

Introduction to QM/MM

Emiliano Ippoliti

Forschungszentrum Jülich

Tutorial with

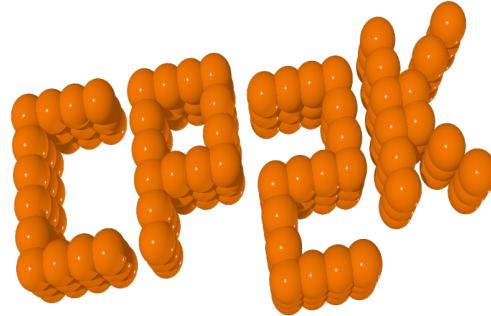
DFT

+

GPW basis set

(Mixed Gaussian and PW)

<https://www.cp2k.org/features>



- **Energy and Forces**
- **Optimisation**
 - Geometry optimisation
 - Nudged elastic band
 - ...
- **Molecular Dynamics**
 - Born-Oppenheimer MD
 - ...
- **Properties**
 - Atomic charges (RESP, Mulliken, ...)
 - Spectra
 - Frequency calculations
 - ...

Outline

First part

- Why QM?
- Elements of Computational Quantum Chemistry
- Why QM/MM?

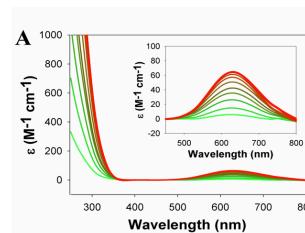
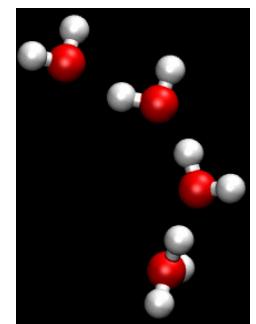
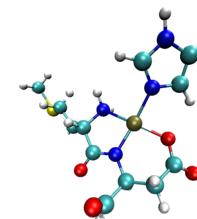
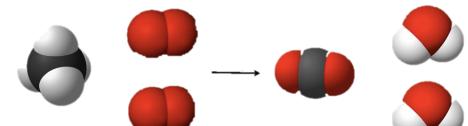
Second part

- QM/MM couplings

QM description

There are cases where the dynamical behavior of the **electrons** of your systems cannot be neglected and a quantum mechanical description is required:

1. Chemical reactions (e.g. enzymatic reactions):
electron transfer, bond breaking and bond formation
2. Systems with metal atoms (e.g. metallo proteins):
no universal parametrization for metal atoms
3. Proton transport (e.g. aerobic generation of ATP and oxygen):
Grotthuss mechanism (charge/topological diffusion vs mass diffusion)
4. Spectroscopic analysis and prediction
(e.g. absorption and fluorescence spectra)



Schödinger Equation (SE)

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t) = \hat{\mathcal{H}}(\mathbf{x}_1, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)$$

- $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)$ = system wavefunction

$|\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)|^2 \propto$ charge density distribution

$$O(t) = \int \dots \int \Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t)^* \hat{O}(\mathbf{x}_1, \dots, \mathbf{x}_N) \Psi(\mathbf{x}_1, \dots, \mathbf{x}_N, t) d\mathbf{x}_1 \dots d\mathbf{x}_N$$

- $\hat{\mathcal{H}}(\mathbf{x}_1, \dots, \mathbf{x}_N)$ = Hamiltonian operator

e.g. $\hat{\mathcal{H}} = -\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{i < j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I} \frac{e^2 Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{I < J} \frac{e^2 Z_I Z_J}{|\mathbf{R}_i - \mathbf{R}_j|}$

\mathbf{r}_i = electronic coordinates, \mathbf{R}_i = nuclear coordinates

Time-dependent and time-independent SE

Time-independent Schrödinger equation

$$\hat{\mathcal{H}}(\mathbf{r}_1, \dots, \mathbf{R}_N) \Psi(\mathbf{r}_1, \dots, \mathbf{R}_N) = E \Psi(\mathbf{r}_1, \dots, \mathbf{R}_N)$$

Solving numerical methods

- Hartree Fock Theory
 - Møller-Plesset Perturbation Theory
 - Coupled Cluster
 - Generalised Valence Bond
 - Complete Active Space SCF
 - Density Functional Theory
- ...

Time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}_1, \dots, \mathbf{R}_N, t) = \hat{\mathcal{H}}(\mathbf{r}_1, \dots, \mathbf{R}_N) \Psi(\mathbf{r}_1, \dots, \mathbf{R}_N, t)$$

Molecular dynamics schemes

- Ehrenfest
 - **Born-Oppenheimer**
 - Car-Parrinello
 - Kühne
- ...

Born-Oppenheimer approximation

$$m_e \ll M_I$$



- Motion of atomic nuclei and electrons can be treated **separately**
- Nuclear motion can be treated **classically**

Born-Oppenheimer MD scheme

time-independent SE, \mathbf{R}_I are here only parameters

$$\left\{ \begin{array}{l} \widehat{\mathcal{H}}_e(\mathbf{r}_1, \dots, \mathbf{R}_N) \psi_0(\mathbf{r}_1, \dots, \mathbf{r}_N) = E_0 \psi_0(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad \mathbf{r}_i = \text{electronic coordinates} \\ M_I \ddot{\mathbf{R}}_I(t) = - \nabla_I \min_{\psi_0} \{ \langle \psi_0 | \widehat{\mathcal{H}}_e | \psi_0 \rangle \} \quad \mathbf{R}_I = \text{nuclear coordinates} \end{array} \right.$$

classical Newton equation: $\min_{\psi_0} \{ \langle \psi_0 | \widehat{\mathcal{H}}_e | \psi_0 \rangle \}$ is the potential felt by the nuclei

$$\begin{aligned} \widehat{\mathcal{H}} &= - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 + \sum_{i < j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I} \frac{e^2 Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{I < J} \frac{e^2 Z_I Z_J}{|\mathbf{R}_i - \mathbf{R}_j|} = \\ &= - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 - \widehat{\mathcal{H}}_e(\mathbf{r}_1, \dots, \mathbf{R}_N) \end{aligned}$$

Algorithm

At each time step:

- Solving the time-independent SE
- Use ψ_0 to find the forces on nuclei
- Move the nuclei

Electronic Structure Methods

Time-independent Schrödinger equation

$$\hat{\mathcal{H}}(\mathbf{r}_1, \dots, \mathbf{R}_N) \Psi(\mathbf{r}_1, \dots, \mathbf{R}_N) = E \Psi(\mathbf{r}_1, \dots, \mathbf{R}_N)$$

Many numerical methods to solve it approximately:

- Hartree Fock Theory
 - Møller-Plesset Perturbation Theory
 - Coupled Cluster
 - Generalised Valence Bond
 - Complete Active Space SCF
 - **Density Functional Theory**
- ...

→ Best compromise
between accuracy and
computational cost



Very suitable for large systems

Density Functional Theory (DFT)

Hohenberg–Kohn Theorems

1. The ground state energy is a unique functional of the electronic density:

$$E = E[\rho]$$

2. The functional for the energy $E[\rho]$ is variational:

$$\min_{\psi_0} \{ \langle \psi_0 | \hat{\mathcal{H}}_e | \psi_0 \rangle \} = \min_{\rho} E[\rho]$$

Benefit:

$\psi_0(\mathbf{r}_1, \dots, \mathbf{r}_n)$ function of $3 \times$ number of electron coordinates (n)

$\rho(\mathbf{r})$ function of 3 coordinates

Drawback:

functional $E[\rho]$ is not known

Kohn-Sham DFT

Kohn and Sham define a fictitious system of **non-interacting** particles with a local potential that generates the same density as the density of the real **fully-interacting** system:

$$\rho(\mathbf{r}) = \sum_i |\varphi_i(\mathbf{r})|^2 \quad \int d\mathbf{r} \varphi_i^*(\mathbf{r}) \varphi_j(\mathbf{r}) = \delta_{ij}$$

$$E[\rho] = E^{KS}[\{\varphi_i\}] =$$

$$= \sum_i \int d\mathbf{r} \varphi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m_e} \nabla^2 \right) \varphi_j(\mathbf{r}) + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) + \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r}) \rho'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{\text{xc}}[\rho]$$

$E_{\text{KIN}}[\rho]$ $E_{\text{ext}}[\rho]$ $E_{\text{H}}[\rho]$

Kohn-Sham equations

$$\min_{\rho} E[\rho]$$



$$\left\{ -\frac{1}{2} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} \right\} \varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r})$$

1 x N -electron
equation



N x one-electron equations

NOT KNOWN

Finding Kohn-Sham orbitals φ_i by iterative procedure

For a given (fixed) ionic configuration \mathbf{R}^N :

1. Solve KS equations for an initial or a previous density ρ :

$$\left\{ -\frac{1}{2} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} \right\} \varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r})$$

2. Determine the updated density:

$$\rho^{\text{new}}(\mathbf{r}) = \sum_i |\varphi_i(\mathbf{r})|^2$$

$$\rho(\mathbf{r}) = \rho^{\text{new}}(\mathbf{r})$$

3. Check if convergence is reached:

$$\int d\mathbf{r} |\rho^{\text{new}}(\mathbf{r}) - \rho(\mathbf{r})| < \text{Threshold}$$

No: back to step 1

Yes: exit

Basis set

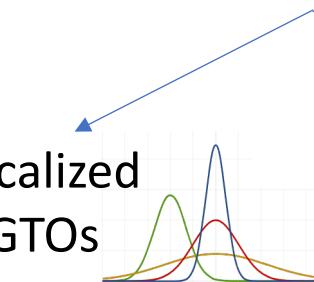
$$\left\{ -\frac{1}{2} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} \right\} \varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r}) \quad \text{KS equations}$$

Discretization: $\varphi_i(\mathbf{r}) \simeq \sum_{k=1}^M a_k g_k(\mathbf{r})$

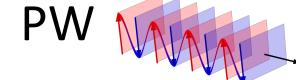
$$\begin{pmatrix} \vdots & \cdots & \ddots & \vdots \\ & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_M \end{pmatrix} = \epsilon_i \begin{pmatrix} a_1 \\ \vdots \\ a_M \end{pmatrix}$$

$\{g_k(\mathbf{r})\}$ = Basis set

Localized
GTOs



Nonlocal
PW



- + Atomic orbital-like
 - + Few functions needed
 - + Analytic integration for many operators
 - + Optimal for regular grids
 - + Finite extent
 - Non-orthogonal
 - Linear dependences for large basis set
 - Complicated to generate
 - Basis set superposition error
- + Orthogonal
 - + Independent from atomic position
 - ± Naturally periodic
 - Many functions needed

Gaussian and Plane Waves method (GPW)

Atom-centered
Gaussian-type basis

$$\varphi_i(\mathbf{r}) \simeq \sum_{k=1}^M a_k g_k(\mathbf{r})$$

+

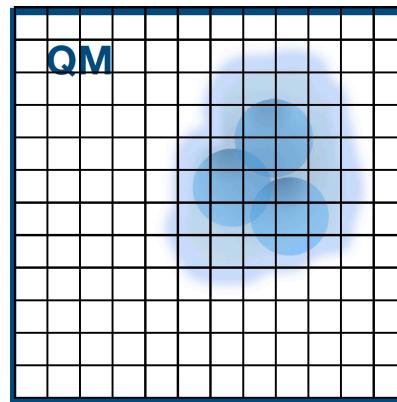
Auxiliary
Plane-wave basis

$$\rho(\mathbf{r}) \simeq \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}=0}^{E_{\text{cut}}} \rho(\mathbf{G}) \exp[i\mathbf{G} \cdot \mathbf{r}]$$



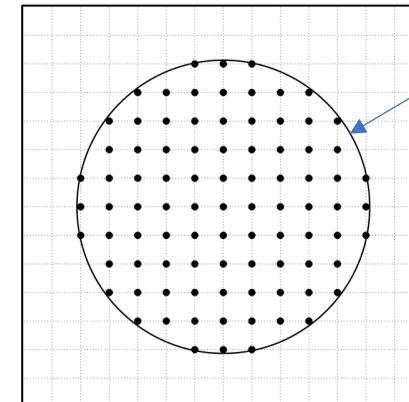
CP2K can scale linearly
with the system size

$$\frac{1}{2} \mathbf{G}^2 < E_{\text{cut}}$$



Real space

Fast Fourier Trasform
(FFT)

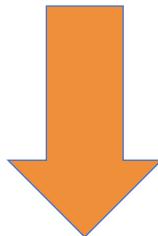


Reciprocal (\mathbf{G}) space

Biological system sizes

Largest systems investigated at full QM level \lesssim 10,000 atoms

Typical biological system sizes \gg 10,000 atoms



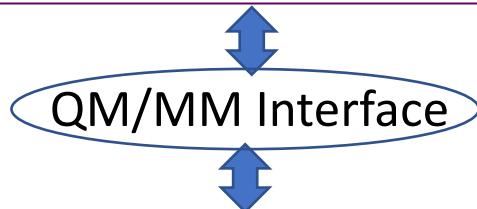
Multiscale approach: QM/MM

Combining different levels of theory and resolutions

QM/MM Approach

The system is separated into two parts:

A small **QM part** comprises the **chemically/photophysically active region** treated by computationally demanding electronic structure methods.



The remainder **MM part** is described efficiently at a lower level of theory by classical force fields.

