Force Fields & Molecular Interactions

Jean-Christophe SOETENS

Department of Chemistry
University of Bordeaux

jean-christophe.soetens@u-bordeaux.fr

http://theo.ism.u-bordeaux.fr/J-C.Soetens/teach.html

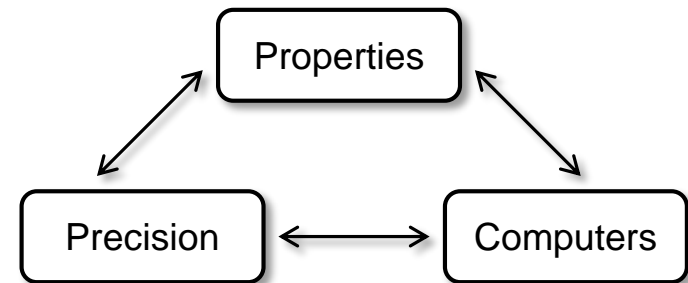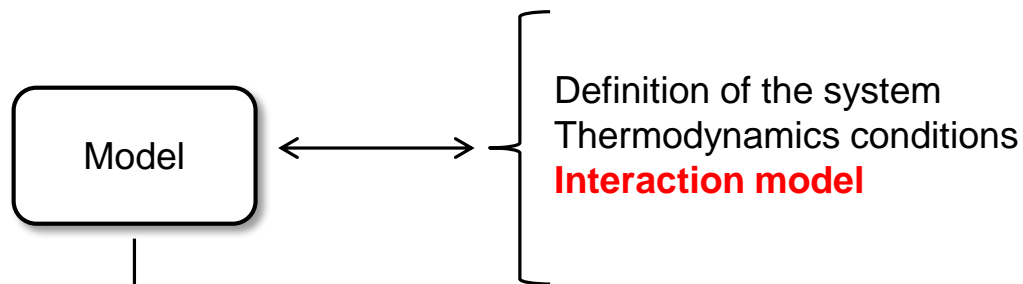***Academic year 2024-2025***

« Force Fields & Molecular Interactions »

I.   Recalling basic ideas about statistical mechanics

II.   General ideas about interaction models

III.   The nature of intermolecular forces

IV.   Representation of the intermolecular potential energy function

V.   Strategies to get a model potential for simulations

VI.   Beyond the pair potential approximation

The machinery of molecular simulations :

Model ←→ {
Definition of the system
Thermodynamics conditions
Interaction model
}

Properties ↔ Precision ↔ Computers

Model → MD or MC → Ensemble P_i or C_i → Properties

High level of theory (quantum mechanics)
      => few particules (electrons),
        short simulations,
        bad statistics, bad precisions

Approximations → classical mechanics
      => large systems,
        long simulations,
        quality of statistical averages

Use the right level of description to catch the phenomena of interest. Don't model bulldozers with quarks. (...) So every good model starts from a question. The modeler should always pick the right level of detail to answer the question.

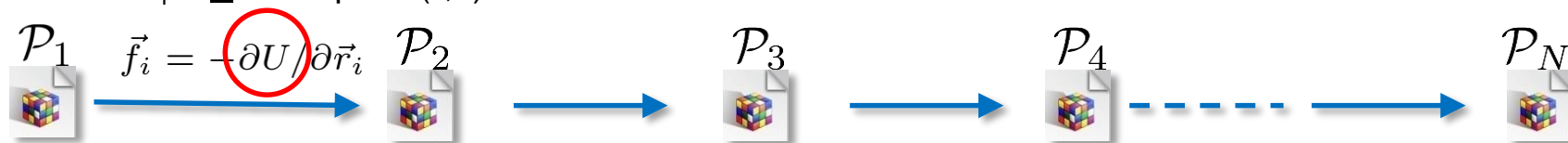*N. Goldenfeld and L.P. Kadanoff, Science, 1999.*

P_i
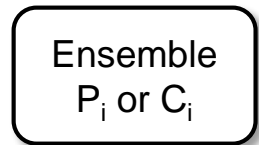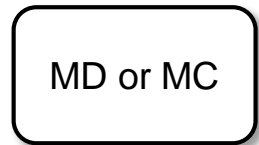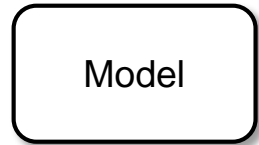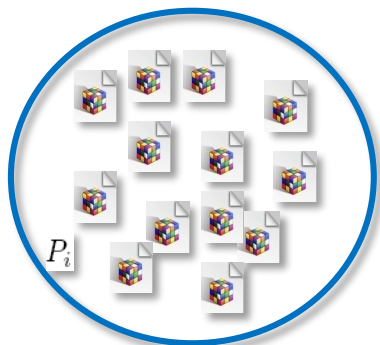
The model is the crucial input of a simulation

« Model » is a complex concept... it includes the technical details of the simulation and the **interaction potential**.

But, when everything is done, every success or failure of the simulation is ascribable to a success or a failure of the model

Molecular Dynamics
$\Rightarrow$ list of P_i in <u>P</u>hase space (**r**,**v**) and ordered in time

$\mathcal{P}_1 \quad \vec{f}_i = -\partial U/\partial \vec{r}_i \quad \mathcal{P}_2 \qquad\qquad \mathcal{P}_3 \qquad\qquad \mathcal{P}_4 \text{-----} \mathcal{P}_N$

Monte Carlo
$\Rightarrow$ list of C_i in <u>C</u>onfiguration space (**r**)
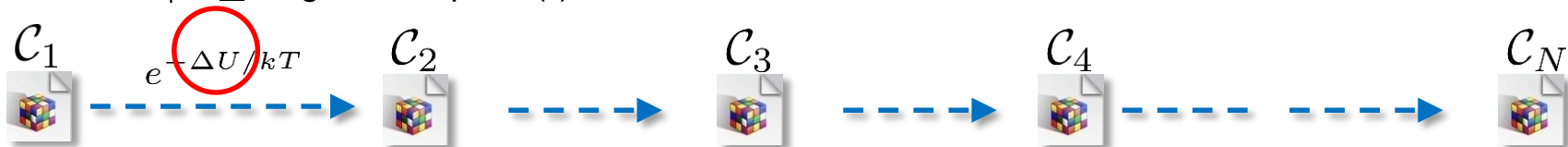
$\mathcal{C}_1 \quad e^{-\Delta U/kT} \quad \mathcal{C}_2 \qquad\qquad \mathcal{C}_3 \qquad\qquad \mathcal{C}_4 \qquad\qquad \mathcal{C}_N$

« Force Fields & Molecular Interactions »

I. Recalling basic ideas about statistical mechanics

II. **General ideas about interaction models**
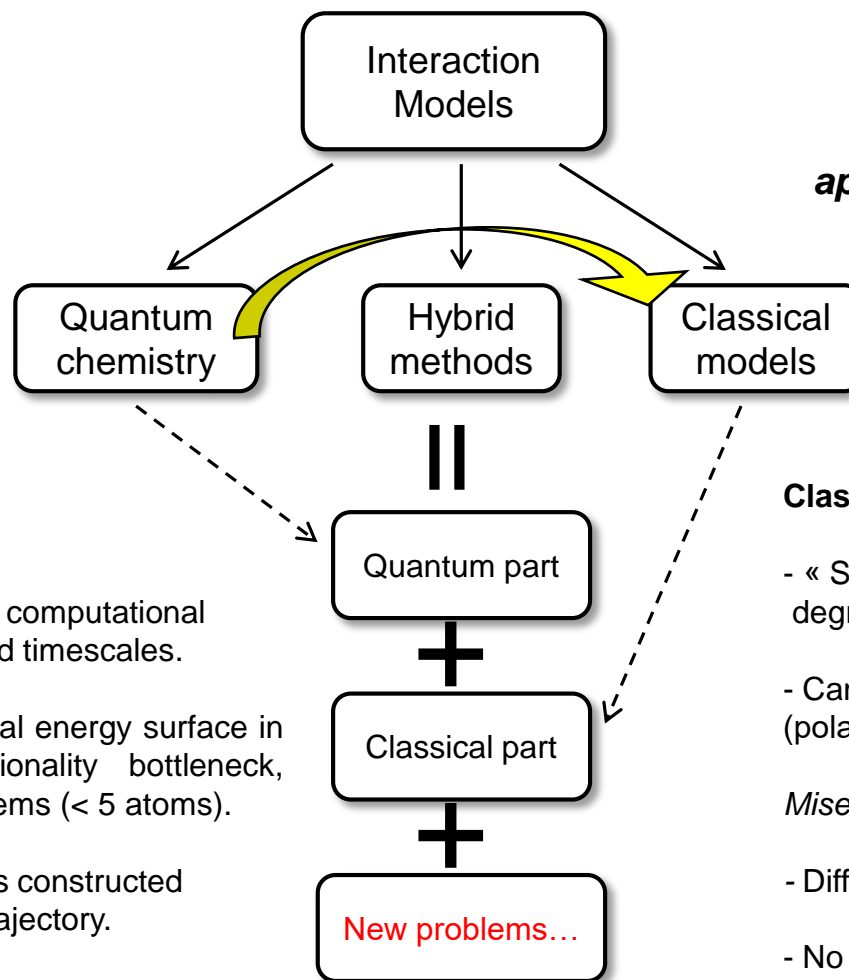
III. The nature of intermolecular forces

IV. Representation of the intermolecular potential energy function

V. Strategies to get a model potential for simulations

VI. Beyond the pair potential approximation

Necessary condition to perform simulation **: Interaction model**

Interaction Models

Quantum chemistry | **Hybrid methods** | **Classical models**

Many simplifications and approximations are necessary

=

Quantum part

+

Classical part

+

New problems…

Potentials and forces from quantum chemistry

- Price of dramatically increased computational costs: much smaller systems and timescales.

- Constructing the whole potential energy surface in advance: exponential dimensionality bottleneck, possibly only for very small systems (< 5 atoms).

- Alternative: on-the-fly potentials constructed along the molecular dynamics trajectory.

Programs for AIMD :
CPMD, CP2K, VASP, NWChem, CASTEP,
CP-PAW, fhi98md,…

Classical models

- « Simple » and cheap : large systems, many degrees of freedom, large timescales.

- Can be refined by including additional terms (polarization, cross intramolecular terms, …).

Misery of empirical force fields:

- Difficult to choose an available model.

- No bond making/breaking – no chemistry !

- Difficult to improve in a systematic way.

- The development of original models is a science in itself.

Born-Oppenheimer approximation:

We start from the Hamiltonian

$$\hat{H}(\{R_I^{\text{nuclei}}\}, \{r_i^{\text{electrons}}\}) = \hat{H}^{\text{nuclei}}(\{R_I^{\text{nuclei}}\}) + \hat{H}^{\text{electrons},\{R_I^{\text{nuclei}}\}}(\{r_i^{\text{electrons}}\})$$

$$\hat{H}^{\text{electrons},\{R_I^{\text{nuclei}}\}}(\{r_i^{\text{electrons}}\}) \qquad \rightarrow \textit{quantum chemistry}$$

$$\hat{H}^{\text{nuclei}}(\{R_I^{\text{nuclei}}\}) \qquad \rightarrow \textit{approximations} \rightarrow \text{Molecular Dynamics} \\ \text{classical mechanics (Newton's equation)}$$

$$\hat{H}^{\text{nuclei}}(\{R_I^{\text{nuclei}}\}) = \hat{T}(\{\dot{R}_I^{\text{nuclei}}\}) + \hat{V}(\{R_I^{\text{nuclei}}\})$$

$$\hat{T}(\{\dot{R}_I^{\text{nuclei}}\})$$

$$\hat{V}(\{R_I^{\text{nuclei}}\}) \leftarrow \text{Potential energy part of the nuclear Hamiltonian : } \textit{model}$$

We shall (mostly) work under the assumption that the motions of atoms and molecules ("particles") can be described using classical mechanics (Newton's equation). I.e. the kinetic energy (temperature) is high enough, the mass is large enough ↔ de Broglie wave length)

Ingredients to develop a model ?
(a) Approximations…?
(b) Results from electronic ab-initio/DFT/..... Calculations ?
(c) Experimental results (IR, Raman, ...) ?
(d) Try and error ?
(e) ....

Problem of writing the potential energy part $\hat{V}(\{R_I^{nuclei}\})$

of our nuclear Hamiltonian $\hat{H}^{nuclei}(\{R_I^{nuclei}\})$

First idea:

One could think of solving the $\hat{H}^{electrons,\{R_I^{nuclei}\}}(\{r_i^{electrons}\})$ problem with very big "single point"

ab-initio (or DFT or ....) <u>quantum chemical calculations for every configurations</u> R_I^{nuclei} we are interested in.

The electronic ground state energies (that we would get from such a calculation) would be points on the (multidimensional) Born-Oppenheimer potential energy surface $\hat{V}(\{R_I^{nuclei}\})$ governing the motions of the nuclei.

Feasibility study : I ≈ 500 to 1000 atoms ≈ 10 000 electrons : impossible (in non-periodic systems).

Better idea :

Put together $\hat{V}(\{R_I^{nuclei}\})$ from smaller pieces.

However, there is a price : approximations must be made.

The most frequently made approximation is the the <u>pair potential approximation.</u>

In other words : the multidimensional function $\hat{V}\left(\{R_I^{nuclei}\}\right)$ is too complex.

The <u>pair potential approximation</u> consists to developp this function over monomer and <u>pair of molecules</u> :

Beyond the pair potential approximation ?

$$\mathcal{V}(\mathbf{r}) = \sum_{i=1}^{N} \mathcal{V}(\mathbf{r}_i) + \sum_{i=1}^{N-1}\sum_{j>i} \mathcal{V}(\mathbf{r}_{ij}) \quad + \sum_{i=1}^{N-1}\sum_{j>i}\sum_{k>j} \mathcal{V}(\mathbf{r}_{ijk}) + \ldots$$

Intramolecular part

Intermolecular part

Three body term

Flexible model

Rigid Model
If = 0

Effective model

and / or

$$+ \mathcal{V}_{induction}$$

ab initio model

empirical model

N - body term
Explicit polarizabilities
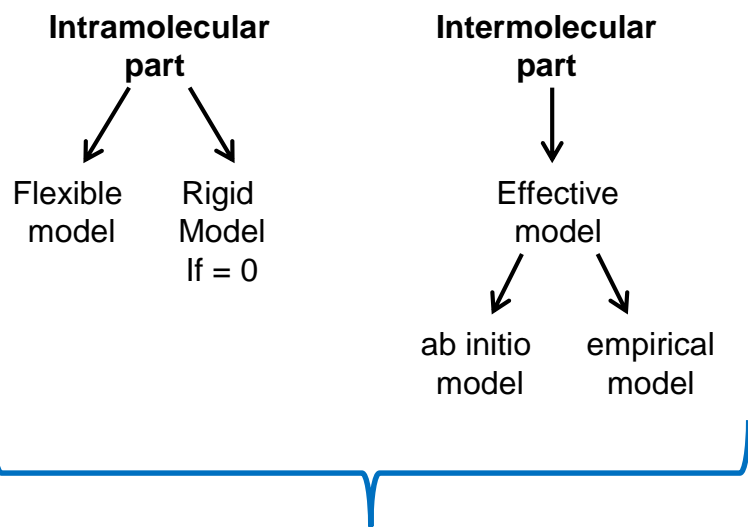
The pair potential approximation is a very drastic assumption

Application of the <u>pair potential approximation</u> :

α

β

2 types of sites

● and 🔵

2 types of sites

🔵 and 🟣
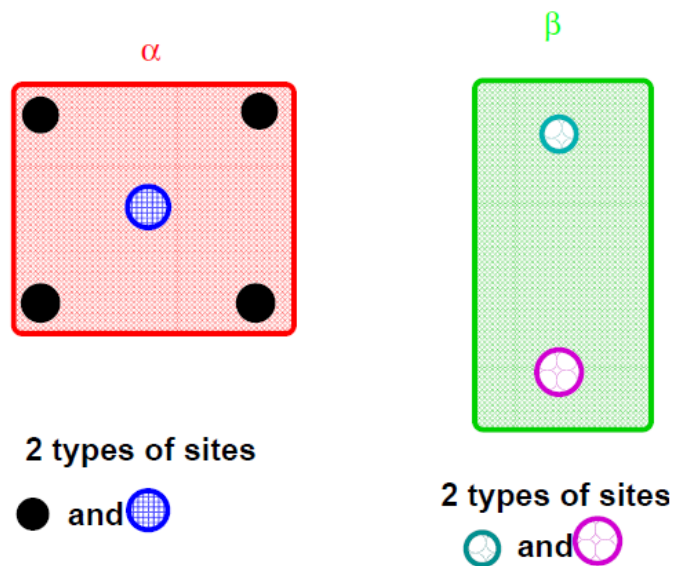
Question : if we want to compute the total energy of an ensemble of 128 molecules of type α (5 sites) and 128 molecules of type β (2 sites)

What is the total number of site/site interactions to compute ?

Answer : α α (5x128) x (5x127)/2 = 203200
α β (5x128) x (2x128)   = 163840
β β (2x128) x (2x127)/2 = 32512

399552

For two molecules :

α α

16 terms
8 terms
1 term

5 x 5 = 25

α β

4 terms
4 terms
1 term
1 term

5 x 2 = 10

β β

1 term
2 terms
1 term

2 x 2 = 4
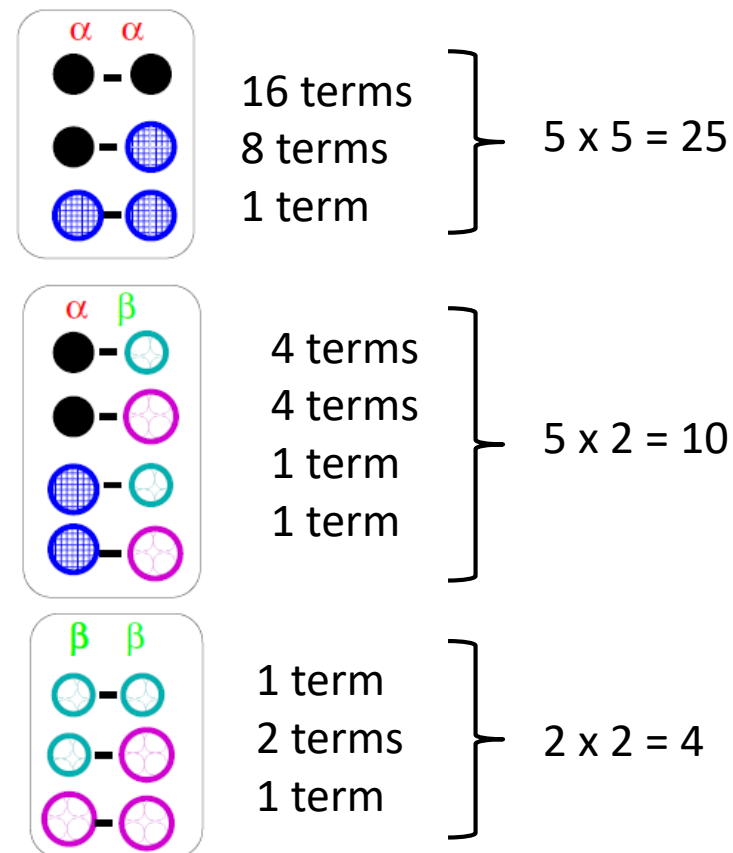
*<u>**Conclusion**</u> : **we still need « simple » mathematical expressions to compute the interactions between sites !***

« Force Fields & Molecular Interactions »

Intermolecular energy :

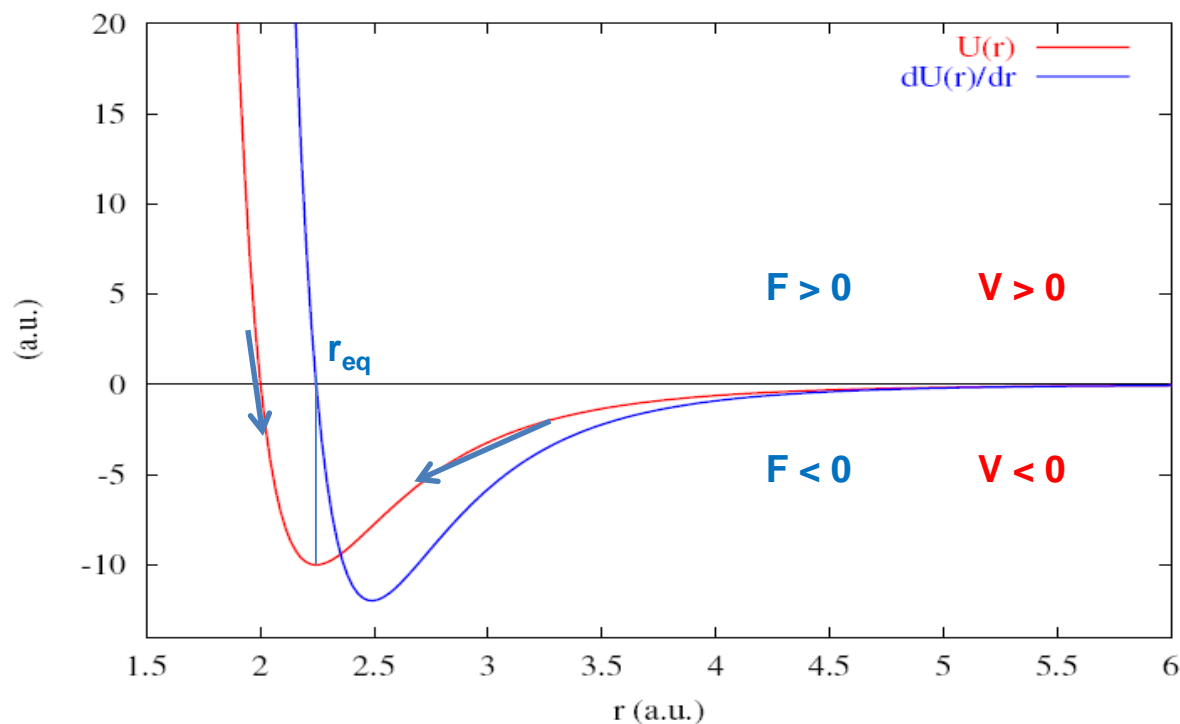Simplest situation : two atoms **a** and **b** infinitely separated :

$$E_{total}(\infty) = E_a + E_b$$

When **a** and **b** are separated by a finite distance **r** :

$$E_{total}(r) = E_a + E_b + V(r)$$

The contribution $V(r)$ to the total energy arising from the interaction is the intermolecular pair potential function

$$V(r) = E_{total}(r) - E_a - E_b$$
$$= E_{total}(r) - E_{total}(\infty)$$

This energy difference is the work done in bringing the two atoms from ∞ distance to r :

$$V(r) = \int_r^{\infty} F(r)dr$$

So that :

$$F(r) = -\frac{dV(r)}{dr}$$

F(r) < 0 : attractive force
F(r) > 0 : repulsive force

Intermolecular energy *vs* sign of the force

Remark on the <u>distance-dependence</u> of intermolecular forces.

According to the dependence of the potential energy V on the separation r, the intermolecular forces can be divided into two classes:

Long - ranged and **short - ranged**

The intermolecular potential energy function is generally modelled by a power law in r :

$$V(r) = \frac{C}{r^\alpha}$$

Typicallly :

Long-ranged interactions: $\alpha \leq 6$ *ex :* coulombic $\pm\ 1\,/\,r$
dispersion $-\ 1\,/\,r^{\,6}$

Short-ranged interactions: $\alpha > 6$ ex : repulsion $1\,/\,r^{\,12}$

Remark on the <u>electric properties</u> of molecules :

The electric charge distribution of a molecule can be seen as multipoles and multipole moments.

Definitions :

Monopole: a point charge (e.g. Na^+, Cl^-).

Dipole: an asymmetric charge distribution in a molecule, where there is no net charge but one end of the molecule is negative (partial charge = $-q$) relative to the other (partial charge = $+q$).

Molecules may possess **higher order electric multipoles**, arising from their non-spherical charge distributions.

Each type of **multipole** has an associated **multipole moment**:

- the monopole moment is the charge of the atom/molecule.

$$Q = \sum_i q_i$$

- the dipole moment is a vector whose magnitude is the product of the charge and distance between the charge centres.

$$\vec{\mu} = \sum_i q_i \vec{R}_i$$

- higher order multipole moments have tensor properties.
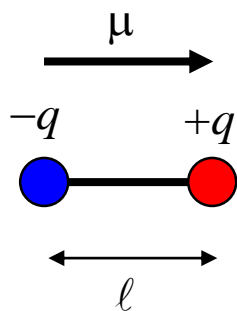
A **polar molecule** is one which possesses a permanent dipole moment: there is an asymmetric charge distribution, with one end of the molecule relatively negative ($-q$) with respect to the other ($+q$).
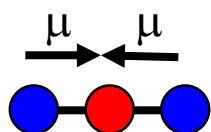
Linear molecules like CO_2 (O=C=O) or planar molecule like benzene (C_6H_6) do not have dipole moments but they have non-zero **quadrupole moments**.

In methane (CH_4), the first non-zero multipole moment is the **octopole moment.**

Atom/Molecule	μ / (10^{-30} C m)	μ / Debye
He	0	0
N_2	0	0
CO_2	0	0
CH_4	0	0
CH_3Cl	6.24	1.87
$CHCl_3$	3.37	1.01
CCl_4	0	0
C_6H_6	0	0
HF	6.37	1.91
HCl	3.60	1.08
H_2O	6.17	1.85
NH_3	4.90	1.47

Some examples of dipole moment (gas phase).

1 Debye = 3.33564 10^{-30} C·m

A molecule which possesses a **permanent dipole moment** is called **polar molecule**.

Fundamental types of intermolecular interactions :

Electrostatic :
interactions between charged atomic or molecular species (ions = monopoles) or between asymmetric charge distributions (dipoles, quadrupoles etc.) in neutral molecules.

Electrostatic interactions can be attractive or repulsive, depending on ionic charges and the orientation of the molecular multipoles.

Induction:
an electric charge (monopole) or higher multipole causes polarization of neighbouring atoms/molecules and induced multipoles. The attractive interaction between the original multipole and the induced multipole gives rise to the induction energy.

Dispersion :
attractive interactions between instantaneous dipoles (and higher multipoles) arising due to fluctuating charge distributions in atoms and molecules.

Repulsion :
At short internuclear separations the electron clouds of the molecules overlap. Repulsive force arise partly from the incompletely screened electrostatic repulsion of the nuclei and partly from the repulsion between the electrons.

Electrostatic energy

The distance-dependence of the interaction energy between an <u>*n*-pole</u> and an <u>*m*-pole</u> is given by:

$$V_{nm}(r) = f(\pm \frac{1}{r^{n+m-1}})$$

where *n* and *m* are the ranks of the multipoles:
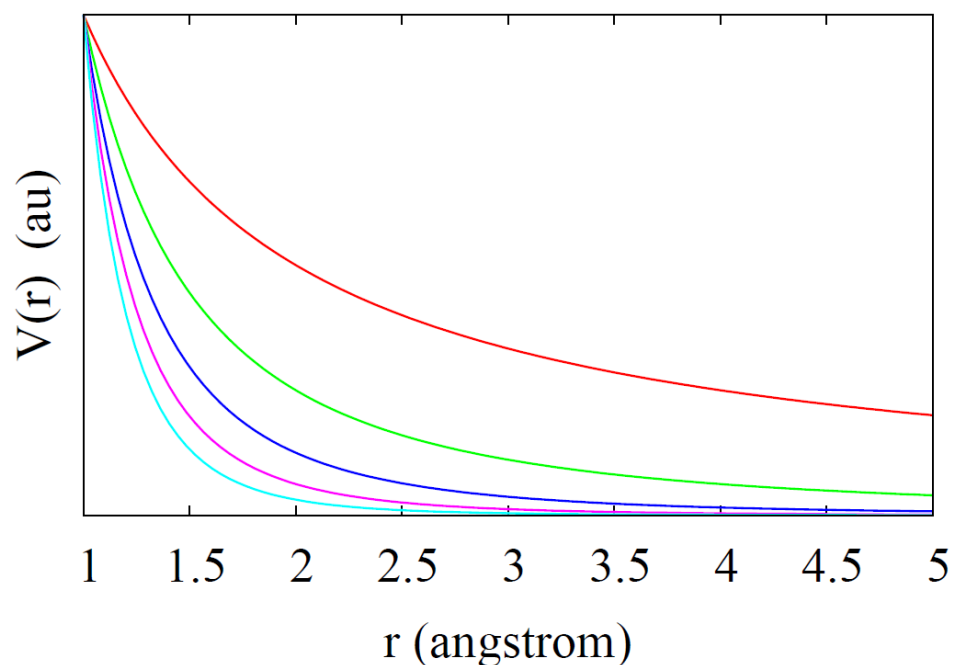*n,m* = 1 (monopole), 2 (dipole), 3 (quadrupole), 4 (octopole) …

Examples:
monopole-monopole   1 / r
monopole-dipole   1 / r^2
dipole-dipole   1 / r^3
dipole-quadrupole   1 / r^4
quadrupole-quadrupole   1 / r^5

Famous coulombic expression :

$$V_{qq'}(r) = \frac{1}{4\pi\epsilon_0}\frac{qq'}{r}$$

Induction energy

Something (a charge or any electric multipole) gives rise to an electric field (*E*) which causes the polarization of neighbouring atoms or molecules.

An induced dipole is thus created, according to the magnitude and orientation of the electric field and to the polarizability (polarizability tensor α) of the atom or molecule.

Without external electric field
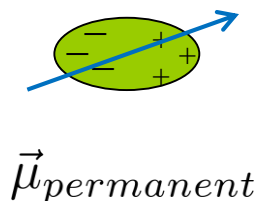
In the presence of an external electric field
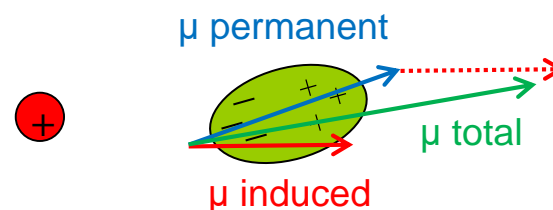
$$\vec{\mu}_{induced} = \bar{\alpha}\vec{E}$$

$$\vec{\mu}_{total} = \vec{\mu}_{permanent} + \vec{\mu}_{induced}$$

The **induction energy** of a molecule A comes from the interaction between the induced moment on the molecule A and the permanent moment of the molecule B who polarized A.

Dispersion energy

The intermolecular forces between nonpolar molecules and closed shell atoms (e.g. rare gas atoms He, Ne, Ar …) is dominated by "London" or dispersion forces.

The dispersion energy contributes to the intermolecular interactions between all pairs of atoms or molecules. It is also generally the dominant contribution, even for polar molecules.

Long range attractive dispersion forces arise from dynamic electron correlation: fluctuations in electron density give rise to instantaneous electronic dipoles (and higher multipoles), which in turn induce dipoles in neighbouring atoms or molecules.

General expansion for the dispersion energy :

$$V_{dispersion}(r) = -\frac{C_6}{r^6} - \frac{C_8}{r^8} - \frac{C_{10}}{r^{10}} - \; \cdots$$

The first term represents the instantaneous dipole–dipole interaction and is dominant.

As we will see later, the higher terms are often omitted when calculating dispersion energies.

« Force Fields & Molecular Interactions »

I.   Recalling basic ideas about statistical mechanics

II. General ideas about interaction models

III. The nature of intermolecular forces

IV. Representation of the intermolecular potential energy function

V.   Strategies to get a model potential for simulations

VI. Beyond the pair potential approximation

Representation of the intermolecular <u>pair potential energy function</u>

$$V(r) = \underbrace{E_{repulsion} + E_{dispersion}}_{\substack{\textit{van der Waals} \\ \textit{interaction}}} + \underbrace{E_{coulomb} + E_{induction}}_{\substack{\textit{electrostatic} \\ \textit{interaction}}}$$

The most frequently used mathematical expression is :

$$V_{ij}(r_{ij}) = \sum_{\alpha \in i} \sum_{\beta \in j} \left(\underbrace{\frac{1}{4\pi\epsilon_0} \frac{q_\alpha q_\beta}{r_{\alpha\beta}}}_{\textit{electrostatic}} + \underbrace{4\, \varepsilon_{\alpha\beta} \left[\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}} \right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}} \right)^{6} \right]}_{\textit{Lennard-Jones potential}} \right)$$

- q_α and q_β are partial electric charges on chosen sites of molecules i and j.
- $\mathcal{E}$ and σ are the Lennard-Jones constants depending on the type of atoms (sites).

For an interaction between two different type of atoms : combining rules

Example of combining rules (OPLS) : $\quad \sigma_{\alpha\beta} = \sqrt{\sigma_\alpha \sigma_\beta} \qquad \epsilon_{\alpha\beta} = \sqrt{\epsilon_\alpha \epsilon_\beta}$

Lennard-Jones potential :

Inside the figure:

$V_{repusion} = 4\,\varepsilon\,(\sigma / r)^{12}$

$V_{lennard\text{-}Jones} = V_{repusion} + V_{dispersion}$

$\sigma\,2^{1/6}$

$-\varepsilon$

$V_{dispersion} = -\,4\,\varepsilon\,(\sigma / r)^{6}$

Axes: V(r) (au) vs r (angstrom), with r values 2, 2.5, 3, 3.5, 4
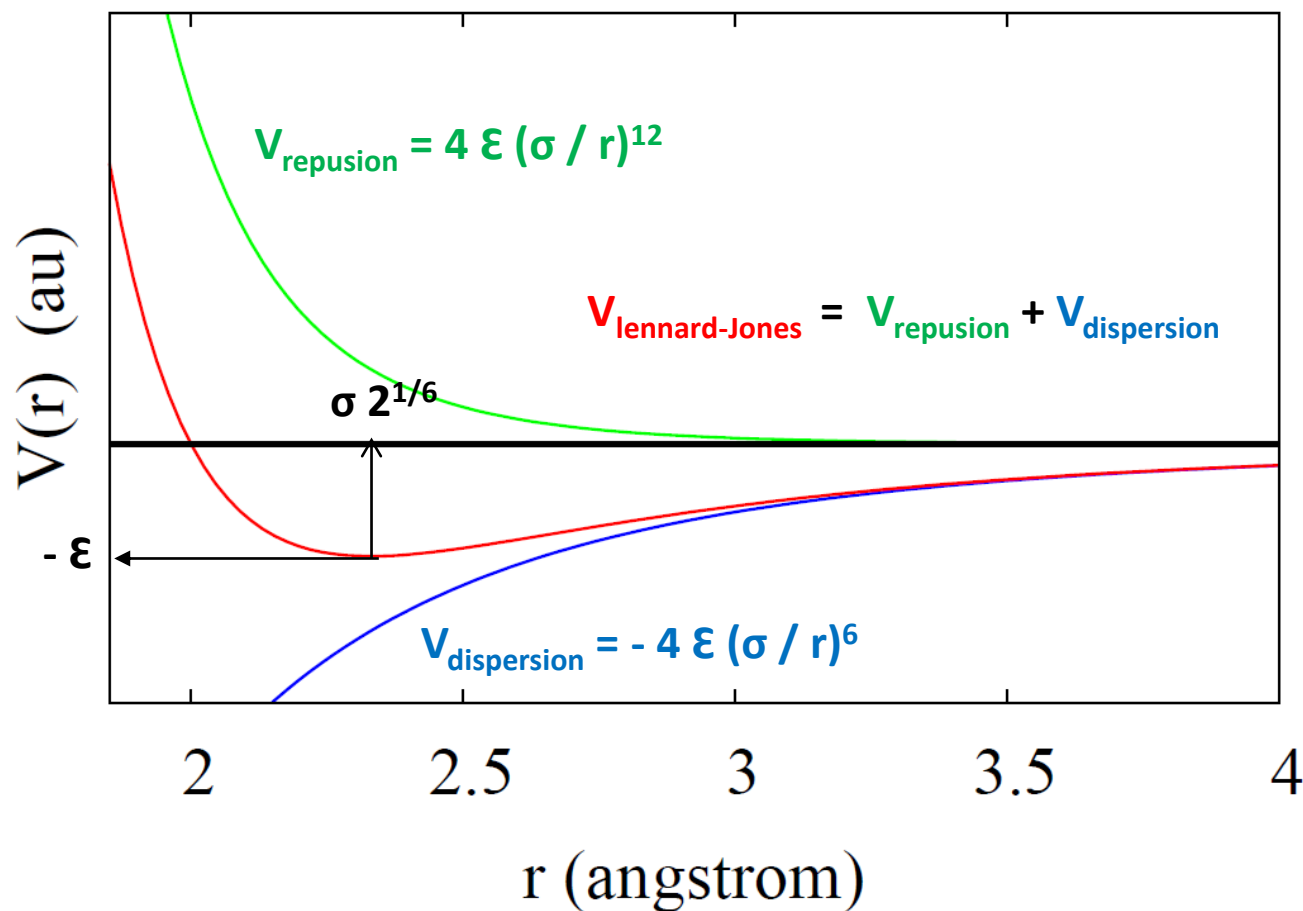
Alternative : Buckingham "exp-6" potential $\qquad D_0\left(exp\left[a\frac{r_0}{r}\right] - b(\frac{r_0}{r})^6\right)$

« Force Fields & Molecular Interactions »

I.   Recalling basic ideas about statistical mechanics

II.   General ideas about interaction models

III.   The nature of intermolecular forces

IV.   Representation of the intermolecular potential energy function

V.   Strategies to get a model potential for simulations

VI.   Beyond the pair potential approximation

```
                        ┌─────────────────────┐
                        │    Choice of a      │
                        │  potential model    │
                        └─────────────────────┘
                         ↙                 ↘
        ┌──────────────────────┐      ┌──────────────────────┐
        │ Some specific models │      │  No available model  │
        │   exist in literature│      │    in literature     │
        └──────────────────────┘      └──────────────────────┘
           ↙            ↘                  ↙            ↘
```

Specific models

- Charasteristics ?
- Quality ?
- Time consuming ?
- Available in codes ?

Force Fields

- Freeware ?
- Commercial ?
- Computable properties ?

Transfer of parameters
from analogous systems ?

- Dangerous approach !
- Verifications ?
- Modifications

Development of a new specific model

- Complex
- Time consuming
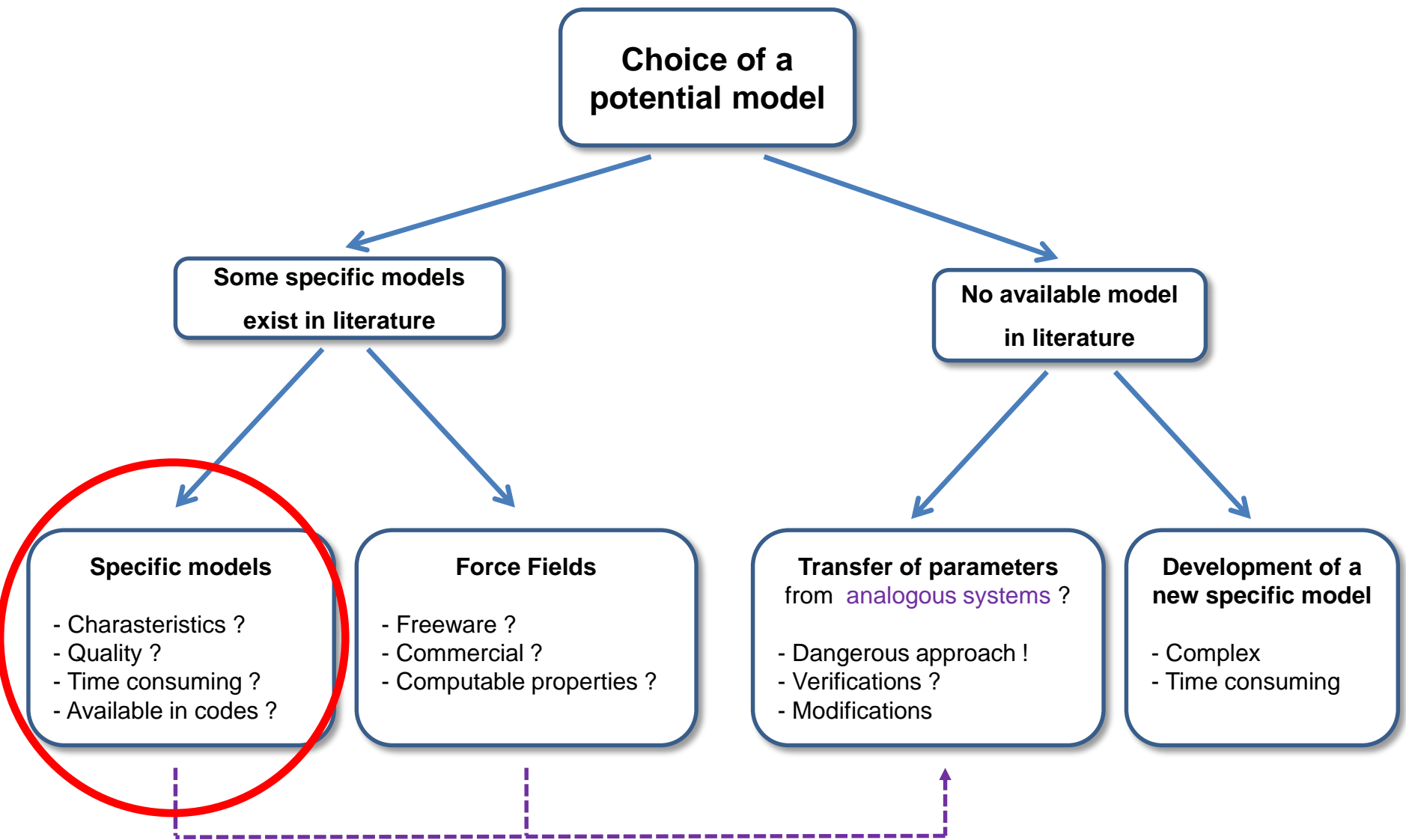
Example of liquid WATER

Just for water (H_2O), there are probably about 50 reasonably good, but quite different models in the literature. They are known by acronyms like BNS, ST2, MCY, TIPS, TIP3P, TIP4P, CF2, CF3, BJH, SPC, SPC/E, etc. etc. etc

Why so many?

Are you lucky ?

Rigid models

Flexible models

Molecular dipoles : changing or not (flexible, polarizable)

Polarizable models

Effective models

Empirical models : tuned to get certain properties right

ab initio models : fitted from quantum chemistry

H_2O

Acronym (date)	Reference	Status	Type	Sites	$\mu_g(D)$	$\mu_l(D)$
BF (1933)	7	empirical	R	4	2.0	2.0
R (1951)	21	empirical	R	5	1.84	1.84
BNS (1971)	22	empirical	R	5	2.17	2.17
ST2 (1973,1993)	23	empirical	R	5	2.35	2.35
CF (1975,1978,1995)	24	empirical	F	3	1.86	1.98
MCY (1976)	25	ab initio	R	4	2.19	2.19
DCF (1978,1980,1993)	26	empirical	F,D,P	3	1.855	-
PE (1979)	27	empirical	P	1	1.855	2.50
SPC (1981)	14	empirical	R	3	2.27	2.27
TIP3P (1981,1983)	15	empirical	R	3	2.35	2.35
RWK (1982)	28	empirical	F	4	1.85	1.89
TIP4P (1983)	15	empirical	R	4	2.18	2.18
BJH (1983)	29	empirical	F	3	1.87	1.99
SPC/F (1985)	30	empirical	F	3	2.27	2.42
MCYL (1986)	31	ab initio	F	4	2.19	2.26
SPC/E (1987)	32	empirical	R	3	2.35	2.35
WK (1989)	33	empirical	R	4	2.60	2.60
SPCP (1989)	34	empirical	P	3	1.85	2.90
CKL (1990)	35	empirical	F,P	4	1.88	2.20
MCHO (1990)	36	ab initio	P	6	2.12	≈3.0
NCC (1990)	37	ab initio	P	6	1.85	2.80
NEMO (1990,1995)	38	ab initio	P	5	2.04	2.89
PTIP4P (1991)	39	empirical	P	4	1.85	2.80
SPC/FP (1991)	40	empirical	F,P	3	1.85	2.44
PSRWK (1991)	41	empirical	P	4	1.88	2.63
KJ (1992)	42	empirical	P	4	1.85	-
NCCvib (1992)	37b	ab initio	F,P	6	1.85	3.11
ASP-W (1992,1998)	43	ab initio	P	3	1.85	2.90
RPOL (1992)	44	empirical	P	3	2.02	2.62
CPMD (1993,1999)	45	DFT+CP	F,D,P	nucl.+el.	1.87	2.95
PPC (1994)	46	ab initio	P	4	2.14	2.51
SPC/FQ (1994)	47	empirical	P	3	1.85	2.83
TIP4P/FQ (1994)	47	empirical	P	4	1.85	2.62
KKY (1994)	48	empirical	F,D	3	2.38	2.21
SQPM (1995)	49	valence bond	P	4	1.85	2.62
SCPDP (1996)	50	empirical	P	4	1.85	2.87
TAB/10D (1998)	51	SCF+MD	P	5	1.85	2.65
NSPCE (1998)	52	empirical	R	3	2.18	2.18
NCF (1998)	53	empirical	F	3	1.85	1.90
MCDHO (2000)	54	ab initio	F, P	4	1.85	3.01
TIP5P (2000)	55	empirical	R	5	2.29	2.29
SPC/HW (2001)	56	empirical	R	3	2.41	2.41
DEC (2001)	57	empirical	R	3	1.85	1.85
SWFLEX (2001)	58	empirical	P	4	1.85	2.59
POLARFLEX (2001)	59	valence bond	F,P	3	1.85	2.55
POL5 (2001)	60	ab initio	P	5	1.85	2.71

*B. Guillot, Journal of Molecular Liquids **101,** 219 (2002)*

Let's look at three points only:

(i) The electric dipole moment

$$\vec{\mu} = \sum_i q_i \vec{R}_i$$

where the **qi** are electric charges and **Ri** their locations (sites). Independent of the origin if $\sum_i q_i = 0$

Take water as an example:

You can read everywhere that the experimental dipole moment of (gas phase) water is 1.86 D (Debye). There is of course an infinite number of point charge distributions (PCD) that will yield this value of 1.86 D.

However:

At large distances (large compared with the extension of the PCD), the interaction between two PCDs will be the same for all PCDs: the dipole-dipole interaction.

At shorter distances, however, the interactions will be very different and depend on the exact locations and magnitudes of the point charges.

Conclusion :

The distribution of point charges is a very (may be the most) important feature that distinguishes these water models.

Static dielectric constant ε_o of liquide water at 298 K
as a function of the average molecular dipole

(Results from MD simulations)
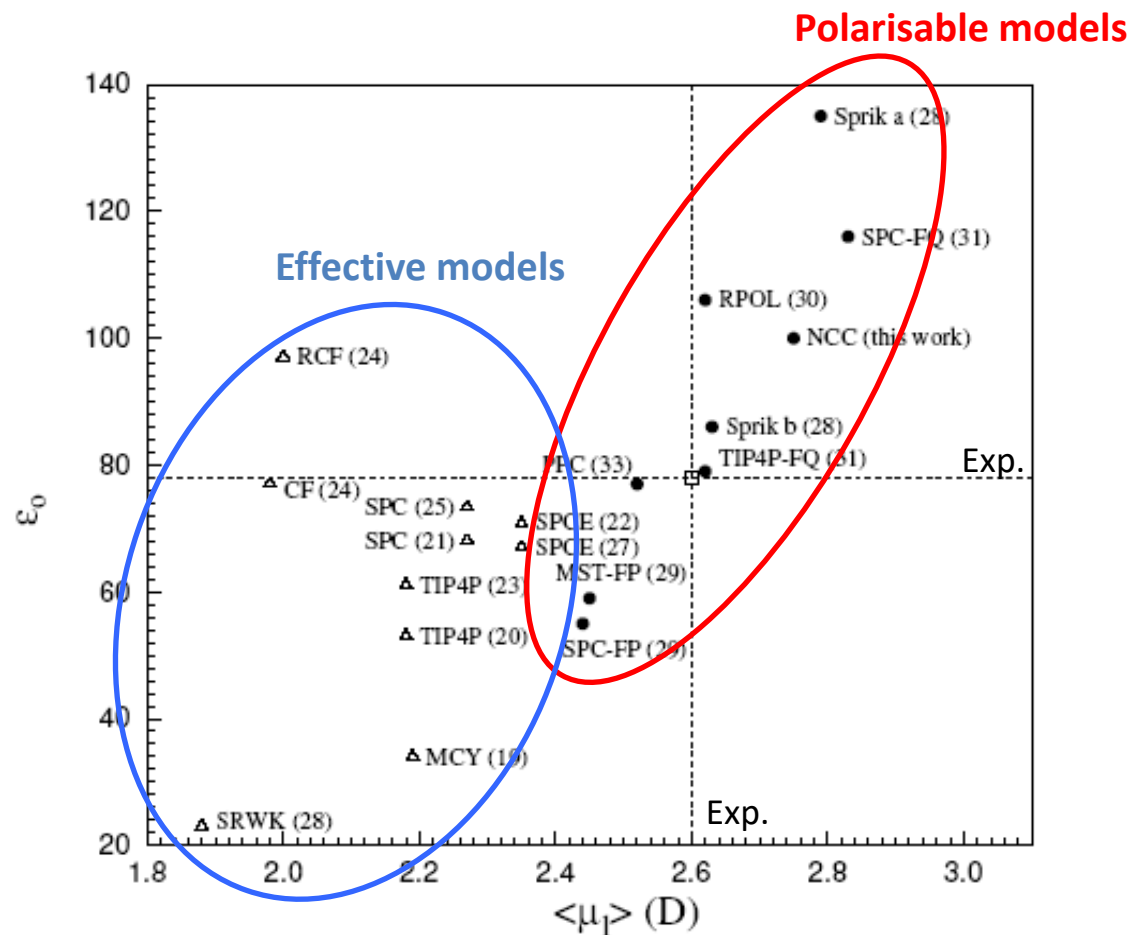
(ii)   « Effective » pair potentials

Aim: we want to stick with the pair potential approximation but tweak it to make it more efficient.

Let's again look at the dipole moment of water as an example:

Distribution of molecular dipole moments in water, idealized   Distribution of molecular dipole moments in water, idealized
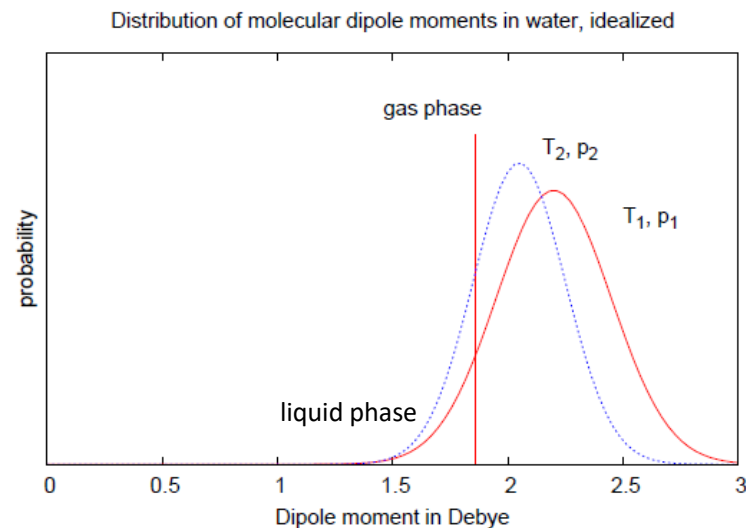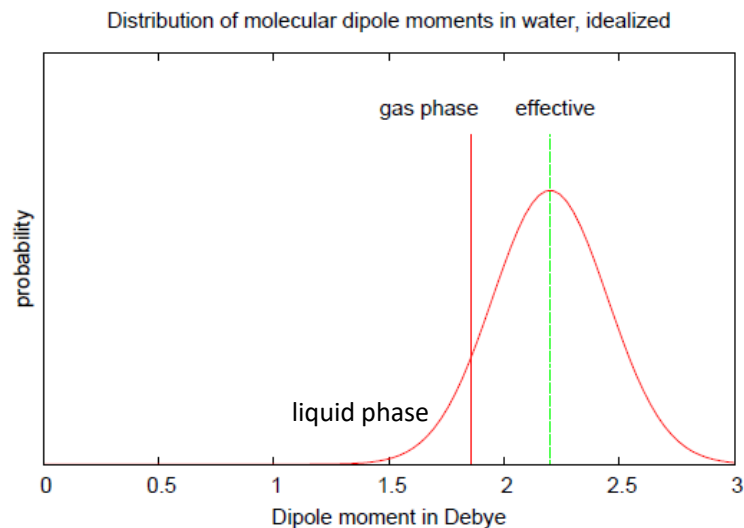
In liquid water, water molecules have a wide distribution of dipole moments

- We try to mimic this distribution by giving our model an enhanced, effective dipole moment.
- Strictly speaking, this effective dipole will depend on the thermodynamic conditions of the system we want to study
- This is contradiction (a somewhat dangerous game).

A microscopic interaction potential must not depend on macroscopic conditions !

(iii) Complexity of the Potential Energy Surface (PES) of the water dimer !

Indices :

0

1(non-planar, C$_S$)
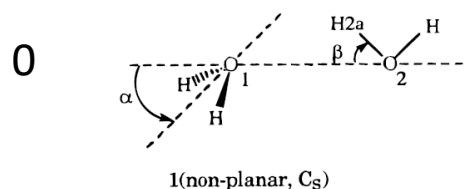
Dimer H$_2$O-H$_2$O : stationary points of the PES obtained by ab initio MP2/6-31G+G(d,p) calculations

1

2 (C$_1$)

4 (cyclic, C$_i$)

9 (bifurcated, C$_{2v}$)

2

3 (planar, C$_S$)

5 (cyclic, C$_2$)

7 (triply-hydrogen-bonded, C$_S$)

10 (planar bifurcated, C$_{2v}$)

3

6 (planar cyclic, C$_{2h}$)

8 (doubly-bifurcated, C$_{2h}$)
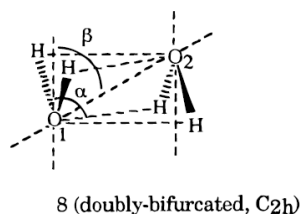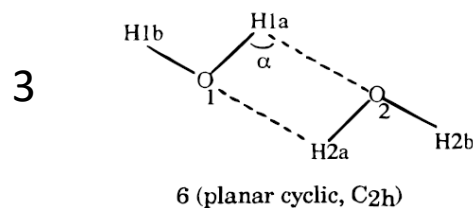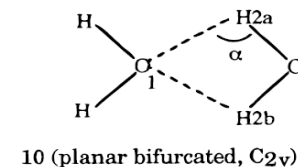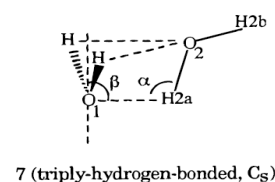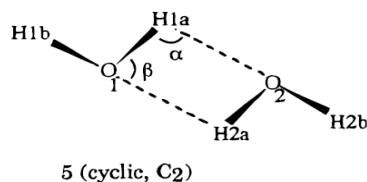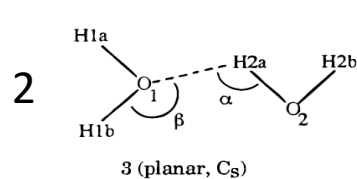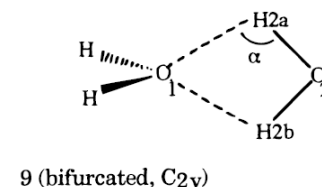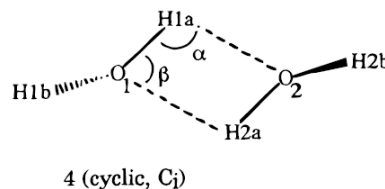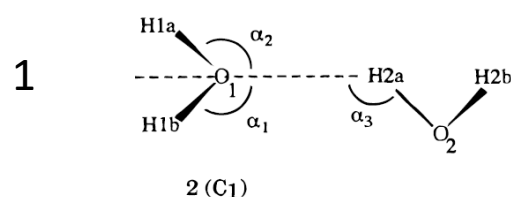
C. Millot et al, J. Phys. Chem. A 102, 754 (1998)

(iii) Complexity of the Potential Energy Surface (PES) of the water dimer !

- real PES is complex
- precision of ab initio methods
- choice of basis sets
- BSSE
- potential model (mathematical)
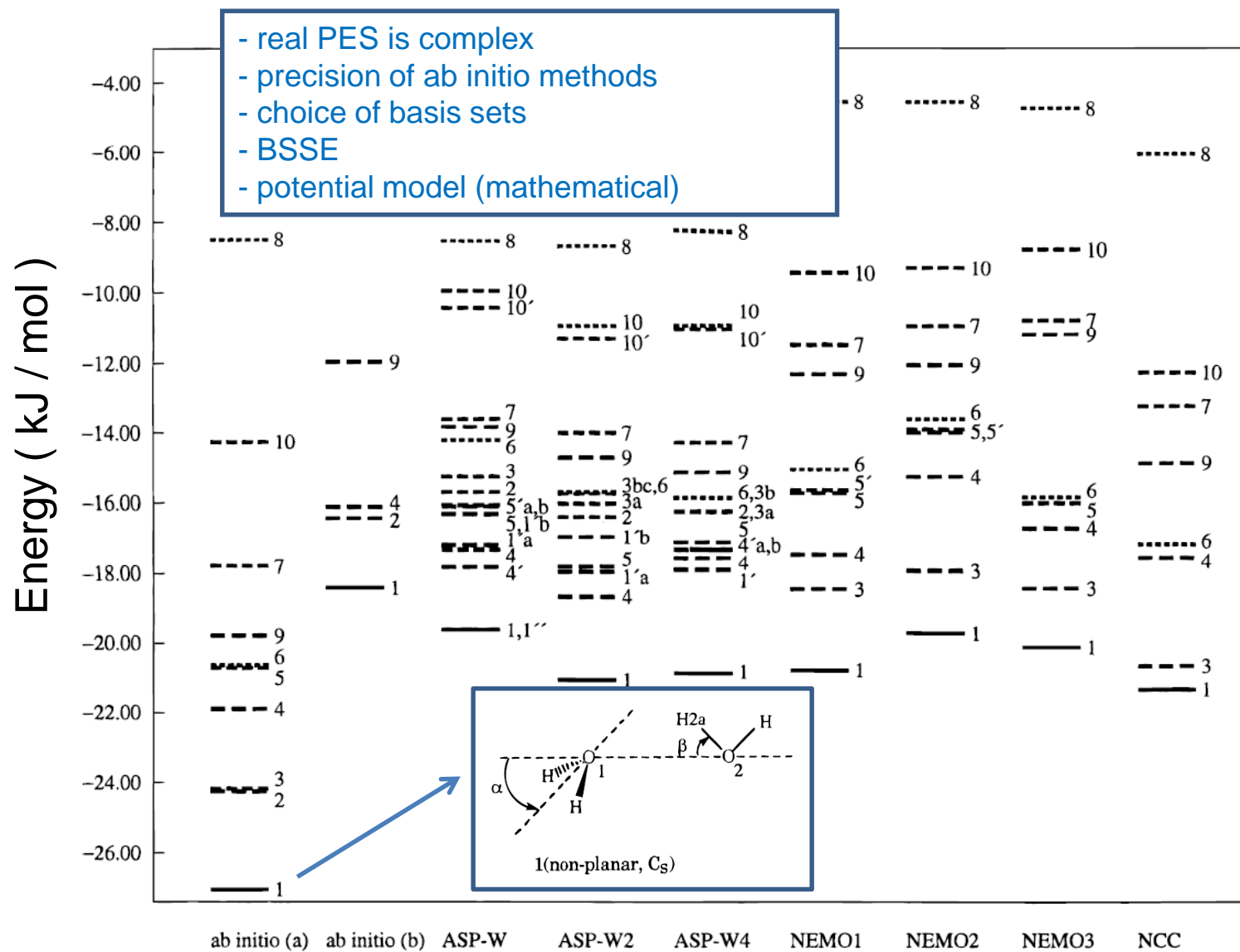
Another example of potential model for liquid AMMONIA

Flexible effective four-site model

$$V_{NN}(r_{NN}) = \frac{14.85}{r_{NN}} + \frac{55719}{r_{NN}^{12}} - \frac{13.6}{r_{NN}^6}$$

$$V_{NH}(r_{NH}) = -\frac{4.95}{r_{NH}} + 0.01042\ exp[-4.6(r_{NH} - 2.4)] - 2\ exp[-2.3(r_{NH} - 2.4)]$$

$$V_{HH}(r_{HH}) = \frac{1.65}{r_{HH}} + 48.64\ exp[-3.7\ r_{HH}]$$

(r in Å and V in 10^{-19} J, q_N = -0.801 e , q_H = 0.267 e)

*S.V. Hannongbua, T. Ishida, E. Spohr, K. Heinzinger, Naturforsch **43a**, 572 (1988)*

$$V^{intra} = \sum_{u=1}^{4} k_u h^{2u} + k_1' h^2 s_1 + k_2' h^4 s_1$$

$$+ \frac{1}{2} \sum_{1=i\neq j}^{5} F_{ij}\ s_i s_j + \sum_{1=i\leq j\leq k}^{5} F_{ijk}\ s_i s_j s_k$$

$$+ \sum_{1=i\leq j\leq k\leq l}^{5} F_{ijkl}\ s_i s_j s_k s_l$$

$$s_1 = \frac{1}{\sqrt{3}}\ (\Delta r_1 + \Delta r_2 + \Delta r_3)$$

$$s_2 = \frac{1}{\sqrt{6}}\ (2\Delta r_1 - \Delta r_2 - \Delta r_3)$$

$$s_3 = \frac{1}{\sqrt{6}}\ (2\Delta\alpha_1 - \Delta\alpha_2 - \Delta\alpha_3)$$

$$s_4 = \frac{1}{\sqrt{2}}\ (\Delta r_2 - \Delta r_3)$$

$$s_5 = \frac{1}{\sqrt{2}}\ (\Delta\alpha_2 - \Delta\alpha_3)$$

*V. Spirko, J. of Molecular Spectroscopy **101**, 30 (1983)*
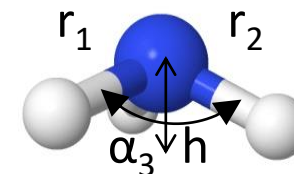
Definition of internal coordinates (r_1, r_2, r_3, α_1, α_2, α_3) and h
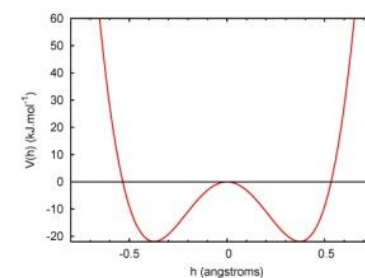
Barrier to planarity $\approx$ 1850 cm^{-1}

Fondamental bands (cm^{-1})	Calc. (this model)		Exp.	
	+	-	+	-
ν_2 (A1)	930	969	932	968
ν_1 (A1)	3342	3343	3336	3337
ν_3 (E)	3456	3457	3443	3444
ν_4 (E)	1631	1632	1626	1627

+ and – refer to the symmetric and antisymmetric states with respect to the inversion motion

Many specific models exists in the littérature :

Water
Methanol
Ethanol
Acetone
CCl_4
$HCCl_3$
CO_2
...

The available MD programs allow to use some of these models.
$\Rightarrow$ Gromos, Charmm, Amber, Sybil, Serius…

But for most of them, it is necessary to program it !

For large molecule and very large molecules (polymers, proteines, etc.),
<u>transferable sets of parameters</u> have been developped : called **Force Fields**.

Choice of a potential model

Some specific models exist in literature

No available model in literature

Specific models

- Charasteristics ?
- Quality ?
- Time consuming ?
- Available in codes ?

Force Fields

- Freeware ?
- Commercial ?
- Computable properties ?

Transfer of parameters
from analogous systems ?

- Dangerous approach !
- Verifications ?
- Modifications

Development of a new specific model

- Complex
- Time consuming

Potential energy funtion
(mathematical equations) $\longrightarrow$ Empirical force field
(equations and parameters)

AMBER   (**Assisted Model Building and Energy Refinement**) - widely used for proteins and DNA.
CHARMM   (**Chemistry at HARvard Molecular Mechanics**) - originally developed at Harvard, widely used for both small molecules and macromolecules
CVFF   broadly used for small molecules and macromolecules.
COSMOS-NMR   hybrid QM/MM forcefield adapted to a variety of inorganic compounds, organic compounds and biological macromolecules, including
               semi-empirical calculation of atomic charges and NMR properties. COSMOS-NMR is optimized for NMR based structure elucidation and implemented in
               COSMOS molecular modelling package.
GROMOS   (**GROningen MOlecular Simulation package**), a general-purpose molecular dynamics computer simulation package for the study of biomolecular systems.
OPLS   (**Optimized Potential for Liquid Simulations**) (variations include OPLS-AA, OPLS-UA, OPLS-2001, OPLS-2005)
               developed by William L. Jorgensen at the Yale University Department of Chemistry.
ENZYMIX   a general polarizable force field for modeling chemical reactions in biological molecules. This force field is implemented with the empirical valence bond (EVB)
               method and is also combined with the semimacroscopic PDLD approach in the program in the MOLARIS package.
ECEPP/2   first force field for polypeptide molecules - developed by F.A. Momany, H.A. Scheraga and colleagues
UFF   (**Universal Force Field**) A general force field with parameters for the full periodic table up to and including the actinoids - developed at Colorado State University.
CFF   (**Consistent Force Field**) - a family of forcefields adapted to a broad variety of organic compounds, includes force fields for polymers, metals, etc.
COMPASS   (**Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies**) - developed by H. Sun at Molecular Simulations Inc,
               parameterized for a variety of molecules in the condensed phase.
MMFF   (**Merck Molecular Force Field**)- developed at Merck, for a broad range of molecules.
MM   MM2 (1977), MM3 (1990), MM4 (1996) - developed by Norman Allinger, parametrized for a broad range of molecules.

Examples of mathematical equations :

Intermolecular (external, nonbonded terms)

$$\sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

Intramolecular parameters

$$\sum_{bonds} K_b(b - b_o)^2 + \sum_{angles} K_\theta(\theta - \theta_o)^2 + \sum_{torsions} K_\phi(1 + \cos(n\phi - \delta))$$
$$+ \sum_{impropers} K_\varphi(\varphi - \varphi_o)^2$$

$$E_T = E_{bond} + E_{angle} + E_{torsion} + E_{cross-terms} + E_{VDW} + E_{elec.} + E_{H-bond} + \ldots$$

- E_{bond}

$$V_{pot} = \tfrac{1}{2} k\, x^2$$

E_{bond}

$R°$
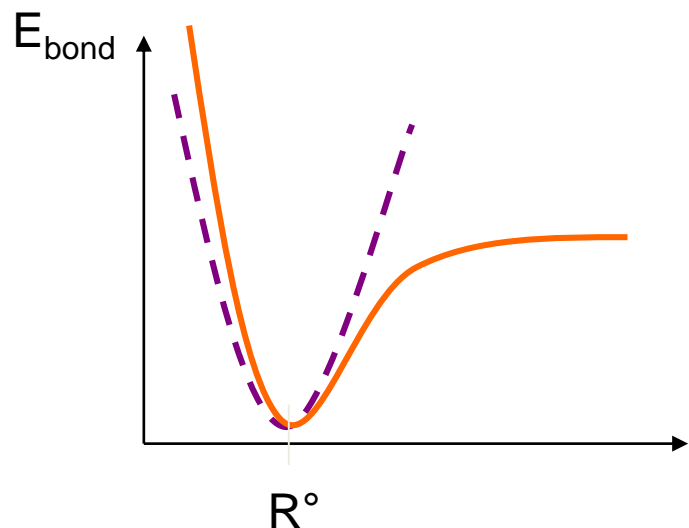
$$E_{liaison} = \tfrac{1}{2} \sum k_L (R - R°)^2$$

Ex: MM2 force fields

Bond	R^O (Å)	k_L (kcal mol^{-1} Å^{-2})
Csp^3-Csp^3	1,523	317
Csp^3-Csp^2	1,497	317
Csp^2=Csp^2	1,337	690
Csp^2=O	1,208	777
Csp^3-Nsp^3	1,438	367
C-N (amide)	1,345	719

• E_{angle}
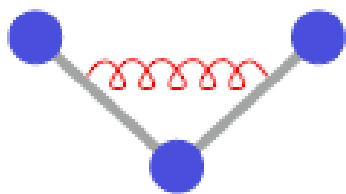
$$E_{angle} = \tfrac{1}{2} \, \Sigma \, k_q \, (q - q°)^2$$

$$E_{angle} = \tfrac{1}{2} \, \Sigma \, k_q \, (q - q°)^2$$

Ex: Amber force fields

Angle	θ^O	k_θ (kcal mol^{-1} deg^{-1})
Csp^3-Csp^3-Csp^3	109.47	0.0099
Csp^3-Csp^3-H	109.47	0.0079
H-Csp^3-H	109.47	0.007
Csp^3-Csp^2-Csp^3	117.2	0.0099
Csp^3-Csp^2=Csp^2	121.4	0.0121
Csp^3-Csp^2=O	122.5	0.0101

• $E_{torsion}$
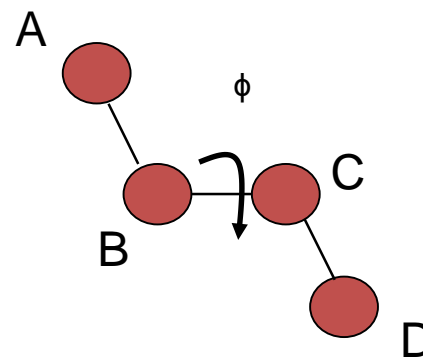
Dihedral angle A-B-C-D

Newman
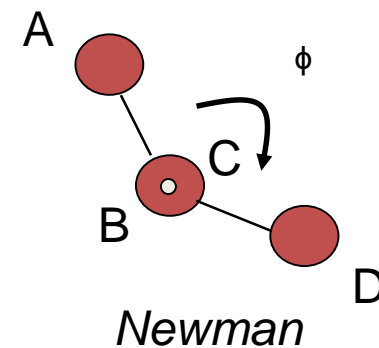
$$E_{torsion} = \tfrac{1}{2} \, \Sigma \, k_\phi \, [\, 1 + \cos (n \, \phi - \tau)]$$

Same as dihedral angle Y-Z-X-A

$\varphi° = 0°$

2515
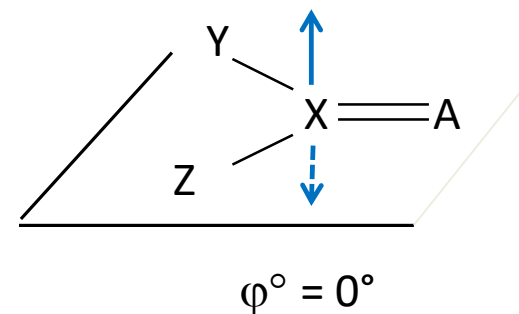
Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules

M. J. Hwang, T. P. Stockfisch, and A. T. Hagler*

Contribution from Biosym Technologies, Inc., 9685 Scranton Road, San Diego, California 92121

$$E = S_b\{\sum_b [{}^2K_b(b-b_0)^2 + {}^3K_b(b-b_0)^3 + {}^4K_b(b-b_0)^4]\}$$

$$+ S_\theta\{\sum_\theta [{}^2K_\theta(\theta-\theta_0)^2 + {}^3K_\theta(\theta-\theta_0)^3 + {}^4K_\theta(\theta-\theta_0)^4]\}$$

$$+ S_\phi\{\sum_\phi [{}^1K_\phi(1-\cos\phi) + {}^2K_\phi(1-\cos 2\phi) + {}^3K_\phi(1-\cos 3\phi)]\}$$

$$+ S_\chi\{\sum_\chi K_\chi\chi^2\} + \sum_{i>j}\frac{q_iq_j}{r_{ij}} + \sum_{i>j}\epsilon\left[2\left(\frac{r^*}{r_{ij}}\right)^9 - 3\left(\frac{r^*}{r_{ij}}\right)^6\right]$$

$$+ S_c\{\sum_b\sum_{b'}K_{bb'}(b-b_0)(b'-b'_0) + \sum_\theta\sum_{\theta'}K_{\theta\theta'}\times$$

$$(\theta-\theta_0)(\theta'-\theta'_0)$$

$$+ \sum_b\sum_\theta K_{b\theta}(b-b_0)(\theta-\theta_0)$$

$$+ \sum_\phi\sum_b (b-b_0)[{}^1K_{\phi b}\cos\phi + {}^2K_{\phi b}\cos 2\phi + {}^3K_{\phi b}\cos 3\phi]$$

$$+ \sum_\phi\sum_{b'} (b'-b'_0)[{}^1K_{\phi b'}\cos\phi + {}^2K_{\phi b'}\cos 2\phi + {}^3K_{\phi b'}\cos 3\phi]$$

$$+ \sum_\phi\sum_\theta (\theta-\theta_0)[{}^1K_{\phi\theta}\cos\phi + {}^2K_{\phi\theta}\cos 2\phi + {}^3K_{\phi\theta}\cos 3\phi]$$

$$+ \sum_\phi\sum_\theta\sum_{\theta'} K_{\phi\theta\theta'}(\theta-\theta_0)(\theta'-\theta'_0)\cos\phi\} \qquad (2)$$

(A) Acyclic Compounds
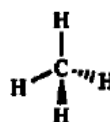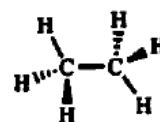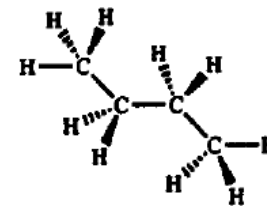
methane   ethane   n-butane   isobutane
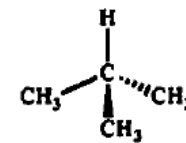
(B) Cyclic Compounds

cyclopropane   cyclobutane   cyclohexane

Table 1. Characteristics of Class II Force Field

I. The <u>Same</u> force field fits properties of:

Isolated small molecules (including structural, thermodynamic, spectroscopic data and dipole moments)

Condensed phases (including crystal lattice vectors, position and orientation of asymmetric units, sublimation energies, liquid heats of vaporization)

Macromolecular systems

II. <u>Fits</u> highly strained molecules including small rings with the same parameters

III. Analytical Form[a]

An anharmonic force field characterized by Morse or quartic bond stretching, quartic angle bending

Well characterized, 1-, 2-, and 3-fold torsion terms

Cross terms

Exponential or 9th power nonbonded repulsion

In summary

Potential   +   set of   +   experimental properties
(equations)   molecules   theoretical properties

FIT

parameters

A. Bond Length

$$E = {}^2K_b(b - b_0)^2 + {}^3K_b(b - b_0)^3 + {}^4K_b(b - b_0)^4$$

bond	b_0 (Å)	${}^2K_b$ (kcal mol^{-1} Å^{-2})	${}^3K_b$ (kcal mol^{-1} Å^{-3})	${}^4K_b$ (kcal mol^{-1} Å^{-4})
H–C	1.111	346.4	−706.6	863.2
C–C	1.535	299.4	−515.8	667.1

B. Bond Angle

$$E = {}^2K_\theta(\theta - \theta_0)^2 + {}^3K_\theta(\theta - \theta_0)^3 + {}^4K_\theta(\theta - \theta_0)^4$$

angle	θ_0 (deg)	${}^2K_\theta$ (kcal mol^{-1} rad^{-2})	${}^3K_\theta$ (kcal mol^{-1} rad^{-3})	${}^4K_\theta$ (kcal mol^{-1} rad^{-4})
H–C–H	107.7	41.7	−7.2	−8.8
H–C–C	110.8	42.7	−8.8	−9.1
C–C–C	112.9	42.3	−9.8	−9.2

C. Torsion Angle

$$E = {}^1K_\phi(1 - \cos\phi) + {}^2K_\phi(1 - \cos 2\phi) + {}^3K_\phi(1 - \cos 3\phi)$$

torsion	${}^1K_\phi$ (kcal mol^{-1})	${}^2K_\phi$ (kcal mol^{-1})	${}^3K_\phi$ (kcal mol^{-1})
H–C–C–H	−0.968	0.010	−0.150
H–C–C–C	−0.968	0.010	−0.150
C–C–C–C	−0.968	0.010	−0.150

Choice of a potential model

Some specific models exist in literature

No available model in literature

Specific models

- Charasteristics ?
- Quality ?
- Time consuming ?
- Available in codes ?

Force Fields

- Freeware ?
- Commercial ?
- Computable properties ?

Transfer of parameters from analogous systems ?

- Dangerous approach !
- Verifications ?
- Modifications

Development of a new specific model

- Complex
- Time consuming

You want to study a system by the use of molecular simulations but :
- *you cannot find a (good) model potential.*
- *you are not confident in the use of parameters coming from analogous systems.*
- *you are interested in the challenging task that consists in the development of an original model.*

Whatever the number of reasons, the question is : how to make a model ?

1. select sites, decide if you need a rigid or a flexible model.

2. be as smart as you can and invent a good (whatever that means) functional form for V with free parameters.

3. determine the free parameters by fitting the V-functions.

   - *do quantum mechanical calculations on a small number of molecules (2, maybe 3 or 4) for as many relative geometries as possible.*
   - *empirically using simulations*

4. test, improve, work hard (sometimes back to step 1 or 2 or 3…), ..., try & error procedures…

Conclusion: *(as already mentionned) the development of models is a science in itself.*
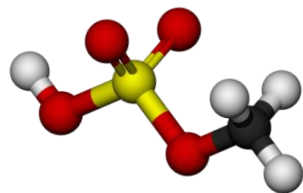
**Development of
a new potential model**

Nature of the model

All atoms, unified atoms,
Flexible or rigid model
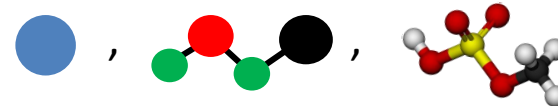
**Mathematical expression for the potential
energy function**
"*Use the right level of description to catch the
phenomena of interest…*"
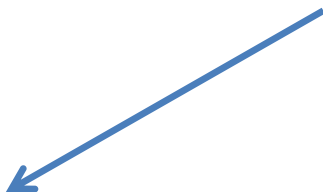
Empirical model

Fit the parameters through
simulations (MD or MC) of some
experimental properties

Ab initiol model

Fit the parameters through
QM calculations

- Geometry
- Frequencies
- PES of dimers, trimers…

Electrostatic model

-Model derived from electrostatic
properties of single molecule.

Fit of the rest (non electrostatic)
of the potential energy
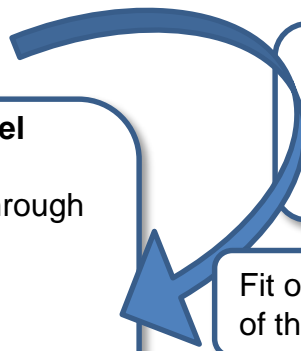
Electrostatic part of a model potential :

$$V(r) = E_{repulsion} + E_{dispersion} + \underline{E_{electrostatic}}$$

$\rho_A(r)$ $\rho_B(r)$

$\{q_i\}$ $\{q_i\}$
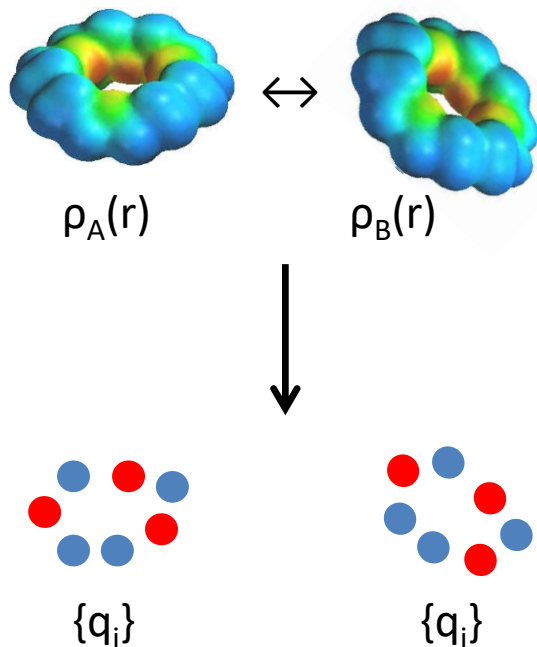
Classical model that simulate the combined electrostatic effects of both the atomic nuclei and the electrons :
- *Set of punctual charges*
Or more complex models like :
- *Distributed multipole moments (q, μ, Θ,…)*
- *Distributed polarizabilites…*

Simplest model : set of atomic charges

Different types of Atomic Charges can be easily obtained from ab initio calculation (ex. Gaussian) :

<u>Mulliken</u> charges and <u>Natural</u> charges (NPA) are both based on orbital occupancies . (*pop=npa*)

<u>AIM</u> (atoms in molecules) charges are based on a division of the molecule into atoms based on the topology of the electron density. (*pop=aim*)

<u>MK</u> (Merz-Singh-Kollman)
and
<u>CHELPG</u> (CHarges from Electrostatic Potentials using a Grid)
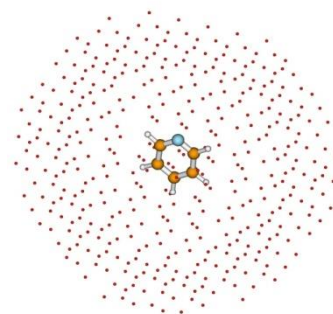
are examples of methods to derived charges by a fit to the molecule's electrostatic potential at a large number of grid points. (*pop=mk and pop=chelpg*)

Applications : various classes of models for CCl_4
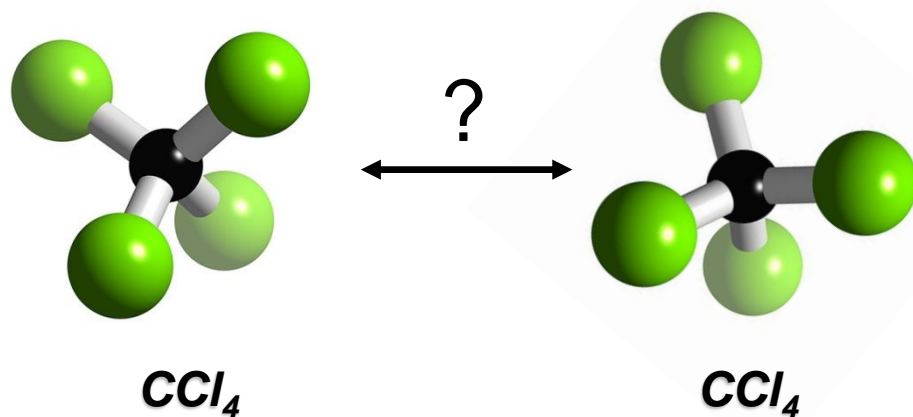
CCl_4 $\longleftrightarrow^{?}$ CCl_4

Questions :

(a) Write the simplest possible intermolecular potential for the CCl_4-CCl_4 interaction.

(b) Write a more sophisticated intermolecular potential assuming that the sites coincide with the atoms.

In each case, find the number of parameters of the model.

Some answers :
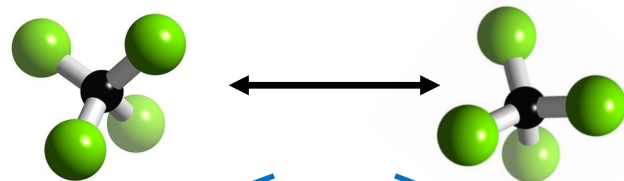
(a) <u>simplest possible</u>
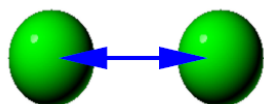
$\mathcal{V}(inter) =$ Lennard-Jones

2 parameters (ε,σ)

(b) <u>more sofisticated</u>

Rigid model
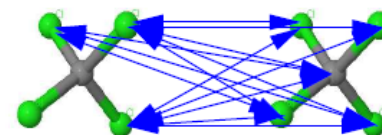
6 parameters : (ε$_C$, σ$_C$, ε$_{Cl}$, σ$_{Cl}$), q$_{Cl}$, q$_C$

Flexible model
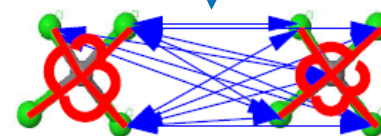
many parameters …

3-body terms, polarisation…
many many parameters …

Example of development of an ab initio model for liquid CCl4
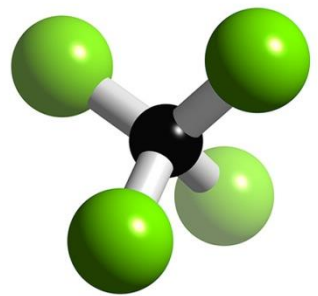
ab initio polarizable model
rigid 5 sites model

9 parameters

$$\mathcal{U} = \frac{1}{2} \sum_{\alpha \in A} \sum_{\beta \in B \neq A} \left(\boxed{\frac{q_\alpha q_\beta}{r_{\alpha\beta}}} + \boxed{exp[-\alpha_{\alpha\beta}(r_{\alpha\beta} - \rho_{\alpha\beta})] - \frac{C_{\alpha\beta}}{r_{\alpha\beta}^6}} \right) + \mathcal{U}_{ind}$$

$$\mathcal{U}_{ind} = \frac{1}{2} \sum_{a \in A} \sum_{b \in B \neq A} \Delta Q_t^a T_{tu}^{ab} \boxed{Q_u^b} \qquad \Delta Q_t^a = - \sum_{b \in B \neq A} \boxed{\alpha_{tt'}^{aa'}} T_{t'u}^{a'b} (Q_u^b + \Delta Q_u^b)$$

CCl$_4$

(I)

ab initio calculations *(MP2/aug-cc-pVDZ and BSSE)*
$\Rightarrow$ molecular geometry (bond length)
$\Rightarrow$ **electrostatic model : multipoles and polarizabilities**
$\Rightarrow$ exploration of the PES of the dimer
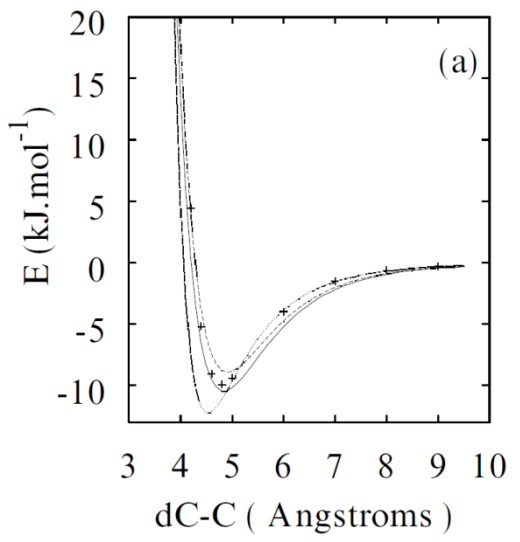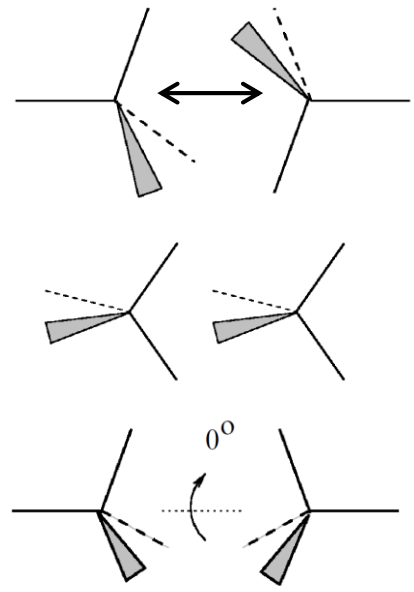
(II)

Fit of the non electrostatic part of the potential:

$$\mathcal{U}_{vdW} = \mathcal{U}_{ab\ initio}^{dimers} - \mathcal{U}_{elect}$$

(a)

E (kJ.mol^{-1}) vs dC-C (Angstroms)
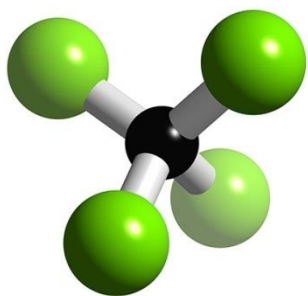
(III)

Test of the model by MD on some condensed phase properties

$\Rightarrow$ thermodynamics
$\Rightarrow$ structural
$\Rightarrow$ dynamical

New compromise in the set of parameters… go back to step II…

CCl₄

2d virial coefficientl B(T)

Model   Exp.
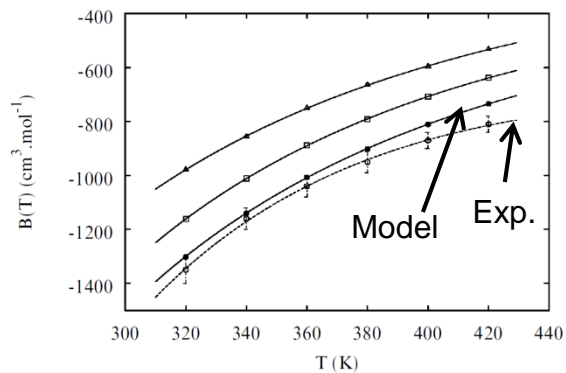
Neutron diffraction
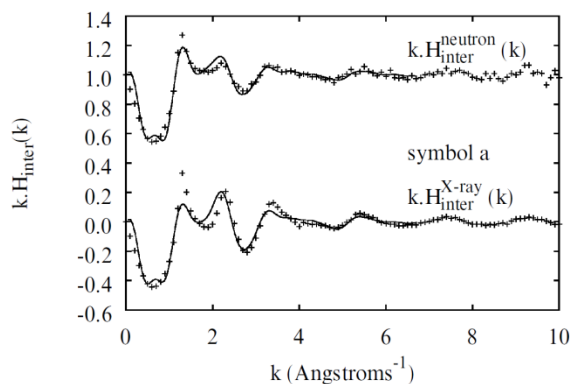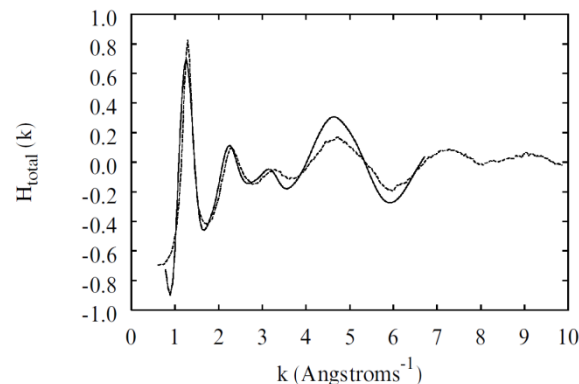
$$G_{\text{inter}}^{\text{neutron}}(r) = 0.02\,g_{CC}(r) + 0.25\,g_{CCl}(r) + 0.73\,g_{ClCl}(r),$$

X-ray diffraction

$$G_{\text{inter}}^{X-\text{ray}}(r) = 0.01\,g_{CC}(r) + 0.15\,g_{CCl}(r) + 0.84\,g_{ClCl}(r),$$

2d virial coefficient ⇔ global property of the PES

⇔ represent the deviations from ideality due to two-molecules interactions

Experiment → $P = \frac{NkT}{V}(1 + \frac{N}{V}B(T) + (\frac{N}{V})^2 C(T) + ...)$

⇕

Model → $B_{cl}(T) = -\frac{1}{2}\int_0^\infty < e^{-U_{12}/kT} - 1 >_{\omega_1 \omega_2} dr$

Dimer H2O - H2O
« Average PES »
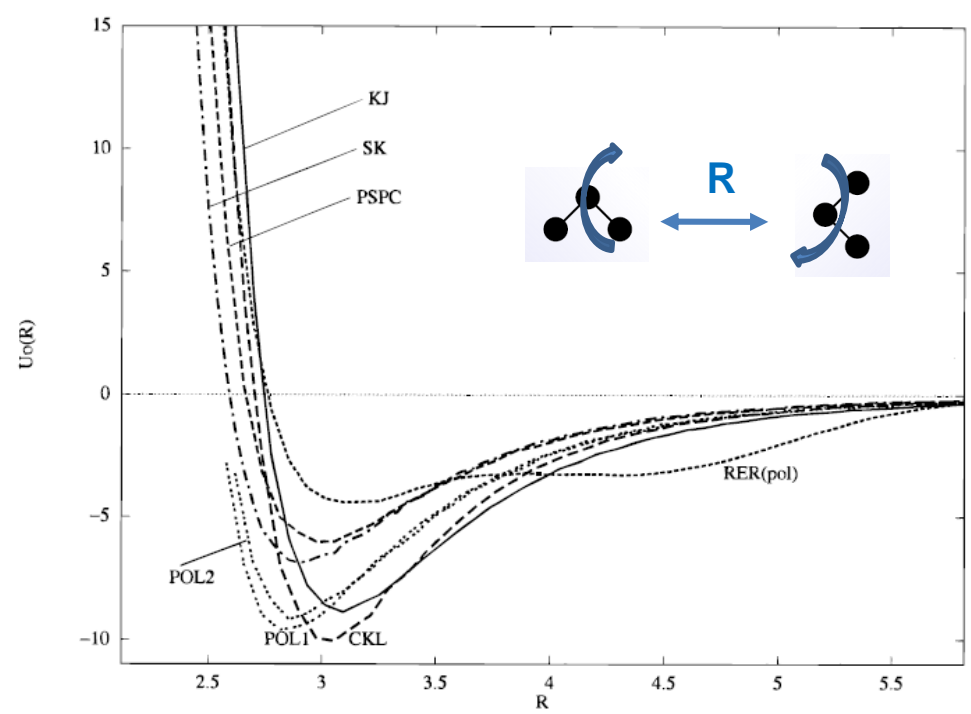
Figure 4. Orientationally averaged pseudopotential $U_0(R)$ (kJ mol^{-1}) (see text for definition) for some polarizable water models vs the center of mass separation R (Å) at 373 K: (a) ASP-W, ASP-W2, ASP-W4, NCC, and NEMO1−3; (b) PSPC, POL1−2, RER(pol), SK, CKL, and KJ.
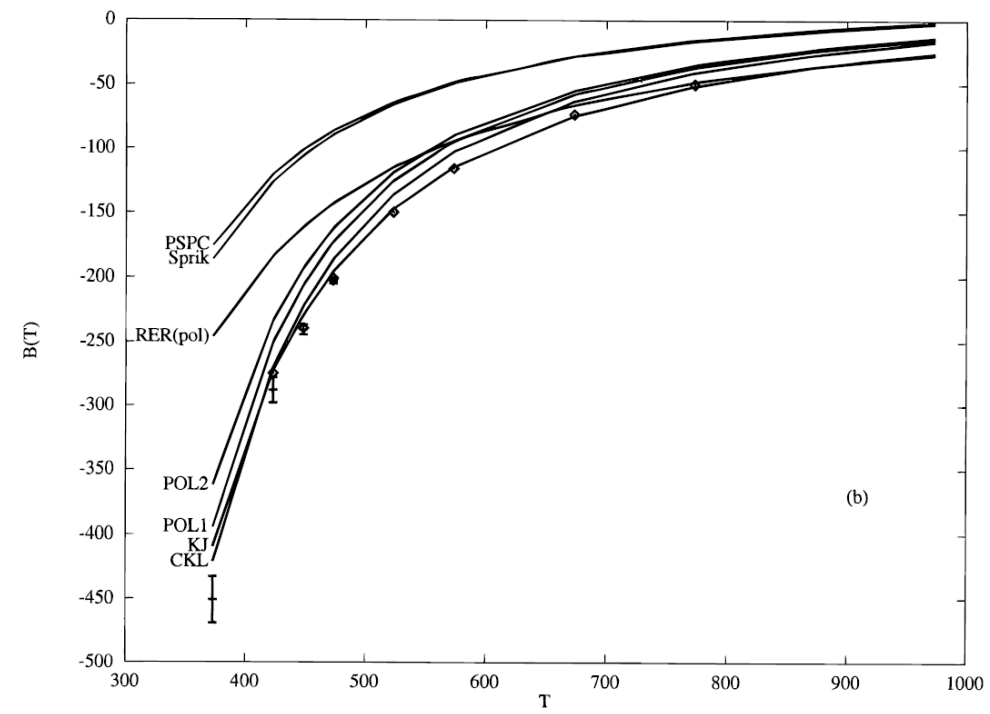
2d virial coefficient ⇔ global property of the PES
⇔ represent the deviations from ideality due to two-molecules interactions

Experiment → $P = \frac{NkT}{V}(1 + \frac{N}{V}B(T) + (\frac{N}{V})^2 C(T) + ...)$

⇕

Model → $B_{cl}(T) = -\frac{1}{2}\int_0^\infty < e^{-U_{12}/kT} - 1 >_{\omega_1\omega_2} dr$

Dimer H2O - H2O
2d virial coefficient B(T)

Figure 3. Second virial coefficient $B(T)$ of water including first-order quantum corrections (cm³ mol⁻¹) for some polarizable intermolecular potentials. The experimental values of ref 87 (set III) are shown by error bars, and those of ref 43 by diamonds. (a) ASP-W, ASP-W2, ASP-W4, NCC, and NEMO1−3; (b) PSPC, POL1−2, RER(pol), SK, CKL, and KJ.

« Force Fields & Molecular Interactions »

I.   Recalling basic ideas about statistical mechanics

II.   General ideas about interaction models

III. The nature of intermolecular forces

IV. Representation of the intermolecular potential energy function

V.   Strategies to get a model potential for simulations

VI. Beyond the pair potential approximation

Recall : **pair potential approximation** = function developed over pair of molecules :

$$\mathcal{V}(\mathbf{r}) \;=\; \sum_{i=1}^{N} \mathcal{V}(\mathbf{r}_i) \;\;+\; \sum_{i=1}^{N-1} \sum_{j>i} \mathcal{V}(\mathbf{r}_{ij})$$

What do we miss ? What do we leave out?

(i) By summing over pair of molecules, you miss the influence that a third particle may have on the (shape, strength, ...) interaction V between A and B.

$$U_{ABC} \neq U_{AB} + U_{BC} + U_{AC}$$
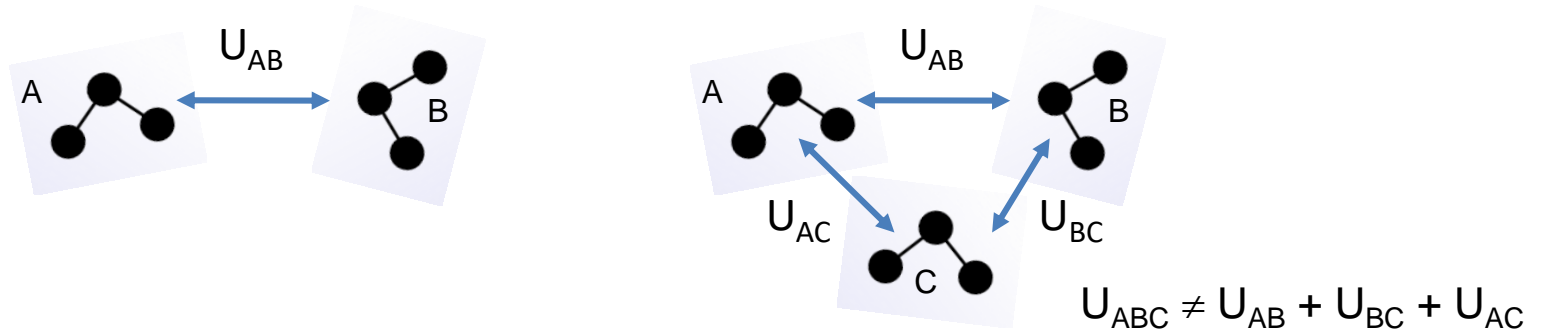
Ex : error in interaction energies in cation–water systems
$\Rightarrow$ 10, 15, 20 % for mono-, di- and tri-valent ions

Solution : *add "3-body" terms in addition to standard "2-body" potential*

$$+ \sum_{i=1}^{N-1} \sum_{j>i} \sum_{k>j} \mathcal{U}(\mathbf{r}_{ijk})$$

(ii) By summing over pair of molecules, you miss the influence
      of the whole system,
      of heterogeneity
      of perturbations

      …

Solution : *use of explicit polarizabilities in the model*

$$\Delta Q_t^a = - \sum_{b \in B \neq A} \alpha_{tt'}^{aa'} \, T_{t'u}^{a'b} (Q_u^b + \Delta Q_u^b)$$

Induced
moments

Permanent   Induced
moments   moments

$\Rightarrow$ self consistency problem (solved by iteration or matrix inversion) : computationally very expensive

$$\mathcal{V}_{ind} = \frac{1}{2} \sum_{a \in A} \sum_{b \in B \neq A} \Delta Q_t^a T_{tu}^{ab} Q_u^b$$

Example :

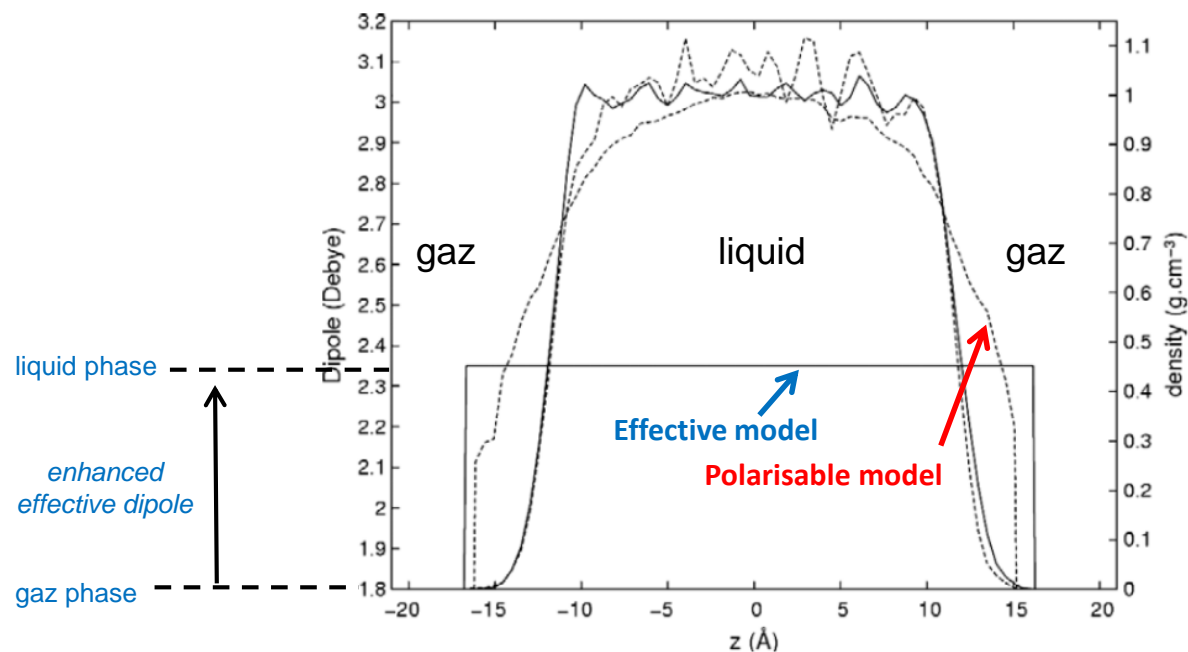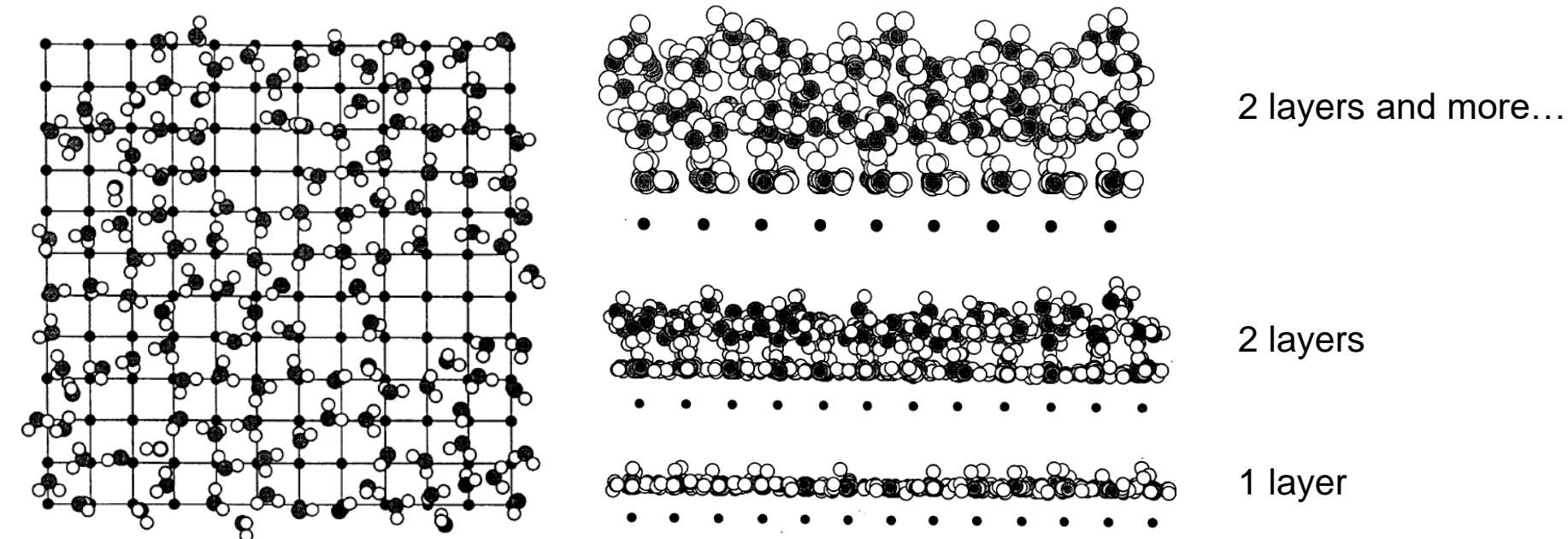Molecular dipole moment of water molecules in a Liquid – Gaz interface
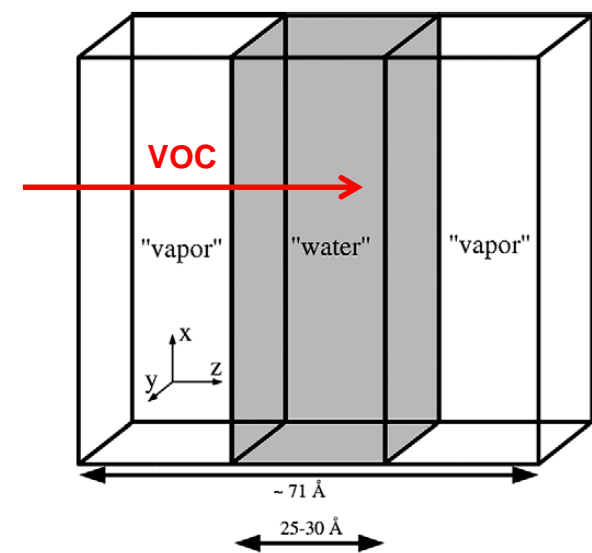

Fig. 2. Simulation cell.

Adsorption of molecules on this interface strongly depend on its properties.
(VOC : volatil organic coumpounds)

Example :

Adsorption of water molecules on an ionic surface (MgO)


2 layers and more…

2 layers

1 layer

There is a competition between the lateral interactions (between water molecules), and the water–surface interaction. The evolution of the molecular dipole of water according to their environment must be taken into account.
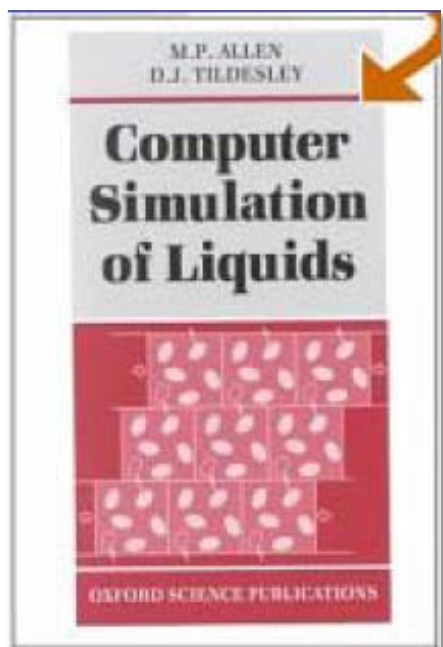
M. P. Allen & D. J. Tildesley
Computer simulation Of liquids
Clarendon Press,
Oxford, 2006.
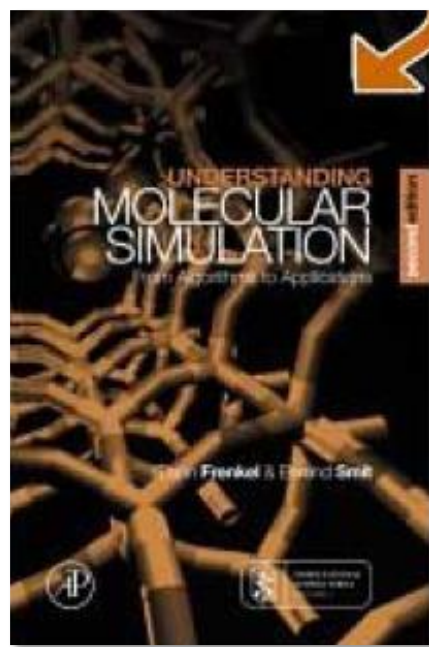
D. Frenkel & B. Smit
Understanding molecular simulation
Academic Press,
New York, 1996
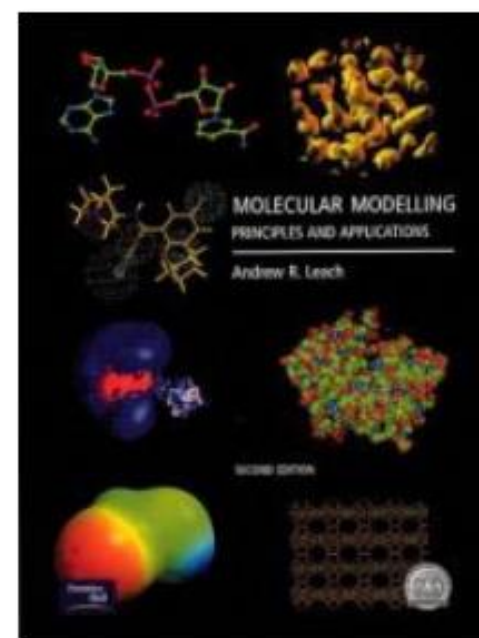
A.R. Leach
Molecular Modelling
Principles and Applications
AIP, Dorchester, 2001.

See also :

J. A. McCammon, S. C. Harvey (1987), Dynamics of Proteins and Nucleic Acids, Cambridge University Press

D. C. Rapaport (1996), The Art of Molecular Dynamics Simulation

J. M. Haile (2001), Molecular Dynamics Simulation: Elementary Methods, Wiley

A. Gerschel (1995), Liaisons intermoléculaires, EDP Sciences, CNRS Edition

Pr. Jean-Christophe Soetens

Institut des Sciences Moléculaires
Groupe Chimie Théorique & Modélisation
UMR CNRS - Université de Bordeaux 5255
351, Cours de la Libération
F-33405 Talence Cedex

(++33) (0) 540 00 22 42

jean-christophe.soetens @u-bordeaux.fr

http://theo.ism.u-bordeaux.fr/J-C.Soetens/