# Electronic structure methods (Hartree-Fock) used in industry from pharmaceuticals, materials and beyond

Ariadni Boziki

University of Luxembourg
Faculty of Science, Technology and Medicine

3$^{\text{rd}}$ of October 2025

# Overview

1. Computational chemistry
2. Importance of computational chemistry methods in industry
3. Hartree-Fock method

# Computational chemistry

**Chemistry on a computer instead of in a laboratory.**

Use of computer calculations to predict the structure and in turn the properties of matter.

Computational chemistry/physics/materials science has become widely used because of:

- Dramatic **increase in computational power/speed** and the
- Design of **efficient quantum chemical algorithms**.

The computer calculations enable us to:

- **explain and rationalize known chemistry**.
- **explore new or unknown chemistry**.

# Why doing chemistry on a computer?

- Calculations **are easy to perform**, while experiments are difficult.
- Calculations **are safe**, whereas many experiments are dangerous.
- Calculations are **less costly**, while experiments can be more expensive.
- Calculations can be performed **in any chemical system**, while experiments are relatively limited.
- Calculations can give **direct information**, whereas there is uncertainty in interpreting experimental observations.
- Calculations give **fundamental information** about the structure-property relationships of materials even at a molecular level.

# When a computational chemistry project is considered successful?

- What is the **property under study**?
- How **accurate should the calculation** be? (inclusion or exclusion of electronic degrees of freedom)
- How **big the system** under study is? (molecule/big system as a membrane-protein in solution/polymeric melt)
- With respect to the size of the system and the property we aim to study, what **level of approximation** we can introduce in the computational protocol?

# When a computational chemistry project is considered successful?

**Golden rule**

Before applying a particular level of theory to an experimentally unknown situation, it is essential to apply the same level of theory to situations where experimental information is available.

It is essential to know how accurate our results have to be with respect to the property under study, for our results to be considered accurate.

"**Chemical accuracy**": within 1 kcal/mol in energy

# The rise of computational chemistry in industry

More than 70% of the leading pharmaceutical companies have integrated specialized computational chemistry teams that drive strategic innovation.

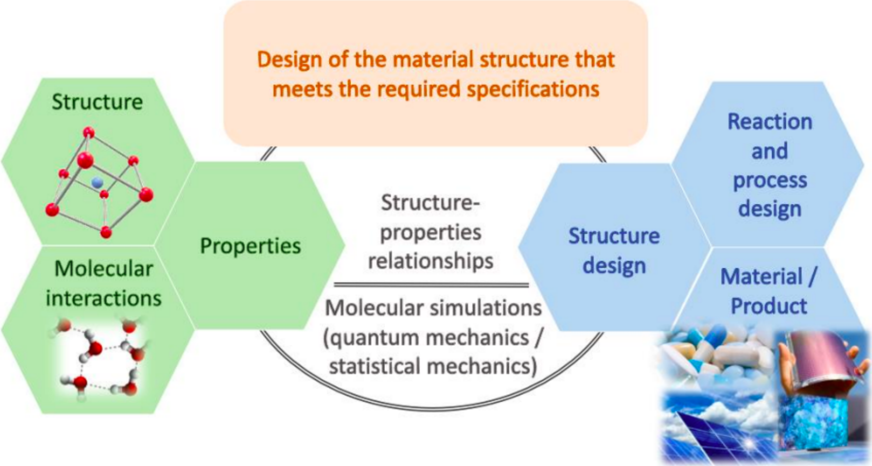
- **Cost efficiency**: Advanced molecular simulations systematically minimize experimental iterations reducing R&D expenditures significantly.
- **Early stage virtual screening**: Computational protocols enable rapid identification of optimal candidates, expediting discovery and filtering ineffective molecules prior to experiment.
- **Enhanced prediction accuracy**: High-precision computational models improve the reliability of property predictions, guiding more effective experimental validation strategies.
- **Accelerated development**: Simulations expedite molecular design cycles, enabling rapid hypothesis testing and shortening product development timelines.

Benefits: **Speed**, **Cost** and **Accuracy**

# Putting computational chemistry to work!

- Drug design
- Medicinal chemistry design
- Consumer packaged food
- Protein/antibody engineering
- Enzyme design
- Organic electronics
- Pharmaceutical formulations
- Catalysis design
- Polymer design
- Surface chemistry
- Energy capture and storage
- Semiconductors
- Peptide design
- Metals, alloys, and ceramics design

# Importance of computational chemistry methods in industry



Design of the material structure that meets the required specifications

Structure

Molecular interactions

Properties

Structure-properties relationships

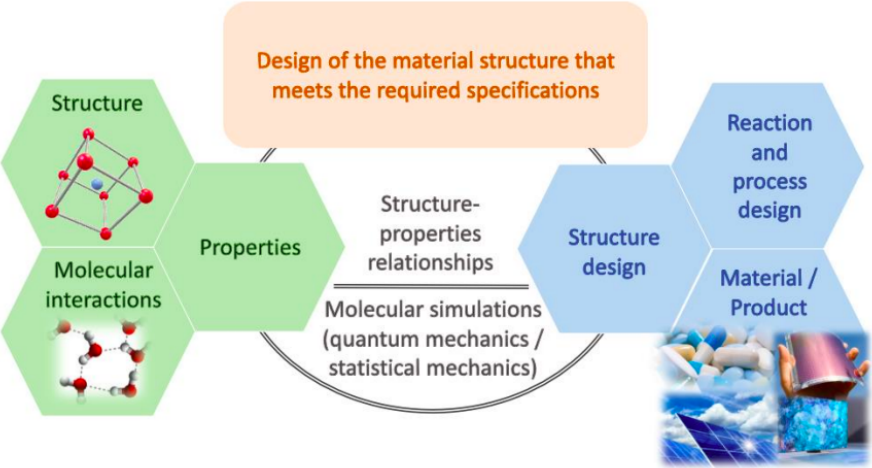
Molecular simulations (quantum mechanics / statistical mechanics)

Structure design

Reaction and process design

Material / Product

# Importance of computational chemistry methods in industry



Design of the material structure that meets the required specifications

Structure

Molecular interactions

Properties

Structure-properties relationships

Molecular simulations (quantum mechanics / statistical mechanics)

Structure design

Reaction and process design

Material / Product

# Postulates of quantum mechanics

1. Postulate 1
    - The state of a quantum system is described by the wavefunction $\Psi(\mathbf{r}, t)$.
    - $\Psi^*(\mathbf{r}, t)\,\Psi(\mathbf{r}, t)\,d\tau$ gives the probability of finding the particle in volume $d\tau$ at $\mathbf{r}$.
    - Normalization: $\int_{-\infty}^{\infty} \Psi^*(\mathbf{r}, t)\,\Psi(\mathbf{r}, t)\,d\tau = 1$
    - Wavefunction must be single-valued, continuous, finite, and normalized (also for many-particle systems).
2. Postulate 2
    - To every observable in classical mechanics there corresponds a linear, Hermitian operator in quantum mechanics.

# Postulates of quantum mechanics

1. Postulate 3
    - Measurement of the observable associated to operator $\hat{A}$ yields only its eigenvalues $a$ (eigenvalue equation):

      $$\hat{A}\Psi(\mathbf{r},t) = a\Psi(\mathbf{r},t)$$

    - If the system is in an eigenstate of $\hat{A}$, with a single eigenvalue $a$, then any measurement will yield $a$.
2. Postulate 4
    - If a system is in a state described by a normalized wavefunction $\Psi$, then the expectation value of the observable corresponding to $\hat{A}$ is given by:

      $$\langle A \rangle = \int_{-\infty}^{\infty} \Psi^*(\mathbf{r},t)\,\hat{A}\,\Psi(\mathbf{r},t)\,d\tau$$

# Postulates of quantum mechanics

1. Postulate 5
    - The wavefunction or state function of a system evolves in time according to the time-dependent Schrödinger equation (the central equation of quantum mechanics):

$$\hat{H}\,\Psi(\mathbf{r},t) = i\hbar\,\frac{\partial}{\partial t}\Psi(\mathbf{r},t)$$

2. Postulate 6
    - The total wavefunction must be antisymmetric with respect to the interchange of all coordinates of one fermion with those of another.

# The molecular Hamiltonian

Quantum chemistry is (typically) based on the **non-relativistic Schrödinger equation** within the Born–Oppenheimer approximation.

$$\hat{H}_{\mathrm{tot}}(\mathbf{R},\mathbf{P},\mathbf{r},\mathbf{p})\,\Psi_{\mathrm{tot}}(\mathbf{R},\mathbf{P},\mathbf{r},\mathbf{p}) = E(\mathbf{R},\mathbf{P})\,\Psi_{\mathrm{tot}}(\mathbf{R},\mathbf{P},\mathbf{r},\mathbf{p})$$

$\mathbf{r},\mathbf{p} = \partial/\partial\mathbf{r}$: electronic collective coordinates.
$\mathbf{R},\mathbf{P} = \partial/\partial\mathbf{R}$: nuclear collective coordinates.
$E$: allowed energy of the system.
$\Psi_{tot}$: function of the positions of all electrons and nuclei.
$\hat{H}_{tot}$: differential operator constructed from the classical Hamiltonian.

# The molecular Hamiltonian

For a system of nuclei and electrons in vacuum with no external fields, neglecting magnetic interactions, using atomic units:

$$\begin{aligned}\hat{H}_{\text{tot}} = &-\tfrac{1}{2}\sum_{I}\frac{1}{M_I}\nabla_I^2 - \tfrac{1}{2}\sum_{n}\nabla_n^2 + \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}\\ &-\sum_{In}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_n|} + \sum_{n<m}\frac{1}{|\mathbf{r}_m - \mathbf{r}_n|}\end{aligned}$$

# The molecular Hamiltonian

Due to Born-Oppenheimer approximation some terms coupling the electrons and nuclei can be neglected.

$$\Psi_{\mathrm{tot}}(\mathbf{R},\mathbf{r}) = \Psi_{\mathrm{nucl}}(\mathbf{R})\,\Psi_{\mathrm{elec}}(\mathbf{r};\mathbf{R})$$

and

$$\hat{H}_{\mathrm{tot}} = \hat{T}_{\mathrm{nucl}}(\mathbf{P},\mathbf{R}) + \hat{H}_{\mathrm{elec}}(\mathbf{p},\mathbf{r};\mathbf{R})$$

ignoring the dependence of $\hat{H}_{elec}$ on the momenta of the nuclei $\mathbf{P}$.

Solution of Schrödinger equation for the electrons, with fixed nuclei, indicated by $(;\mathbf{R})$:

$$\hat{H}_{\mathrm{elec}}(\mathbf{p},\mathbf{r};\mathbf{R})\,\Psi_{\mathrm{elec}}(\mathbf{r};\mathbf{R}) = E(\mathbf{R})\,\Psi_{\mathrm{elec}}(\mathbf{r};\mathbf{R})$$

$V(\mathbf{R})$, potential energy surface: collection of all possible nuclear configurations, $\mathbf{R}$ together with the associated energies.

# The molecular Hamiltonian

Approximate Schrödinger equation for the nuclei, integrating over all electron positions **r** and ignoring any incovenient term.

$$\langle \Psi_{\mathrm{elec}}(\mathbf{r},\mathbf{R}) | \hat{H}_{\mathrm{tot}} | \Psi_{\mathrm{elec}}(\mathbf{r},\mathbf{R}) \rangle \approx \hat{H}_{\mathrm{nucl}} = \hat{T}_{\mathrm{nucl}}(\mathbf{P},\mathbf{R}) + V(\mathbf{R})$$

with

$$\left( \hat{T}_{\mathrm{nucl}}(\mathbf{P},\mathbf{R}) + V(\mathbf{R}) \right) \Psi_{\mathrm{nucl}}(\mathbf{R}) = E_{\mathrm{nucl}} \Psi_{\mathrm{nucl}}(\mathbf{R})$$

# The nuclear Schrödinger equation

$V(\mathbf{R})$ is usually expanded to second order $\mathbf{R}$ about a stationary point $\mathbf{R}_0$:

$$V(\mathbf{R}) \approx V(\mathbf{R}_0) + \frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 V(\mathbf{R})}{\partial \mathbf{R}_i \partial \mathbf{R}_j}\right)(\mathbf{R}_i - \mathbf{R}_{0,i})(\mathbf{R}_j - \mathbf{R}_{0,j})$$

No need to solve the electronic Schrödinger equation at many $\mathbf{R}$'s. Only required:

- locate stationary point $\mathbf{R}_0$.
- evaluate energy and Hessian (second derivatives) at that $\mathbf{R}_0$.

# The electronic Schrödinger equation

How can one compute the required potential $V(\mathbf{R})$ that acts on the nuclei at a given geometry $\mathbf{R}$?

$$\hat{H}_{\text{elec}}(\mathbf{p},\mathbf{r};\mathbf{R})\,\Psi_{\text{elec}}(\mathbf{r};\mathbf{R}) = E(\mathbf{R})\,\Psi_{\text{elec}}(\mathbf{r};\mathbf{R})$$

in vaccum, in the absence of fields, and neglecting magnetic effects:

$$\begin{aligned}\hat{H}_{\text{elec}}(\mathbf{R}) = & -\tfrac{1}{2}\sum_{n}\nabla_n^2 + \sum_{I<J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \\ & -\sum_{In}\frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_n|} + \sum_{n<m}\frac{1}{|\mathbf{r}_m - \mathbf{r}_n|}\end{aligned}$$

# The electronic Schrödinger equation

**Pauli exclusion principle**: Because electrons are indistinguishable fermions, any permutation of two electrons must change the sign of the wavefunction $\Psi_{elect}(\mathbf{r};\mathbf{R})$.

Slater determinant many-electron wavefunctions satisfy the Pauli principle.

$$\begin{aligned}\Psi_{\text{el}}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N) = \sum_{m_1,m_2,\ldots,m_N} & C_{m_1,m_2,\ldots,m_N} \\ & \times \left|\phi_{m_1}(\mathbf{r}_1)\,\phi_{m_2}(\mathbf{r}_2)\,\cdots\,\phi_{m_N}(\mathbf{r}_N)\right|\end{aligned}$$

where

$$\left|\phi_{m_1}(\mathbf{r}_1)\,\phi_{m_2}(\mathbf{r}_2)\,\cdots\,\phi_{m_N}(\mathbf{r}_N)\right| = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_{m_1}(\mathbf{r}_1) & \phi_{m_2}(\mathbf{r}_1) & \cdots & \phi_{m_N}(\mathbf{r}_1) \\ \phi_{m_1}(\mathbf{r}_2) & \phi_{m_2}(\mathbf{r}_2) & \cdots & \phi_{m_N}(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m_1}(\mathbf{r}_N) & \phi_{m_2}(\mathbf{r}_N) & \cdots & \phi_{m_N}(\mathbf{r}_N) \end{vmatrix}$$

# The electronic Schrödinger equation

$\phi_{m_i}(\mathbf{r}_i)$: one-electron molecular orbitals which are usually given as an expansion in "atomic orbitals", $\chi_n$:

$$\phi_m(\mathbf{r}, s) = \sum_n D_{mn}\, \chi_n(\mathbf{r}) \otimes s$$

Coefficients $D$ and $C$ fully characterize the solution of the electronic Schrödinger equation for atoms and molecules.

# Main topic of electronic structure methods

**The development of approximated methods for the solution of the Schrödinger equation for the electrons (given by the coefficients $D$ and $C$) for obtaining the potential for the nuclei dynamics $V(\mathbf{R})$.**

Translation of this problem into a formulation suited for computation: choose an appropriate set of basis functions.

# Basis functions in quantum chemistry

One or many-electron molecular wavefunctions are described in the basis of atom centered functions.

Simplified atomic orbitals: sum of Gaussians centered on the atoms times a polynomial, $P_l$, in the electron coordinates relative to that atom:

$$\chi_n(\mathbf{r}) = \sum_l N_n^l \exp\big( - \alpha_{nl} \, |\mathbf{r} - \mathbf{R}_I^n|^2 \big) \, P_l\big(\mathbf{r} - \mathbf{R}_I^n\big)$$

**Basis sets**: conventional sets of atomic orbitals that cover the polynomials up to a certain order with certain choices of "$\alpha$".

Typical names: `6-31G`, `TZ2P` and `cc-pVQZ`.

Algorithmic procedure for the choice of basis sets: choose a basis, then vary coefficients $D$ and $C$ to approximate $\Psi_{elec}$ that best solves the Schrödinger equation.

# Slater type orbitals (STOs)

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N\, Y_{l,m}(\theta,\varphi)\, r^{\,n-1}\, e^{-\zeta r}$$

$N$: normalization constant
$Y_{l,m}$: spherical harmonic functions
exponential dependence on the distance between the nucleus and the electron: mirrors the exact decay behavior of the orbitals for the hydrogen atom.

STOs are primarily used for atomic and diatomic systems where high accuracy is required, and in semiempirical methods where all three- and four-center integrals are neglected.

# Gaussian type orbitals (GTOs)

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N\,Y_{l,m}(\theta,\varphi)\,r^{\,2n-2-l}\,e^{-\zeta r^2}$$

$$\chi_{\zeta,l_x,l_y,l_z}(x,y,z) = N\,x^{l_x}\,y^{l_y}\,z^{l_z}\,e^{-\zeta r^2}$$

where the sum of $l_x$, $l_y$ and $l_z$ determines the type of orbital (for example $l_x + l_y + l_z = 1$ is a p-orbital).

# Comparison between STO and GTO

GTOs are widely used in electronic structure calculations, but their $r^2$ exponential form introduces some limitations compared with STOs:

At the nucleus, a GTO has zero slope, lacking the “cusp” (discontinuous derivative), behavior of an STO, and thus **does not represent the near-nuclear region as accurately**. Far from the nucleus, a GTO decays too rapidly, so **the tail of the wavefunction is poorly described**.

Both STOs and GTOs can form complete basis. To reach the same accuracy, roughly three times as many GTOs are needed compared with STOs. This increase in size is compensated by the fact that integrals with GTOs can be calculated much more easily. **For reasons of computational efficiency, GTOs are therefore used almost universally in practice**.

# Hartree-Fock method

Hartree-Fock (HF) provides a mean-field approximation:

- Electrons do not interact with each other. The effect of other electrons is accounted for in a mean-field theory context.
- Wavefunction represented as a single Slater determinant made of one-electron molecular orbitals.

# Simplified notation for the Hamiltonian

$$\hat{H}_{\mathrm{el}} = \sum_i \hat{h}(i) + \sum_{i<j} \hat{v}(i,j) + V_{NN}$$

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}$$

$$\hat{v}(i,j) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

- $\hat{h}(i)$: one-electron operator.
- $\hat{v}(i,j)$: two-electron operator.
- $V_{NN}$: constant for fixed set of nuclei $\mathbf{R}$.

# Variational principle

For a time-independent Hamiltonian operator, any trial wavefunction will have an energy expectation value that is greater than or equal to the true ground state wavefunction.

- HF energy is defined variationally:

$$E_{\mathrm{HF}}[\{\phi_i\}] = \frac{\langle \Psi | \hat{H}_{\mathrm{el}} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0$$

- Energy functional minimized with respect to orbitals $\phi_i$.

# Enforcing orthonormality

To ensure that the variational procedure leaves the orbitals orthonormal, Lagrange multipliers are introduced:

$$L[\{\phi_i\}] = E_{\mathrm{HF}}[\{\phi_i\}] - \sum_{ij} \epsilon_{ij} \left( \langle \phi_i | \phi_j \rangle - \delta_{ij} \right)$$

- $\epsilon_{ij}$ are the undetermined Langrange multipliers.
- $\langle i | j \rangle$ is the overlap between spin orbitals $i$ and $j$.

# Hartree-Fock equations

$$
\begin{aligned}
h(x_1)\phi_i(x_1) + \sum_{j\neq i} \left[ \int dx_2\, |\phi_j(x_2)|^2 r_{12}^{-1} \right] \phi_i(x_1) \\
- \sum_{j\neq i} \left[ \int dx_2\, \phi_j^*(x_2)\phi_i(x_2) r_{12}^{-1} \right] \phi_j(x_1) = \epsilon_i \phi_i(x_1)
\end{aligned}
$$

where $\epsilon_i$ is the energy eigenvalue associated with orbital $\phi_i$.

- First term: one-electron Hamiltonian $h(x_1)$.
- Second: Coulomb interaction of an electron in spin orbital $\phi_i$ with the average charge distribution of the other electrons. (**HF: a mean field theory**).
- Third: exchange term (it arises from the antisymmetry requirement of the wavefunction).

# Hartree-Fock operator

$$f(x_1) = h(x_1) + \sum_j \big(J_j(x_1) - K_j(x_1)\big)$$

$$f(x_1)\phi_i(x_1) = \epsilon_i \phi_i(x_1)$$

- Self-consistent eigenvalue problem.
- Orbitals $\phi_i$ are eigenfunctions of $f$.

# Matrix representation of the Hartree-Fock equation

Introducing a basis set transforms the HF equations into the Roothaan equations. Denoting the atomic orbital basis functions as $\chi$, we have the expansion:

$$\phi_i = \sum_{\mu=1}^{K} D_{\mu i}\, \chi_\mu$$

for each spin orbital $i$. This leads to:

$$f(x_1) \sum_\nu D_{\nu i} \chi_\nu(x_1) = \epsilon_i \sum_\nu D_{\nu i} \chi_\nu(x_1)$$

# Projecting the Fock equation to a basis

Left multiplying by $\chi_\mu(\chi_1)$ and integrating yields a matrix equation:

$$\sum_\nu D_{\nu i} \int dx_1 \chi_\mu^*(x_1) f(x_1) \chi_\nu(x_1) = \epsilon_i \sum_\nu D_{\nu i} \int dx_1 \chi_\mu^*(x_1) \chi_\nu(x_1)$$

Introduce matrix element notation:

$$S_{\mu\nu} = \int dx_1 \chi_\mu^*(x_1) \chi_\nu(x_1), \qquad F_{\mu\nu} = \int dx_1 \chi_\mu^*(x_1) f(x_1) \chi_\nu(x_1)$$

# Self-consistent-field procedure

In matrix form:

$$\sum_{\nu} F_{\mu\nu} D_{\nu i} = \epsilon_i \sum_{\nu} S_{\mu\nu} D_{\nu i}$$

More simple as matrices:

$$\mathbf{F}\mathbf{D} = \mathbf{S}\mathbf{D}\varepsilon$$

$\varepsilon$ is a diagonal of the orbital energies $\epsilon_i$.

- Generalized eigenvalue problem: $\mathbf{F}\mathbf{D} = \mathbf{S}\mathbf{D}\varepsilon$
- To simplify: transform to an orthogonal basis ($\mathbf{S} \rightarrow I$).
- Then it reduces to a standard eigenvalue problem for $\mathbf{F}$.
- But: $\mathbf{F}$ depends on the orbitals (and thus on $\mathbf{D}$).
- $\Rightarrow$ Must solve **iteratively** until self-consistency is reached.

**Solution of Hartree-Fock-Roothaan equations are often called the self-consistent-field procedure**.

# Self-consistent-field procedure

1. Build an initial density $P^{(0)}$.
2. **Loop until convergence**:
    1. Form $F^{(k)}$ from $P^{(k)}$.
    2. Solve $F^{(k)}C^{(k+1)} = SC^{(k+1)}\varepsilon^{(k+1)}$.
    3. Update $P^{(k+1)}$ from occupied columns of $C^{(k+1)}$.
    4. Check $\Delta E$, $\|\Delta P\|$.
3. Report $E_{\text{el}}$ and orbitals.