

From the Schroedinger Equation to Molecular Dynamics (the notion of molecular structure, atomistic simulations, direct/ inverse problems)

Andrea Giansanti

Dipartimento di Fisica, Sapienza Università di Roma

Andrea.Giansanti@roma1.infn.it

CB_22_23 LECTURE N. 2 and 3, Rome 26th and 28th Sep 2023

Slides either inspired or based on friendly shared suggestions by guest lecturers in the previous courses: Giuseppe D'Adamo and Giovanni Ciccotti

DIPARTIMENTO DI FISICA



SAPIENZA
UNIVERSITÀ DI ROMA

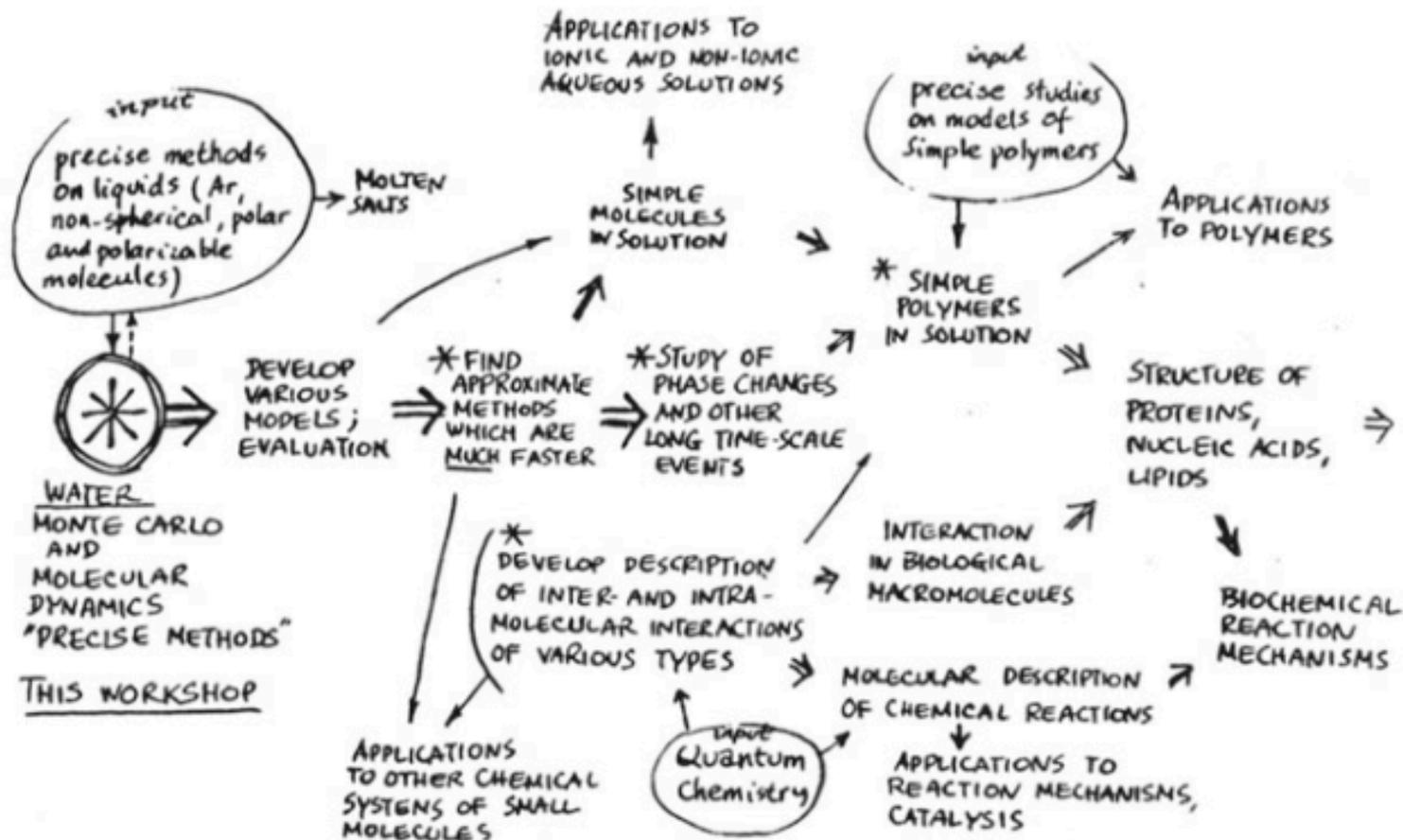
outline lectures 2 and 3

- **FROM THE SCHROEDINGER EQUATION TO MOLECULAR DYNAMICS**
-
- **-Molecular structures and the Born-Oppenheimer approximation**
- **-Basic principles of Ab-initio Molecular Dynamics: Born-Oppenheimer surfaces**
- **-Hellman-Feynman Theorem**
- **MODELLING $\epsilon_0(\mathbf{R})$ FORCE FIELDS**
- **Morse potential as the simplest force field**
- **Force field as a network (just the concept)**
- **Pairwise additive approximation**
- **General form of a force fields: bonded terms/non bonded terms**

- **BIBLIOGRAPHY PACK_2**
- **slides**
- **Tuckerman_Born-Oppenheimer [S]**
- **Griebel2007_Numerical Simulation in Molecular Dynamics_Chap2.pdf [S]**
- **Feynman1939.pdf [R]**

Architecture of the course

- Atomistic simulation (deterministic/stochastic)
- Data analysis (dimensional reduction)
- Data analysis bayesian inference
- **Direct/inverse** modelling
- Stochastic methods in computational biology(evolution and the space of biological sequences)
- Programming



4.8.72

Herman Berendsen's prevision for the future developments of molecular simulation, that he presented at the 1972 CECAM workshop on "Molecular Dynamics and Monte Carlo calculations of water"

Equilibrium (classical) Statistical Mechanics THE DIRECT PROBLEM

$$O = \overline{\hat{O}(R, P)} \equiv \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau dt \hat{O}(R(t), P(t)) \quad \text{micro} \quad (1)$$

Temporal average of an observable along a trajectory generated by the newtonian dynamics

$$\simeq \lim_{\mathcal{N} \rightarrow \infty} \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \hat{O}(R(ih), P(ih)) \quad (2)$$

Estimator of (1)

Relative probability of state l

$$= \lim_{\mathcal{N} \rightarrow \infty} \sum_{\ell \in \text{visited states}} \left[\frac{n_\ell}{\mathcal{N}} \right] \hat{O}_\ell \quad (3)$$

Estimator of (1) as an expected value over a probability scheme attached to “states”, to be considered as statistically independent bunches of snapshots, similar to **meso-states**

$$\simeq \int dRdP \rho(R, P) \hat{O}(R, P) = \langle \hat{O}(R, P) \rangle \quad (4)$$

Ensemble average as in MonteCarlo: (1) and (4) are the same if the dynamics is ergodic

General remarks, to partly recapitulate the previous lecture

Biology is an **evolutionary science**, i.e. *the living matter is the result of an historical experience, through random mutations, selection and self-organization.*

The big difference between physics and biology is that the former treats living matter without an evolutionary point of view (...Galilei's subtraction of the animal ...)

One of the aims of our course is to make a bridge from the physical approach, based on the B-O approach to the symbolic space of biological sequences and to integrative modelling.

[the archives of biology (Pauling and Zuckerkandl 1965) made by texts written in a 20 letters alphabet (proteins, amino acids) and in a 4 letter alphabet (nucleic acids, nucleotides). See the parallel courses in Biochemistry and Molecular biology. From the atomistic point of view: everything is made of 4 basic atoms, namely; C, H, N, O.

Our target is succeed in interpreting this language through machine learning probabilistic methods aimed at letting the information and information fluxes emerge (Statistical Biophysics--->see Theoretical Biophysics)

- Problems of Reductionism and the search for fundamental dofs.
(e.g. free-electron model of metal conductivity)
- P.W.Anderson opened the season of “complexity” in 1972
(P.W. Anderson, *More is different*. Science, 177, 393-396 (1972)).

In a nutshell:

- THERE IS NO FUNDAMENTAL SCALE
- AT EACH SCALE THERE IS NEW PHYSICS
- EACH SCALE HAS ITS OWN LEVEL OF COMPREHENSION

ALLRIGHT! But, then if a theory is scale invariant then it should be robust against crossing the scales see- e-g- the problem of building

Portable, scale invariant, coarse-grained energy functions

(see e.g. D’Adamo, Pelissetto, Pierleoni)

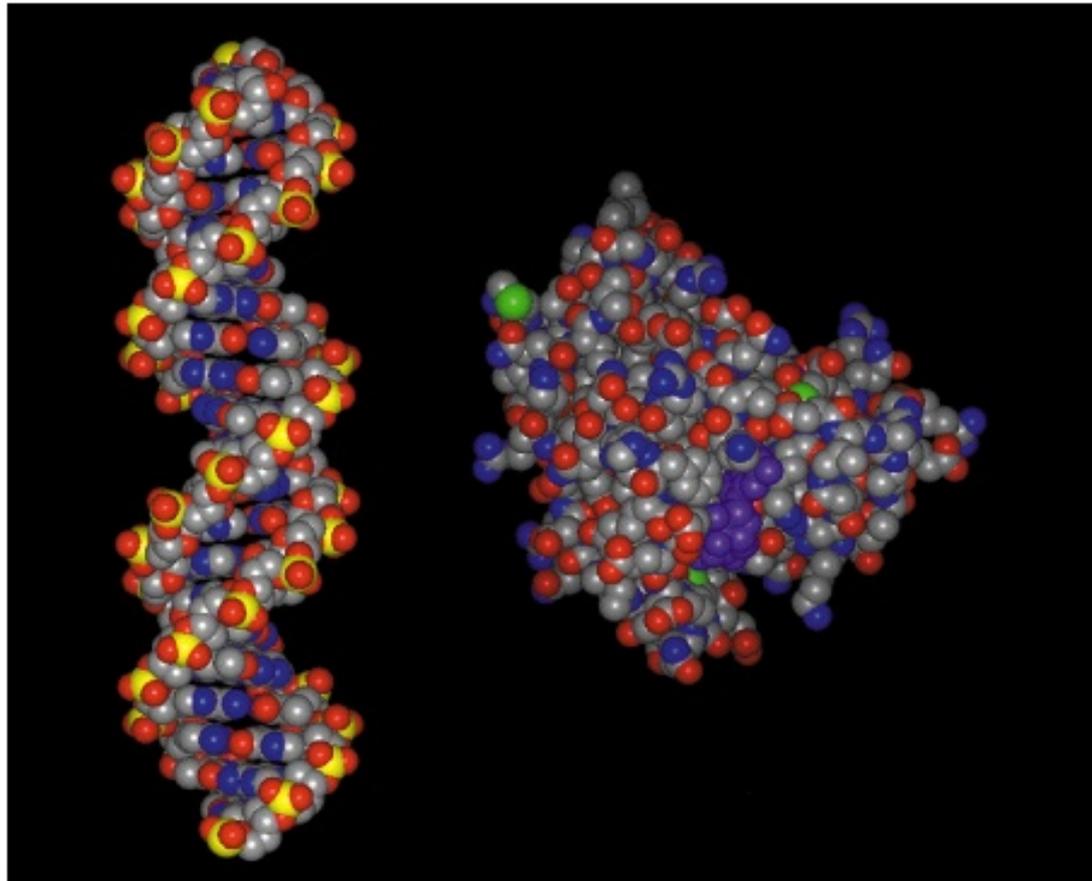
TO BE USED IN ATOMISTIC SIMULATIONS

HOMEWORK: INSTALL CHIMERA!

<https://www.cgl.ucsf.edu/chimera/>

<https://www.cgl.ucsf.edu/chimerax/>

crystal



Nucleic acid
(DNA)

Protein
(Ras protein)

1 nm

Amorphous,
glass



MOLECULAR STRUCTURE

(...an intrinsically classical notion)

- Mechanical equilibrium configurations
- No net force condition (what is a force in QM?)
- Minima of the potential energy landscape

MOLECULAR DYNAMICS WHAT FOR

◆ a classical system of *quantum* particles interacting via an effective interaction potential,

is obtained ab initio, from quantum mechanics, (AIMD) by some suitable fitting procedure → *phenomenological* model, e.g. with a Pairwise Additive Potential,

◆ using models of PAP, the equations of motion are numerically integrable for a number of particles finite but large enough to study, by statistical approach, the thermal properties of matter

Also dynamical (time dependent) features emerge from the simulation if the model is reliable

For the students who are curious about the history of what we are talking about you could see the freshly published book...

The “fundamental” law at atomic scale

$$i\hbar \frac{\partial \tilde{\Psi}(r, R, t)}{\partial t} = \mathcal{H} \tilde{\Psi}(r, R, t) \Leftrightarrow \mathcal{H} \Psi_n = E_n \Psi_n$$

Non relativistic time dependent Schroedinger Equation with the “matter hamiltonian”:

$$\mathcal{H}(r, R; p, P) = K_N(P) + \underbrace{K_e(p) + V(r, R)}_{H_e(r, p|R)}$$

Coulomb term

Electronic hamiltonian

Born-Oppenheimer approximation

Nuclear masses exceed electronic ones by 2-3 orders of magnitude: $m_N \gg m_e$

$$H_e(r, p | R) \Phi_s(r | R) = E_s(R) \Phi_s(r | R)$$

We split the molecular Schroedinger. Equation into two: electronic and nuclear

Electronic wave-functions are more delocalized

$$|\nabla_R \Phi_s| \ll |\nabla_r \Phi_s|$$

$$\Psi(r, R; t) = \sum_s \chi_s(R; t) \Phi_s(r | R) \simeq \chi_0(R; t) \Phi_0(r | R)$$

adiabatic approximation: B. O. 1927

The evolution of the nuclear approximated w.f. $\chi_0(R; t)$ is given by:

$$\begin{aligned} i\hbar \frac{\partial}{\partial t} \chi_0(R; t) &= \mathcal{H}_N(R, P) \chi_0(R; t) \\ &\equiv [K_N(P) + E_0(R)] \chi_0(R; t) \end{aligned}$$

The dynamics of the nuclei, apparently independent from the electrons, is driven by the B.O. surface: $E_{BO}(R)$ as the interaction potential (a mean field, modelizable, no more Coulomb!)-->The eigenvalue of the ground state electronic function parametrized by R.

In the rigid adiabatic approx. no excited electronic states, no jumps are considered.

The force fields of protein MD come from **models** of E_{BO}

If nuclei are heavy enough and/or temperature is high enough so that the thermal quantum length is negligible w.r. to nuclear separations:

$$\Lambda = \frac{h}{\sqrt{m_N k_B T}} \ll \text{internuclear } r$$

Then the dynamics is no more quantum, but
Newton: and nuclei move classically in the
effective potential built up by electronic glue

$$m_N \ddot{R} = -\nabla E_0(R)$$

I. THE BORN-OPPENHEIMER APPROXIMATION

The next few lectures will treat the problem of quantum chemistry, a subfield of quantum mechanics also known as molecular quantum mechanics. The idea of quantum chemistry is to use only the simple facts molecules and, indeed, all of ordinary matter, can be viewed as composed only of positively charged nuclei and negatively charged electrons. This *universal* description is then subject to a quantum mechanical treatment from which the properties of the system are derived or computed.

We begin our discussion of AIMD by considering a system of N nuclei described by coordinates, $\mathbf{R}_1, \dots, \mathbf{R}_N \equiv \mathbf{R}$, momenta, $\mathbf{P}_1, \dots, \mathbf{P}_N \equiv \mathbf{P}$, and masses M_1, \dots, M_N , and N_e electrons described by coordinates, $\mathbf{r}_1, \dots, \mathbf{r}_{N_e} \equiv \mathbf{r}$, momenta, $\mathbf{p}_1, \dots, \mathbf{p}_{N_e} \equiv \mathbf{p}$, and spin variables, $s_1, \dots, s_{N_e} \equiv s$. The Hamiltonian of the system is given by

$$\begin{aligned} H &= \sum_{I=1}^N \frac{\mathbf{P}_I^2}{2M_I} + \sum_{i=1}^{N_e} \frac{\mathbf{p}_i^2}{2m} + \sum_{i>j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I>J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{R}_I - \mathbf{r}_i|} \\ &\equiv T_N + T_e + V_{ee}(\mathbf{r}) + V_{NN}(\mathbf{R}) + V_{eN}(\mathbf{r}, \mathbf{R}) \end{aligned}$$

where m is the mass of the electron, and $Z_I e$ is the charge on the I th nucleus. In the second line, T_N , T_e , V_{ee} , V_{NN} , and V_{eN} represent the nuclear and electron kinetic energy operators and electron-electron, electron-nuclear, and nuclear-nuclear interaction potential operators, respectively. Note that this Hamiltonian is universal in that it describes all of everyday matter from biological macromolecules such as proteins, enzymes and nucleic acids, to metals and semiconductors to synthetic materials such as plastics. Thus, if we could solve for the eigenvalues and eigenfunctions of this Hamiltonian, we could, predict any property we wished of a given system. This fact lead the

$$[T_{\text{N}} + T_{\text{e}} + V_{\text{ee}}(\mathbf{r}) + V_{\text{NN}}(\mathbf{R}) + V_{\text{eN}}(\mathbf{r}, \mathbf{R})] \Psi(\mathbf{x}, \mathbf{R}) = E\Psi(\mathbf{x}, \mathbf{R})$$

Molecular eigen-value problem

$$\Psi(\mathbf{x}, \mathbf{R}) = \phi(\mathbf{x}, \mathbf{R})\chi(\mathbf{R})$$

Separation of scales, adiabatic factorization... $\nabla_I\chi(\mathbf{R}) \gg \nabla_I\phi(\mathbf{x}, \mathbf{R})$

$$\frac{[T_{\text{e}} + V_{\text{ee}}(\mathbf{r}) + V_{\text{eN}}(\mathbf{r}, \mathbf{R})] \phi(\mathbf{x}, \mathbf{R})}{\phi(\mathbf{x}, \mathbf{R})} = E - \frac{[T_{\text{N}} + V_{\text{NN}}(\mathbf{R})] \chi(\mathbf{R})}{\chi(\mathbf{R})}$$

$$[T_{\text{e}} + V_{\text{ee}}(\mathbf{r}) + V_{\text{eN}}(\mathbf{r}, \mathbf{R})] \phi(\mathbf{x}, \mathbf{R}) = \varepsilon(\mathbf{R})\phi(\mathbf{x}, \mathbf{R}) \quad \text{Electronic eq}$$

$$[T_{\text{N}} + V_{\text{NN}}(\mathbf{R}) + \varepsilon_n(\mathbf{R})] \chi(\mathbf{R}) = E\chi(\mathbf{R}) \quad \text{nuclear}$$

Moreover, each electronic eigenvalue, $\varepsilon_n(\mathbf{R})$, will give rise to an electronic *surface*, and these surfaces are known as *Born-Oppenheimer surfaces*. Thus, the full internuclear potential for each electronic surface is given by $V_{\text{NN}}(\mathbf{R}) + \varepsilon_n(\mathbf{R})$. On each Born-Oppenheimer surface, the nuclear eigenvalue problem can be solved, which yields a set of levels

Nuclear motions are ruled by the time dependent Schroedinger Eq.

$$[T_N + V_{NN}(\mathbf{R}) + \varepsilon_n(\mathbf{R})] X(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} X(\mathbf{R}, t)$$

$$[T_e + V_{ee}(\mathbf{r}) + V_{eN}(\mathbf{r}, \mathbf{R})] \phi_0(\mathbf{x}, \mathbf{R}) = \varepsilon_0(\mathbf{R}) \phi_0(\mathbf{x}, \mathbf{R})$$

$$[T_N + \varepsilon_0(\mathbf{R}) + V_{NN}(\mathbf{R})] X(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} X(\mathbf{R}, t)$$

Coupling the nuclear dynamics with electronic on the flight solutio of the electronic eigen-value problem is **Ab-Initio Car-Parrinello MD**

Moreover, if nuclear quantum effects can be neglected, then we may arrive at classical nuclear evolution by assuming $X(\mathbf{R}, t)$ is of the form

$$X(\mathbf{R}, t) = A(\mathbf{R}, t)e^{iS(\mathbf{R}, t)/\hbar} \quad (11)$$

and neglecting all terms involving \hbar , which yields an approximate equation for $S(\mathbf{R}, t)$:

$$H_N(\nabla_1 S, \dots, \nabla_N S, \mathbf{R}_1, \dots, \mathbf{R}_N) + \frac{\partial S}{\partial t} = 0 \quad (12)$$

which is just the classical Hamiltonian-Jacobi equation with

$$H_N(\mathbf{P}_1, \dots, \mathbf{P}_N, \mathbf{R}_1, \dots, \mathbf{R}_N) = \sum_{I=1}^N \frac{\mathbf{P}_I^2}{2M_I} + V_{\text{NN}}(\mathbf{R}) + \varepsilon_0(\mathbf{R}) \quad (13)$$

denoting the classical nuclear Hamiltonian. The Hamilton-Jacobi equation is equivalent to classical motion on the ground-state surface, $E_0(\mathbf{R}) = \varepsilon_0(\mathbf{R}) + V_{\text{NN}}(\mathbf{R})$ given by

$$\begin{aligned} \dot{\mathbf{R}}_I &= \frac{\mathbf{P}_I}{M_I} \\ \dot{\mathbf{P}}_I &= -\nabla_I E_0(\mathbf{R}) \end{aligned} \quad (14)$$

Note that the force $-\nabla_I E_0(\mathbf{R})$ contains a term from the nuclear-nuclear repulsion and a term from the derivative of the electronic eigenvalue, $\varepsilon_0(\mathbf{R})$. Because of the Hellman-Feynman theorem, the latter can be expressed as

$$\nabla_I \varepsilon_0(\mathbf{R}) = \langle \phi_0(\mathbf{R}) | \nabla_I H_e(\mathbf{R}) | \phi_0(\mathbf{R}) \rangle$$

II. PROOF OF THE HELLMAN-FEYNMAN THEOREM

Consider a system with a Hamiltonian $H(\lambda)$ that depends on some parameters λ . Let $|\psi(\lambda)\rangle$ be an eigenvector of $H(\lambda)$ with eigenvalue $E(\lambda)$

$$H(\lambda)|\psi(\lambda)\rangle = E(\lambda)|\psi(\lambda)\rangle \quad (16)$$

We further assume that $|\psi(\lambda)\rangle$ is normalized so that

$$\langle\psi(\lambda)|\psi(\lambda)\rangle = 1 \quad (17)$$

The Hellman-Feynman theorem states that

$$\frac{dE}{d\lambda} = \left\langle\psi(\lambda)\left|\frac{dH}{d\lambda}\right|\psi(\lambda)\right\rangle \quad (18)$$

The proof of the Hellman-Feynman theorem is straightforward. We begin with the fact that

$$E(\lambda) = \langle\psi(\lambda)|H(\lambda)|\psi(\lambda)\rangle \quad (19)$$

Differentiating both sides yields

$$\frac{dE}{d\lambda} = \left\langle\psi(\lambda)\left|\frac{dH}{d\lambda}\right|\psi(\lambda)\right\rangle + \left\langle\frac{d\psi}{d\lambda}\left|H(\lambda)\right|\psi(\lambda)\right\rangle + \left\langle\psi\left|H(\lambda)\right|\frac{d\psi}{d\lambda}\right\rangle \quad (20)$$

Since $|\psi(\lambda)\rangle$ is an eigenvector of $H(\lambda)$, this can be written as

$$\begin{aligned} \frac{dE}{d\lambda} &= \left\langle\psi(\lambda)\left|\frac{dH}{d\lambda}\right|\psi(\lambda)\right\rangle + E(\lambda)\left\langle\frac{d\psi}{d\lambda}\left|\psi(\lambda)\right\rangle + E(\lambda)\left\langle\psi\left|\frac{d\psi}{d\lambda}\right\rangle \right. \\ &\left. \left\langle\psi(\lambda)\left|\frac{dH}{d\lambda}\right|\psi(\lambda)\right\rangle + E(\lambda)\left[\left\langle\frac{d\psi}{d\lambda}\left|\psi(\lambda)\right\rangle + \left\langle\psi\left|\frac{d\psi}{d\lambda}\right\rangle\right]\right. \end{aligned}$$

However, since $|\psi(\lambda)\rangle$ is normalized, we have, from the normalization condition:

$$\begin{aligned} \langle\psi(\lambda)|\psi(\lambda)\rangle &= 1 \\ \left\langle\frac{d\psi}{d\lambda}\left|\psi(\lambda)\right\rangle + \left\langle\psi\left|\frac{d\psi}{d\lambda}\right\rangle &= 0 \end{aligned}$$

Hence, the term in square brackets vanishes, and we have

$$\frac{dE}{d\lambda} = \left\langle\psi(\lambda)\left|\frac{dH}{d\lambda}\right|\psi(\lambda)\right\rangle \quad (21)$$

which is just the Hellman-Feynman theorem.

So, we have grounded the notion of molecular structure and molecular dynamics on quantum mechanics. Moreover, we have given, through Hellmann-Feynman's theorem a meaning to the notion of Molecular forces. **E scusate se è poco**

□ INTERACTIONS FROM FIRST PRINCIPLES

- We try to solve the Schrödinger equation under the Born-Oppenheimer approximation

$$\left(\hat{K}_e + \hat{V} \right) \Phi_e(\mathbf{r}_e; \mathbf{R}_n) = E_{BO}(\mathbf{R}_n) \Phi_e(\mathbf{r}_e; \mathbf{R}_n)$$

Practical methods (DFT) to solve the electronic problem scales with N_e^3 !!

- The Born-Oppenheimer surface represents an effective potential between the nuclei mediated by the electronic degrees of freedom.

- It embeds many-body nuclear effects...

$$E_{BO}(\mathbf{R}_n) \equiv \sum_{i < j} U_2(R_{ij}) + \sum_{i < j < k} U_3(R_{ij}, R_{jk}, R_{ik}, \Theta_{ij}, \Theta_{jk}) + \dots$$

Typical many-body series for simple atomic systems

So, we have grounded the notion of molecular structure and molecular dynamics on quantum mechanics. Moreover, we have given, through Hellmann-Feynman's theorem a meaning to the notion of Molecular forces. **E scusate se è poco**

EMPIRICAL/V.S. STATISTICAL POTENTIAL

AND NOW ARTIFICIAL INTELLIGENCE JUMPS ON THE STAGE
Exercise go to JCP and google for AI...

□ INVERSE PROBLEM: FROM "ENSEMBLE AVERAGES" TO MOLECULAR MECHANICAL MODELS

- Most (Classical) simulations use a set of simple classical functions to represent the energy, adjusting the potential's shape to optimize agreement with experimental data and quantum simulations of small molecules.

$$E_{BO}(\mathbf{R}) = \sum_{\xi} \sum_{\lambda} U_{\xi}(\psi(\{\mathbf{R}_{\lambda}\}))$$

Interaction type
(Bonded, non-bonded,
angular interactions...)

Set of particles involved
in a specific interactions

Scalar variable, function of
the nuclear coordinates
(distance, bending angle...)

□ Pop-Philosophies:

- 1) Analytical potentials with unknown parameters (Large macromolecules);
- 2) Potentials with free-functional forms (Simple liquids, Coarse-graining).

- We tune the potentials in order to reproduce some target properties: Structural parameters, vibrational properties, Charge density, Scattering functions, Basic thermodynamics, Detailed structural informations...

□ MOLECULAR MECHANICAL MODELS FOR MACROMOLECULES

$$\sum_{\xi} \sum_{\lambda} U_{\xi}(\psi_{\xi}(\{\mathbf{R}_{\lambda}\})) = \sum_{\{\text{bonds}\}} \frac{k_{\alpha}^d}{2} (d_{\alpha} - d_{\alpha}^0)^2 + \sum_{\{\text{angles}\}} \frac{k_{\alpha}^{\theta}}{2} (\theta_{\alpha} - \theta_{\alpha}^0)^2 + \sum_{\{\text{dihedrals}\}} \frac{k_{\alpha}^{\phi}}{2} (1 + \cos(n\phi_{\alpha} - \phi_{\alpha}^0)) + \sum_{\{\text{pairs}\}} 4\epsilon_{\alpha,\beta} \left[\left(\frac{\sigma_{\alpha,\beta}}{R_{\alpha,\beta}} \right)^{12} - \left(\frac{\sigma_{\alpha,\beta}}{R_{\alpha,\beta}} \right)^6 \right] + \frac{q_{\alpha}q_{\beta}}{4\pi\epsilon_0 R_{\alpha,\beta}} + \dots$$

CLASS I FORCE-FIELDS:

AMBER, GROMOS, GROMACS etc... (Standard MD Packages...)

- Huge number of adjustable parameters for a SPECIFIC force field:

$$U(\{R\}; \{d^0, k^d, \theta^0, k^{\theta}, n, \phi_0, k^{\phi}, \epsilon, \sigma, q\})$$

- Force-field parameters are typically tuned to reproduce properties from a combination of experiments/electronic structure calculations on small systems
- N parameters \leftrightarrow N experimental/theoretical inputs
- Additivity Hypothesis: We assumed that the molecular energy can be expressed as a sum of different terms arising from different physical mechanism (Dispersion forces, bond stretching, Coulomb interactions...)

□ MOLECULAR MECHANICAL MODELS FOR MACROMOLECULES (3)

FINAL REMARKS:

IN THE DERIVATION OF FORCE FIELDS WE ASSUME (AMONG OTHER ASSUMPTIONS) THE TRANSFERABILITY HYPOTHESIS:

1) POTENTIALS DERIVED FOR REPRESENTATIVE STRUCTURE CAN BE APPLIED (SUCCESSFULLY?) TO THE PREDICTIONS OF LARGE MACROMOLECULES WITH THE SAME UNITS (Lego Approach: With small blocks you can TRY to build everything)

2) IF THE INPUT PARAMETERS SHOW A SIGNIFICANT (NOT NEGLIGIBLE) DEPENDENCE ON THE THERMODYNAMIC STATE, PLEASE PAY ATTENTION: YOU CANNOT TRANSFER YOUR FORCE FIELDS TO ANY THERMODYNAMIC STATE

GOOD REFS:

1) L. Monticelli and E. Salonen, *Biomolecular Simulations: Methods and Protocols*; Springer (2013)

2) Tamar Schlick-*Molecular Modeling and Simulation: An interdisciplinary Guide* (Chapt 9), Springer (2012)!