

The N-body quantum problem in quantum chemistry

Michel Caffarel

CNRS-Lab. de Chimie et Physique Quantiques,

Université de Toulouse

caffarel@irsamc.ups-tlse.fr

A lot of warnings ...

General remarks

- Here, **non-relativistic electronic** N-body quantum problem
- Another N-body quantum problem in quantum chemistry (QC) : **ro-vibrational Hamiltonian** and infrared spectra
- **Relativistic effects** may play an important role

Generalities

- 1) the N-body quantum Hamiltonian of QC
- 2) The paradigm of dynamical/static correlation energy in QC

Methods

- 1) Density functional theory
- 2) Wave function theory
 - Configuration Interaction-type (CI) approaches
 - Selected CI (deterministic and stochastic)
 - New representations of the CI wavefunction: Matrix Product (Tensor Network) States (DMRG and generalizations)
- 3) Quantum Monte Carlo
- 4) GW-type approaches
- 5) Machine Learning in quantum chemistry

The Hamiltonian

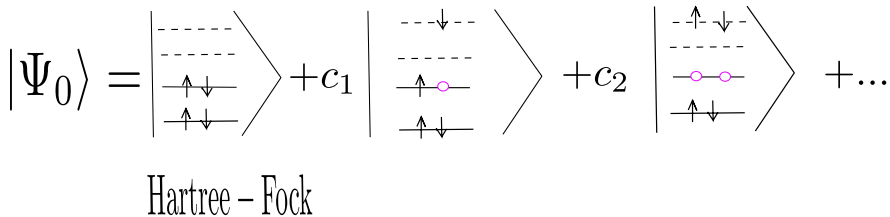
Ab initio molecular Hamiltonian

$$H = \sum_{i,j\sigma} \langle \phi_i | h | \phi_j \rangle a_{j\sigma}^+ a_{i\sigma} + \sum_{i,j,k,l,\sigma,\sigma'} \langle \phi_i \phi_j | \frac{1}{r_{12}} | \phi_k \phi_l \rangle a_{l\sigma'}^+ a_{k\sigma}^+ a_{j\sigma'} a_{i\sigma}$$

- $\phi_i(r)$ basis of one-body functions = molecular orbitals. In practice, a finite number \rightarrow *finite basis set effects*
- The **Hamiltonian is essentially exact**
- Strong attractive nucleus centers
- **Combinatorial explosion** of the size of the Hilbert space (number of determinants). However, **not specific to QC**

The paradigm of dynamical/static correlation energy

- Used to rationalize the various contributions beyond mean-field (Hartree-Fock)
- At the root of classification methods in QC



Definition of the **correlation energy**

$$E_c = E_0 - E_{HF}$$

- Chemistry is not interested in total energies but into **differences of total energies**
- "Chemical accuracy": $\Delta E \sim 1 \text{ kcal/mole} \sim 0.0016 \text{ a.u}$

Fundamental point: $|\Delta E| \ll |E_c| \ll |E_0|$

- No Quantum Chemistry method is able to reach the required accuracy except for very few-electron systems
- Quantum Chemistry is the **science of compensation of errors** based on physically relevant and coherent approximations

Decomposition of the correlation energy

$$E_c = E_{static} + E_{dynamic} + E_{stat-dyn}$$

1) **Static or non-dynamical correlation energy**

If unoccupied molecular orbitals are close in energy to the highest occupied one, the one-determinant Hartree-Fock wavefunction is qualitatively wrong.

The static correlation energy is the energetic contribution directly resulting from this multi-determinant character →

MultiReference (MR) methods

In practice, **three important cases:**

- 1) Breaking of bonds
- 2) System with open shell transition metal atoms
- 3) Excited-states

Correlation energy

2) **Dynamical correlation energy**

The expansion in determinants (antisymmetrized product of one-body MO's) is not the appropriate basis for describing $1/r_{ij} \rightarrow$ very slow convergence of CI expansion

Very short range electron-electron interaction

The kinetic term $\frac{1}{2} \frac{\nabla^2 \Psi}{\Psi}$ is expected to diverge as $1/r_{ij}$ when two-electron approaches (electron-electron cusp conditions).

Short-range interaction: depends strongly on the two-body matrix (probability of finding one-electron around an other electron).

Correlation energy

3) **Coupling static/dynamical correlation**

It can be important or not.

It is the case when the relative weight of the main determinants related to static correlation are significantly affected by dynamical correlation effects

Density Functional Theory (DFT)

- DFT for ground-state properties
- Time-Dependent DFT (TD-DFT) for transition energies

Unique feature of DFT : **From $\Psi(1, 2, \dots, N)$ to $\rho(\mathbf{r})$ in an exact manner**

Theory: Electrons moving in a one-body effective potential V_{xc} depending only on ρ . For each electron i

$$h_i = -\frac{1}{2}\nabla_i^2 + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} + V_{e-N}(\mathbf{r}) + V_{xc}[\rho(\mathbf{r})]$$

where

$$V_{xc}[\rho(\mathbf{r})] = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}$$

Challenge: Find accurate representations of the exchange-correlation energy functional $E_{xc}(\rho)$

Plethora of approximations (The "Functional zoo" of Perdew)

In practice, decomposition into an exchange and correlation part

$$E_{xc}(\rho) = E_x(\rho) + E_c(\rho)$$

A large variety of approximations for E_x and E_c .

- Approximate forms built from known limits (uniform electron gas, small electron-electron distance regime, etc.).
- Use of exact mathematical constraints for the unknown exact functional
- Making DFT Functionals depending on other variable besides the density: $\nabla\rho$, $\nabla^2\rho$, kinetic energy density (orbital-dependent!), Hartree-Fock exchange (orbital-dependent), etc.
- Mixture with wavefunction methods: hybrid functional (introduction of a part of the Hartree-Fock exchange), double hybrid (a part of the correlation energy computed by perturbation (MP2), range-separation approach (mixture of wavefunction and DFT without double counting of correlation energies), etc.
- Brute force approach. Parametrization and optimization, for example, by machine learning

In practice, a variety of benchmark sets (G3, S22, etc.) are introduced and performances of various functionals are compared to experimental or accurate theoretical results.

Some important difficulties and challenges

- 1) **Practical challenge:** Improve (both in accuracy and domain of applications) current DFT functionals without losing the usefulness of DFT (reasonable cost for large systems, pb. of the scaling of the cost).
- 2) Static correlation energy
- 4) Self-interaction and delocalization errors (self-interaction tends to make Kohn-Sham orbitals too delocalized, use of Hartree-Fock, etc.)
- 5) Weak interactions (van der Waals)

Wave-function based methods

Configuration Interaction (CI) expansion

$$\Psi = D_{HF} + \sum_{I>0} c_I D_I$$

The remaining contribution can be seen as resulting from the residual potential

$$V = H - H_{HF}$$

- **Perturbation approach:** Møller-Plesset expansion MP2,...MPn
- **Diagonalization approach:** CIS (diagonalization of H in the space all single excitations), CISD (diagonalisation of H in the space all double excitations) and so on up to the Full CI (FCI) limit (all possible excitations)

Taking *explicitly* into account near-degeneracy (static correlation energy)

Choice of a (complete) active space made up of the most relevant low-lying orbitals (may include lot of heuristic)

CAS-CI (orbitals HF)

CAS-SCF (orbitals are optimized self-consistently)

- **Perturbation** CAS-PT2 (or better NEV-PT2)
- **Diagonalization:** MRCI

The separation static-dynamic correlation energy may severely break down. Internally decontracted approaches: perturbation method can be used for revisiting the multireference.

An important aspect for chemistry (reactivity): size consistency (strict separability)

When breaking apart a molecule the wavefunction must factorize into fragment wavefunctions. It is realized using an exponential structure of the wavefunction.

- Coupled Cluster approaches (from nuclear physics!)

$$\Psi = e^{T_A + T_B} \Psi_0$$

The so-called **"golden" standard** of QC = CCSD(T). Efficient method for obtaining the "chemical accuracy" but, unfortunately, scaling in N^7

- MRCC= yet another story...

Some developments:

- *F12* – *methods* to reduce the finite-size effects
- Linear-scaling approaches for treating large systems

Fighting the exponential increase... Selected CI approaches

Recent strong revival of a **very old idea**: Only a very tiny part of the CI determinants plays a significant role, we need to select them

They have become **state-of-the art approaches for multireference problems** (together with DMRG). For example, Zimmerman (2017) FCI space=(118o,32e), $N_{det} \sim 10^{38}$.

Common idea to all variants:

Starting from $|D_0\rangle$, the important determinants connected to it by H may be selected using as criterium (perturbative argument)

$$\frac{|\langle D|H|D_0\rangle|^2}{\langle D|H|D\rangle - E_0} \text{ large}$$

A number of such determinants are selected, and the process is iterated.

The selection of determinants can be done

- **Stochastically**: the FCI-QMC approach of Alavi et al.
- **Deterministically** : CIPSI and all other recent variants (HBCI, ACI, etc.)

Adding dynamical correlation energy: Second-order perturbation (PT2) treatment is particularly efficient but very expensive due to the large number of selected determinants (up to $\sim 10^9$) \Rightarrow use of hybrid stochastic-deterministic approaches

Important: Selection method can be used for any CI-type approach, for example in the case of CASSCF: Allow to reach CAS(50o,50e)!!

New representations of the CI wavefunction: Matrix Product (Tensor Network) States (DMRG and generalizations)

It has become [state-of-the-art approaches for large multireference problems](#). For example, Li et al. (2019), FCI space including $\sim 3.610^{35}$ dets.

Standard Configuration Interaction (CI) representation:

$$\Psi = \sum_{n_1, n_2, \dots, n_M} c_{n_1 n_2 \dots n_M} |n_1 n_2 \dots n_M\rangle \quad n_i = 0, 1$$

Coefficients c obtained from diagonalization of the H matrix

Matrix Product State representation:

$$\Psi = \sum_{n_1, n_2, \dots, n_M} \text{Tr}[A_1^{n_1} \dots A_M^{n_M}] |n_1 n_2 \dots n_M\rangle$$

Matrix elements of matrices A_i 's are the variational parameters of the representation. Sizes of A_i 's

$$(1, 2^1) \times (2^1, 2^2) \times \dots \times (2^{\frac{M}{2}}, 2^{\frac{M}{2}+1}) \times (2^{\frac{M}{2}+1}, 2^{\frac{M}{2}}) \times \dots \times (2^2, 2) \times (2^1, 1)$$

Key point: How to make efficiently the **optimization** of the very numerous parameters. This form allows to optimize the wavefunction step by step by **sweeping on matrices A** .

Density Matrix Renormalization Group (DMRG) is one way of doing that.

- Adapted to static correlation energy
- Adding on top dynamical correlation energy

{ DMRG + PT2
DMRG + short ranged DFT
etc.

Quantum Monte Carlo

Once $t \rightarrow it$ (which does not change the time-independent quantities) the time-dependent Schrödinger equation can be interpreted as a **diffusion equation** for electrons

$$\frac{\partial \psi(x, t)}{\partial t} = \frac{1}{2} \nabla^2 \psi(x, t) - V(x) \psi(x)$$

Introducing a (good approximation) of the wavefunction, the **trial wavefunction** $\psi_T(x)$ we get

$$\frac{\partial f}{\partial t} = \frac{1}{2} \nabla^2 f - \nabla(\mathbf{b}f) - E_L(x)f$$

Simulation: Electrons are **diffused**, **drifted** with $\mathbf{b}(x) = \frac{\nabla \psi_T}{\psi_T}$ and **duplicated/killed** proportionally to the local energy $E_L(x) = \frac{H\psi_T}{\psi_T}$.

At large time, the distribution $f(x, t)$ evolves toward $\phi_0(x)$

QMC II

Exact approach, good scaling, trivial to implement, massively parallel

However

- electrons are blocked by the zeroes of $\Psi_T \Rightarrow$ **fixed-node error**
- State-of-the-art of total energies. Problem of the **control of the fixed-node error**, in particular in the energy differences

GW-type approaches

- State-of-the-art approach to compute energy gaps in solid-state physics.
- Growing interest in quantum chemistry: Calculation of electronic excitations (a priori better but more expensive than TD-DFT) but also ground-state correlation energies

Theory based on 1-, 2-particle,... Green's functions, $G_i(t)$

For example, $G_1(t)$ = overlap between the initial state $|\Phi(t=0)\rangle = a^+|\Psi_0(N)\rangle$ with itself after some time of evolution t

$$|\Phi(t)\rangle = e^{-itH_{N+1}t} a^+|\Psi_0(N)\rangle$$

$$G(t) \sim \langle \Phi(t) | \Phi(t=0) \rangle$$

Physically, it gives **a measure of how the electron keeps or not its integrity as an electron** in presence of interaction with other electrons.

Equations obeyed by G_k , $k=1,N$

Full N -particle Green's function easily obtained

$$G_N(t) \sim e^{-itH_N} \quad \text{with} \quad i\frac{\partial G_N}{\partial t} = H_N G_N$$

Fourier transform : $G_N(\omega) \sim \frac{1}{\omega - H_N}$.

$$\frac{1}{\omega - H_N} = \frac{1}{\omega - H_N^0} + \frac{1}{\omega - H_N^0} (H_N - H_N^0) \frac{1}{\omega - H_N}$$

\Rightarrow **N-particle Dyson eq.:** $G_N(\omega) = G_N^0(\omega) + G_N^0(\omega) V_N G_N(\omega)$

Hierarchy of coupled Dyson-like equations for the Green's functions G_k from $k = N$ to $k = 1$

Equations II

For $k = 2$:

Dyson-like equation for $G_2 \Rightarrow =$ **Bethe-Salpeter equation**

For $k = 1$:

Dyson eq. for G_1

$$G_1(\omega) = G_1^0(\omega) + G_1^0(\omega)\Sigma(\omega)G_1(\omega)$$

\Rightarrow Exact formalism going from a N-particle Dyson equation involving the **known bare potential V** to a one-particle Dyson equation **involving the unknown "renormalized" potential $\Sigma(\omega)$** encoding all interactions

GW approximation

The self-energy can be expressed as

$$\Sigma = i \int GW(1 + K)$$

where K is a contribution connected to the 2-particle Green's function (itself obeying the Bethe-Salpeter equation). Taking $K = 0$ decouples the hierarchy of Green's functions and leads to an autonomous (Dyson) equation for G , this is the **GW approximation**

$$\Sigma \sim i \int GW$$

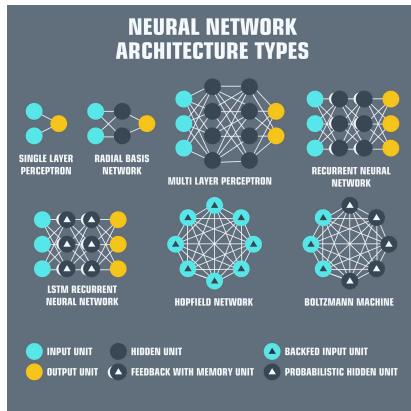
GW IV

Here, W can be seen as a **screened coulombic** interaction, it can be expressed in function of the polarization (response) function χ

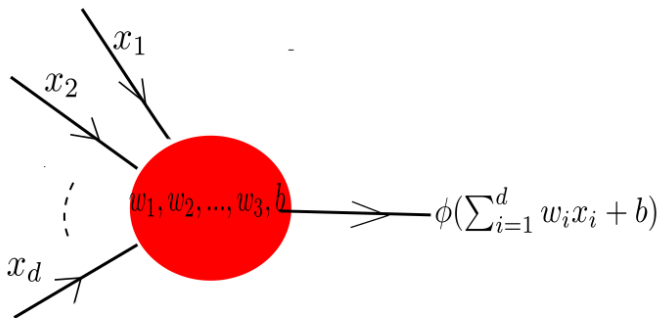
In practice, G and χ are expressed using the KS orbitals and a self-consistent scheme is used or not (G_0W_0 variant) to get the Green's function.

Machine Learning in QC

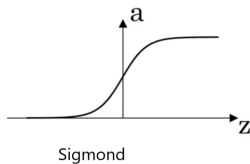
Machine learning in QC: mainly supervised neural networks



Formal neuron

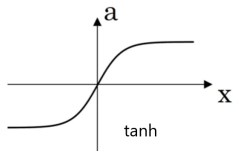


Examples of activation functions $\phi(x)$



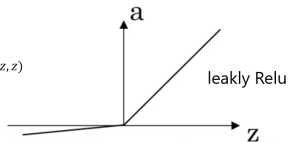
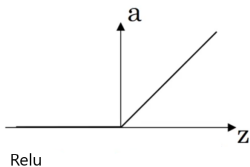
$$\text{Sigmoid: } a = \frac{1}{1 + e^{-z}}$$

$$\text{tanh: } a = \frac{e^x - e^{-x}}{e^{-x} + e^x}$$

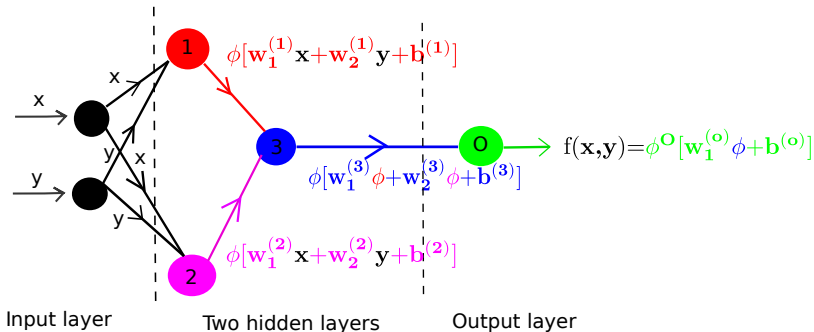


$$\text{Relu: } a = \max(0, z)$$

$$\text{Leaky Relu: } a = \max(0.01z, z)$$



Combining neurons: Neural Networks



$$f(x,y) = \phi^{(O)} \left[w_1^{(O)} \phi \left[w_1^{(3)} \phi \left[w_1^{(1)} x + w_2^{(1)} y + b^{(1)} \right] w_2^{(3)} \phi \left[w_1^{(2)} x + w_2^{(2)} y + b^{(2)} \right] + b^{(3)} \right] + b^{(O)} \right]$$

The NN function is **highly-nonlinear** but **analytically known**
Difficult to visualize what kind of fit function it is!

BUT: Universal approximation theorem: Cybenko (1989),
Hornik (1991)

However, (remarks borrowed from S. Mallat, lecture notes of
Collège de France (2019))

1. Fit is not supposed to work
2. If it does work (good choice of the hyperparameters), results may be extraordinary
3. We do not really know **when,how**, and **why** it works

Example: fit of DFT data

1) Learning step

Training set = a large set of DFT calculations for various geometries \mathbf{R}_i , $E_{DFT}(\mathbf{R}_i)$, $i = 1, N$

Optimization of the parameters of the NN of each neuron,

$$\mathbf{p} = [(w^1, b^1), \dots, (w^k, b^k), \dots]$$

For example, by minimization of

$$\chi^2(\mathbf{p}) = \sum_{i=1, N} [E_i^{DFT} - f(\mathbf{p}, \mathbf{R}_i)]^2$$

2) **Exploitation step** Use of the NN for arbitrary values for arbitrary geometries \mathbf{R} , **restitution time much smaller than the cost of one DFT calculation.**

- Quantum chemistry is the **science of compensation of errors**
- Is it also **the science of acronyms?**

Only in this presentation:

CI, DMRG, QMC, GW, HF, DFT, TD-DFT, MP2, CIS, CISD, FCI, FCI-QMC, CASCI, CASSCF, CASPT2, NEVPT2, CIPSI, HBCI, ACI, MRCI, CCSD(T), PT2

