Psi_i \rangle + \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[\rho] \\ + \int \rho(\mathbf{r})v(\mathbf{r})d\mathbf{r} + \frac{1}{2} \sum_{I,J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

- minimization can be performed by means of a *fictitious dynamics* of the Ψ_i
- simplest possibility is *steepest-descent* dynamics in fictitious time variable t

$$\dot{\Psi}_i(\mathbf{r}, t) = -\frac{1}{2} \frac{\delta E}{\delta \Psi_i^*(\mathbf{r}, t)}$$

$$\frac{\delta E}{\delta \Psi_i^*(\mathbf{r}, t)} = 2H_{KS}\Psi_i(\mathbf{r}, t)$$

Dynamical approach to energy functional minimization

- *orthonormality constraints* have to be added

$$\int \Psi_i^*(\mathbf{r})\Psi_j(\mathbf{r}) d\mathbf{r} = \delta_{ij}$$

- for a given ionic configuration $\{\mathbf{R}_I\}$ the functional $E[\{\Psi_i\}, \{\mathbf{R}_I\}]$ has a single minimum
- the steepest-descent procedure would lead to absolute minimum of $E[\{\Psi_i\}, \{\mathbf{R}_I\}]$ *without a need for explicit diagonalization*
- more powerful techniques than steepest-descent can be used (e.g. conjugate gradients)
- to get equilibrium ionic structure, we want to minimize $E[\{\Psi_i\}, \{\mathbf{R}_I\}]$ also with respect to the ionic coordinates $\{\mathbf{R}_I\}$

$$E_0 = \min_{\{\mathbf{R}_I\}} \{ \min_{\{\Psi_i\}} E[\{\Psi_i\}, \{\mathbf{R}_I\}] \}$$

- at each step the electronic problem is minimized to perfect self-consistency

Minimization in the coupled space of electrons and ions

$$E_0 = \min_{\{\mathbf{R}_I\}} \left\{ \min_{\{\Psi_i\}} E[\{\Psi_i\}, \{\mathbf{R}_I\}] = \min_{[\{\mathbf{R}_I\}, \{\Psi_i\}]} E[\{\Psi_i\}, \{\mathbf{R}_I\}] \right\}$$

- generalized scheme: *electrons and ions can be optimized at the same time*

$$\begin{aligned} \bar{\mu}_i \dot{\Psi}_i(\mathbf{r}, t) &= -\frac{1}{2} \frac{\delta E}{\delta \Psi_i^*(\mathbf{r}, t)} + \text{constraints} \\ \bar{M}_I \dot{\mathbf{R}}_I &= -\frac{\partial E}{\partial \mathbf{R}_I(t)} \end{aligned}$$

- “masses” $\bar{\mu}_i, \bar{M}_I$ are introduced to bring both problems to the same time scale
- *electrons and ions can be relaxed at the same time*
- perfect electronic self-consistency *is not needed at every step* of the ionic minimization and is only reached at the end of the procedure when the minimum is found
- the scheme allows to optimize large systems

Car-Parrinello Lagrangian Car, Parrinello (1985) [20]

- instead of local optimization one prefers to perform *global optimization* (there can be more minima as function of $\{\mathbf{R}_I\}$)
- *second-order Newton dynamics* for all degrees of freedom generated by an *extended Lagrangian*

$$\begin{aligned} \mathcal{L} &= 2 \sum_i^{occ} \int d\mathbf{r} \mu_i |\dot{\Psi}_i(\mathbf{r})|^2 + \frac{1}{2} \sum_I M_I \dot{\mathbf{R}}_I^2 - E[\{\Psi_i\}, \{\mathbf{R}_I\}] \\ &+ 2 \sum_{ij} \Lambda_{ij} \left(\int d\mathbf{r} \Psi_i^*(\mathbf{r}) \Psi_j(\mathbf{r}) - \delta_{ij} \right) \end{aligned}$$

M_I are physical ionic masses and μ_i are arbitrary parameters

- two classical kinetic energy terms

$$\begin{aligned} K_e &= 2 \sum_i^{occ} \int d\mathbf{r} \mu_i |\dot{\Psi}_i(\mathbf{r})|^2 \text{ electrons} \\ K_I &= \frac{1}{2} \sum_I M_I \dot{\mathbf{R}}_I^2 \text{ ions} \end{aligned}$$

Car-Parrinello Lagrangian

- equations of motion

$$\mu \ddot{\Psi}_i(\mathbf{r}, t) = -\frac{1}{2} \frac{\delta E}{\delta \Psi_i^*(\mathbf{r}, t)} + \sum_j \Lambda_{ij} \Psi_j(\mathbf{r}, t)$$
$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E}{\partial \mathbf{R}_I(t)}$$

- sampling of the parameter space can be performed with MD
- analogy to simulated annealing (Kirkpatrick, Gelatt and Vecchi 1983)
- a global minimum can be found (unlike steepest-descent dynamics)

Car-Parrinello Molecular Dynamics

- the Newtonian dynamics can be used to simulate quasi Born-Oppenheimer trajectories for the ions
- the correct MD would be obtained from the equations of motion

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial V(\{\mathbf{R}_I\})}{\partial \mathbf{R}_I}$$

- the ionic trajectories generated by the CP Lagrangian would coincide with the true BO ones only when $E[\{\Psi_i\}, \{\mathbf{R}_I\}]$ is all the time in the instantaneous minimum
- if the fictitious electronic masses μ_i are chosen in such way that the time scale of electronic motion is much smaller than that of the ionic motion, there will be *very little energy transfer between the electrons and ions*
- if the electrons are at the beginning in the minimum, they will remain close to the minimum for a long time, *close to the instantaneous BO surface*
- the electrons must be “cold” regardless of the ionic temperature
- the system must remain in a *metastable* two-temperatures regime
- this regime of CPMD is efficient in keeping $E[\{\Psi_i\}, \{\mathbf{R}_I\}]$ close to the minimum *without explicit minimization*

Constants of motion

- the Car-Parrinello equations of motion can be integrated to perform ab initio MD
- the Lagrangian is time-independent and has a *conserved quantity* (with no direct physical significance)

$$H = K_e + K_I + E[\{\Psi_i\}, \{\mathbf{R}_I\}]$$

- as long as “fictitious” kinetic energy K_e remains negligible, the physical hamiltonian H_I is approximately conserved

$$H_I = K_I + E[\{\Psi_i\}, \{\mathbf{R}_I\}] \approx \text{const}$$

- the CP simulation can be seen as simulation of the ionic system in a micro-canonical ensemble
- usual statistical mechanics averages for the ions can be calculated along the trajectory
- K_e measures the deviation from the Born-Oppenheimer surface
- K_e should not have an appreciable drift during the MD run
- $\dot{\Psi}_i(\mathbf{r}, t = 0) = 0$ is chosen as initial condition for the wavefunctions

Example 9. Toy system of 8 Si atoms in the diamond structure (Pastore, Smargiassi, Buda 1991) [21]

- system is an insulator with direct gap of 2.24 eV at Γ
- H is constant
- H_I is practically constant
- K_e oscillates within $\sim 3 \times 10^{-5}$ Hartree, the equipartition value would be 9.5×10^{-3} Hartree
- total of 20000 time steps
- no systematic drift of K_e
- at the end of the simulation, the departure from the Born-Oppenheimer surface was $< 10^{-5}$ Hartree

Example 10. Toy system of 8 Si atoms in the diamond structure (Pastore, Smargiassi, Buda 1991)

- two different time scales are clearly visible on the evolution of the K_e
- the slower one corresponds to the ionic motion
- the faster (and smaller) component is due to the intrinsic dynamics of the KS orbitals
- the fast component allows the electrons to follow *adiabatically* the slowly changing ground state

Dynamics of electrons

The dynamics of the electrons is determined by the KS eigenvalues

- for small deviations from the ground state the dynamics corresponds to set of harmonic oscillators with frequencies

$$\omega_{ij} = [2(\epsilon_j - \epsilon_i)/\mu]^{\frac{1}{2}}$$

where ϵ_j (ϵ_i) is the KS eigenvalue of an empty (occupied) state

- the lowest frequency that can appear in the system is

$$\omega_{min} = \left(\frac{2E_g}{\mu} \right)^{\frac{1}{2}}$$

where E_g is the energy gap

Control of adiabaticity

- the high-frequency motion of the electrons serves *two purposes*:
 1. allows the electrons to follow the slow motion of the ions
 2. makes the irreversible exchange of energy between the fast and slow variables very small
- $\omega_{min}^{el} \gg \omega_{max}^{ion}$ is the condition of the adiabatic separation
- ω_{max}^{ion} and gap E_g are physical properties of the system
- we can only *control the adiabaticity* by the fictitious electron mass μ

- decreasing μ shifts the electronic spectrum upwards, but also stretches the whole spectrum
- the maximum electronic frequency scales as

$$\omega_{max}^{el} \propto [E_{cutoff}/\mu]^{\frac{1}{2}}$$

- the maximum MD time step is proportional to the *inverse highest frequency in the system* $\Delta t_{max} \propto [\mu/E_{cutoff}]^{\frac{1}{2}}$
- compromise between adiabaticity and time step has to be made
- typical values are $\mu \sim 500 - 1500$ a.u.
- typical time step is $\Delta t \sim 5 - 10$ a.u. (0.12 - 0.24 fs)

The case of metals

CPMD works well for the systems with *large gap*

- in metals the *gap is close to zero*
- the *violation of adiabaticity* causes *drift of the electronic energy* which causes problems
- Nosé-Hoover thermostats applied separately to the electrons and ions (Blöchl and Parrinello 1992) can keep the two temperatures different
- for metals it is certainly better to use a true Born-Oppenheimer MD
- another source of non-adiabaticity is level crossing of the electronic levels

CP forces and Born-Oppenheimer forces Pastore, Smargiassi, Buda (1991) [21]

- the deviations of the CP forces from the true Born-Oppenheimer ones are *small and oscillatory*
- *average* CPMD forces reproduce the BO ones with high accuracy

CP equations of motion in the plane wave representation

- CP Lagrangian

$$\begin{aligned} \mathcal{L} = & \mu \sum_i^{occ} \sum_{\mathbf{G}} |\dot{c}_i(\mathbf{G})|^2 + \frac{1}{2} \sum_I M_I \dot{\mathbf{R}}_I^2 - E[\{\mathbf{G}\}, \{\mathbf{R}_I\}] \\ & + \sum_{ij} \Lambda_{ij} \left(\sum_{\mathbf{G}} c_i^*(\mathbf{G}) c_j(\mathbf{G}) - \delta_{ij} \right) \end{aligned}$$

- the orthonormality constraint does not depend on the nuclear positions
- equations of motion

$$\begin{aligned} \mu \ddot{c}_i(\mathbf{G}, t) &= -\frac{\partial E}{\partial c_i^*(\mathbf{G})} + \sum_j \Lambda_{ij} c_j(\mathbf{G}) \\ M_I \ddot{\mathbf{R}}_I &= -\frac{\partial E}{\partial \mathbf{R}_I} \end{aligned}$$

- the 2 sets of equations are coupled through the KS functional
- EOM can be integrated by the *velocity Verlet/RATTLE algorithm*

Integration of the CP equations of motion

$$\begin{aligned}
\dot{\tilde{\mathbf{R}}}_I(t + \delta t) &= \dot{\mathbf{R}}_I(t) + \frac{\delta t}{2M_I} \mathbf{F}_I(t) \\
\mathbf{R}_I(t + \delta t) &= \mathbf{R}_I(t) + \delta t \dot{\tilde{\mathbf{R}}}_I(t + \delta t) \\
\dot{\tilde{\mathbf{c}}}_i(t + \delta t) &= \dot{\mathbf{c}}_i(t) + \frac{\delta t}{2\mu} \mathbf{f}_i(t) \\
\tilde{\mathbf{c}}_i(t + \delta t) &= \mathbf{c}_i(t) + \delta t \dot{\tilde{\mathbf{c}}}_i(t + \delta t) \\
\mathbf{c}_i(t + \delta t) &= \tilde{\mathbf{c}}_i(t + \delta t) + \sum_j \mathbf{X}_{ij} \mathbf{c}_j(t) \text{ where } \mathbf{X}_{ij} = \frac{\delta t^2}{2\mu} \Lambda_{ij}^P \\
\text{calculate } & \mathbf{F}_I(t + \delta t) \\
\text{calculate } & \mathbf{f}_i(t + \delta t) \\
\dot{\mathbf{R}}_I(t + \delta t) &= \dot{\tilde{\mathbf{R}}}_I(t + \delta t) + \frac{\delta t}{2M_I} \mathbf{F}_I(t + \delta t) \\
\dot{\mathbf{c}}'_i(t + \delta t) &= \dot{\tilde{\mathbf{c}}}_i(t + \delta t) + \frac{\delta t}{2\mu} \mathbf{f}_i(t + \delta t) \\
\dot{\mathbf{c}}_i(t + \delta t) &= \dot{\mathbf{c}}'_i(t + \delta t) + \sum_j \mathbf{Y}_{ij} \mathbf{c}_j(t + \delta t) \text{ where } \mathbf{Y}_{ij} = \frac{\delta t}{2\mu} \Lambda_{ij}^v
\end{aligned}$$

Imposing the orthogonality constraints

- Lagrange multipliers for positions Λ_{ij}^P and velocities Λ_{ij}^v are treated as independent variables

$$\begin{aligned}
\mathbf{C}^\dagger(t + \delta t) \mathbf{C}(t + \delta t) - \mathbf{I} &= 0 \\
\mathbf{X} \mathbf{X}^\dagger + \mathbf{X} \mathbf{B} + \mathbf{B}^\dagger \mathbf{X}^\dagger &= \mathbf{I} - \mathbf{A} \\
A_{ij} &= \tilde{c}_i^\dagger(t + \delta t) \tilde{c}_j(t + \delta t) \\
B_{ij} &= \tilde{c}_i^\dagger(t) \tilde{c}_j(t + \delta t)
\end{aligned}$$

- the equation can be solved iteratively, starting from $\mathbf{X}^{(0)} = \frac{1}{2}(\mathbf{I} - \mathbf{A})$

$$\mathbf{X}^{(n+1)} = \frac{1}{2} [\mathbf{I} - \mathbf{A} + \mathbf{X}^{(n)}(\mathbf{I} - \mathbf{B}) + (\mathbf{I} - \mathbf{B})\mathbf{X}^{(n)} - (\mathbf{X}^{(n)})^2]$$

- orthogonality condition on the orbital velocities

$$\dot{c}_i^\dagger(t + \delta t) c_j(t + \delta t) + c_i^\dagger(t + \delta t) \dot{c}_j(t + \delta t) = 0$$

is applied to the trial states $\dot{\mathbf{C}}' + \mathbf{Y} \mathbf{C}$

- the rotation matrix \mathbf{Y} is obtained without iteration

$$\mathbf{Y} = -\frac{1}{2}(\mathbf{Q} + \mathbf{Q}^\dagger), \text{ where } Q_{ij} = c_i^\dagger(t + \delta t)c_j^\dagger(t + \delta t)$$

Scaling of the CPMD method

- N is number of occupied orbitals
- M is number of plane waves
- kinetic energy is diagonal so its action on Ψ is $O(NM)$
- action of local potential on Ψ is $O(NM \log M)$ (FFT)
- action of non-local terms (pseudopotential) on Ψ is $O(N^2M)$
- orthogonalization procedure is also $O(N^2M)$
- if $N \ll M$, the big dimension M enters only via M or $M \log M$
- significant improvement over $O(M^3)$ scaling of a direct diagonalization

Practical aspects of CPMD simulation

1. for the initial ionic configuration the wavefunction must be well optimized
2. equations of motion are integrated
3. the kinetic energy of the electrons has to be monitored
4. if the electrons become hot, the wavefunction has to be quenched to the BO surface

Example 11. Importance of the choice of the fictitious mass μ in study of liquid water (Grossman, Schwegler, Draeger, Gygi and Galli 2004) [22]

- water is a wide-gap insulator with $\epsilon_{HOMO} - \epsilon_{LUMO} \sim 4$ eV
- vibrational spectra of the electronic degrees of freedom are calculated from

$$\nu(\omega) = \int_0^\infty dt \cos(\omega t) \langle \sum_i \dot{\Psi}_i(t) \dot{\Psi}_i(0) \rangle$$

- the highest ionic mode is the O-H stretch mode
- values of $\frac{\mu}{M}$ up to $\sim \frac{1}{5}$ are appropriate
- values of $\frac{\mu}{M} \sim \frac{1}{3}$ are too large and result in a strong deviation in structural properties

A different possibility - Ehrenfest MD

- there are various ways how to make the electrons *follow adiabatically the ground state* as the ions move
- *Ehrenfest MD* - Newton dynamics for ions and *time-dependent Schrödinger equation* for electrons

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_{\mathbf{R}_I} \langle \Psi_0(t) | \hat{H}_e | \Psi_0(t) \rangle$$
$$i \frac{\partial}{\partial t} |\Psi_0(t)\rangle = \hat{H}_e(\{\mathbf{R}_I\}, \{\mathbf{r}_i\}) |\Psi_0(t)\rangle$$

- if the Born-Oppenheimer approximation is valid, electrons follow adiabatically and remain in the ground state (no diagonalization needed)
- propagation of the wavefunction is unitary and so the norm and orthogonality is preserved
- Trotter-Suzuki formula can be used to derive the integrator

$$\exp(-i\hat{H}\delta t) \simeq \exp(-i\hat{T}\frac{\delta t}{2}) \exp(-i\hat{V}\delta t) \exp(-i\hat{T}\frac{\delta t}{2})$$

- practically, integration of the Schrödinger equation is difficult and requires extremely small time step
- CPMD *combines advantages* of Born-Oppenheimer and Ehrenfest MD

Example 12. Simulation of thermally induced conversion of carbon from diamond to graphite (De Vita, Galli, Canning, Car 1996) [23]

- diamond is a high-pressure phase of carbon and is metastable at ambient conditions
- the most stable phase is hexagonal graphite
- the barrier for conversion to graphite is very high and diamond is *stable even on a geological time scale*
- important for chemical vapour deposition growth of diamond films which may contain graphitic islands
- experiments suggested that the transition is *surface-induced*
- CPMD study on a diamond slab containing 12 (111) atomic layers, separated by a vacuum slab of 9 Å

- slab is initially terminated by (2x1) reconstructed surfaces
- 192 atom slab was heated to $T \sim 2500$ K
- cutoff of 33 and 99 Ry used for orbitals and density, respectively
- only Γ point used in the calculations
- total simulation time about 15 ps

Example 13. Simulation of thermally induced conversion of carbon from diamond to graphite (De Vita, Galli, Canning, Car 1996)

- original reconstructed slab at 0 K
- initial stage of graphitization at 2500 K
 - nucleation of a graphitic seed
 - initial breaking of bonds
 - flattening of double layer
 - change of hybridization from sp^3 to sp^2
- fast penetration of the graphite phase into the diamond slab proceeding via correlated breaking of nearest neighbour bonds along the z -direction
- end of the phase transition, defect-free graphitic planes

Example 14. Simulation of thermally induced conversion of carbon from diamond to graphite (De Vita, Galli, Canning, Car 1996)

- the seed of the transition from previous simulation was taken and simulated at 1000 K
- *a stable interface was formed*
- to reduce finite-size effects, a larger 288 atom system was prepared from the interface and simulated in the range 0 - 1000 K, *interface still stable*
- networks of graphitic sheets are coherently connected to the buckled diamond (111) double layers
- the interface reconstruction maximizes the number of 6-membered rings at the phase boundary

7 Extensions of the CPMD technique

Pulay stress Froyen and Cohen 1986

[24]

- the use of an *incomplete plane-wave basis set* has some consequences
- N_{PW} at constant cutoff E_c depends on the volume $N_{PW} \sim \frac{\sqrt{2}}{3\pi^2} \Omega E_c^{\frac{3}{2}}$
- if E^{tot} is not completely converged, $E^{tot} = E^{tot}(N_{PW}, \Omega)$
- $E^{tot}(\Omega)$ curves at constant N_{PW} and constant E_c are *different*
- in lack of full convergence it is better to keep $E_c = const$ because this guarantees constant spatial resolution
- stress tensor is typically calculated with $N_{PW} = const$
- *Pulay stress* arises from the volume dependence of N_{PW}

$$\frac{dE^{tot}}{d\Omega} = \left. \frac{\partial E^{tot}}{\partial \Omega} \right|_{N_{PW}} + \frac{\partial E^{tot}}{\partial N_{PW}} \frac{\partial N_{PW}}{\partial \Omega} = \left. \frac{\partial E^{tot}}{\partial \Omega} \right|_{N_{PW}} + \frac{\partial E^{tot}}{\partial E_c} \frac{\partial E_c}{\partial N_{PW}} \frac{\partial N_{PW}}{\partial \Omega}$$

$$p_{E_c} = p_{N_{PW}} - \frac{2}{3} \frac{E_c}{\Omega} \left. \frac{\partial E^{tot}}{\partial E_c} \right|_{\Omega}$$

- the correction term can be easily determined from few calculations with different E_c (should be taken into account in constant-pressure calculations)

Constant-pressure Car-Parrinello MD Focher, Chiarotti, Bernasconi, Tosatti, Parrinello (1994) [25]

- various *extended Lagrangian* techniques can be combined
- combination of Car-Parrinello and Parrinello-Rahman Lagrangians
- the original wavefunctions depend on the cell so they have to be expressed in the scaled variables \mathbf{s}

$$\mathcal{L} = \mu \sum_i^{occ} \int d\mathbf{s} |\dot{\Psi}_i(\mathbf{s})|^2 + \frac{1}{2} \sum_i M_I (\dot{\mathbf{S}}_I^T \mathbf{G} \dot{\mathbf{S}}_I) - E[\{\Psi_i\}, \{\mathbf{h}\mathbf{S}_I\}]$$

$$+ \sum_{ij} \Lambda_{ij} \left(\int d\mathbf{s} \Psi_i^*(\mathbf{s}) \Psi_j(\mathbf{s}) - \delta_{ij} \right) + \frac{1}{2} W \text{Tr} \dot{\mathbf{h}}^T \dot{\mathbf{h}} - PV$$

where $\mathbf{h} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$, $\mathbf{r} = \mathbf{h}\mathbf{s}$, $V = \det \mathbf{h}$, $\mathbf{G} = \mathbf{h}^T \mathbf{h}$

- as the cell fluctuates, the \mathbf{G} vectors change and the number of plane waves within given E_{cutoff} fluctuates too
- allows study of structural transitions in solids *without empirical potentials*
- important to study matter at extreme conditions (temperatures, pressures), e.g. geoscience, planetary science

Example 15. Simulation of pressure-induced conversion of carbon from graphite to diamond (Scandolo, Bernasconi, Chiarotti, Focher, Tosatti 1995) [26]

- at pressure of ~ 15 GPa and $T > 1000$ K graphite converts to the high-pressure phase, *diamond*
- process of *great technological and fundamental importance*
- two kinds of diamond structure, *cubic* and *hexagonal* (Lonsdaleite)
- in shock wave experiments, formation of hexagonal diamond was observed
- three kinds of graphite - *hexagonal*, *orthorhombic* and *rhombohedral*
- constant-pressure CPMD study, cutoff of 35 Ry, only Γ point
- starting structure hexagonal diamond, 48 or 64 atoms supercell

Example 16. Simulation of pressure-induced conversion of carbon from graphite to diamond (Scandolo, Bernasconi, Chiarotti, Focher, Tosatti 1995) Possible geometric paths for graphite to diamond transformation

- rhombohedral graphite to cubic diamond
- orthorhombic graphite to cubic diamond
- orthorhombic graphite to hexagonal diamond
 - *stacking of graphite determines the orientation of the final diamond phase*
 - shock-wave experiment is consistent with the mechanism (b)

Example 17. Simulation of pressure-induced conversion of carbon from graphite to diamond (Scandolo, Bernasconi, Chiarotti, Focher, Tosatti 1995) Dynamics of the transformation from graphite to diamond in a 48 atom cell

- compression to 30 GPa at 200 K, sliding of the graphitic planes between hex and orth

- heating to 1000 K
- fast increase of pressure, sliding of the graphitic planes between orth and hex
- at 90 GPa (much higher than the exp. value of 15 GPa) *orthorhombic graphite converted to cubic diamond*
- mechanism corresponds to the path (b), no formation of hex diamond observed

Example 18. Simulation of pressure-induced conversion of carbon from graphite to diamond (Scandolo, Bernasconi, Chiarotti, Focher, Tosatti 1995) Transformation from graphite to diamond in a 64 atom cell

- in a bigger cell more paths are possible
- simulation performed at 150 and 1500 K
- sliding of graphite to orth stacking observed again
- at 1500 K hex diamond was found
- at 150 K structure with *mixed hexagonal and cubic diamond arrangement was found*
- role of the orthorhombic graphite confirmed

Combination of CPMD and path integrals Marx and Parrinello 1994 [27]

- when light atoms (hydrogen, ...) are present, *quantum effects* on the ionic motion are important
- static quantum effects can be calculated within the *path-integral formalism*
- path-integrals can be sampled by either MC or MD
- extended Lagrangian *combining CPMD and path-integrals*

$$\begin{aligned}
\mathcal{L} = & \frac{1}{P} \sum_{s=1}^P \left\{ \sum_i^{occ} \mu \langle \dot{\psi}_i^{(s)} | \dot{\psi}_i^{(s)} \rangle - E[\{\psi_i^{(s)}\}, \{\mathbf{R}_I^{(s)}\}] \right. \\
& + \left. \sum_{ij} \Lambda_{ij}^{(s)} \left(\langle \psi_i^{(s)} | \psi_j^{(s)} \rangle - \delta_{ij} \right) \right\} \\
& + \sum_{s=1}^P \left\{ \sum_I \frac{1}{2} M_I' (\dot{\mathbf{R}}_I^{(s)})^2 - \sum_I \frac{1}{2} M_I \omega_P^2 (\mathbf{R}_I^{(s)} - \mathbf{R}_I^{(s+1)})^2 \right\}
\end{aligned}$$

- P is the Trotter number, $M'_I \neq M_I$ is *fictitious* mass for the ions, $\omega_P^2 = P/\beta^2$

Combination of CPMD and path integrals

- path integral is sampled by MD, the dynamics is *unphysical* (just for sampling the configuration space)
- *Problems:*
 - because of the harmonic springs, path integral representation behaves as collection of nearly harmonic oscillators with $\omega_P \propto \sqrt{P}$
 - MD sampling of the ionic system presents strong *non-ergodicity*, for large P potential energy $\frac{1}{P}U(\{\mathbf{R}\})$ is just a small perturbation
 - because of the *high ionic frequencies* the adiabaticity condition of CPMD may be violated, use of small μ would require too small time step
- *Solution:*
 - can be solved by diagonalization of the harmonic part of the Lagrangian - *staging* which brings all modes to the same scale (Tuckerman, Marx, Klein, Parrinello 1996) [28]
 - application of Nosé-Hoover *chain* thermostat to the ions (massive thermostating - one chain for each Cartesian component of ionic coordinates)
 - application of Nosé-Hoover *chain* thermostat to the electrons
- method parallelizes trivially - each Trotter slice runs on separate node, little communication

Example 19. Tunnelling and zero-point motion in high-pressure ice Benoit, Marx, Parrinello (1998) [29]

- normally the configuration O – H ... O is *highly asymmetric* (one covalent and one hydrogen bond)
- upon high pressure the bonds get *shorter*
- the normal configuration can transform into one where the *proton stays in the middle* between two oxygens
- transition from ice VII to ice X - *non-molecular* phase of ice

- *quantum tunnelling effects crucial* - constant-pressure path-integral CPMD

Example 20. Tunnelling and zero-point motion in high-pressure ice Benoit, Marx, Parrinello (1998)

- the symmetrization transition occurs in quantum system at ~ 72 GPa while in the classical one at ~ 102 GPa
- two forms of ice X: proton-disordered symmetric ice (c) and proton-ordered symmetric ice (d)

Metadynamics and CPMD Iannuzzi, Laio, Parrinello (2003) [30]

- Lagrangian metadynamics - useful to study chemical reactions
- large barriers can be crossed in very short time (few ps) and a lot of energy is released to degrees of freedom associated with the collective variable
- set of collective variables $s_\alpha(\{\mathbf{R}_I\})$, $\alpha = 1, \dots, d$
- Lagrangian includes the *dynamical* collective variables $\tilde{s}_\alpha(\{\mathbf{R}_I\})$ with fictitious mass M_α

$$\mathcal{L} = \mathcal{L}_0 + \sum_{\alpha} \frac{1}{2} M_{\alpha} \dot{\tilde{s}}_{\alpha}^2 - \sum_{\alpha} \frac{1}{2} k_{\alpha} [s_{\alpha}(\{\mathbf{R}_I\}) - \tilde{s}_{\alpha}]^2 + V(t, \tilde{\mathbf{s}})$$

\mathcal{L}_0 is the usual CPMD Lagrangian

- the instantaneous values of the collective variables $S_{\alpha}(\{\mathbf{R}_I\})$ fluctuate around the corresponding \tilde{S}_{α}
- the scheme allows for *explicit control of the dynamics and temperature of the collective variable* \tilde{s}_{α} via thermostat, avoiding possible instabilities

Metadynamics and CPMD

- if the mass M_{α} is large, the dynamics of the collective variables is slow and *adiabatically separated* from that of ions and electrons
- M_{α} is tuned to obtain a smooth evolution of \tilde{s}_{α}
- the spring constants k_{α} must be chosen in such a way that the typical value of the difference $|s_{\alpha}(\{\mathbf{R}_I\}) - \tilde{s}_{\alpha}|$ is smaller than the length on which the free energy varies of $\sim k_B T$

- the history-dependent potential can be constructed in form of a d-dimensional Gaussian tube accumulating infinitesimal tube slices along the trajectory of $\tilde{\mathbf{s}}$

$$V(t, \tilde{\mathbf{s}}) = \int_0^t dt' |\dot{\tilde{\mathbf{s}}}(t')| w \exp \left\{ -\frac{(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(t'))^2}{2\delta s^2} \right\} \delta \left(\frac{\dot{\tilde{\mathbf{s}}}(t')}{|\dot{\tilde{\mathbf{s}}}(t')|} \cdot (\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(t')) \right)$$

δs is the size in the orthogonal direction

Dehydrogenation process of a Si_6H_8 cluster, collective coordinates are Si-Si, Si-H, and H-H coordination numbers

Example 21. Study of Azulene-to-Naphthalene Rearrangement Stirling, Iannuzzi, Laio, Parrinello (2004) [31]

- complex potential energy surface
- various *C-C and C-H coordination numbers* used as collective variables to explore various proposed mechanisms
- several possible mechanisms were found with different barriers

Hybrid Schemes Quantum Mechanics/Molecular Mechanics (QM/MM)

Literature:

P. Sherwood, in *Modern Methods and Algorithms of Quantum Chemistry* NIC Ser. Vol. 1, John von Neumann Institute for Computing, Juelich 2000, pp. 257 - 277

- despite all progress in ab initio methods, it is not possible to treat large enough systems
- in chemistry (catalysis), biochemistry (study of enzymes), materials science etc., often many thousands of atoms participate in the processes
- *partitioning of the system to the QM and MM regions* - multiscale approach
- QM is used to model more important processes in the *inner region* (small)
- MM (force field) is used for the less important processes in the *outer* (large) region (environment)
- the *boundary* must be sufficiently far from the *reaction center* (requires chemical intuition and testing)
- ideally, the boundary should not cut molecules (not always possible)

- if the boundary cuts a *covalent bond*, the interaction between the two systems is *strong* and the QM calculation in the QM region alone does not make sense

Approaches to QM termination

- treatment of *bonds that connect different sides of the boundary* is the most difficult aspect
- some way of *termination or treating the boundary* is necessary

1. link atoms

- *additional centers* are added to the QM calculation which are *invisible* to the MM calculation
- *positions of the link atoms* are either independent variables or function of positions of atoms in the (I) and (O) regions

2. boundary region

- subset of centers of the systems which are *present in both QM and MM calculations*

Construction of the QM/MM energy expression

Two ways to construct the total energy of the system (E = entire, I = inner, O = outer, L = link)

Additive schemes

- the QM and MM energies are considered *complementary*
- the case of link atoms

$$E_{QM/MM} = E_{MM}^O + E_{QM}^{I,L} + E_{QM-MM}^{I,O} - E_{corr}^{I,L}$$

- the term $E_{QM-MM}^{I,O}$ includes all terms that couple the two regions (e.g. van der Waals interactions)
- mainly used in biomolecular modeling
- the case of boundary-region based methods

$$E_{QM/MM}^E = E_{MM}^{O,B} + E_{QM}^{I,B} + E_{QM/MM}^{I,O,B}$$

- the boundary atoms are treated by both QM and MM so the classical energy must be modified to avoid multiple counting of the interactions

Subtractive schemes

- the entire system is treated by MM and the energy of the I region is also evaluated by MM
- for the link atoms schemes

$$E_{QM/MM}^E = E_{MM}^E + E_{QM}^{I,L} - E_{MM}^{I,L}$$

- *no coupling term is required*
- the difference between the QM and MM forces in the link region should be small
- it is useful if the forcefield is designed to reproduce the forces of the particular QM approximation used for the inner region
- not suitable when the electronic structure of the QM region is significantly perturbed by the environment
- simpler to implement, mainly used in chemistry

Choice of QM and MM models

- QM - typically DFT (CPMD)
- sometimes even a higher level method is necessary (*QM/QM/MM*)
- choice of MM depends on whether additive or subtractive scheme is used
- in additive schemes, the type of MM model has large influence on the treatment of the boundary
- two kinds of force fields:
 - *valence force fields* - bond stretching terms, angle bending terms, torsions, etc.
 - *ionic force fields* based on Coulomb and short range forces (shell models)
- in additive schemes based on link atoms, it is easier to use the valence force fields - more easy to separate the different contributions and leave out what is treated by QM

Treatment of QM/MM electrostatic terms

- A *mechanical embedding* - QM calculation is performed in the gas phase
- B *electrostatic embedding* - classical partition appears as external charge distribution, polarization of the QM part due to the MM charges occurs as part of the QM calculation
- C *polarised embedding* polarisation of the MM region due to charges in the QM region is also included
- D extension of C where the QM and MM polarizations are brought to *self-consistency*, e.g. iteratively

Termination of the QM region

- to terminate a broken covalent bond the simplest choice is to add a *hydrogen link atom* - “capping” of the dangling bond
- the *position of the link atom* is an extra degree of freedom
- *optimised link atoms* - the link atoms are allowed to move
- *constrained link atoms* - link atom coordinates are expressed as function of real atom coordinates
- the force acting on the link atom has to be distributed to the real atoms

$$\frac{dE}{d\mathbf{R}_I} = \frac{\partial E}{\partial \mathbf{R}_I} + \frac{\partial E}{\partial \mathbf{R}_L} \frac{\partial \mathbf{R}_L}{\partial \mathbf{R}_I}$$

- the constrained approach is better for MD - no spurious vibrational frequencies are introduced

Example 22. QM/MM study of defect formation in silica Zipoli, Laino, Laio, Bernasconi, Parrinello (2006) [33]

- a defect in a solid implies a change in electronic configuration as well as relaxation of the lattice
- if the fully QM technique is used, the size limitation (few hundred atoms) would imply extremely high concentration of defects and strong defect-defect interaction
- in QM/MM, the defect can be treated by QM and the bulk part by MM
- the QM subsystem is embedded in a classical cluster composed of 508 SiO₂ units

8 Minimization techniques for the KS functional

Literature:

M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 - 1097 (1992)

Two different approaches to optimization of KS orbitals

- Fixed-point methods (*iterative diagonalization*)
 - we do not need all eigenvectors and eigenvalues, only the occupied ones ($N \ll M$)
 - there are methods which allow to do this and do not require storage of the full matrix, just the *calculation of $\hat{H}|\Psi\rangle$*
 - Lanczos (1950), Davidson (1975), ...
 - used together with iterative improvement of the density (mixing)
 - robust, well suited for systems with no or small gap (metals)
 - for review, see Kresse and Furthmüller (1996) [34]
- Direct non-linear minimization methods
 - steepest descent
 - conjugate gradients
 - convergence acceleration schemes
 - orbital rotation schemes
 - ...
- in both cases *preconditioning* is useful

Steepest descent

- function $F(\mathbf{x})$ to be minimized in multidimensional space \mathbf{x}
- the simplest possibility is to *follow the direction of negative gradient*
- gradient operator G

$$\mathbf{g}^1 = - \left. \frac{\partial F}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^1} = -G\mathbf{x}^1$$

- one moves from the point \mathbf{x}^1 in the direction \mathbf{g}^1 to the point $\mathbf{x}^2 = \mathbf{x}^1 + b^1 \mathbf{g}^1$ where $F(\mathbf{x}^1 + b^1 \mathbf{g}^1)$ has minimum

$$\mathbf{g}^1 \cdot \mathbf{g}^2 = 0$$

- the new search direction is *orthogonal* to the previous one
- the process converges but often needs a large number of steps, especially in case of strongly anisotropic valley
- at each step the search direction is chosen using only information from the present point

Conjugate gradients

- a better technique would combine the information from all the sampling points and search in direction *orthogonal* to previous searches
- we start from point \mathbf{x}_i and perform line-minimization of the function $F(\mathbf{x}_i + \lambda_i \mathbf{d}_i)$ along the direction \mathbf{d}_i
- the new minimum is \mathbf{x}_{i+1}
- the search directions \mathbf{d}_i are obtained from

$$\begin{aligned} \mathbf{d}_m &= \mathbf{g}_m + \gamma_m \mathbf{d}_{m-1} \\ \gamma_m &= \frac{\mathbf{g}_m \cdot \mathbf{g}_m}{\mathbf{g}_{m-1} \cdot \mathbf{g}_{m-1}}, \quad \gamma_1 = 0 \end{aligned}$$

- the directions \mathbf{d}_i are *conjugate*
- minimizations along conjugate directions are *independent*
- subsequent step does not spoil the previous one

Preconditioned conjugate gradients scheme

- for efficient optimization in many dimensions it is useful to bring all degrees of freedom onto the same length scale - *preconditioning*
- for large \mathbf{G} vectors the KS matrix $\mathbf{H}_{\mathbf{G},\mathbf{G}'}$ is dominated by kinetic energy which is diagonal - simple preconditioner

$$\begin{aligned} \mathbf{K}_{\mathbf{G},\mathbf{G}'} &= \mathbf{H}_{\mathbf{G},\mathbf{G}} \delta_{\mathbf{G},\mathbf{G}'} \text{ if } |\mathbf{G}| \geq G_c \\ \mathbf{K}_{\mathbf{G},\mathbf{G}'} &= \mathbf{H}_{\mathbf{G}_c,\mathbf{G}_c} \delta_{\mathbf{G},\mathbf{G}'} \text{ if } |\mathbf{G}| \leq G_c \end{aligned}$$

where G_c is a free parameter (one can take $G_c = 0.5$ Hartree for each system)

Conjugate gradients for plane-wave wavefunction optimization

- update scheme

$$\begin{aligned}
 c_i(\mathbf{G}) &\leftarrow c_i(\mathbf{G}) + \alpha h_i(\mathbf{G}) \\
 h_i^n(\mathbf{G}) &= \begin{cases} g_i^n(\mathbf{G}) & n = 0 \\ g_i^n(\mathbf{G}) + \gamma^{n-1} h_i^{n-1}(\mathbf{G}) & n = 1, 2, 3, \dots \end{cases} \\
 g_i^n(\mathbf{G}) &= \mathbf{K}_{\mathbf{G},\mathbf{G}}^{-1} \psi_i^n(\mathbf{G}) \\
 \gamma^n &= \frac{\sum_i \langle g_i^{n+1}(\mathbf{G}) | g_i^{n+1}(\mathbf{G}) \rangle}{\langle g_i^n(\mathbf{G}) | g_i^n(\mathbf{G}) \rangle}
 \end{aligned}$$

where $\psi_i^n(\mathbf{G})$ is the wavefunction gradient

- it is efficient to perform the update band by band (Teter, Payne, Allan 1989) [35]

Direct inversion in the iterative subspace (DIIS) Pulay (1980) [36], Hutter, Lüthi, Parrinello (1994) [37]

- *convergence acceleration* for iterative sequences, using information from m previous iterations
- we have a sequence of m vectors $\{\mathbf{x}_i\}$
- assume that for each vector we can guess the difference to the stationary point \mathbf{x}_0

$$\mathbf{e}_i = \mathbf{x}_i - \mathbf{x}_0$$

- ansatz: find the best linear combination

$$\mathbf{x}_{m+1} = \sum_{i=1}^m c_i \mathbf{x}_i$$

ideally we would have $\sum_{i=1}^m c_i \mathbf{x}_i = \mathbf{x}_0$

$$\sum_{i=1}^m c_i \mathbf{x}_0 + \sum_{i=1}^m c_i \mathbf{e}_i = \mathbf{x}_0$$

Direct inversion in the iterative subspace (DIIS)

- this would be satisfied if

$$\sum_{i=1}^m c_i = 1 \text{ and } \sum_{i=1}^m c_i \mathbf{e}_i = \mathbf{e}_{M+1} = 0$$

- the last condition is in practice satisfied only in the *mean-square sense*
- we introduce a Lagrange multiplier λ and matrix $b_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$ and minimize

$$\frac{d}{dc_i} \left(\sum_{i,j=1}^m c_i c_j \mathbf{e}_i \cdot \mathbf{e}_j - \lambda \sum_{i=1}^m c_i \right) = \frac{d}{dc_i} \left(\sum_{i,j=1}^m c_i b_{ij} - \lambda \sum_{i=1}^m c_i \right) = 0$$

- system of $(m + 1)$ linear equations is solved

$$\begin{pmatrix} b_{11} & b_{12} & \dots & b_{1m} & 1 \\ b_{21} & b_{22} & \dots & b_{2m} & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mm} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \\ -\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Direct inversion in the iterative subspace (DIIS)

- the error vectors \mathbf{e}_m are not known, but close to the minimum \mathbf{x}_0 the function is approximately quadratic $F(\mathbf{x}) \approx \frac{1}{2} \mathbf{x} \mathbf{H} \mathbf{x} - \mathbf{a} \mathbf{x}$ (\mathbf{H} is Hessian, assumed to be constant)

$$G(\mathbf{x}) \approx \mathbf{H} \mathbf{x} - \mathbf{a} = \mathbf{H}(\mathbf{x} - \mathbf{x}_0 + \mathbf{x}_0) - \mathbf{a} = \mathbf{H}(\mathbf{x} - \mathbf{x}_0) \\ \mathbf{e}_m = \mathbf{H}^{-1} G(\mathbf{x}_m)$$

- Hessian H can be *approximated by the diagonal part* of the KS Hamiltonian

$$H_{\mathbf{G}, \mathbf{G}'}^{KS} = \frac{1}{2} G^2 + V_{\mathbf{G}, \mathbf{G}'}^H + V_{\mathbf{G}, \mathbf{G}'}^{XC} + V_{\mathbf{G}, \mathbf{G}'}^i$$

or by the preconditioner matrix $K_{\mathbf{G}, \mathbf{G}'}$

- the new estimate for \mathbf{x}_0 is

$$\mathbf{x}_0 \approx \mathbf{x}_{M+1} - \mathbf{e}_{M+1}$$

Born-Oppenheimer MD and CPMD

- due to progress in minimization and iterative diagonalization techniques it is possible to perform also Born-Oppenheimer MD
- for systems with slow ionic dynamics, large time step can be used
- BO might be even faster than CPMD but with a large time step the energy conservation decreases
- in systems with stiff covalent bonds, even in BO a shorter time step has to be used
- BO is preferred for *metals* where CPMD has problems

Part III

Special topics

9 Tight-binding methods

The tight-binding method

Literature: C. M. Goringe, D. R. Bowler and E. Hernández, Rep. Prog. Phys. **60**, 1447 (1997)

- TB is a simplified QM scheme using a series of approximations
- TB is between fully ab initio schemes and empirical models
- TB uses a minimal set of *localized basis functions*
- the basis set is not explicitly constructed but has *the same symmetry properties as the atomic orbitals*
- instead of exact many-body Hamiltonian operator a *parametrized Hamiltonian matrix* is used
- number of basis states is much smaller compared to plane wave DFT
- e.g. modeling carbon, only 2s and 2p orbitals are considered (4 states)
- origin - linear combination of atomic orbitals (LCAO) Method for the Periodic Potential Problem - Slater and Koster (1954)

The tight-binding method

- Hamiltonian matrix elements are written as

$$H_{i\alpha j\beta}(\mathbf{k}) = \sum_{\mathbf{R}_j, J} \exp[i\mathbf{k} \cdot (\mathbf{R}_j - \mathbf{R}_i)] h_{\alpha\beta J}(|\mathbf{R}_j - \mathbf{R}_i|) G_{\alpha\beta J}(l, m, n)$$

sum goes over periodic images \mathbf{R}_j , α, β correspond to atomic orbitals, k, l, m are the direction cosines of the vector $\mathbf{R}_j - \mathbf{R}_i$

- J is the angular momentum of the bond
 - if either α or β is s , $J = 0$ (σ)
 - if both α and β are at least p , $J = 0$ or 1 (π) etc.
- $G_{\alpha\beta J}(l, m, n)$ is the angular dependence tabulated in the paper by Slater and Koster (1954)
- often the function $h_{\alpha\beta J}(r)$ is short-ranged and a cutoff is used (e.g. at second neighbour distance)
- the whole scheme is described by the functions $h_{\alpha\beta J}(r)$, parametrized analytically or numerically (e.g. splines)
- the quality of a TB scheme is determined by the parametrization

The tight-binding method

- the Hamiltonian can be solved by direct diagonalization

$$\hat{H}|\Psi_i\rangle = \epsilon_i|\Psi_i\rangle$$

$$E_{tot} = E_{band} + E_{rep} = 2 \sum_{i=1}^N \epsilon_i + \sum_{i \neq j} U_{ij}$$

- E_{band} is the *band structure term*
- pairwise additive E_{rep} accounts for all terms which are not included in E_{band}
- the \hat{H} matrix is *sparse*
- TB is *non self-consistent* scheme so no iteration is performed, energy is obtained in one step for each atomic configuration

- if charge transfer is a problem, *local charge neutrality* can be imposed as a constraint
- Hubbard U term can also be used (cost for charging an atom)
- by differentiating E_{band} (Hellman - Feynman theorem) the forces on atoms can be found and TB MD can be performed

The tight-binding method

- parameters of TB schemes are either determined from ab-initio calculations or fitted to experimental results
- compared to *full ab initio* methods TB is *2 to 3 orders of magnitude faster*, but less transferable
- compared to *classical* potentials, TB is *2 to 3 orders of magnitude slower*, but QM nature of bonding is retained
- in TB the angular nature of bonding is correctly described even in configurations far from equilibrium structure
- useful in situations where QM effects are important but fully ab initio treatment of the large system is impossible
- TB is often used for simulations of carbon structures - fullerenes, nanotubes, onions etc.
- large scale applications - defects (point and extended), dislocations, grain boundaries, study of brittle fracture, crack propagation etc.

Example 23. Spanning the continuum to quantum length scales in a dynamic simulation of brittle fracture Abraham, Broughton, Bernstein, Kaxiras (1998) [39]

- rapid brittle fracture of a Si slab flawed by a microcrack at its center, under uniaxial tension
- multiscale simulation with 3 regions (2 *dynamical* boundaries):
 - continuum (finite elements) - “far-field” region
 - atomistic (classical MD) - region around the crack (large strain gradients but no bond rupture)
 - quantum - (TB) in the region of *bond failure* at the crack tip

10 Linear scaling methods

Literature:

Stefan Goedecker, Rev. Mod. Phys. **71**, 1085 - 1123 (1999)
D R Bowler, T Miyazaki and M J Gillan, J. Phys.: Condens. Matter **14**, 2781 (2002)

Scaling of electronic structure calculations

- basic operation of all methods is $\hat{H}\Psi$
- due to use of FFT $\hat{H}\Psi$ scales as $NM \ln M$ where N is the number of atoms and M is a number of plane waves (compared to $O(NM^2)$ with the straight-forward approach)
- bottleneck is the *orthogonalization step* which is $O(N^2M)$
- $M \sim N$, for large enough systems we get $O(N^3)$ scaling
- with larger system there are *more basis functions, more wavefunctions* and each wavefunction has to be *orthogonal to all the others*
- in classical simulations we have typically $O(N)$
- it is possible to achieve $O(N)$ scaling $t_{CPU} = c_1N$ also in electronic structure calculations but the prefactor c_1 is very large
- crossover where the $O(N)$ methods win over $O(N^3)$ methods is still at large N
- with faster computers the importance of $O(N)$ methods will *grow* because we will get beyond the crossover point

Locality in quantum mechanics

- QM is *local* - properties of a certain region are only weakly influenced by factors which are far away from the region
- this is not taken into account in the theories based on *wavefunctions which extend over the whole system*

- single-particle density matrix in terms of KS orbitals $\psi_i(\mathbf{r})$

$$F(\mathbf{r}, \mathbf{r}') = \sum_i^{occ} f_i \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}') \quad \rho(\mathbf{r}) = F(\mathbf{r}, \mathbf{r})$$

$$T = -\frac{1}{2} \int d\mathbf{r}' \nabla_{\mathbf{r}}^2 F(\mathbf{r}, \mathbf{r}') \Big|_{\mathbf{r}=\mathbf{r}'}$$

$$V = \int d\mathbf{r} F(\mathbf{r}, \mathbf{r}) V(\mathbf{r})$$

- as $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$, $F(\mathbf{r}, \mathbf{r}') \rightarrow 0$, *exponentially in insulators and algebraically in metals*
- in principle it should be possible to achieve $O(N)$ scaling
- concept of *localization region* - quantities are evaluated only inside the region and neglected outside

Strategies for $O(N)$ scaling

- typically $O(N)$ algorithms are based either *directly on the density matrix* or its representation in terms of *localized* (Wannier) functions
- finite temperature density matrix is the Fermi operator

$$F = f[\beta(\hat{H} - \mu)] \text{ where } f(x) = \frac{1}{1 + \exp(x)}$$

- e.g. the *Fermi operator expansion method* (Goedecker and Colombo 1994) approximates the function $f(x)$ by a polynomial over the interval between the smallest and largest eigenvalue of \hat{H}

$$F = \sum_{p=0}^n c_p \hat{H}^p$$

- problems with *metals where the gap is zero*
- other methods are density matrix minimization, orbital minimization, divide-and-conquer methods, penalty functionals, ...
- $O(N)$ methods are typically used in connection with TB schemes where the prefactor is smaller due to much smaller number of basis states compared to DFT

Linear scaling method suitable also for metals Krajewski and Parinello 2006 [40]

- generic expression for the total energy in theories that can be formulated in an effective single particle form (e.g. Kohn-Sham, tight binding)

$$E = 2 \sum_{i=1}^N \epsilon_i + V_r$$

- first term is the so-called *band structure term*, ϵ_i are eigenvalues of a *single-particle Hamiltonian* H
- V_r corrects for double counting and accounts for the direct ion-ion interaction, can be calculated in $O(N)$ operations
- the band structure term has an apparent $O(N^3)$ complexity
- it is convenient to use *grand-canonical ensemble*
- grand canonical potential for independent fermions

$$\Omega = -\frac{2}{\beta} Tr \ln (1 + e^{\beta(\mu - \mathbf{H})}) = -\frac{2}{\beta} \ln \det (1 + e^{\beta(\mu - \mathbf{H})})$$

where the identity $Tr \ln \mathbf{A} = \ln \det \mathbf{A}$ was used

Linear scaling method suitable also for metals

- factorization

$$1 + e^{-\beta(\mu - \mathbf{H})} = \prod_{l=1}^{P/2} (\mathbf{M}_l^\dagger \mathbf{M}_l)$$

where P is an even integer and

$$\mathbf{M}_l = 1 + e^{i\pi(2l-1)/P} e^{\frac{\beta}{P}(\mu - \mathbf{H})}$$

- $\mathbf{M}_l^\dagger \mathbf{M}_l$ is a positive definite operator and the identity holds

$$\det (\mathbf{M}_l^\dagger \mathbf{M}_l)^{-\frac{1}{2}} = \frac{\int \mathcal{D}[\Phi_l] e^{-\Phi_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \Phi_l}}{\int \mathcal{D}[\Phi_l] e^{-\Phi_l^\dagger \Phi_l}}$$

- *field theoretical formulation*

$$\Omega = \frac{4}{\beta} \sum_{l=1}^{P/2} \ln \int \mathcal{D}[\Phi_l] e^{-\Phi_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \Phi_l} + \text{const.}$$

- the auxiliary field Φ_l has the *dimension of the full Hilbert space*

Linear scaling method suitable also for metals

- *no sign problem*
- observables can be calculated as derivatives of Ω , e.g. $N_e = -\frac{\partial}{\partial \mu} \Omega$
- the low temperature limit provides the band structure term

$$E^{\text{band}} = 2 \sum_{i=1}^N \epsilon_i = \lim_{\beta \rightarrow \infty} \frac{\partial}{\partial \beta} (\beta \Omega) + \mu N_e$$

- force on particle I at position \mathbf{R}_I coming from the band term

$$\mathbf{F}_I^{\text{band}} = -\nabla_{\mathbf{R}_I} \Omega$$

- general expression for the derivatives is

$$\frac{\partial \Omega}{\partial \lambda} = \frac{4}{\beta} \sum_{l=1}^{P/2} \frac{\int \mathcal{D}[\Phi_l] \left[\Phi_l^\dagger \left(\frac{\partial \mathbf{M}_l^\dagger}{\partial \lambda} \mathbf{M}_l + \mathbf{M}_l^\dagger \frac{\partial \mathbf{M}_l}{\partial \lambda} \right) \Phi_l \right] e^{-\Phi_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \Phi_l}}{\int \mathcal{D}[\Phi_l] e^{-\Phi_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \Phi_l}}.$$

- so far no approximation was made

Linear scaling method suitable also for metals

- high-temperature expansion can be made on $e^{\frac{\beta}{P}(\mu - \mathbf{H})}$ in \mathbf{M}_l

$$\mathbf{M}_l = 1 + e^{\pi(2l-1)/P} \left[1 + \frac{\beta}{P} (\mu - \mathbf{H}) \right] + \mathcal{O} \left(\left(\frac{\beta}{P} \right)^2 \right)$$

- we assume the case of *tight-binding* where \mathbf{H} is sparse
- the operator \mathbf{M}_l has the same *sparsity* of \mathbf{H} which leads to linear scaling
- no assumption was made on the character of the spectrum ϵ_i so the scheme is usable also for metals
- sampling of the integral can be done by drawing a sequence of normal distributed random numbers Ψ_l and computing Φ_l
- solving the equations $\mathbf{M}_l \Phi_l = \Psi_l$ is $O(N)$
- no diagonalization necessary
- the quantities calculated are *noisy - statistical error*
- in case of metals $\mathbf{M}_{l \approx \frac{P}{2}}$ can have eigenvalues close to zero

Sampling the auxiliary fields by Langevin dynamics

- the distribution $e^{-\Phi_l^\dagger \mathbf{M}_l^\dagger \mathbf{M}_l \Phi_l}$ can be sampled by *Langevin dynamics*

$$m_l \ddot{\Phi}_l = -\mathbf{M}_l^\dagger \mathbf{M}_l \Phi_l - \gamma_e \dot{\Phi}_l + \xi_l$$

m_l is a fictitious mass, γ_e is friction term

- the components α of the white random noise vector ξ_l obey the relation

$$\langle \xi_l^\alpha(0) \xi_l^\alpha(t) \rangle = 2m_l \gamma_e \delta(t)$$

- friction and random noise simulate collisions with the heat bath
- trajectory of the *stochastic differential equation* samples the *canonical ensemble*
- using the Liouville formalism the integration algorithms can be found
- no matrix inversion is necessary, the expensive operation is the matrix-vector product $\mathbf{M}_l \Phi_l$ which is $O(N)$ because \mathbf{M}_l is sparse
- since the inter-atomic forces are noisy, energy-conserving MD cannot be performed

Using the scheme for MD simulation

- the same Langevin dynamics can be used to sample also the ions

$$M \ddot{\mathbf{R}}_I = \mathbf{F}_I - \gamma_I \dot{\mathbf{R}}_I + \Xi_I$$

- the random force obeys the relations

$$\begin{aligned} \langle \Xi_I(0) \Xi_I(t) \rangle &= 6k_B T M \gamma_I \delta(t) \\ \langle \mathbf{F}_I(0) \Xi_I(t) \rangle &= 0 \end{aligned}$$

- from the electrons Langevin dynamics an approximation $\mathbf{F}_I^L = \mathbf{F}_I + \Xi_I^L$ to the exact forces \mathbf{F}_I is found
- assumption: the statistical error Ξ_I^L is a white noise

$$\begin{aligned} \langle \Xi_I^L(0) \Xi_I^L(t) \rangle &\cong 6k_B T M \gamma_I^L \delta(t) \quad \text{and} \\ \langle \mathbf{F}_I(0) \Xi_I^L(t) \rangle &\cong 0 \end{aligned}$$

- the noises Ξ_I^L and Ξ_I add together

$$M \ddot{\mathbf{R}}_I = \mathbf{F}_I - (\gamma_I + \gamma_I^L) \dot{\mathbf{R}}_I + \Xi_I + \Xi_I^L$$

Using the scheme for MD simulation

- the value of γ_I^L is unknown but can be tuned until the equipartition theorem $\langle \frac{1}{2} M \dot{\mathbf{R}}_I^2 \rangle = \frac{3}{2} k_B T$ is satisfied
- once the value of γ_I^L is determined it has to be kept constant during the simulation
- the assumptions about Ξ_I^L were explicitly verified for a small system and found to be satisfied
- even with noisy forces a *correct Boltzmann sampling* can be performed
- after each ionic displacement the fields Φ_l evolve under the action of the new $M_l^\dagger M_l$ until the distribution is equilibrated
- the fields Φ_l are used to calculate the ionic forces for the next integration step
- the chemical potential μ is continuously adjusted during the simulation to keep the number of electrons constant
- the algorithm can be *parallelized* easily

Example 24. Comparison of results for liquid and crystalline Si obtained by the linear scaling method and standard diagonalization

- crossover point for Si is at ~ 500 atoms
- 64 Si atoms, $P = 200$ was sufficient to make the approximation valid
- the crystalline Si is at 300 K (semiconductor)
- the liquid Si is at 3000 K (*metallic*)

11 Density functional perturbation theory

Literature: S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001)

Density functional perturbation theory (DFPT) Baroni, Giannozzi, Testa (1987) [41]

- we want to calculate also derivatives of the total energy with respect to *structural parameters* or *external perturbations*
- *linear response* in periodic systems
- structural parameters can be particle positions or unit cell
- response to microscopic particle displacements is related to *phonons*
- response to macroscopic strain is related to *elastic constants*
- external perturbation can be e.g. static or time-dependent electric field (polarization, dielectric constant)
- in principle, one can calculate the response via finite differences of energy
- for *long-wavelength* perturbations a large supercell would be needed
- the case of *external electric field* is particular
- we consider only *static* perturbations

General strategy

- the external potential $V(\{\lambda_i\})$ is a differentiable function of a set of parameters $\{\lambda_i\}$
- Hellman - Feynman theorem

$$\frac{\partial E}{\partial \lambda_i} = \int \frac{\partial V_\lambda(\mathbf{r})}{\partial \lambda_i} n_\lambda(\mathbf{r}) d\mathbf{r}$$

$$\frac{\partial^2 E}{\partial \lambda_i \partial \lambda_j} = \int \frac{\partial^2 V_\lambda(\mathbf{r})}{\partial \lambda_i \partial \lambda_j} n_\lambda(\mathbf{r}) d\mathbf{r} + \int \frac{\partial V_\lambda(\mathbf{r})}{\partial \lambda_i} \frac{\partial n_\lambda(\mathbf{r})}{\partial \lambda_j} d\mathbf{r}$$

- we need the electron-density response $\frac{\partial n_\lambda(\mathbf{r})}{\partial \lambda_i}$

$$\Delta n(\mathbf{r}) = 4 \operatorname{Re} \sum_{i=1}^{occ} \psi_i^*(\mathbf{r}) \Delta \psi_i(\mathbf{r}) = 4 \sum_{i=1}^{occ} \psi_i^*(\mathbf{r}) \Delta \psi_i(\mathbf{r})$$

because for a real potential $\psi_i(\mathbf{r})$ and $\psi_i^*(\mathbf{r})$ are degenerate

- finite-difference operator

$$\Delta^\lambda F = \sum_i \frac{\partial F_\lambda}{\partial \lambda_i} \Delta \lambda_i$$

Straightforward approach

- first-order perturbation theory correction to a given eigenfunction of the Schrödinger equation is

$$\Delta\psi_n(\mathbf{r}) = \sum_{m \neq n} \psi_m(\mathbf{r}) \frac{\langle \psi_m | \Delta V^{KS} | \psi_n \rangle}{\epsilon_n - \epsilon_m}$$

- the charge-density response

$$\Delta n(\mathbf{r}) = 4 \sum_{i=1}^{occ} \sum_{m \neq n} \psi_n^*(\mathbf{r}) \psi_m(\mathbf{r}) \frac{\langle \psi_m | \Delta V^{KS} | \psi_n \rangle}{\epsilon_n - \epsilon_m}$$

- contributions from occupied states cancel each other
- $n(\mathbf{r})$ does not respond to perturbations which only mix the occupied states *within the occupied-state manifold*
- a knowledge of the *full spectrum* of H^{KS} is required (no iterative diagonalization techniques)
- *extensive summations over empty conduction bands* have to be performed

Density functional perturbation theory approach

- variation of the KS orbitals and energies in the *first-order perturbation theory*

$$\begin{aligned} (H^{KS} - \epsilon_n) |\Delta\psi_n\rangle &= -(\Delta V^{KS} - \Delta\epsilon_n) |\psi_n\rangle \\ \Delta\epsilon_n &= \langle \psi_n | \Delta V^{KS} | \psi_n \rangle \end{aligned}$$

(Sternheimer equation)

- first-order correction to the self-consistent potential

$$\Delta V^{KS}(\mathbf{r}) = \Delta V(\mathbf{r}) + \int \frac{\Delta n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \left. \frac{dv_{xc}(n)}{dn} \right|_{n=n(\mathbf{r})} \Delta n(\mathbf{r})$$

- equations form a set of *self-consistent* equations for the perturbed system ($\Delta V^{KS}(\mathbf{r})$ and $\Delta n(\mathbf{r})$ must be selfconsistent)
- all equations are *coupled* because each $\Delta\psi_n$ depends through $\Delta n(\mathbf{r})$ on all other $\Delta\psi_m$, $m \neq n$
- *generalized linear problem*, can be solved iteratively, similar to standard KS problem

Density functional perturbation theory approach

- the operator $(H^{KS} - \epsilon_n)$ is singular because it has a *null eigenvalue*
- only the *projection on the empty-state manifold* is important
- we introduce projectors on occupied (v) and empty (c) KS states

$$P_v = \sum_{v\mathbf{k}} |\psi_v^{\mathbf{k}}\rangle \langle \psi_v^{\mathbf{k}}|$$
$$P_c = 1 - \sum_{v\mathbf{k}} |\psi_v^{\mathbf{k}}\rangle \langle \psi_v^{\mathbf{k}}|$$

- we make the operator nonsingular by adding αP_v

$$(H^{KS} + \alpha P_v - \epsilon_n) |\Delta\psi_n\rangle = -P_c \Delta V^{KS} |\psi_n\rangle$$

- we only discuss the case of *finite gap* (insulators, semiconductors)
- the case of metals requires modification

Monochromatic perturbation

- the perturbing potential can be decomposed into Fourier components

$$\Delta V^{KS}(\mathbf{r}) = \sum_{\mathbf{q}} \Delta v_{KS}^{\mathbf{q}}(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}}$$

- one can show that responses to perturbations of different wavelengths are *decoupled*
- for each wavevector \mathbf{q} a set of self-consistent equations can be solved in terms of *lattice-periodic functions* $u_{\mathbf{k}}(\mathbf{r})$ only
- the workload for each calculation is of the same order as that for the unperturbed system
- phonons can be calculated without the use of supercell

Response to particle displacement - Phonons

- we assume a *crystalline solid* with atoms at regular positions
- \mathbf{R}_μ is unit cell, τ_s is equilibrium position of the atom s within the unit cell and \mathbf{u}_s^μ is displacement

$$\mathbf{R}_s^\mu = \mathbf{R}_\mu + \tau_s + \mathbf{u}_s^\mu$$

- harmonic force-constants

$$\Phi_{\alpha\beta}^{ss_1}(\mathbf{R}_\mu, \mathbf{R}_{\mu_1}) = \frac{\partial^2 E^{tot}}{\partial \mathbf{u}_{s\alpha}^\mu \partial \mathbf{u}_{s_1\beta}^{\mu_1}} = \Phi_{\alpha\beta}^{ss_1}(\mathbf{R}_\mu - \mathbf{R}_{\mu_1})$$

- $\Phi_{\alpha\beta}^{ss_1}(\mathbf{R}_\mu - \mathbf{R}_{\mu_1})$ depend only on the distance $\mathbf{R}_\mu - \mathbf{R}_{\mu_1}$ (translational invariance)
- Fourier transform of $\Phi_{\alpha\beta}^{ss_1}(\mathbf{R}_\mu - \mathbf{R}_{\mu_1})$

$$\tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q}) = \sum_{\mathbf{R}} e^{-i\mathbf{q}\cdot\mathbf{R}} \Phi_{\alpha\beta}^{ss_1}(\mathbf{R}) = \frac{1}{N_c} \frac{\partial^2 E}{\partial \mathbf{u}_{s\alpha}^*(\mathbf{q}) \partial \mathbf{u}_{s_1\beta}(\mathbf{q})}$$

- distortion pattern

$$\mathbf{R}_s^\mu[\mathbf{u}_s(\mathbf{q})] = \mathbf{R}_\mu + \tau_s + \mathbf{u}_s(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{R}_\mu}$$

Response to particle displacement - Phonons

- $\tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q})$ has an electronic and ionic contribution

$$\begin{aligned} \tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q}) &= {}^{el} \tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q}) + {}^{ion} \tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q}) \quad \text{where} \\ {}^{el} \tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q}) &= \frac{1}{N_c} \left[\int \left(\frac{\partial n(\mathbf{r})}{\partial \mathbf{u}_{s\alpha}^*(\mathbf{q})} \right)^* \frac{\partial V^{ion}(\mathbf{r})}{\partial \mathbf{u}_{s_1\beta}(\mathbf{q})} d\mathbf{r} + \int n(\mathbf{r}) \frac{\partial^2 V^{ion}(\mathbf{r})}{\partial \mathbf{u}_{s\alpha}^*(\mathbf{q}) \partial \mathbf{u}_{s_1\beta}(\mathbf{q})} d\mathbf{r} \right] \\ V^{ion}(\mathbf{r}) &= \sum_{\mu s} v_s[\mathbf{r} - \mathbf{R}_\mu - \tau_s - \mathbf{u}_s^\mu] \end{aligned}$$

- $v_s[\mathbf{r}]$ is the ionic (pseudo-) potential
- the term ${}^{ion} \tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q})$ is independent on the electronic structure (calculated from Ewald summation)

$$\frac{\partial V^{ion}(\mathbf{r})}{\partial \mathbf{u}_{s_1\beta}(\mathbf{q})} = - \sum_{\mu} \frac{\partial v_s[\mathbf{r} - \mathbf{R}_\mu - \tau_s]}{\partial \mathbf{r}} e^{i\mathbf{q}\cdot\mathbf{R}_\mu}$$

- phonon frequencies are solutions of the *secular equation*

$$\det \left| \frac{1}{\sqrt{M_s M_t}} \tilde{\Phi}_{\alpha\beta}^{ss_1}(\mathbf{q}) - \omega^2(\mathbf{q}) \right| = 0$$

Response to particle displacement - Phonons

- the case of long-wavelength vibrations in polar materials has to be treated in a special way (LO-TO splitting)
- the accuracy of ab initio lattice dynamics calculations can compete with that of absorption or diffraction spectroscopies
- DFPT can provide predictions for systems that have not yet been synthesized

Response to external electric field

- dielectric constant is defined as

$$\epsilon_{\infty}^{\alpha\beta} = 1 + 4\pi \frac{\partial P_{\alpha}}{\partial E_{\beta}}$$
$$\mathbf{P} = \frac{-e}{V} \int d\mathbf{r} \mathbf{r} n(\mathbf{r})$$
$$\epsilon_{\infty}^{\alpha\beta} = 1 - 4\pi \frac{e}{V} \int d\mathbf{r} r_{\alpha} \frac{\partial n(\mathbf{r})}{\partial E_{\beta}}$$

- we need to calculate

$$\Delta^{\mathbf{E}} n(\mathbf{r}) = 4 \sum_{n=1}^{occ} \psi_n^*(\mathbf{r}) \Delta^{\mathbf{E}} \psi_n(\mathbf{r})$$

- QM is formulated in terms of *potentials*, not fields
- potential of the homogeneous electric field $V_{\mathbf{E}}(\mathbf{r}) = e\mathbf{E} \cdot \mathbf{r}$ is *unbounded from below* and *incompatible* with the periodic boundary conditions

$$\Delta V^{KS} = e\mathbf{E} \cdot \mathbf{r} + \Delta V_H + \Delta V_{XC}$$

Response to external electric field

- position operator \mathbf{r} and its matrix elements between states satisfying PBC are *ill-defined* (boundary dependent)

- to get the linear charge-density response we need only *off-diagonal* matrix elements

$$\Delta\psi_n(\mathbf{r}) = \sum_{m \neq n} \psi_m(\mathbf{r}) \frac{\langle \psi_m | \Delta V^{KS} | \psi_n \rangle}{\epsilon_n - \epsilon_m}$$

which can be calculated via

$$\langle \psi_m | \mathbf{r} | \psi_n \rangle = \frac{\langle \psi_m | [H^{KS}, \mathbf{r}] | \psi_n \rangle}{\epsilon_m - \epsilon_n}, \quad \forall m \neq n$$

- if V^{KS} is local, $[H^{KS}, \mathbf{r}]$ is proportional to the *momentum operator*

$$[H^{KS}, \mathbf{r}] = -\frac{\hbar^2}{m} \frac{\partial}{\partial \mathbf{r}} \quad (\text{in normal units})$$

- can be generalized also to non-local pseudopotential

12 Berry phase theory of macroscopic polarization in crystalline dielectrics

Literature: R. Resta, Rev. Mod. Phys. **66**, 899 (1994)

Berry phase Berry (1984) [42]

Geometric phase in quantum systems

- quantum Hamiltonian with a parametric dependence

$$\hat{H}(\xi) |\Psi(\xi)\rangle = E(\xi) |\Psi(\xi)\rangle$$

- parameter ξ can be quite general
- $|\Psi(\xi)\rangle$ is non-degenerate ground state for any ξ
- phase difference between two states *does not have any physical meaning*

$$e^{-i\Delta\phi_{12}} = \frac{\langle \Psi(\xi_1) | \Psi(\xi_2) \rangle}{|\langle \Psi(\xi_1) | \Psi(\xi_2) \rangle|}$$

$$\Delta\phi_{12} = -\text{Im} \ln \langle \Psi(\xi_1) | \Psi(\xi_2) \rangle$$

because the phase factors of the states are *arbitrary*

Berry phase

- total phase difference *along a closed path is gauge-invariant*

$$\begin{aligned}\gamma &= \Delta\phi_{12} + \Delta\phi_{23} + \Delta\phi_{34} + \Delta\phi_{41} \\ &= -\text{Im} \ln \langle \Psi(\xi_1) | \Psi(\xi_2) \rangle \langle \Psi(\xi_2) | \Psi(\xi_3) \rangle \langle \Psi(\xi_3) | \Psi(\xi_4) \rangle \langle \Psi(\xi_4) | \Psi(\xi_1) \rangle\end{aligned}$$

because *all arbitrary phases cancel in pairs*

- assume a continuous closed circuit C in the parameter space

$$\begin{aligned}e^{-i\Delta\phi_{12}} &= \frac{\langle \Psi(\xi) | \Psi(\xi + \Delta\xi) \rangle}{|\langle \Psi(\xi) | \Psi(\xi + \Delta\xi) \rangle|} \\ -i\Delta\phi &\approx \langle \Psi(\xi) | \nabla_{\xi} \Psi(\xi) \rangle \cdot \Delta\xi\end{aligned}$$

- in the continuum limit we get a *circuit integral* of the *Berry connection*

$$\gamma = \sum_{s=1}^N \Delta\phi_{s,s+1} \rightarrow \oint_C \langle \Psi(\xi) | \nabla_{\xi} \Psi(\xi) \rangle \cdot d\xi$$

- a gauge-invariant quantity is potentially a *physical observable*
- the phase γ cannot be expressed in terms of eigenvalues of any operator

Berry phase theory of macroscopic polarization

King-Smith, Vanderbilt (1993) [43], Resta (1993) [44]

- polarization of a periodic charge distribution is *ill-defined* (Martin 1974)
- changes of polarization $\Delta\mathbf{P}$ can be *unambiguously* defined (related to current)

$$\Delta\mathbf{P} = \int_0^1 d\lambda \frac{d\mathbf{P}}{d\lambda}$$

Berry phase theory of macroscopic polarization

- we decompose \mathbf{P} into ionic and electronic contributions

$$\begin{aligned}\Delta\mathbf{P} &= \Delta\mathbf{P}_{el} + \Delta\mathbf{P}_{ion} \\ \Delta\mathbf{P}_{el} &= \frac{1}{V} \int d\mathbf{r} \mathbf{r} \Delta\rho(\mathbf{r})\end{aligned}$$

- Wannier transformation to *localized* wavefunctions

$$a_n^{(\lambda)}(\mathbf{r}) = \frac{\sqrt{V}}{(2\pi)^3} \int d\mathbf{q} \psi_n^{(\lambda)}(\mathbf{q}, \mathbf{r})$$

- periodic charge difference $\Delta\rho = \rho^{(1)} - \rho^{(0)}$

$$\Delta\rho(\mathbf{r}) = 2e \sum_{n=1}^{occ} \sum_l [|a_n^{(1)}(\mathbf{r} - \mathbf{R}_l)|^2 - |a_n^{(0)}(\mathbf{r} - \mathbf{R}_l)|^2] \quad \mathbf{R}_l \text{ are lattice sites}$$

- $\Delta\rho$ is decomposed into sum of *localized* and *neutral* charge distributions
- dipole moment per cell

$$\Delta\mathbf{P}_{el} = \frac{2e}{V} \sum_{n=1}^{occ} \int d\mathbf{r} \mathbf{r} [|a_n^{(1)}(\mathbf{r})|^2 - |a_n^{(0)}(\mathbf{r})|^2]$$

Berry phase theory of macroscopic polarization

- the phases of the Bloch functions are arbitrary so it has to be demonstrated that $\Delta\mathbf{P}_{el}$ is unique
- periodic part u of the Bloch orbitals

$$u_n^{(\lambda)}(\mathbf{q}, \mathbf{r}) = e^{-i\mathbf{q}\cdot\mathbf{r}} \psi_n^{(\lambda)}(\mathbf{q}, \mathbf{r}) = \sqrt{V} e^{-i\mathbf{q}\cdot\mathbf{r}} \sum_l e^{i\mathbf{q}\cdot\mathbf{R}_l} a_n^{(\lambda)}(\mathbf{r} - \mathbf{R}_l)$$

- basic identity

$$\begin{aligned} \frac{i}{(2\pi)^3} \int d\mathbf{q} \langle u(\mathbf{q}) | \nabla_{\mathbf{q}} u(\mathbf{q}) \rangle &= \frac{i}{(2\pi)^3} \int d\mathbf{q} \int d\mathbf{r} u^*(\mathbf{q}, \mathbf{r}) \nabla_{\mathbf{q}} u(\mathbf{q}, \mathbf{r}) \\ &= i \frac{V}{(2\pi)^3} \int d\mathbf{q} \int d\mathbf{r} \sum_l e^{-i\mathbf{q}\cdot(\mathbf{R}_l - \mathbf{r})} a^*(\mathbf{r} - \mathbf{R}_l) \nabla_{\mathbf{q}} \sum_m e^{i\mathbf{q}\cdot(\mathbf{R}_m - \mathbf{r})} a(\mathbf{r} - \mathbf{R}_m) \\ &= i \frac{V}{(2\pi)^3} \int d\mathbf{r} \sum_{lm} a^*(\mathbf{r} - \mathbf{R}_l) a(\mathbf{r} - \mathbf{R}_m) i(\mathbf{R}_m - \mathbf{r}) \int d\mathbf{q} e^{-i\mathbf{q}\cdot(\mathbf{R}_l - \mathbf{r})} e^{i\mathbf{q}\cdot(\mathbf{R}_m - \mathbf{r})} \\ &= - \sum_m \int d\mathbf{r} (\mathbf{R}_m - \mathbf{r}) |a(\mathbf{r} - \mathbf{R}_m)|^2 = \int d\mathbf{r} \mathbf{r} |a(\mathbf{r})|^2 \end{aligned}$$

Berry phase theory of macroscopic polarization

- $u_n^{(\lambda)}(\mathbf{q})$ are eigenstates of the Hamiltonian

$$H^{(\lambda)}(\mathbf{q}) = \frac{1}{2m}(\mathbf{p} + \hbar\mathbf{q})^2 + V^{(\lambda)}(\mathbf{r})$$

- $u_n^{(\lambda)}(\mathbf{q})$ satisfy the relation $u_n^{(\lambda)}(\mathbf{q} + \mathbf{G}, \mathbf{r}) = \exp(-i\mathbf{G} \cdot \mathbf{r})u_n^{(\lambda)}(\mathbf{q}, \mathbf{r})$ where \mathbf{G}_i are reciprocal lattice vectors
- we define $\mathbf{q} = \xi_1 \mathbf{G}_1 + \xi_2 \mathbf{G}_2 + \xi_3 \mathbf{G}_3$ and $\xi_4 = \lambda$

$$\mathbf{G}_j \int d\mathbf{r} \mathbf{r} |a_n^{(\xi_4)}(\mathbf{r})|^2 = i \int d\xi_1 d\xi_2 d\xi_3 \langle u_n(\xi) | \frac{\partial}{\partial \xi_j} u_n(\xi) \rangle \quad j = 1, 2, 3$$

where $|u_n(\xi)\rangle = |u_n^{(\lambda)}(\mathbf{q})\rangle$ and the integral is performed over the unit cube

- we consider the *parametric Hamiltonian* $\hat{H}(\xi)$ in the 4D space of ξ
- we define the Berry connection

$$\chi(\xi) = i \sum_{n=1}^{occ} \langle u_n(\xi) | \nabla_{\xi} u_n(\xi) \rangle$$

Berry phase theory of macroscopic polarization

- we want to express the change of polarization $\Delta\mathbf{P}_{el}$ as a Berry phase

$$\gamma(C) = - \oint_C d\phi = - \oint_C \chi(\xi) \cdot d\xi$$

- we choose C to be a contour parallel to ξ_3 and ξ_4 axes at given values of ξ_1 and ξ_2
- the two sides parallel to ξ_4 give zero contribution since the segments differ by \mathbf{G}_3
- the two remaining sides give contribution

$$\gamma(C) = - \oint_C d\phi = \oint_{C_0} d\phi - \oint_{C_1} d\phi$$

where C_0 is $\xi = (\xi_1, \xi_2, x, 0)$ and C_1 is $\xi = (\xi_1, \xi_2, x, 1)$, $0 \leq x \leq 1$

- the component of $\Delta\mathbf{P}_{el}$ along \mathbf{G}_3 is

$$\mathbf{G}_3 \cdot \Delta\mathbf{P}_{el} = \frac{2e}{V} \int d\xi_1 d\xi_2 \gamma(C)$$

where the double integral is over a unit square

Berry phase theory of macroscopic polarization

- macroscopic polarization is a property of the *phase* of the wavefunctions and not of the periodic charge densities (where phase is lost)
- the Berry phase formalism is *useful for simulations* because it requires only the wavefunctions of the initial $\xi_4 = 0$ and final $\xi_4 = 1$ state
- applies only to the periodic case ($\mathbf{E} = 0$), cannot be used to calculate the static dielectric constant
- calculation of *infrared spectra* (Bernasconi, Silvestrelli, Parrinello 1998) [45]

$$\epsilon_2(\omega) = \frac{2\pi\omega}{3Vk_B T} \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle \mathbf{M}(t) \mathbf{M}(0) \rangle$$

where \mathbf{M} is the total dipole moment of the sample and the electronic contribution is calculated via the Berry phase scheme

- more modern version (Resta 1998) [46] *QM Position Operator in Extended Systems*

$$\langle X \rangle = \frac{L}{2\pi} \text{Im} \ln \langle \Psi_0 | e^{i\frac{2\pi}{L} \hat{X}} | \Psi_0 \rangle \quad \text{where} \quad \hat{X} = \sum_{i=1}^N x_i$$

Part IV

Path Integral Quantum Monte Carlo methods

13 Quantum vs. classical simulation

Why do we do classical simulations ?

Because we don't know how to perform the integration on high-dimensional phase space. In a classical simulation the *probability distribution* to be sampled on the phase space is known analytically.

$$p = \exp(-\beta E(\mathbf{R}, \mathbf{P})) \quad (1)$$

Simulation (e.g. *Monte Carlo* or *molecular dynamics* or other methods) just performs the *sampling* of the distribution in the multi-dimensional space.

Why do we do quantum simulations ?

Because in the quantum case, we *don't* even know analytically the distribution, just the underlying hamiltonian.

$$\hat{\rho} = \exp(-\beta\hat{H}(\hat{\mathbf{R}}, \hat{\mathbf{P}})) \quad (2)$$

The density matrix in the above expression is an operator which is *not diagonal* in $\hat{\mathbf{R}}$ and thus *cannot* be sampled directly. Simulation has to account for both the *function* to be sampled and the *sampling* itself.

QMC is more difficult.

Finite-temperature methods

- Path Integral QMC

calculate a trace over the thermal density matrix $\exp(-\beta H)$

Zero-temperature methods

- Variational QMC
- Projector Methods

calculate properties of a single wavefunction Ψ

these techniques can be applied to both continuum and lattice models

14 Imaginary time density matrix

Consider a non-relativistic system of N particles with only *translational* degrees of freedom interacting via a pair potential (for simplicity, not an essential condition) described by the Hamiltonian

$$H = -\frac{\hbar^2}{2m} \sum_{i=1}^N \nabla_i^2 + \sum_{i<j} v(r_{ij}) = K + V, \quad (3)$$

We want to study properties of this system in equilibrium at a finite temperature T . This means calculating averages of operators corresponding to physical quantities. Let $\{\Phi_i\}$ be the complete system of eigenfunctions of the hamiltonian with corresponding eigenvalues $\{E_i\}$, $H\Phi_i = E_i\Phi_i$. For the beginning we ignore the statistics of the particles and assume they are *distinguishable*.

Prescription of statistical mechanics

For inverse temperature $\beta = 1/k_B T$ we have density matrix ρ and partition function Z and we can calculate the thermal average of an operator \mathcal{O}

$$\begin{aligned}\rho &= \exp(-\beta H) \\ Z &= \text{Tr} \rho = \sum_i \exp(-\beta E_i) \\ \langle \mathcal{O} \rangle &= \frac{1}{Z} \text{Tr}[\rho \mathcal{O}] = \frac{1}{Z} \sum_i \langle \Phi_i | \mathcal{O} | \Phi_i \rangle \exp(-\beta E_i)\end{aligned}$$

However, this form is not very useful - we don't know neither $\{\Phi_i\}$ nor $\{E_i\}$ and there's no easy way to calculate them for non-trivial many body systems.

Note: Path integrals we will deal with from now on are sometimes called also imaginary time path integrals. This originates from the relation between the time evolution operator in quantum mechanics $U = \exp(-\frac{i}{\hbar} H t)$ and the equilibrium density matrix in statistical mechanics $\rho = \exp(-\beta H)$. Substituting for the real time t in the expression for U the imaginary time $t \rightarrow -i\beta\hbar$ (Wick rotation) we obtain ρ .

Let's try a different route.

$$\begin{aligned}\langle \mathcal{O} \rangle &= \frac{1}{Z} \int d\mathbf{R} \int d\mathbf{R}' \langle \mathbf{R} | \exp(-\beta H) | \mathbf{R}' \rangle \langle \mathbf{R}' | \mathcal{O} | \mathbf{R} \rangle \\ &= \frac{1}{Z} \int d\mathbf{R} \int d\mathbf{R}' \rho(\mathbf{R}, \mathbf{R}'; \beta) \langle \mathbf{R}' | \mathcal{O} | \mathbf{R} \rangle ,\end{aligned}$$

where we introduced the position representation of the density matrix $\rho(\mathbf{R}, \mathbf{R}'; \beta) = \langle \mathbf{R} | \exp(-\beta H) | \mathbf{R}' \rangle$. Product of two density matrices is again a density matrix.

$$\exp[-(\beta_1 + \beta_2)H] = \exp(-\beta_1 H) \exp(-\beta_2 H) \quad (4)$$

This can be translated into a convolution property for matrix elements of the density matrix

$$\rho(\mathbf{R}_1, \mathbf{R}_3; \beta) = \int d\mathbf{R}_2 \rho(\mathbf{R}_1, \mathbf{R}_2; \beta_1) \rho(\mathbf{R}_2, \mathbf{R}_3; \beta_2) \quad (5)$$

In order to obtain a path-integral representation, we now write $\beta = P\tau$, where $\tau = \beta/P$ is the *time step*. The integer number P is the *Trotter number*. The operator identity

$$\begin{aligned}\exp(-\beta H) &= (\exp(-\tau H))^P \\ &= \exp(-\tau H) \exp(-\tau H) \dots \exp(-\tau H)\end{aligned}$$

now can be translated into the following *exact* convolution property for the density matrix

$$\begin{aligned} \rho(\mathbf{R}_0, \mathbf{R}_P; \beta) &= \int \dots \int d\mathbf{R}_1 d\mathbf{R}_2 \dots d\mathbf{R}_{P-1} \\ &\times \rho(\mathbf{R}_0, \mathbf{R}_1; \tau) \rho(\mathbf{R}_1, \mathbf{R}_2; \tau) \dots \rho(\mathbf{R}_{P-1}, \mathbf{R}_P; \tau) \end{aligned}$$

For P finite we have a discrete time-path $\{\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_P\}$, taking the limit $P \rightarrow \infty$ we get a continuous path.

15 Path Integral Monte Carlo (PIMC)

What did we gain ?

Quite a lot. The problem is how to evaluate an exponential of an operator consisting of two parts which do not commute.

$$\begin{aligned} H &= K + V, [K, V] \neq 0 \Rightarrow \\ \exp(-\beta H) &= \exp(-\beta(K + V)) \neq \exp(-\beta K) \exp(-\beta V) \end{aligned}$$

But the density matrices $\rho(\mathbf{R}_{i-1}, \mathbf{R}_i; \tau)$ in the representation (6) are at inverse temperature $\tau = \beta/P$, and therefore correspond to temperature PT which is P times *higher* than the true temperature T .

At high temperatures quantum mechanics becomes equivalent to the classical mechanics !

For large P , $\tau = \beta/P$ is small enough and we can apply the *short time approximation* for $\rho(\mathbf{R}_{i-1}, \mathbf{R}_i; \tau)$ which consists of writing

$$\exp(-\tau(K + V)) \approx \exp(-\tau K) \exp(-\tau V) . \quad (6)$$

The convergence of this approximation for large P is guaranteed by the Trotter formula

$$\exp(-\beta(K + V)) = \lim_{P \rightarrow \infty} [\exp(-\tau K) \exp(-\tau V)]^P \quad (7)$$

This is the simplest approximation possible and is called *the primitive approximation*. Inserting the decomposition of the identity operator $\mathbf{1} = \int d\mathbf{R}'_1 |\mathbf{R}'_1\rangle \langle \mathbf{R}'_1|$ in the position representation we get for the short time density matrix

$$\begin{aligned} \rho(\mathbf{R}_1, \mathbf{R}_2; \tau) &= \langle \mathbf{R}_1 | \exp(-\tau H) | \mathbf{R}_2 \rangle \approx \\ &\int d\mathbf{R}'_1 \langle \mathbf{R}_1 | \exp(-\tau K) | \mathbf{R}'_1 \rangle \langle \mathbf{R}'_1 | \exp(-\tau V) | \mathbf{R}_2 \rangle \end{aligned}$$

The potential energy operator is diagonal in the position representation

$$\langle \mathbf{R}'_1 | \exp(-\tau V) | \mathbf{R}_2 \rangle = \exp(-\tau V(\mathbf{R}_2)) \delta(\mathbf{R}'_1 - \mathbf{R}_2) , \quad (8)$$

therefore

$$\rho(\mathbf{R}_1, \mathbf{R}_2; \tau) \approx \langle \mathbf{R}_1 | \exp(-\tau K) | \mathbf{R}_2 \rangle \exp(-\tau V(\mathbf{R}_2)) . \quad (9)$$

The matrix element of the exponential of the kinetic energy operator is just the *free particle density matrix*. For a single particle we can calculate it e.g. by inserting twice the resolution of identity operator in the momentum representation (another possibility is to solve the Bloch equation)

$$\begin{aligned} & \langle \mathbf{r}_1 | \exp(-\tau K) | \mathbf{r}_2 \rangle \\ &= \int d\mathbf{p} \int d\mathbf{p}' \langle \mathbf{r}_1 | \mathbf{p} \rangle \langle \mathbf{p} | \exp(-\tau K) | \mathbf{p}' \rangle \langle \mathbf{p}' | \mathbf{r}_2 \rangle \\ &= \int d\mathbf{p} \int d\mathbf{p}' \langle \mathbf{r}_1 | \mathbf{p} \rangle \langle \mathbf{p} | \mathbf{p}' \rangle \langle \mathbf{p}' | \mathbf{r}_2 \rangle \exp(-\tau \frac{\mathbf{p}^2}{2m}) \\ &= \frac{1}{(2\pi\hbar)^3} \int d\mathbf{p} \exp\left(\frac{i}{\hbar} \mathbf{p} \cdot (\mathbf{r}_1 - \mathbf{r}_2) - \tau \frac{\mathbf{p}^2}{2m}\right) \\ &= \left(\frac{Pm}{2\pi\beta\hbar^2}\right)^{\frac{3}{2}} \exp\left[-\frac{Pm}{2\beta\hbar^2}(\mathbf{r}_1 - \mathbf{r}_2)^2\right] \end{aligned}$$

Using these expressions we obtain the discrete path-integral representation of the density matrix in the primitive approximation

$$\begin{aligned} \rho_P(\mathbf{R}_0, \mathbf{R}_P; \beta) &= \left(\frac{Pm}{2\pi\beta\hbar^2}\right)^{\frac{3NP}{2}} \int d\mathbf{R}_1 \dots \int d\mathbf{R}_{P-1} \\ &\times \exp\left(-\sum_{i=1}^P \left[\frac{Pm}{2\beta\hbar^2}(\mathbf{R}_{i-1} - \mathbf{R}_i)^2 + \frac{\beta}{P}V(\mathbf{R}_i)\right]\right) \end{aligned}$$

Using this form, we get for the approximate partition function

$$\begin{aligned} Z_P &= \int d\mathbf{R}_0 \rho_P(\mathbf{R}_0, \mathbf{R}_0; \beta) = \left(\frac{Pm}{2\pi\beta\hbar^2}\right)^{\frac{3NP}{2}} \int d\mathbf{R}_1 \dots \\ &\int d\mathbf{R}_P \exp\left(-\sum_{i=1}^P \left[\frac{Pm}{2\beta\hbar^2}(\mathbf{R}_{i-1} - \mathbf{R}_i)^2 + \frac{\beta}{P}V(\mathbf{R}_i)\right]\right) \\ &= \left(\frac{Pm}{2\pi\beta\hbar^2}\right)^{\frac{3NP}{2}} \int d\mathbf{R}_1 \dots \int d\mathbf{R}_P \exp(-\beta H_{eff}^P) , \end{aligned}$$

where we set $\mathbf{R}_0 = \mathbf{R}_P$ (periodic boundary condition along the Trotter direction). It has a form of a *classical configurational integral*.

The last expression represents a mapping from a quantum to a classical system with a temperature dependent effective hamiltonian (or action) given by

$$H_{eff}^P = \sum_{i=1}^P \left[\frac{Pm}{2\beta^2\hbar^2} (\mathbf{R}_{i-1} - \mathbf{R}_i)^2 + \frac{1}{P} V(\mathbf{R}_i) \right] \quad (10)$$

Therefore we can use in principle *all* simulation techniques developed for classical systems to sample the path integral representation of the quantum system.

Trotter approximation is the *only* approximation used in the derivation of the path integral representation. For $P \rightarrow \infty$ it becomes again exact and equal to the Feynman path integral over the space of continuous paths. For finite P it is *under control*, since for any inverse temperature β we can make the time step τ arbitrarily small by increasing P and thus improve the accuracy of the path-integral representation to any extent (in principle).

The price we have to pay for the mapping quantum \rightarrow classical are the additional integrations. We have to deal with P “replicas” of the original system and altogether with $3NP$ degrees of freedom instead of original $3N$. The configuration $\{\mathbf{R}_i\} = \{\mathbf{r}_i^1, \mathbf{r}_i^2, \dots, \mathbf{r}_i^N\}$ will be referred to as the i -th *Trotter slice*.

Few remarks on the Trotter number P

- P depends on the properties of the system and temperature
- the lower is the temperature T , the higher is the value of P necessary for a given accuracy. It is therefore impossible to reach zero temperature. *Path integral is not a ground state method !*
- the stronger are the quantum effects, the higher is the required value of P .
- for $P = 1$, we obtain the classical system.
- in practical applications, P is typically in the range 10 - 100.
- we come back to this point later in more detail.

Properties and physical meaning of the mapping[47]

Each particle \mathbf{r}^i of the system is now represented by P particles

$$\mathbf{r}^i \rightarrow \{\mathbf{r}_k^i\}_{k=1}^P = \{\mathbf{r}_1^i, \mathbf{r}_2^i, \dots, \mathbf{r}_P^i\}. \quad (11)$$

This can be interpreted as a chain of “beads”, or a polymer ring. The effective classical hamiltonian H_{eff}^P consists of two parts.

- interactions

$$\sum_{i=1}^P \left[\frac{Pm}{2\beta^2\hbar^2} (\mathbf{R}_{i-1} - \mathbf{R}_i)^2 \right] = \frac{Pm}{2\beta^2\hbar^2} \sum_{i=1}^P \sum_{k=1}^N (\mathbf{r}_{i-1}^k - \mathbf{r}_i^k)^2 \quad (12)$$

can be regarded as *harmonic* springs acting between “beads” representing the same particle in successive Trotter slices (nearest neighbours along the Trotter direction). The higher P , the *stronger* become the springs.

- interactions $\sum_{i=1}^P \frac{1}{P} V(\mathbf{R}_i)$ act *only between “beads” in the same Trotter slice* and are equal to the potential energy of the quantum system *divided by* P . The higher P , the *weaker* become these interactions.

The chain of beads connected with springs thus can be regarded as *polymer*. Because of the periodic boundary condition imposed along the Trotter direction, $\mathbf{R}_0 = \mathbf{R}_P$, the chain must come back to its initial point after P steps, so it is a *ring polymer*. The polymers interact with each other in a particular way - only within the same Trotter slice.

The path integral defines an isomorphism between the quantum system and the classical system of *interacting ring polymers*. This provides us with a method to understand many properties of quantum systems entirely in terms of classical statistical mechanics.

How to calculate the averages of observables ?

The path-integral mapping allows us to easily calculate the average of any operator which is *diagonal* in the position representation.

$$\begin{aligned} \langle \mathcal{O} \rangle_P &= \frac{1}{Z_P} \int d\mathbf{R} \rho_P(\mathbf{R}, \mathbf{R}; \beta) \langle \mathbf{R} | \mathcal{O} | \mathbf{R} \rangle \\ &= \frac{1}{Z_P} \left(\frac{Pm}{2\pi\beta\hbar^2} \right)^{\frac{3NP}{2}} \int d\mathbf{R}_1 \dots \int d\mathbf{R}_P \langle \mathbf{R}_P | \mathcal{O} | \mathbf{R}_P \rangle \\ &\times \exp\left(-\sum_{i=1}^P \left[\frac{Pm}{2\beta\hbar^2} (\mathbf{R}_{i-1} - \mathbf{R}_i)^2 + \tau V(\mathbf{R}_i) \right]\right) \end{aligned}$$

All Trotter slices are equal, therefore we can take the average of the operator \mathcal{O} in the last expression in all slices and divide by P .

$$\begin{aligned} \langle \mathcal{O} \rangle_P &= \frac{1}{Z_P} \left(\frac{Pm}{2\pi\beta\hbar^2} \right)^{\frac{3NP}{2}} \int d\mathbf{R}_1 \dots \int d\mathbf{R}_P \\ &\frac{1}{P} \sum_{j=1}^P \langle \mathbf{R}_j | \mathcal{O} | \mathbf{R}_j \rangle \exp\left(-\sum_{i=1}^P \left[\frac{Pm}{2\beta\hbar^2} (\mathbf{R}_{i-1} - \mathbf{R}_i)^2 + \tau V(\mathbf{R}_i) \right]\right) \end{aligned}$$

In this way we can calculate any correlation functions, e.g. structure-related quantities. We simply calculate the quantity on each Trotter slice like in a classical simulation, and then take the average over all Trotter slices.

How to get an estimator for internal energy ?

Internal energy contains apart from potential also the kinetic energy contribution, which is not diagonal in the position representation. The easiest way to derive the estimator is via thermodynamics.

$$E_P = -\frac{1}{Z_P} \frac{\partial Z_P}{\partial \beta} \quad (13)$$

The effective hamiltonian itself is temperature dependent. Taking the derivative we get

$$E_P = \frac{3NP}{2\beta} + \frac{1}{Z_P} \left(\frac{Pm}{2\pi\beta\hbar^2} \right)^{\frac{3NP}{2}} \int d\mathbf{R}_1 \dots \int d\mathbf{R}_P \\ \times \left[\sum_{i=1}^P \left(-\frac{Pm}{2\beta^2\hbar^2} (\mathbf{R}_{i-1} - \mathbf{R}_i)^2 + \frac{1}{P} V(\mathbf{R}_i) \right) \right] \exp(-\beta H_{eff}^P)$$

The expression in the square bracket can be seen as *estimator*, or quantity which has to be averaged over the probability distribution $\exp(-\beta H_{eff}^P)$ in order to get the internal energy of the system. The last equation can thus be written as

$$E_P = \frac{3NP}{2\beta} + \left\langle \frac{1}{P} \sum_{i=1}^P \left[-\frac{m}{2} \sum_{k=1}^N \left(\frac{\mathbf{r}_{i-1}^k - \mathbf{r}_i^k}{\hbar\tau} \right)^2 + V(\mathbf{R}_i) \right] \right\rangle$$

Clearly, the expression

$$K = \frac{3NP}{2\beta} - \left\langle \frac{1}{P} \sum_{i=1}^P \left[\frac{m}{2} \sum_{k=1}^N \left(\frac{\mathbf{r}_{i-1}^k - \mathbf{r}_i^k}{\hbar\tau} \right)^2 \right] \right\rangle$$

corresponds to kinetic energy (primitive estimator). At first look it might look strange, because of the *negative sign* in front of the energy of the springs. The “faster” the particle moves along the path between the slice $i-1$ and i , the lower is the kinetic energy. The paradox is resolved when one realizes that we are dealing with *imaginary time* path integral, so the sign is reverted. In fact, it’s consistent with the uncertainty principle: the more delocalized the particle is, the lower the kinetic energy. When the particle is confined to stay within a small region of space, the kinetic energy is large.

While the last expression has a clear physical interpretation, it is not very useful for practical computation. Kinetic energy arises as a *difference* between

two quantities which both diverge as P . A statistical error in the average will thus be amplified by this factor and for large P serious problems show up.

How to get a good estimator for internal energy ?

A solution to the above problem was provided in Ref.[48]. The idea is to extract the kinetic energy directly from the potential energy, without explicit reference to the energy of the harmonic springs. In classical statistical mechanics, the *virial* theorem connects the two quantities. Applying the theorem to path integrals, one finds a new estimator for kinetic energy

$$K = \left\langle \frac{1}{2P} \sum_{i=1}^P \mathbf{R}_i \cdot \frac{\partial V(\mathbf{R}_i)}{\partial \mathbf{R}_i} \right\rangle \quad (14)$$

which is *exact* and has much smaller variance than the primitive estimator. It is often used in real simulations.

Specific heat can be either calculated from internal energy via finite differences, or one can derive a fluctuation formula directly from the path integral representation.

What happens when P is large ?

Computationally, the workload of PIMC scales linearly with P . The higher P , the smaller is the error introduced by the Trotter approximation. Thus one might think it's easy to improve the accuracy by using high values of P .

However, this is not true,

because of the particular properties of the path-integral action. The harmonic term grows linearly with P , while the potential energy term decreases as $\frac{1}{P}$. For large P therefore the harmonic term becomes completely *dominant* while the potential energy term becomes just a *small perturbation*. In a perfectly harmonic system, there is *no exchange of energy* between the normal modes. If the perturbation on the harmonic term is small, which is just the case here for large P , the energy exchange between the normal modes will be small. The system will enter the so called KAM (Kolmogorov - Arnold - Moser) regime, characterized by *loss of ergodicity* [49]. If one used molecular dynamics to sample the action, it would fail completely to sample the phase space, since the time average *will not agree* with the phase space average. With Monte Carlo, the problem is to some extent less severe, since there's no deterministic dynamics. However, even here the stiff springs cause a trouble since they allow only small displacements of the single particle moves resulting in a large auto-correlation times and non-efficient sampling.

Convergence of the path integral representation in P

For any finite P , the path integral representation is approximate. How does it converge to exact quantum results when $P \rightarrow \infty$?

In case of the primitive approximation the answer for the partition function Z and average of a Hermitian operator \mathcal{O} is provided by the formulas (Fye 1986) [50]

$$\begin{aligned} Z_P &= Z + O(\tau^2) = Z + O\left[\left(\frac{\beta}{P}\right)^2\right] = Z + O(P^{-2}) \\ \langle \mathcal{O} \rangle_P &= \langle \mathcal{O} \rangle + O(\tau^2) = \langle \mathcal{O} \rangle + O\left[\left(\frac{\beta}{P}\right)^2\right] \\ &= \langle \mathcal{O} \rangle + O(P^{-2}) \end{aligned}$$

The last result is extremely useful, since it explicitly provides the asymptotic scaling form as a function of P of the approximate averages. It is therefore not necessary to reach the regime where the correction is negligible. Instead it is sufficient to reach the asymptotic $O(P^{-2})$ regime. In practice, one usually performs several simulations with different values of P . If the results for a given quantity are equal for different values of P within the statistical error, the quantity can be considered converged and its value equal to the exact quantum value. If the values for different P differ considerably, one can try to plot them against $O(P^{-2})$ and if they collapse on a straight line, extrapolate to $P \rightarrow \infty$.

More specifically, it can be proved [51] that the expansion of averages is an even function of $1/P$, and one can write

$$\langle \mathcal{O} \rangle_P = \langle \mathcal{O} \rangle + \frac{a}{P^2} + \frac{b}{P^4} + \frac{c}{P^6} + \dots \quad (15)$$

so even when the asymptotic $O(P^{-2})$ regime is not reached, one can try to fit and extrapolate the simulation data provided they are available for several values of P .

16 Strategies to solve the “large P problem”

There are several possibilities aimed at different aspects of the problem.

1. Improving on the action
2. Improving on the sampling
3. Improving on the extrapolation to $P \rightarrow \infty$

Improving on the action (how to improve the accuracy of the path integral representation while keeping P low)

The decomposition of the hamiltonian in K and V we used (primitive approximation) is the simplest possible, and usually does not provide the best accuracy fo a given P . It is useful to include in the action as many of its exact properties as possible.

In the case of liquid He, the primitive approximation would require a Trotter number as large as $P = 1000$ in order to reach the temperature of the λ transition. With improved action, the actual simulation was performed with $P = 20$ [52].

Some methods to improve on the action

- renormalization of the pair density matrix [52]
- matrix squaring method [52]
- higher order approximations [51, 53]

Improving on the sampling of the path integral

In principle, we have to with ordinary multidimensional integration. It should therefore be possible to apply any standard sampling method like Monte Carlo or molecular dynamics. However, due to the already mentioned problem with ergodicity, sampling might not always be efficient.

How to remedy the ergodicity problem ?

The origin of the stiffness of the springs is the very form of the free particle propagator.

$$\begin{aligned} & \langle \mathbf{r}_1 | \exp(-\tau K) | \mathbf{r}_2 \rangle \\ &= \left(\frac{Pm}{2\pi\beta\hbar^2} \right)^{\frac{3}{2}} \exp \left[-\frac{Pm}{2\beta\hbar^2} (\mathbf{r}_1 - \mathbf{r}_2)^2 \right] \\ &= \left(\frac{1}{2\pi\lambda^2} \right)^{\frac{3}{2}} \exp \left[-\frac{1}{2\lambda^2} (\mathbf{r}_1 - \mathbf{r}_2)^2 \right] \end{aligned}$$

where $\lambda^2 = \frac{\beta\hbar^2}{Pm} = \frac{\tau\hbar^2}{m}$ is the *thermal wavelength* corresponding to the inverse temperature τ . For $|\mathbf{r}_1 - \mathbf{r}_2| \gtrsim \lambda$, the value of the propagator drops to zero very fast. Since $\lambda \rightarrow 0$ as P grows, the single particle moves will be constrained to smaller and smaller displacements.

In the classical limit $\hbar \rightarrow 0$, $\lambda \rightarrow 0$ and the springs will become infinitely stiff, constraining the beads representing a given atom to have the same position in *all* Trotter slices (chain will collapse). There will be no quantum fluctuations, but one would still like to sample the *classical phase space*. The only way to do so is to move all the beads representing a given atom in *all* Trotter slices *simultaneously*. Such move is, however, a *multislice* or collective move. We recover back the usual Monte Carlo sampling; we have P beads each feeling the potential U/P , which are constrained to move simultaneously.

Even when we are not in the classical limit, it is *extremely* useful to employ such *multislice* moves. Moving all the beads representing a given atom in all Trotter slices simultaneously, one does *not* deform the springs; $(\mathbf{r}_i + \Delta) - (\mathbf{r}_{i+1} + \Delta) = \mathbf{r}_i - \mathbf{r}_{i+1}$, so the internal structure of the path is not changed. Therefore the displacements of these *multislice* moves are *not directly constrained* by the thermal wavelength λ , but rather by the structure of the potential $U(\mathbf{R})$ and usually can be much larger. Such translational moves of whole chains can be interpreted as sampling the *classical* degrees of freedom of the center of mass of the chains. On the other hand, the local moves where each bead is displaced independently, alter the structure of the path and can be interpreted as sampling the internal, or *quantum* degrees of freedom of the chain. Both kinds of moves can be mixed either at random, or on a regular basis; e.g. after performing a local move on each bead of a chain, one performs a collective move on the whole chain. The presence of collective moves considerably improves the sampling and decreases the autocorrelation time of the random walk.

Fourier space and normal-mode sampling.

It's possible to further generalize the above idea. The moves of a whole chain actually represent moves of the $k = 0$ Fourier mode of the chain. One can decompose the total kinetic action (springs) in normal modes and instead of acting on the beads act directly on *all* the normal modes. A different approach is to express the *exact* path integral representation (in the limit $P \rightarrow \infty$) as integrals over the Fourier components. In this way one obtains an *infinite* number of modes; an approximation is introduced by truncating the number of modes in order to make the problem tractable. Instead of discretizing the path in M time steps, one truncates the number of modes, so the problem is how to approximate the effect of the neglected modes. Partial averaging [54].

Quite generally, in MC we have the freedom to choose the moves, or even invent special ones, provided we satisfy the detailed balance. Several sophisticated methods have been designed, in particular in connection with liquid He simulation.

- multilevel Metropolis method [52]
- bisection method [52]
- staging [55]
- PIMD with Nose Hoover thermostat. We will come back to this again in connection with ab-initio path integrals.

Improved extrapolation to $P \rightarrow \infty$

For systems with stiff degrees of freedom, like covalent bonds in molecules or solids, at low temperatures it might be difficult to reach the P^{-2} scaling regime

unless a very large value of P is used. Remarkably, the Trotter approximation is not exact even for a harmonic oscillator, so if we wanted to simulate a stiff one at a low temperature, we would have to use a large P in order to recover a trivial quantum mechanical result. However, for a harmonic oscillator the exact results for a *finite* P can be calculated analytically.

This fact is a basis of the improved extrapolation method of Ref.[56]. Often the dominant part of the pronounced systematic P dependence of the path integral results comes from hard, high frequency modes of the system, which are almost harmonic. Provided a harmonic approximation of the system is known, it is possible to perform its finite P evaluation and correct for the systematic P dependence of the raw PIMC data. The P dependence of the corrected data is then considerably less pronounced and the extrapolation to $P \rightarrow \infty$ can be performed with much lower values of P . For strongly anharmonic systems, it is possible to use the *self-consistent harmonic approximation*.

17 Techniques for treating rotational degrees of freedom

Techniques presented so far discretized the full kinetic energy in cartesian coordinates. This is suitable for systems where all degrees of freedom are translational, which is in principle always applicable in condensed matter systems consisting of atoms. Sometimes, however, it is useful to treat some subsystems of the system as *rigid objects* and express the hamiltonian of the system also in terms of rotational degrees of freedom. Typical case are molecular crystals or molecular adsorbates on surfaces. Usually the intramolecular degrees of freedom are very stiff, with high vibration frequencies corresponding to temperatures of the order of 10^3 K. Even at room temperature these are completely frozen and thus *decoupled* from translations and rotations of the molecule. Under such conditions, it is well justified to treat the molecules as *rigid rotors* characterized apart from mass also by *moment of inertia* which can be fully represented by the coordinates of the centre of mass and a set of angles specifying the orientation.

Apart from physical motivation, there's a lot of technical reasons for using rigid molecule models. Neglecting the stiff vibrational degrees of freedom, the Trotter number can be considerably reduced with respect to the values necessary to treat fully flexible models in Cartesian coordinates. This brings a possibility of much longer runs within a given amount of CPU time and results in much better statistics. Moreover, these systems often exhibit phase transitions involving the orientational degrees of freedom. These are typically induced by relatively weak *dipolar or quadrupolar* interactions and therefore the ordering can only

take place at a low temperature, where the quantum effects on the rotations also become important. Such orientational phase transitions involve long correlation lengths, and therefore exhibit pronounced finite-size effects. In order to extract information about the nature of the transition, it is often important to be able to simulate *large* systems and perform finite-size scaling. This has only been possible due to use of the rigid rotor representation of the molecules.

We will assume that the system is described by the hamiltonian

$$H = T^{tr} + T^{rot} + V \quad (16)$$

where

$$T^{rot} = \sum_{i=1}^{d^{rot}} \frac{L_i^2}{2\Theta_{ii}} \quad (17)$$

and L_i are the components of the angular momentum operator and Θ_{ii} are moment(s) of inertia of the rigid molecule. d^{rot} is the number of angles necessary to specify the orientation of the molecule.

Now the decomposition of the identity operator can be written as $\mathbf{1} = \int d\mathbf{R}d\omega |\mathbf{R}\omega\rangle\langle\mathbf{R}\omega|$, where ω represents the angles specifying the orientation. We can write the short-time approximation

$$\begin{aligned} \exp(-\tau(T^{tr} + T^{rot} + V)) = \\ \exp(-\frac{\tau}{2}V) \exp[-\tau(T^{tr} + T^{rot})] \exp(-\frac{\tau}{2}V) . \end{aligned}$$

where we treated the potential energy operator in a more symmetric way. In the rigid rotor approximation, $[T^{tr}, T^{rot}] = 0$ and so we can write

$$\begin{aligned} \rho(\mathbf{R}_1\omega_1, \mathbf{R}_2\omega_2; \tau) &= \langle\mathbf{R}_1\omega_1| \exp(-\tau H) |\mathbf{R}_2\omega_2\rangle \approx \\ &\langle\mathbf{R}_1\omega_1| \exp[-\frac{\tau}{2}V(\mathbf{R}_1)] \exp(-\tau T^{tr}) \exp(-\tau T^{rot}) \\ &\exp[-\frac{\tau}{2}V(\mathbf{R}_2)] |\mathbf{R}_2\omega_2\rangle = \exp[-\frac{\tau}{2}(V(\mathbf{R}_1) + V(\mathbf{R}_2))] \\ &\langle\mathbf{R}_1| \exp(-\tau T^{tr}) |\mathbf{R}_2\rangle \langle\omega_1| \exp(-\tau T^{rot}) |\omega_2\rangle \end{aligned}$$

The term $\langle\omega_1| \exp(-\tau T^{rot}) |\omega_2\rangle$ is *new* and represents a propagator for rotational motion, or free *rotor* density matrix (sometimes also called kernel). Generally, this is more difficult to evaluate than the free *particle* density matrix (propagator for translational motion). For simplicity of the notation, we will now concentrate on a single free rotor without loss of generality. It is useful to treat separately the three cases depending on the dimensionality of the rotation.

One-dimensional rotation (linear molecule confined to rotate in a plane)

$$d^{rot} = 1$$

There's just one angle $0 \leq \phi < 2\pi$ specifying the orientation of the molecule. The hamiltonian reads

$$T^{rot} = -B \frac{\partial^2}{\partial \phi^2}, \quad (18)$$

where $B = \frac{\hbar^2}{2\Theta}$ is the *rotational constant* and Θ is the moment of inertia of the rotor.

Note. Formally the same hamiltonian appears in the context of Josephson junctions, where the particle number operator conjugate to the superconducting phase operator $\hat{\phi}$ is $\hat{n} = -i \frac{\partial}{\partial \phi}$. The electrostatic charging energy then reads $E = \frac{Q^2}{2C} = \frac{(en)^2}{2C} = -E_c \frac{\partial^2}{\partial \phi^2}$, where $E_c = \frac{e^2}{2C}$.

The hamiltonian (18) has an orthonormal set of eigenfunctions $\langle \phi | m \rangle = \frac{1}{\sqrt{2\pi}} e^{im\phi}$ with eigenvalues $E_m = Bm^2$. We can evaluate the free *rotor* density matrix as follows

$$\begin{aligned} \langle \phi_1 | \exp(-\tau T^{rot}) | \phi_2 \rangle &= \\ \sum_{m_1, m_2} \langle \phi_1 | m_1 \rangle \langle m_1 | \exp(-\tau T^{rot}) | m_2 \rangle \langle m_2 | \phi_2 \rangle &= \\ \frac{1}{2\pi} \sum_{m_1=-\infty}^{\infty} \exp[-B\tau m_1^2 - i(\phi_1 - \phi_2)m_1] & \end{aligned}$$

Now we can use the Poisson summation formula

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{M=-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) e^{2\pi i M x} dx \quad (19)$$

and transform the last equation into

$$\begin{aligned} \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \exp[-B\tau m^2 - im(\phi_1 - \phi_2)] &= \\ \left(\frac{1}{4\pi B\tau} \right)^{\frac{1}{2}} \sum_{M=-\infty}^{\infty} \exp\left[-\frac{1}{4B\tau} (\phi_1 - \phi_2 + 2\pi M)^2\right] & \end{aligned}$$

With the last form we get the path integral representation of the partition function of a free rotor

$$Z_P = \left(\frac{P}{4\pi B\beta} \right)^{\frac{P}{2}} \prod_{i=1}^P \left[\sum_{M_i=-\infty}^{\infty} \int_0^{2\pi} d\phi_i \right] \quad (20)$$

$$\exp \left[-\beta \sum_{j=1}^P \frac{P}{4B\beta^2} (\phi_j - \phi_{j+1} + 2\pi M_j)^2 \right] \quad (21)$$

Apart from the *continuous* angular variables ϕ_j , we have also the *discrete* integer variables M_j . In order to fully specify the path in imaginary time, we thus need the full set $\{\phi_1, \dots, \phi_P, M_1, \dots, M_P\}$. Both sets of variables therefore have to be sampled in the Monte Carlo simulation. In principle, this could be done, but it is more convenient to perform a further transformation to get a more tractable expression.

The winding number representation

One can define the following set of new variables

$$\begin{aligned}\tilde{\phi}_1 &= \phi_1 \\ \tilde{\phi}_2 &= \phi_2 - 2\pi M_1 \\ \tilde{\phi}_3 &= \phi_3 - 2\pi M_1 - 2\pi M_2 \\ &\vdots \\ \tilde{\phi}_P &= \phi_P - 2\pi \sum_{j=1}^{P-1} M_j\end{aligned}$$

in which the $2\pi M_j$ terms successively cancel for $j = 1, \dots, P - 1$, since $\tilde{\phi}_j - \tilde{\phi}_{j+1} = \phi_j - \phi_{j+1} + 2\pi M_j$. For $j = P$, instead, all the terms sum up and we get $\phi_P - \phi_{P+1} + 2\pi M_P = \tilde{\phi}_P - \tilde{\phi}_1 + 2\pi \sum_{j=1}^P M_j = \tilde{\phi}_P - \tilde{\phi}_1 + 2\pi n$.

The sum $n = \sum_{j=1}^P M_j$ is called the *winding number* of the path. The path is now specified by the set $\{\tilde{\phi}_1, \dots, \tilde{\phi}_P, n\}$, and except for $\tilde{\phi}_1 \in [0, 2\pi)$, the other angles $\tilde{\phi}_j \in (-\infty, \infty)$ for $j = 2, \dots, P$.

The winding number counts how many times a given path winds around the circle $[0, 2\pi)$ and thus represents a *topological* property of the path which is invariant with respect to the small (local) deformations. A purely local Monte Carlo algorithm is thus incapable of sampling the winding number, and *global* moves updating at the same time all P angles $\{\tilde{\phi}_1, \dots, \tilde{\phi}_P\}$ are necessary. One has to cut the path, change the winding number and reconnect it again.

The decoupled winding number representation

In the previous scheme the angular mismatch due to the non-zero winding number was concentrated in one of the Trotter slices. A further simplification is obtained by applying the transformation

$$\bar{\phi}_j = \tilde{\phi}_j - 2\pi n \frac{j-1}{P}, \quad j = 1, \dots, P \quad (22)$$

which distributes the mismatch uniformly through all the slices. One can show

this leads to the representation

$$Z_P = \left(\frac{P}{4\pi B\beta} \right)^{\frac{P}{2}} \sum_{n=-\infty}^{\infty} \int_0^{2\pi} d\bar{\phi}_1 \prod_{i=2}^P \left[\int_{-\infty}^{\infty} d\bar{\phi}_i \right] \exp \left[-\beta \left\{ \sum_{j=1}^P \frac{P}{4B\beta^2} (\bar{\phi}_j - \bar{\phi}_{j+1})^2 + \frac{(2\pi n)^2}{4\beta^2 B} \right\} \right]$$

where the winding number is *decoupled* from the angular degrees of freedom. The sum over n , however, cannot be performed explicitly, since for non-free rotors the transformation (22) has to be applied also in the potential energy propagator (not included in the above expression). Nevertheless, the winding number n can be sampled directly as a random variable from a Gaussian distribution, which is convenient for the simulation.

Moreover, this representation provides some physical insight into the distribution of winding numbers. It's gaussian and thus only paths with $\frac{(2\pi n)^2}{2\lambda^2} \lesssim 1$, where $\lambda^2 = 2\beta B = \frac{\hbar^2 \beta}{\Theta}$ can be interpreted as angular thermal wavelength, can contribute significantly. The lower the temperature, the more probable become paths with larger winding numbers.

Two-dimensional rotation (linear molecule rotating in 3D space)

$$d^{rot} = 2$$

The orientation of the molecule is defined by two angles, $\omega = (\vartheta, \phi)$. The kinetic energy operator

$$T^{rot} = -B \left[\frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial^2}{\partial \varphi^2} \right] \quad (23)$$

is proportional to the square of the angular momentum operator. The orthonormal eigenstates of T^{rot} are the spherical harmonics $\langle \vartheta, \phi | lm \rangle = Y_{lm}(\vartheta, \phi)$ with eigenvalues $E_l = Bl(l+1)$. We again calculate the free *rotor* density matrix by inserting the decomposition of the identity operator in the angular momentum representation

$$\langle \vartheta_1 \phi_1 | \exp(-\tau T^{rot}) | \vartheta_2 \phi_2 \rangle = \sum_l \sum_{m=-l}^l \langle \vartheta_1 \phi_1 | lm \rangle \langle lm | \vartheta_2 \phi_2 \rangle \exp(-\tau Bl(l+1))$$

Using the addition theorem for spherical harmonics we can explicitly sum over m and get

$$\langle \vartheta_1 \phi_1 | \exp(-\tau T^{rot}) | \vartheta_2 \phi_2 \rangle = \sum_l \frac{2l+1}{4\pi} P_l(\cos \gamma_{12}) \exp(-\tau Bl(l+1)),$$

where P_l is the Legendre polynomial of degree l and $\cos \gamma_{12} = \cos \vartheta_1 \cos \vartheta_2 + \sin \vartheta_1 \sin \vartheta_2 \cos(\phi_1 - \phi_2)$ is the relative angle between the two orientations (ϑ_1, ϕ_1) and (ϑ_2, ϕ_2) .

The sum over l in the last expression cannot be performed analytically. In principle, one could also sample the discrete variable l by Monte Carlo, but because of the oscillating sign of the Legendre polynomials, this would lead to a severe *sign problem*. It is much more convenient to perform the sum over l numerically and store the resulting kernel function of $\cos \gamma_{12}$ in a table for simulation.

Three-dimensional rotation (arbitrary molecule rotating in 3D space)

$$d^{rot} = 3$$

In this case we need 3 Euler angles to specify the orientation of the molecule. The approach is the same as above just the formulas are more complicated.

Coupling to nuclear spins

So far we ignored the symmetry requirements imposed on the wavefunction of the rotor by the statistics of the nuclei. The spins of the nuclei implicitly constrain the symmetry of the rotational state and have to be taken into account in the construction of the kernel (e.g. not all values of the angular momentum may be allowed). Several subtle points arise related to different kinds of averaging (*quenched* or *annealed*) over the molecular species with different symmetry (a well-known case is ortho vs. para hydrogen). In some cases, the sign problem is present. For discussion, see Ref.[57] and references therein.

18 Lattice systems with discrete degrees of freedom

So far we discussed only systems with continuous degrees of freedom. The PIMC methodology can, however, be applied also to lattice systems with discrete degrees of freedom. We will illustrate the procedure here on the well-known case of Ising model in transverse field [51]. The hamiltonian of the system reads

$$H = -\Gamma \sum_i \sigma_i^x - J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z = K + U$$

$$K = -\Gamma \sum_i \sigma_i^x, \quad U = -J \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z,$$

where we assume that the spins occupy sites of a d -dimensional cubic lattice and $\sigma_i^x, \sigma_i^y, \sigma_i^z$ are Pauli matrices corresponding to spin on lattice site i . Clearly, $[K, U] \neq 0$. We proceed analogously to the case of continuous degrees of free-

dom.

$$\begin{aligned}
Z &= \text{Tr} e^{-\beta H} = \text{Tr} (e^{-\beta H/P})^P = \text{Tr} (e^{-\beta(K+U)/P})^P \\
&= \sum_{s^1} \dots \sum_{s^P} \langle s^1 | e^{-\beta(K+U)/P} | s^2 \rangle \langle s^2 | e^{-\beta(K+U)/P} \dots | s^P \rangle \\
&\quad \times \langle s^P | e^{-\beta(K+U)/P} | s^1 \rangle
\end{aligned}$$

Here, s^k denotes a configuration of all spins in k -th Trotter slice. *in z components*
So far exact, now we again apply the Trotter decomposition.

$$\begin{aligned}
e^{-\beta(K+U)/P} &\approx e^{-\beta K/P} e^{-\beta U/P} \\
Z_P &= \sum_{s^1} \dots \sum_{s^P} \langle s^1 | e^{-\beta K/P} e^{-\beta U/P} | s^2 \rangle \langle s^2 | \dots \\
&\quad \times e^{-\beta K/P} e^{-\beta U/P} | s^P \rangle \langle s^P | e^{-\beta K/P} e^{-\beta U/P} | s^1 \rangle
\end{aligned}$$

We have to evaluate the expression $\langle s^k | e^{-\beta K/P} e^{-\beta U/P} | s^{k+1} \rangle$.

$$\begin{aligned}
&\langle s^k | e^{-\beta K/P} e^{-\beta U/P} | s^{k+1} \rangle \\
&= \langle s^k | e^{-\beta K/P} \sum_{\tilde{s}^k} |\tilde{s}^k\rangle \langle \tilde{s}^k | e^{-\beta U/P} | s^{k+1} \rangle \\
&= \langle s^k | e^{-\beta K/P} | s^{k+1} \rangle e^{-\beta U(s^{k+1})/P}
\end{aligned}$$

Spin operators corresponding to different spins commute. Therefore

$$\begin{aligned}
\langle s^k | e^{-\beta K/P} | s^{k+1} \rangle &= \langle s^k | \exp\left(\frac{\beta\Gamma}{P} \sum_{i=1}^N \sigma_i^x\right) | s^{k+1} \rangle \\
&= \prod_{i=1}^N \langle s_i^k | \exp\left(\frac{\beta\Gamma}{P} \sigma_i^x\right) | s_i^{k+1} \rangle
\end{aligned}$$

It's easy to show that

$$\begin{aligned}
\langle \uparrow | e^{a\sigma_x} | \uparrow \rangle &= \langle \downarrow | e^{a\sigma_x} | \downarrow \rangle = \cosh a \\
\langle \uparrow | e^{a\sigma_x} | \downarrow \rangle &= \langle \downarrow | e^{a\sigma_x} | \uparrow \rangle = \sinh a
\end{aligned}$$

which can be written as an Ising-like interaction (s, s' now mean single spins)

$$\langle s | e^{a\sigma_x} | s' \rangle = C e^{J' s s'} \tag{24}$$

with $J' = -\frac{1}{2} \ln \tanh a$, $C^2 = \frac{1}{2} \sinh a$. We get for our matrix elements

$$\begin{aligned} \langle s^k | e^{-\beta K/P} | s^{k+1} \rangle &= \prod_{i=1}^N \langle s_i^k | \exp\left(\frac{\beta \Gamma}{P} \sigma_i^x\right) | s_i^{k+1} \rangle \\ &= \prod_{i=1}^N C e^{J' s_i^k s_i^{k+1}} = C^N \prod_{i=1}^N e^{J' s_i^k s_i^{k+1}} \\ J' &= -\frac{1}{2} \ln \tanh \frac{\Gamma \beta}{P}, C = \left(\frac{1}{2} \sinh \frac{2\Gamma \beta}{P} \right)^{\frac{1}{2}} \end{aligned}$$

which can be interpreted as an Ising-like coupling between the Trotter replicas of the same spin which are *nearest-neighbours* along the Trotter dimension.

The expression $\langle s^k | e^{-\beta K/P} e^{-\beta U/P} | s^{k+1} \rangle$ now can be written as follows

$$\langle s^k | e^{-\beta K/P} e^{-\beta U/P} | s^{k+1} \rangle = C^N \prod_{i=1}^N e^{J' s_i^k s_i^{k+1}} \prod_{\langle ij \rangle} e^{\frac{\beta J}{P} s_i^k s_j^k} \quad (25)$$

For the full partition function we thus get

$$Z_P = C^{NP} \sum_{s^1} \dots \sum_{s^P} \prod_{k=1}^P \prod_{i=1}^N e^{J' s_i^k s_i^{k+1}} \prod_{\langle ij \rangle} e^{\frac{\beta J}{P} s_i^k s_j^k}$$

which represents a partition function of a $(d+1)$ -dimensional *anisotropic* Ising model. It has couplings $\frac{\beta J}{P}$ in the d -dimensional hyperplanes and J' along the extra dimension. The β -dependent term C^{NP} just adds a trivial contribution to the free energy. The Trotter decomposition has thus introduced an *additional* dimension into the problem. However, even when we take the thermodynamic limit $N \rightarrow \infty$, the size of the system along the extra dimension remains finite and equal to the Trotter number P .

The mapping of a d -dimensional *quantum* system onto a $(d+1)$ -dimensional *classical* system provides a practical tool for simulation. Almost all the properties (including difficulties for large P) mentioned for continuous degrees of freedom remain. Some new peculiar properties appear:

- sampling - in discrete systems we cannot adjust the acceptance ratio by varying the size of the moves, we have to accept what comes.
- it's not always clear how to decompose the hamiltonian in order to apply the Trotter formula, there may be many different possibilities. Checkerboard decomposition.

- sometimes it's necessary to invent sophisticated multiparticle moves in order to satisfy the conservation laws imposed by a particular decomposition of the hamiltonian
- one has to carefully derive estimators for response functions (e.g. susceptibilities) related to operators which do not commute with the hamiltonian, as they are not a simple translation of classical expressions.
- for many examples and details, see Ref.[51]

What happens as $T \rightarrow 0$?

If we fix $\tau = \beta/P$ and let $P \rightarrow \infty$, we effectively perform the limit $\beta \rightarrow \infty$ which corresponds to zero temperature. The size of the system along the extra dimension becomes also infinite and thus we have a truly $(d+1)$ dimensional system. It can be shown in this way that the ground state of the model in d dimensions is equivalent to a $(d + 1)$ -dimensional model where Γ plays the role of temperature. This has an intrinsic theoretical importance beyond simulations; e.g. in the Ising model in transverse field, the quantum phase transition driven at $T = 0$ by change of Γ has the universality class and critical exponents corresponding to the temperature driven transition in $(d + 1)$ -dimensional Ising model.

References

- [1] F. Nogueira, A. Castro, M. A. L. Marques, *A Tutorial on the Density Functional Theory*. Lecture Notes in Physics **620**, 218 - 256, (2003) Springer Verlag
- [2] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 - 1497 (1979)
- [3] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 - 4228 (1982)
- [4] Steven G. Louie, Sverre Froyen and Marvin L. Cohen, Phys. Rev. B **26**, 1738 - 1742 (1982)
- [5] N. Troullier and José Luriaas Martins, Phys. Rev. B **43**, 1993 - 2006 (1991)
- [6] Leonard Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 - 1428 (1982)
- [7] David Vanderbilt, Phys. Rev. B **41**, 7892 (1990)

- [8] Broyden, C. G., *Mathematics of Computation* (American Mathematical Society) **19**, 577 - 593 (1965)
- [9] L. J. Sham and M. Schlüter, *Phys. Rev. Lett.* **51**, 1888 - 1891 (1983)
- [10] Lars Hedin, *Phys. Rev.* **139**, A796 - A823 (1965)
- [11] M. T. Yin and Marvin L. Cohen, *Phys. Rev. B* **26**, 5668 - 5687 (1982)
- [12] O. Gunnarsson, B. I. Lundqvist, *Phys. Rev. B* **13**, 4274 - 4298 (1976)
- [13] A. D. Becke, *Phys. Rev. A* **38**, 3098 - 3100 (1988)
- [14] Chengteh Lee, Weitao Yang, and Robert G. Parr, *Phys. Rev. B* **37**, 785 - 789 (1988)
- [15] John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 - 3868 (1996)
- [16] John P. Perdew, Stefan Kurth, Aleš Zupan, Peter Blaha, *Phys. Rev. Lett.* **82**, 2544 - 2547 (1999)
- [17] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003).
- [18] S. Grimme, *J. Comput. Chem.* **27**, 1787 (2006).
- [19] Joost VandeVondele, Fawzi Mohamed, Matthias Krack, Jürg Hutter, Michiel Sprik, and Michele Parrinello, *J. Chem. Phys.* **122**, 014515 (2005)
- [20] R. Car, M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 - 2474 (1985)
- [21] G. Pastore, E. Smargiassi, F. Buda, *Phys. Rev. A* **44**, 6334 - 6347 (1991)
- [22] Jeffrey C. Grossman, Eric Schwegler, Erik W. Draeger, François Gygi, and Giulia Galli, *J. Chem. Phys.* **120**, 300 (2004)
- [23] A. DeVita, G. Galli, A. Canning and R. Car, *Nature*, **379**, 523 (1996)
- [24] Sverre Froyen and Marvin L Cohen, *J. Phys. C: Solid State Phys.* **19** (1986) 2623 - 2632
- [25] P. Focher, G. L. Chiarotti, M. Bernasconi, E. Tosatti, and M. Parrinello, *Europhys. Lett.* **26**, 345 (1994).
- [26] S. Scandolo, M. Bernasconi, G. L. Chiarotti, P. Focher, and E. Tosatti, *Phys. Rev. Lett.* **74**, 4015 - 4018 (1995)

- [27] D. Marx and M. Parrinello, *Z. Phys. B: Condens. Matter* **95**, 143 (1994)
- [28] Mark E. Tuckerman, Dominik Marx, Michael L. Klein, and Michele Parrinello, *J. Chem. Phys.* **104**, 5579 (1996)
- [29] Magali Benoit, Dominik Marx and Michele Parrinello, *Nature* **392**, 258 - 261 (1998)
- [30] Marcella Iannuzzi, Alessandro Laio, and Michele Parrinello, *Phys. Rev. Lett.* **90**, 238302 (2003)
- [31] A. Stirling, M. Iannuzzi, A. Laio, and M. Parrinello, *ChemPhysChem*, **5**, 1558, (2004)
- [32] P. Sherwood, in "Modern Methods and Algorithms of Quantum Chemistry" NIC Ser. Vol. 1, John von Neumann Institute of Computing, Juelich 2000, pp. 257 - 277
- [33] Federico Zipoli, Teodoro Laino, Alessandro Laio, Marco Bernasconi, and Michele Parrinello, *J. Chem. Phys.* **124**, 154707 (2006)
- [34] G. Kresse, J. Furthmüller, *Phys. Rev. B* **54**, 11169 - 11186 (1996); *Comp. Mat. Sci.* **6**, 15 - 50 (1996)
- [35] Michael P. Teter, Michael C. Payne, Douglas C. Allan, *Phys. Rev. B* **40**, 12255 - 12263 (1989)
- [36] P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980)
- [37] Jürg Hutter, Hans Peter Lüthi and Michele Parrinello, *Comp. Mat. Sci.* **2**, 244 (1994)
- [38] C M Goringe, D R Bowler and E Hernández, *Rep. Prog. Phys.* **60**, 1447 (1997)
- [39] F. F. Abraham, J. Q. Broughton, N. Bernstein, and E. Kaxiras, *Comput. Phys.* **12**, 538 (1998)
- [40] Florian R. Krajewski and Michele Parrinello, *Phys. Rev. B* **71**, 233105 (2005)
- [41] Stefano Baroni, Paolo Giannozzi, Andrea Testa, *Phys. Rev. Lett.* **58**, 1861 - 1864 (1987)
- [42] M. V. Berry, *Proc. Roy. Soc. Lond. A* **392**, 45 (1984)

- [43] R. D. King-Smith and David Vanderbilt, Phys. Rev. B **47**, 1651 - 1654 (1993)
- [44] R. Resta, Europhys. Lett. **22**, 133 (1993)
- [45] M. Bernasconi, P. L. Silvestrelli, and M. Parrinello, Phys. Rev. Lett. **81**, 1235 - 1238 (1998)
- [46] R. Resta, Phys. Rev. Lett. **80**, 1800 - 1803 (1998)
- [47] D. Chandler, P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).
- [48] M.F. Herman, E.J. Bruskin and B.J. Berne, J. Chem. Phys. **76**, 5150 (1982).
- [49] R.W. Hall, B.J. Berne, J. Chem. Phys. **81**, 3641 (1984).
- [50] R.M. Fye, Phys. Rev. B, **33**, 6271 (1986).
- [51] Progs. Theor. Phys. **56**, 1454 (1976), *Quantum Monte Carlo Methods in Equilibrium and Nonequilibrium Systems*, Proceedings of the Ninth Taniguchi International Symposium, Susono, Japan, 1986, ed. by M. Suzuki (Springer-Verlag Berlin Heidelberg 1987)
- [52] D. M. Ceperley, Rev. Mod. Phys., **67**, 279 (1995).
- [53] M. Takahashi, M. Imada, J. Phys. Soc. Jpn, **53**, 3765 (1984).
- [54] J.D. Doll, R.D. Coalson, D.L. Freeman, Phys. Rev. Lett., **55**, 1 (1985).
- [55] M. Sprik, M.L. Klein, D. Chandler, Phys. Rev. B, **31**, 4234 (1985).
- [56] A. Cuccoli, A. Macchi, G. Pedrolli, V. Tognetti, R. Vaia, Phys. Rev. B, **51**, 12369 (1995).
- [57] Marx D., Muser M.H., *Path integral simulations of rotors: theory and applications* J PHYS-CONDENS MAT 11: (11) R117-R155 (1999).