

Statistical Mechanics & Simulations

Jean-Christophe SOETENS

Department of Chemistry
University of Bordeaux

jean-christophe.soetens@u-bordeaux.fr

Academic year 2024-2025

« *Statistical Mechanics & Simulations* »

- *Where ?* Here... A10 – Lavoisier
- *What ?* Course, without overdoing things...
Practice : a lot...
- *Examination :* Final exam (3h) x 0.7 (and 2d chance)
 Intermediate x 0.3 (0.15 + 0.15) (definitive)

Course materials : <http://theo.ism.u-bordeaux.fr/J-C.Soetens/teach.html>

*The purpose of this course : « **Statistical Mechanics & Simulations** »*

- To become familiar with the basic concepts underlying molecular simulations
- To understand the possibilities and limitations of (present day) simulations
- To understand the machinery of the two main methods :

Molecular Dynamics simulations

Monte Carlo simulations

- To be able to understand a scientific article dealing with molecular simulation and to discuss with experts in the field
- To give you some ideas on how to start your own studies
- To become familiar with a numerical scientific environment (software, programming, graphics, etc)

PART I : « Statistical Mechanics & Simulations »

- I. Overview of Statistical Mechanics & Molecular Simulations
- II. Molecular Dynamics Simulations
- III. Monte Carlo methods
- IV. « Outputs » : extracting properties from simulations
- V. Initiation to statistical thermodynamics

PART II :« 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

PRACTICE

Molecular Simulation

Practice

Academic year 2024-2025

University of Bordeaux

J-C. Soetens

jean-christophe.soetens@u-bordeaux.fr

« Personal part »
From scratch

P0: Numerical environment

- P0-A) Connexion to gaudi (`gaudi.ism.u-bordeaux.fr`)
- P0-B) OS (Unix), text editor, graphic tools
- P0-C) Start programming in FORTRAN

« Tools »

P1: From statistical mechanics to simulation methods.

- P1-A) Ising model
- P1-B) The Boltzmann distribution

« Classic part »

P2: Molecular Dynamics simulations : how it works ?

P3: Molecular Dynamics simulations : first applications.

- P3-A) Liquid water under normal conditions.
- P3-B) Aqueous ionic solution.

Objectives
Protocols
Calculations
Analyzes
...

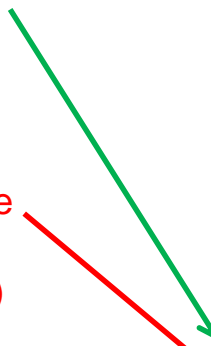
P4: Molecular interactions.

- P4-A) *ab initio* calculations of electrostatic potential derived charges.
- P4-B) Exploration of a PES and fit of an intermolecular interaction model.

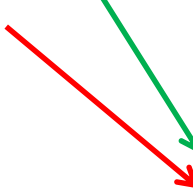
P5: To go further...

- P5-A) Read, analyze, summarize and present a scientific article.
- P5-B) Programming (in FORTRAN) to develop your own tools.

P5-A) Oral presentation
(intermediate x 0.15)



P5-B) Unix, Fortran language
Gnuplot, VMD, etc.
(intermediate x 0.15)



Course materials : <http://theo.ism.u-bordeaux.fr/J-C.Soetens/teach.html>

Theoretical Chemistry & Modeling Group



Theoretical Chemistry & Modeling Group

Home page	Theoretical chemistry	Department of Chemistry	Map
-----------	-----------------------	-------------------------	-----

Jean-Christophe SOETENS, Teaching

Contact: Bât A12, 3^{ème} étage Est, #B13

☎ (+33) (0) 540 00 22 42 ✉ jean-christophe.soetens@u-bordeaux.fr

Computational chemistry (4TCH914U)

Part of the Physical Chemistry & Chemical Physics (PCCP) International master program

Planning :

Tuesday 8:00-12:20 - weeks 36-43 & 45-51 - A10 / Lavoisier

Course content :

Basic concepts : Elementary classical statistical mechanics. Ensembles and fluctuations, Molecular interactions, Molecular mechanics, Equilibrium molecular dynamics simulations, Monte Carlo methods, Calculations of properties.

Advanced methods : Overview of various software, Methods of energy minimization, Free energy calculations, Computer graphics visualization, Elaborated interaction models (electrostatics and polarizability), Ewald sum and reaction field methods for treating long-range electrostatic interactions.

Course Material :

Course : [part I](#)

Course : [part II](#)

Articles : [Mynam, J. Chem. Phys. 2021](#)

[Metropolis, J. Chem. Phys. 1953](#)

[Rahman, Phys. Rev. 1964](#)

Practice

Topics

Application P1-B : [Excel file](#)

Unix OS : [french](#), [english](#)

Programming language : [Fortran](#) & <http://www.idris.fr/formations/fortran/>

[Connections to computers](#)

Useful software

[Gnuplot](#) - graphing utility (free)

[VMD](#) - Visual Molecular Dynamics (free)

[Gaussian 16](#) - electronic structure modeling

[Molden](#) - displaying results from ab Initio packages (free)

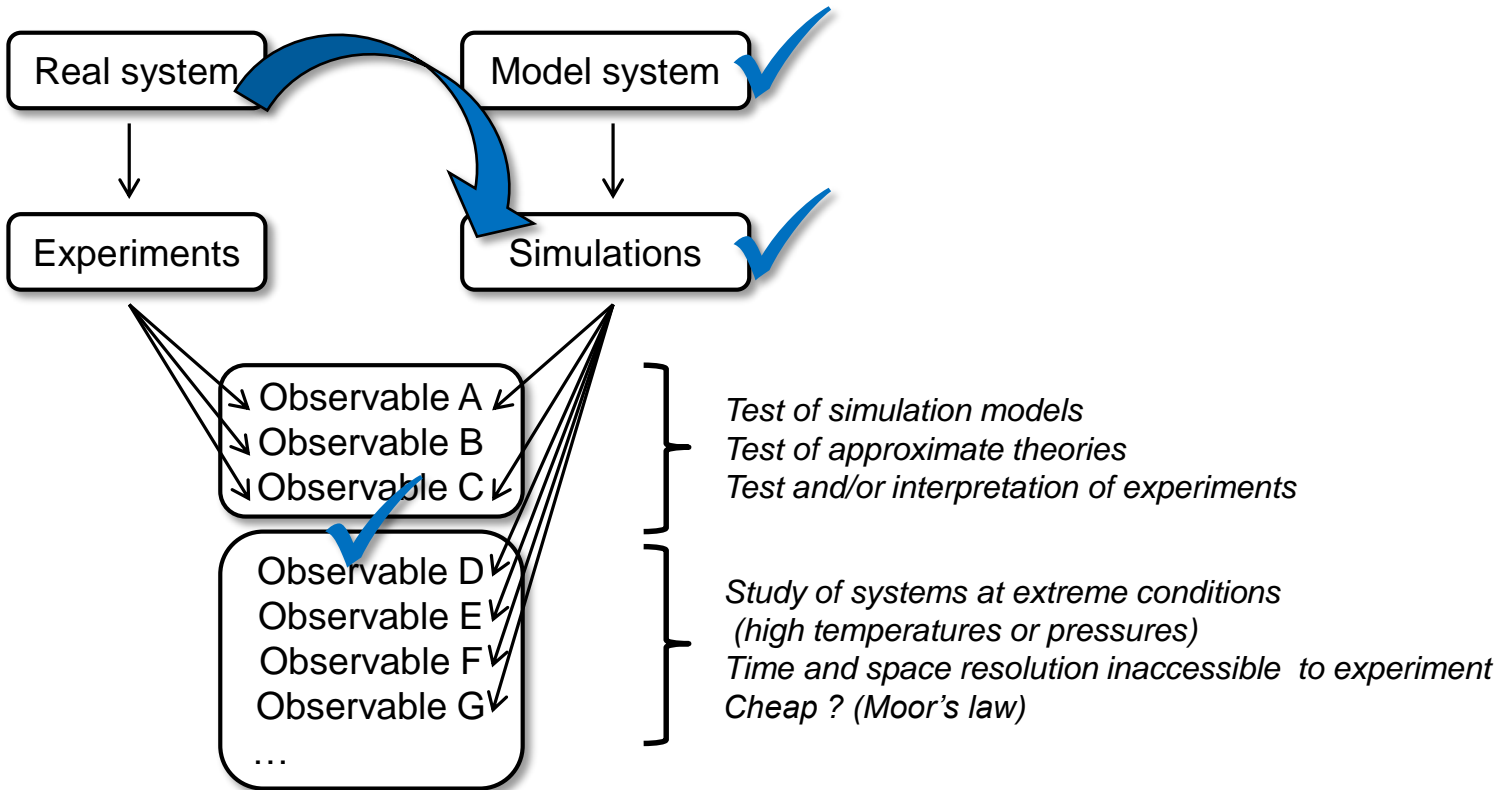
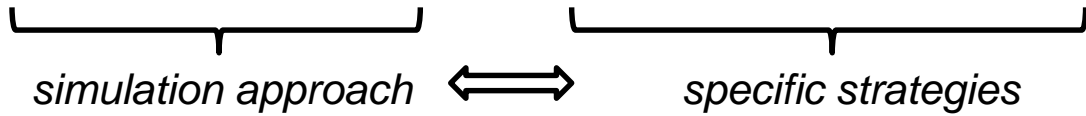
Examples of past exams :

[2021 \(supplementary material\)](#)

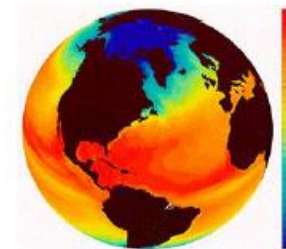
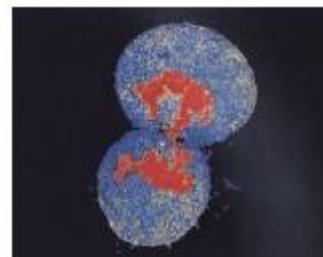
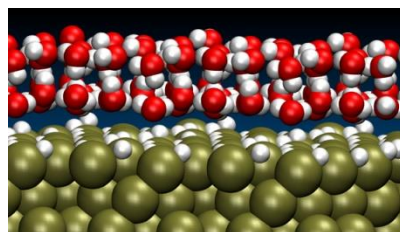
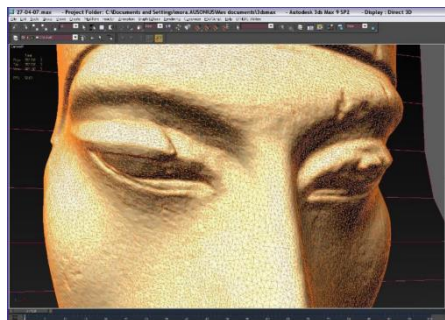
[2022 \(supplementary material\)](#)

[2023 \(supplementary material\)](#)

Simulations of Molecular Systems ?



Some words about « simulations »



➤ **Meteorology, climatology, Earth Sciences**

Measure warming and climate change

Anticipate weather events

➤ **Astrophysics, particle physics and plasma physics**

Understand the evolution of galaxies

Explore quantum theories (QCD, ITER)

➤ **Materials science, chemistry, nanoscience**

Development of smart materials

Better understanding of the properties

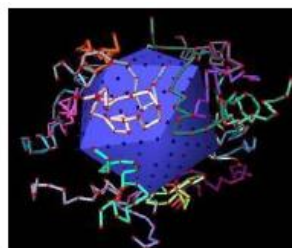
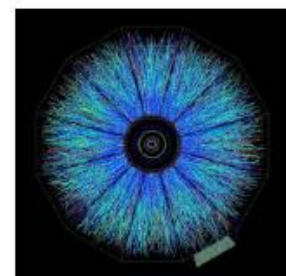
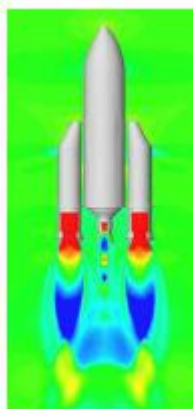
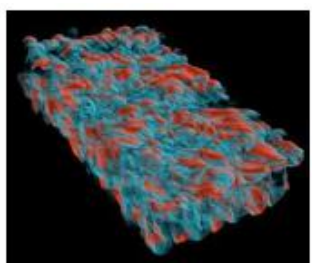
➤ **Life Sciences, human sciences**

Exploring biological systems, brain

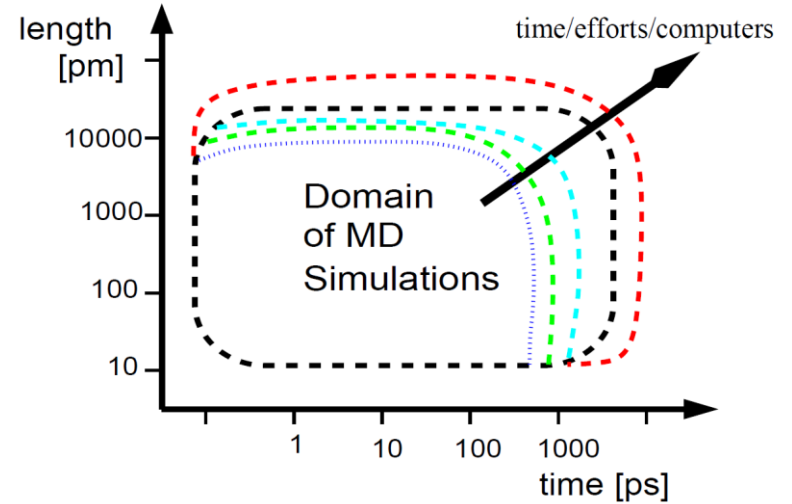
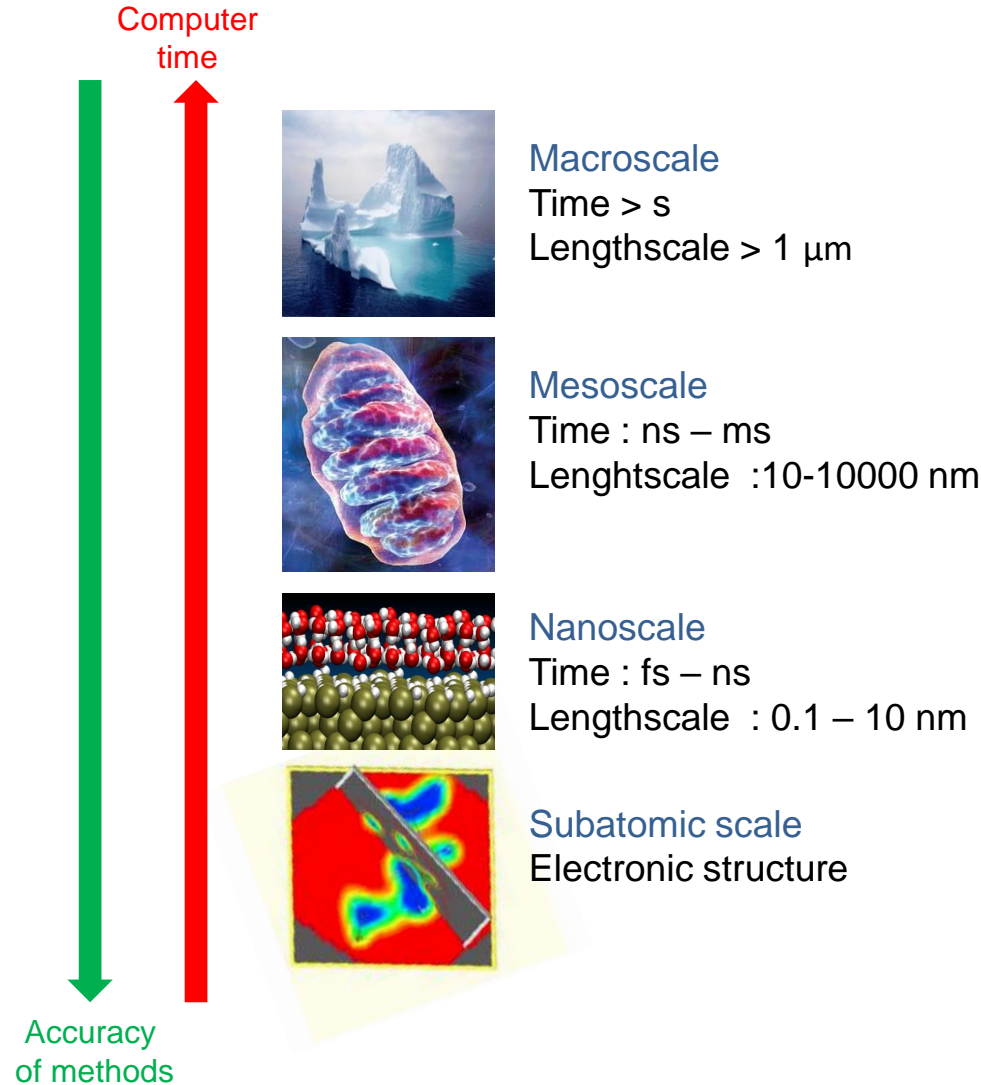
Develop targeted drug molecules

➤ **Engineering**

➤ **Etc...**



Time and length scales



Phase space
Statistics
Free energy, entropy
Molecular Dynamics, Monte Carlo

ab initio
Electrons
Potential Energy Surface

*The purpose of this course : « **Statistical Mechanics & Simulations** »*

- To become familiar with the basic concepts underlying molecular simulations
- To understand the possibilities and limitations of (present day) simulations
- To understand the machinery of the two main methods :

Molecular Dynamics simulations

Monte Carlo simulations

- **To be able to understand a scientific article dealing with molecular simulation and to discuss with experts in the field**
- To give you some ideas on how to start your own studies
- To become familiar with a numerical scientific environment (software, programming, graphics, etc)

Exercise : let's look at a typical molecular simulation article in physical chemistry to see what we need to learn before understanding such an article !?

Effect of temperature on concentrated electrolytes for advanced lithium ion batteries

Cite as: J. Chem. Phys. 154, 214503 (2021); doi: [10.1063/5.0049259](https://doi.org/10.1063/5.0049259)

Submitted: 3 March 2021 • Accepted: 17 May 2021 •

Published Online: 2 June 2021



View Online



Export Citation



CrossMark

Mahesh Mynam,^{a)}  Surbhi Kumari, Bharath Ravikumar,  and Beena Rai 

AFFILIATIONS

TCS Research, Tata Research Development and Design Centre, 54B, Hadapsar Industrial Estate, Pune 411013, India

^{a)} Author to whom correspondence should be addressed: mahesh.mynam@tcs.com

ABSTRACT

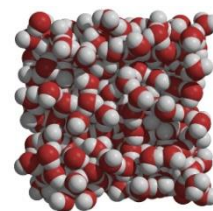
Salt-concentrated electrolytes are emerging as promising electrolytes for advanced lithium ion batteries (LIBs) that can offer high energy density and improved cycle life. To further improve these electrolytes, it is essential to understand their inherent behavior at various operating conditions of LIBs. Molecular dynamics (MD) simulations are extensively used to study various properties of electrolytes and explain the associated molecular-level phenomena. In this study, we use classical MD simulations to probe the properties of the concentrated electrolyte solution of 3 mol/kg lithium hexafluorophosphate (LiPF_6) salt in the propylene carbonate solvent at various temperatures ranging from 298 to 378 K. Our results reveal that the properties such as ionic diffusivity and molar conductivity of a concentrated electrolyte are more sensitive to temperature compared to that of dilute electrolytes. The residence time analysis shows that temperature affects the Li^+ ion solvation shell dynamics significantly. The effect of temperature on the transport and dynamic properties needs to be accounted carefully while designing better thermal management systems for batteries made with concentrated electrolytes to garner the advantages of these electrolytes.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0049259>

« Simulations » in chemical physics

Macroscopic point of view


Real system
 N_A molecules or particles
 described by few variables
 such as N, V, p, T...

Microscopic point of view


Model system
 classical → microstates
 $\mathcal{P}_i(\vec{r}_i, \vec{v}_i)$

One Macroscopic state ⇔ **Many compatible microscopic states**

⇒ for a classical system of N particles at time t : $\mathcal{P}_i(\vec{r}_{i1}, \vec{r}_{i2}, \vec{r}_{i3}, \dots, \vec{v}_{i1}, \vec{v}_{i2}, \vec{v}_{i3}, \dots)$

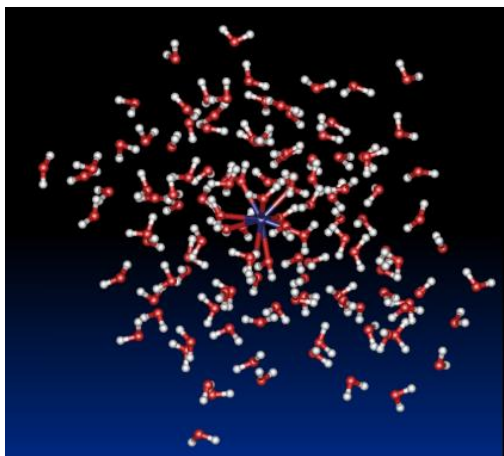
⇒ to compute the thermodynamic average of a quantity \mathcal{A}

$$\langle \mathcal{A} \rangle = \frac{1}{Z} \int \int \mathcal{A}(\vec{r}_i, \vec{v}_i) \mathcal{W}(\vec{r}_i, \vec{v}_i) dr dp \quad \text{with} \quad Z = \int \int \mathcal{W}(\vec{r}_i, \vec{v}_i) dr dp$$

we thus have to know for each point of the phase space :

$\mathcal{A}(\vec{r}_i, \vec{v}_i)$ value of the property in the microstate $\mathcal{P}_i(\vec{r}_i, \vec{v}_i)$
 $\mathcal{W}(\vec{r}_i, \vec{v}_i)$ weight (probability) of the microstate

Microscopic point of view : how many compatible microscopic states ?



Application:

system of 36 water molecules \leftrightarrow 108 atoms

7 spatial coordinates considered for each atom :

- original position (x_o, y_o, z_o)

- 2 moves in each of the 3 directions $(x_o \pm \delta, y_o \pm \delta, z_o \pm \delta)$

Work to do :

remember :

$$\langle \mathcal{A} \rangle = \frac{1}{Z} \int \int \mathcal{A}(\vec{r}_i, \vec{v}_i) \mathcal{W}(\vec{r}_i, \vec{v}_i) dr dp$$

\Rightarrow Generate all corresponding configurations C_i

\Rightarrow Compute each $A(C_i)$ and $W(C_i)$

(potential energy function for example, 1 ns on a laptop)

\Rightarrow Result : $\langle A \rangle$

Question : computer time necessary to do this work ?

Answer :

\Rightarrow Generate all corresponding configurations C_i \rightarrow Nb of configurations = ?

\Rightarrow Compute each $A(C_i)$ and $W(C_i)$ \rightarrow Total computer time = ?

\Rightarrow Result : $\langle A \rangle$

Laptop Supercomputer (1.5 million processors) \rightarrow ?

Quiz : 1 day on this supercomputer \leftrightarrow how many H_2O ? \rightarrow ? molecules

Summary up to there

- We cannot use, for practical reasons, the formulae that we find in statistical mechanics textbooks in a straightforward manner to compute averages.

- What we need / what we can do :
 - 1- to construct a list of microstates compatible with the macrostate we are interesting in.
 - 2- to compute the properties for each element of the list.
 - 3- to take the averages over the ensemble of elements.

- Warning : this list will always be finite !

Take home message...

Real condensed phase ?

- phase space $\Leftrightarrow \mathcal{P}_i(\vec{r}_i, \vec{v}_i)$
- large number of molecules
- complex potential energy functions



**“Brute force” approach is infeasible
and
Extremely inefficient**

we know only at the end which microstates
contribute to the integral

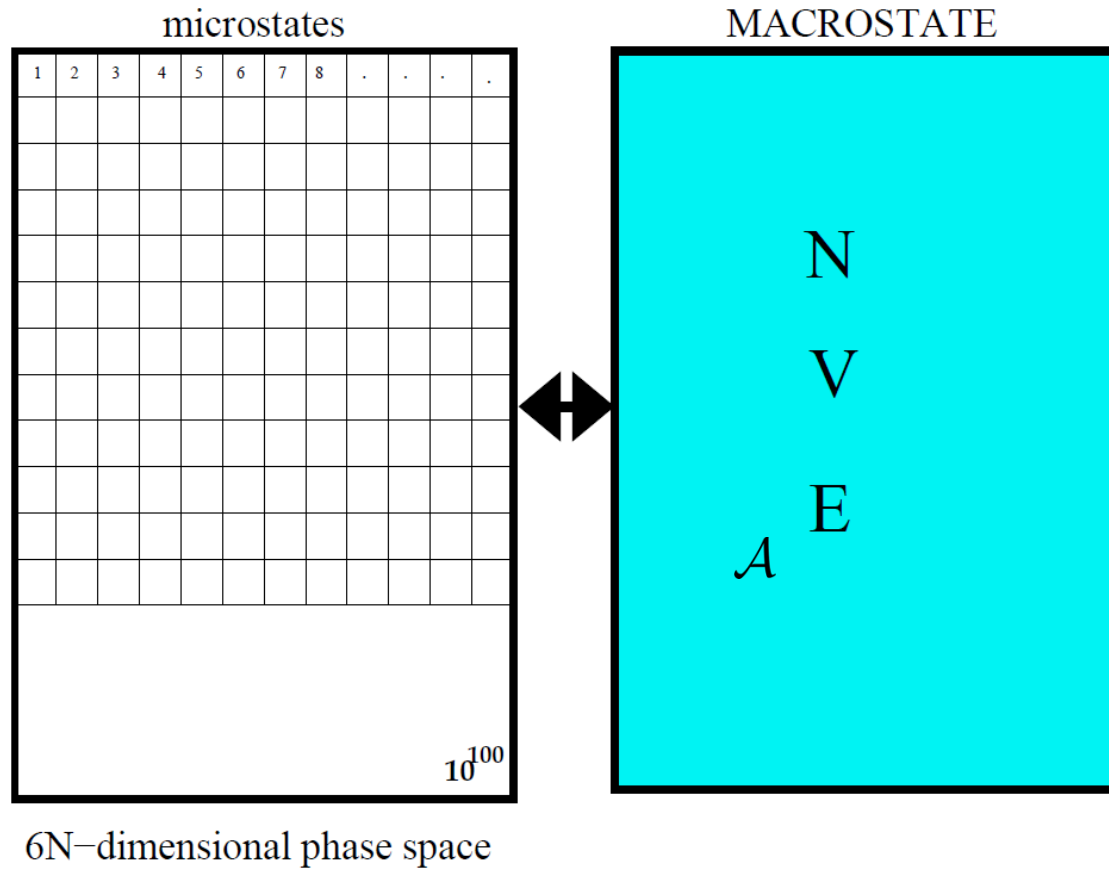
How to generate the microstates efficiently ?

SOLUTION = SIMULATION

Simulations = *Methods to explore the dominant regions of the phase space*

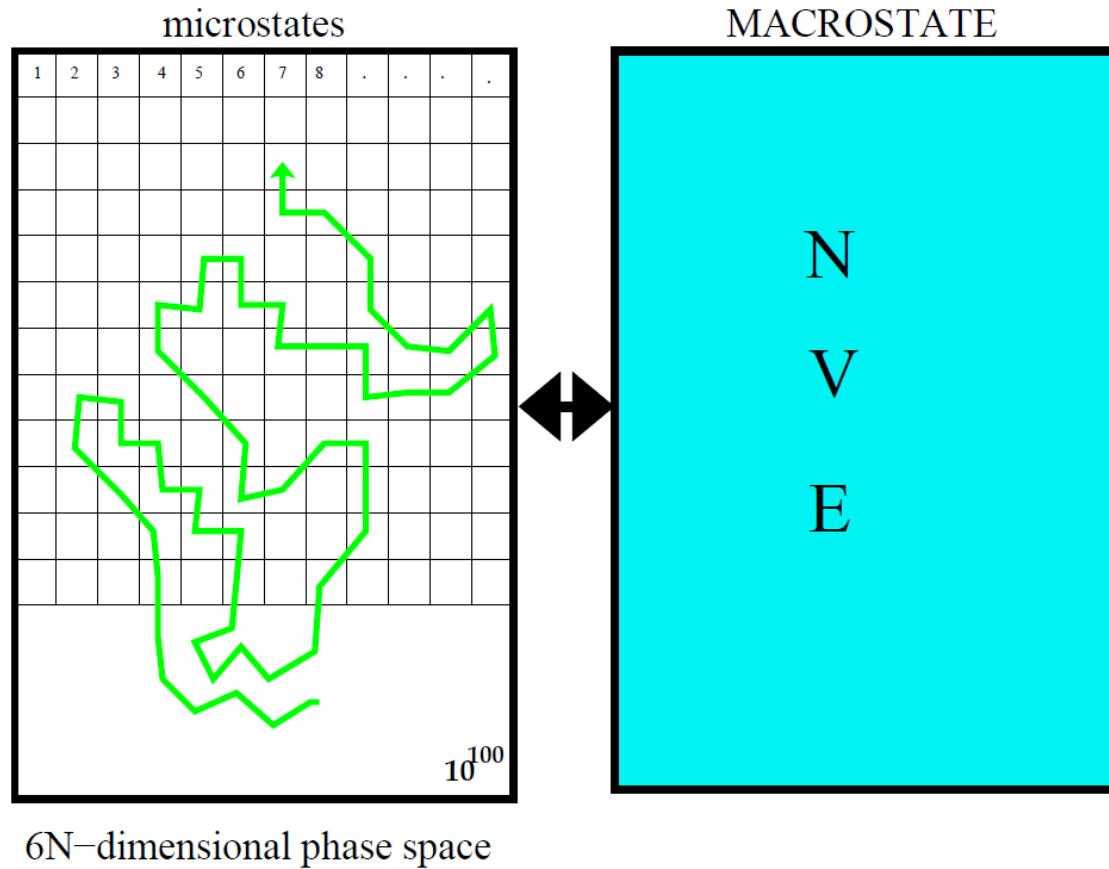
= *Methods to generate the dominant contributions to the integral*

Phase space :

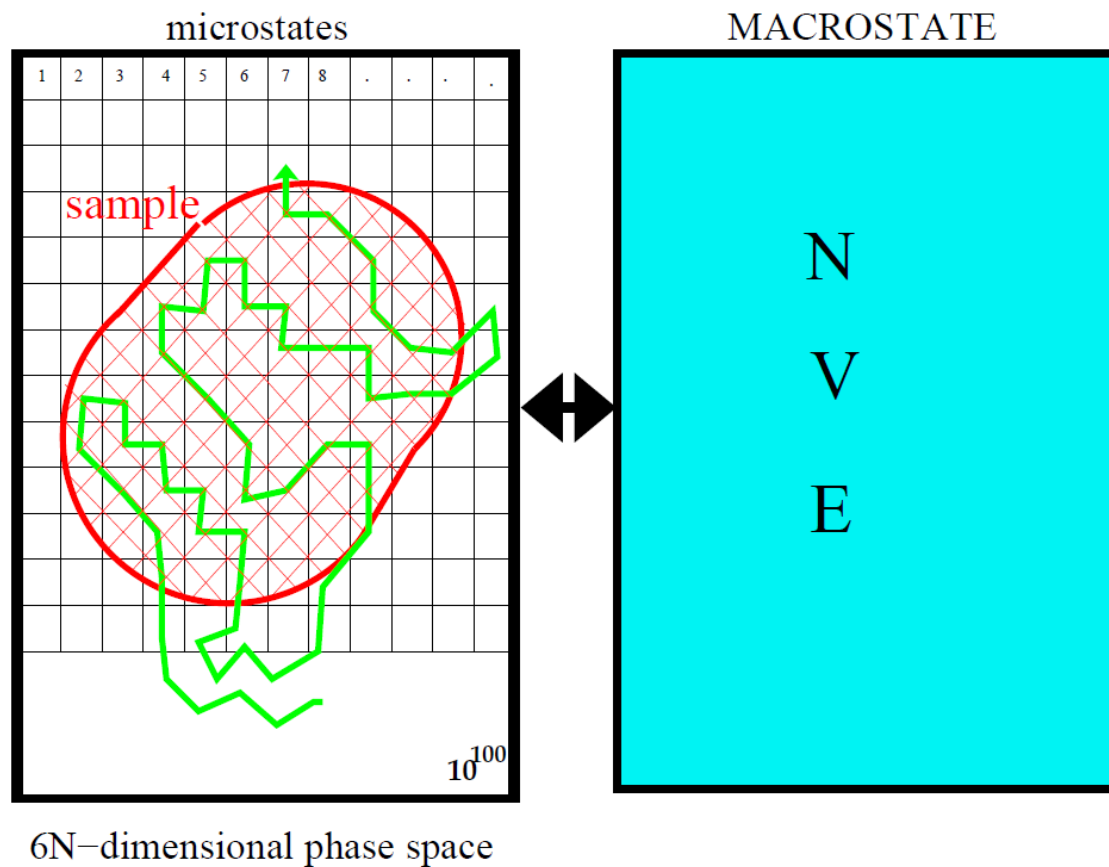


Classical system of N particles at time t : $\mathcal{P}_i(\vec{r}_{i1}, \vec{r}_{i2}, \vec{r}_{i3}, \dots, \vec{v}_{i1}, \vec{v}_{i2}, \vec{v}_{i3}, \dots)$

Exploration of the dominant regions of the phase space :



Finite sample of microstates:



Simulation approach :

⇒ Construct a finite list of microscopic states which are representative and compatible with some set of macroscopic variables (T, P, ρ...)



Ensemble of N microstates \mathcal{P}_i

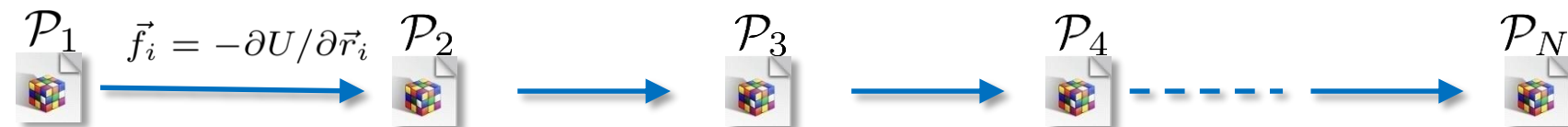
$$\langle \mathcal{A}(\mathcal{P}_i) \rangle = \frac{1}{N} \sum_{i=1}^N \mathcal{A}(\mathcal{P}_i)$$

Representative ?

$$\langle \mathcal{A}(\mathcal{P}_i) \rangle_N \approx \langle \mathcal{A}(\mathcal{P}_i) \rangle_{N \rightarrow \infty}$$

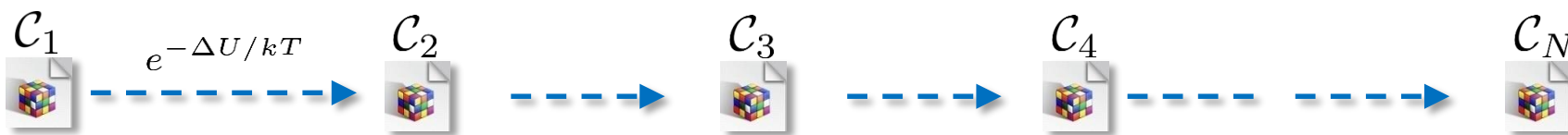
Molecular Dynamics

⇒ list of \mathcal{P}_i in Phase space (\mathbf{r}, \mathbf{v}) and ordered in time



Monte Carlo

⇒ list of \mathcal{C}_i in Configuration space (\mathbf{r})



Molecular Dynamics : MD

Generates Ps characterized by :

- *Constant number of particules, N*
- *Constant volume, V*
- *Constant total energy, E*

This ensemble is known as the NVE ensemble
(**microcanonic ensemble** of statistical mechanics)

Added values

MD generates the Ps in a meaningful order
The order parameter is the time, t

Consequences

MD generates a trajectory

MD allow to compute dynamical properties

Monte Carlo : MC

Generates only the position part of the Ps
(Cs) characterized by :

- *Constant number of particules, N*
- *Constant volume, V*
- *Constant temperature, T*

This ensemble is known as the NVT ensemble
(**canonic ensemble** of statistical mechanics)

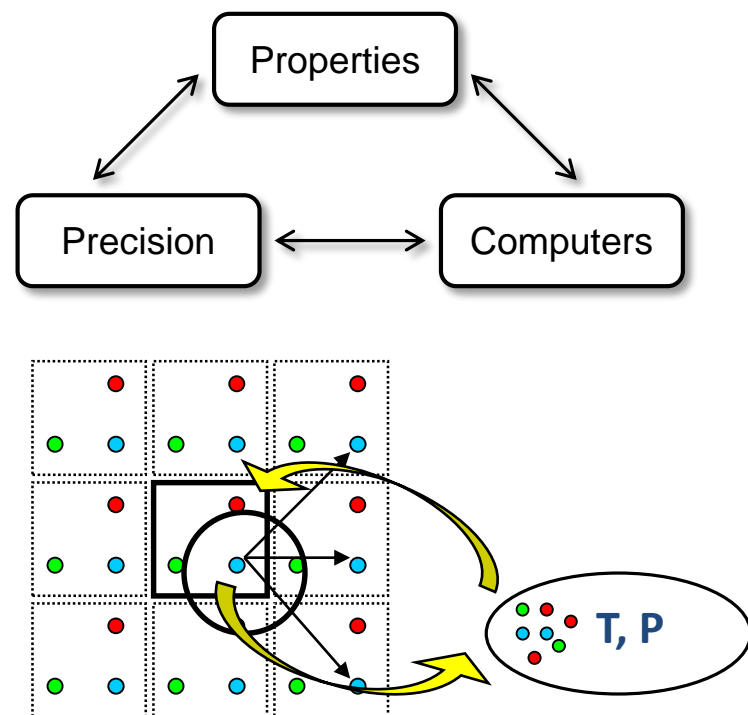
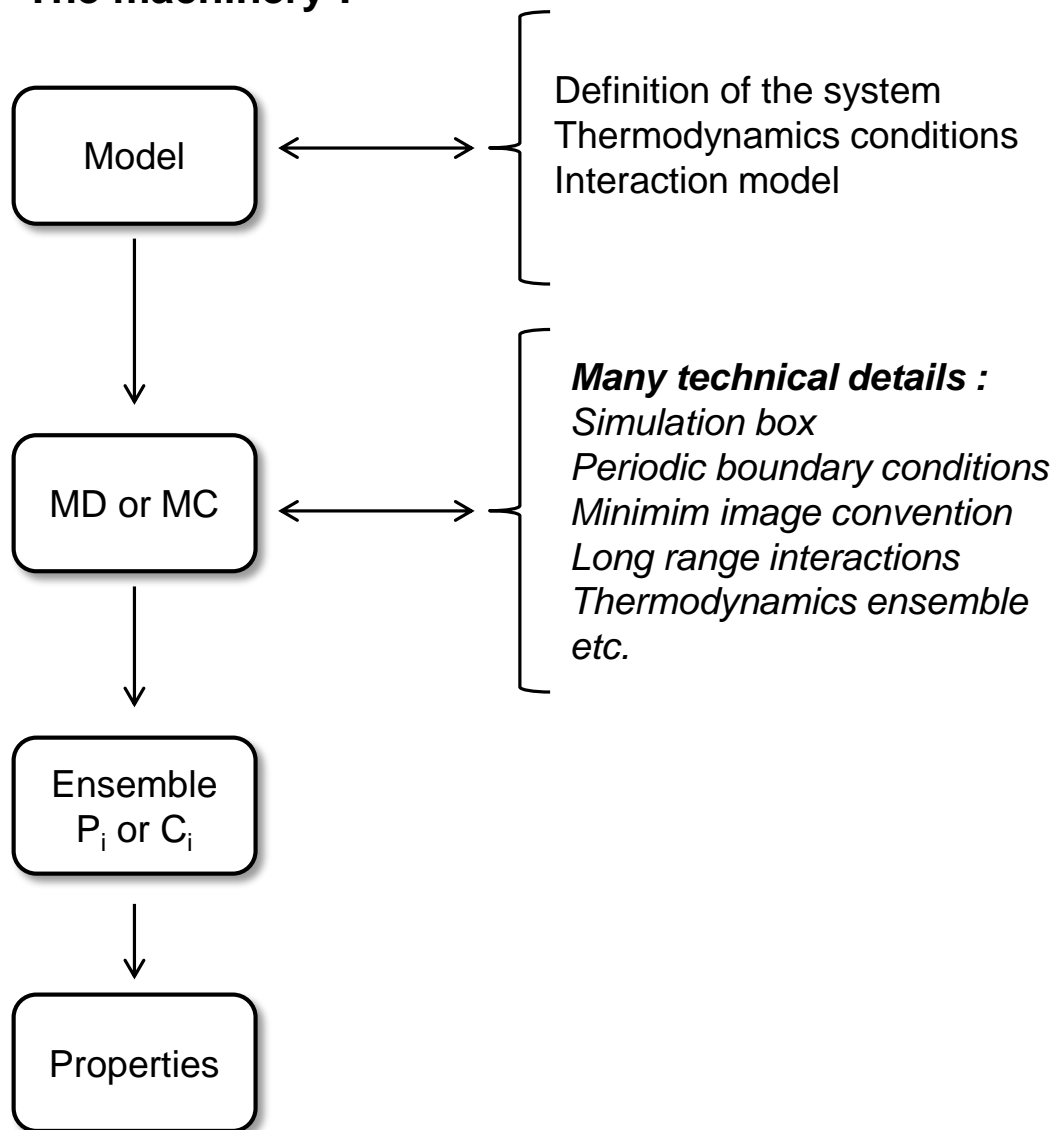
MC generates the Cs in a random order

Consequences

No trajectories, just a package of Cs

Dynamical properties inaccessible

The machinery :





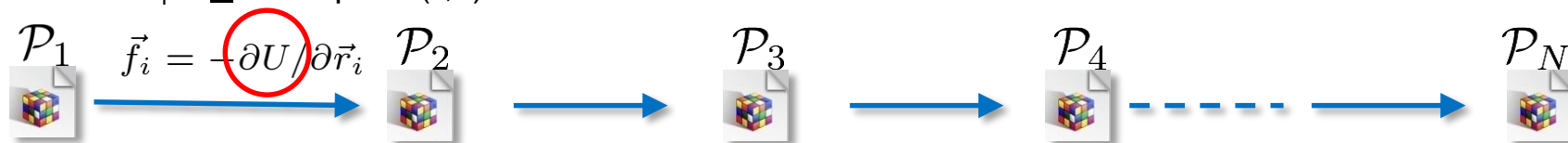
The model is the crucial input of a simulation

« 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.

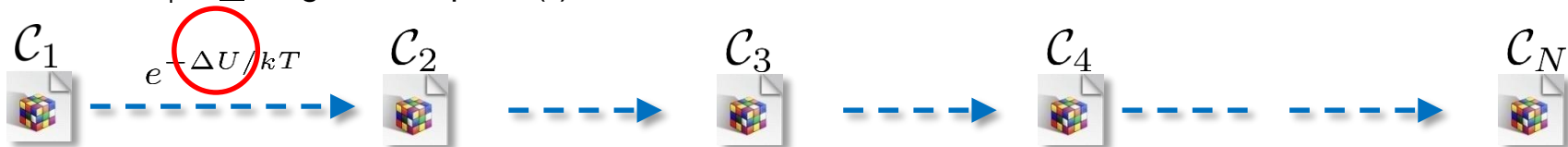
Molecular Dynamics

⇒ list of P_i in Phase space (\mathbf{r}, \mathbf{v}) and ordered in time

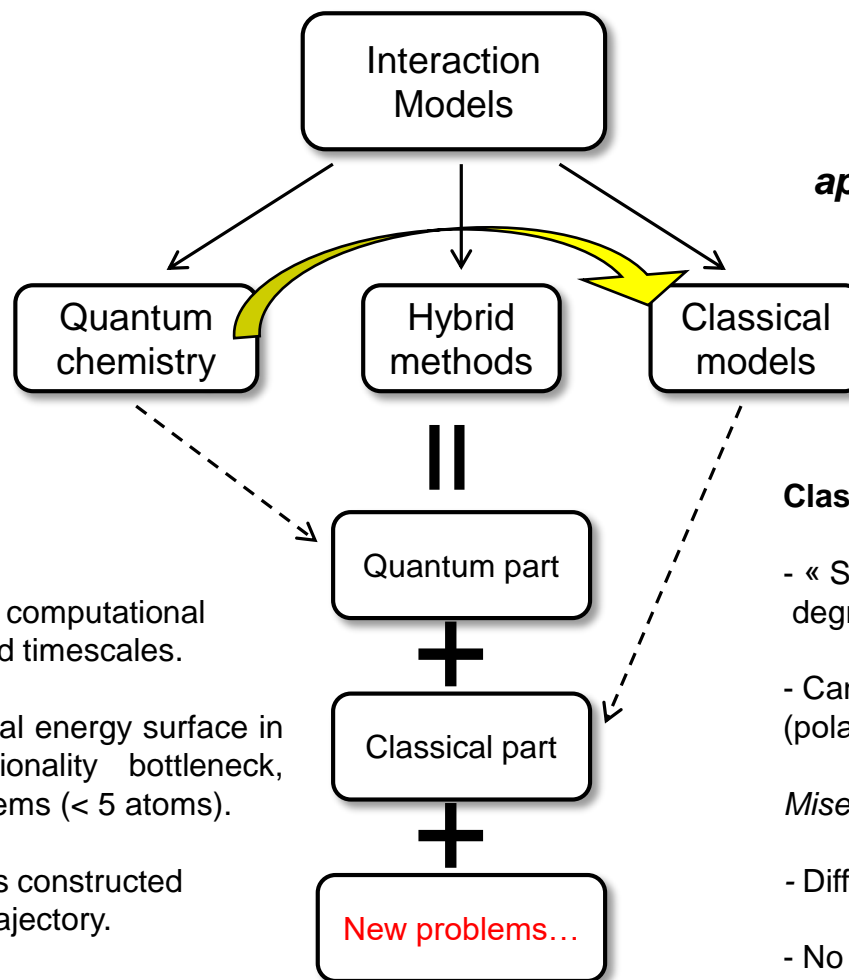


Monte Carlo

⇒ list of C_i in Configuration space (\mathbf{r})



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



Many simplifications and approximations are necessary

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.

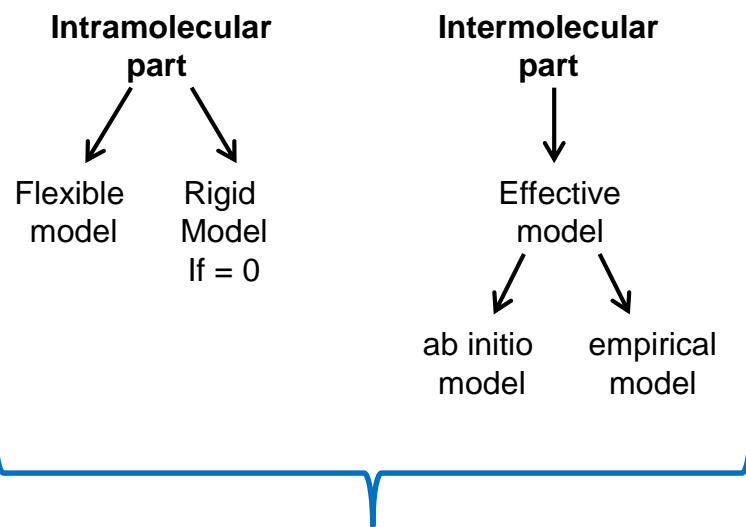
Classical models

A multidimensional function $\hat{V}(\{R_I^{nuclei}\})$ is still too complex.

The pair potential approximation consists to developp this function over monomers and pair of molecules :

Beyond the pair potential approximation ?

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



The pair potential approximation is a very drastic assumption

Three body term

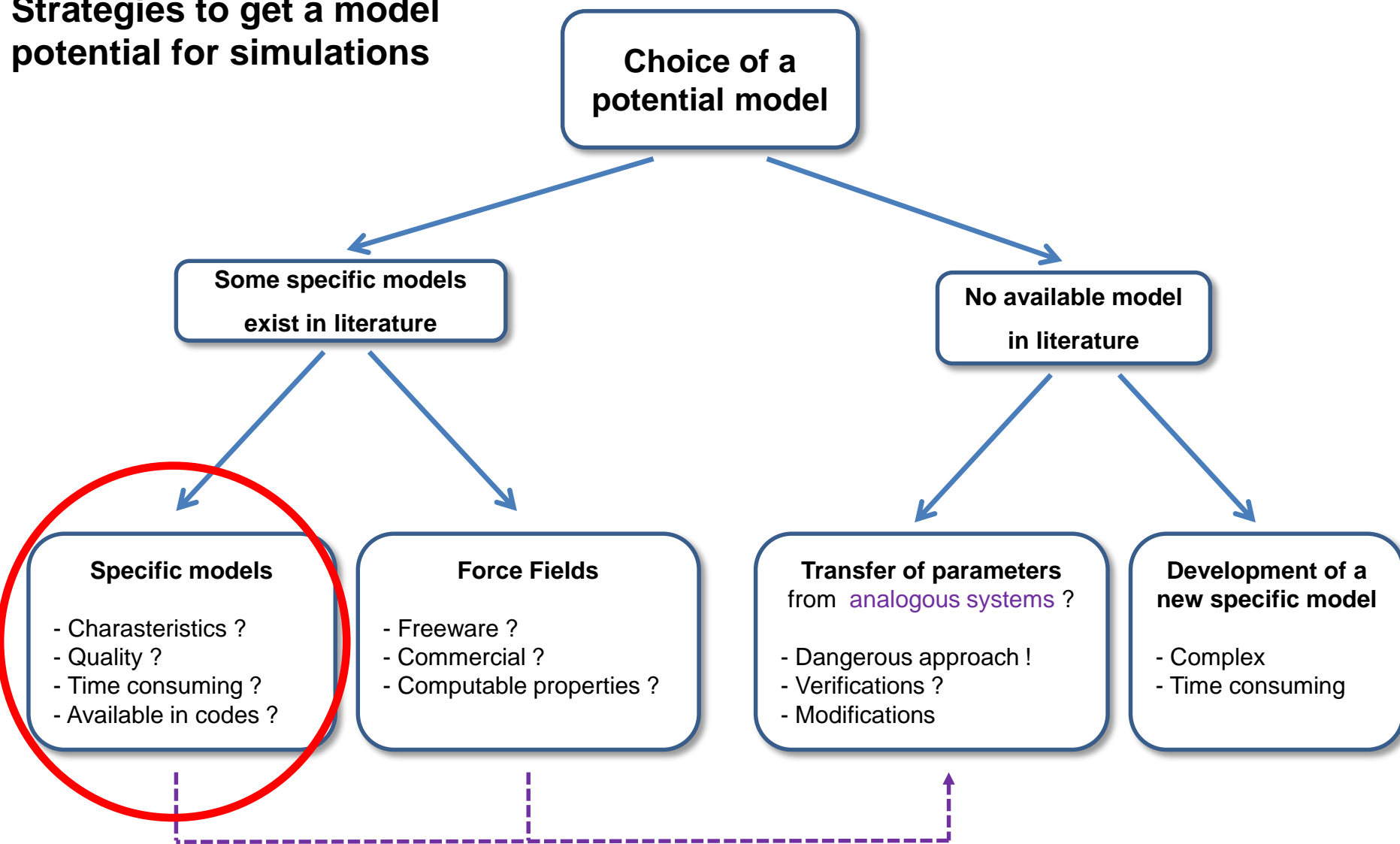
and / or

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

N - body term

Explicit polarizabilities

Strategies to get a model potential for simulations



Example of liquid WATER

Just for water (H₂O), 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

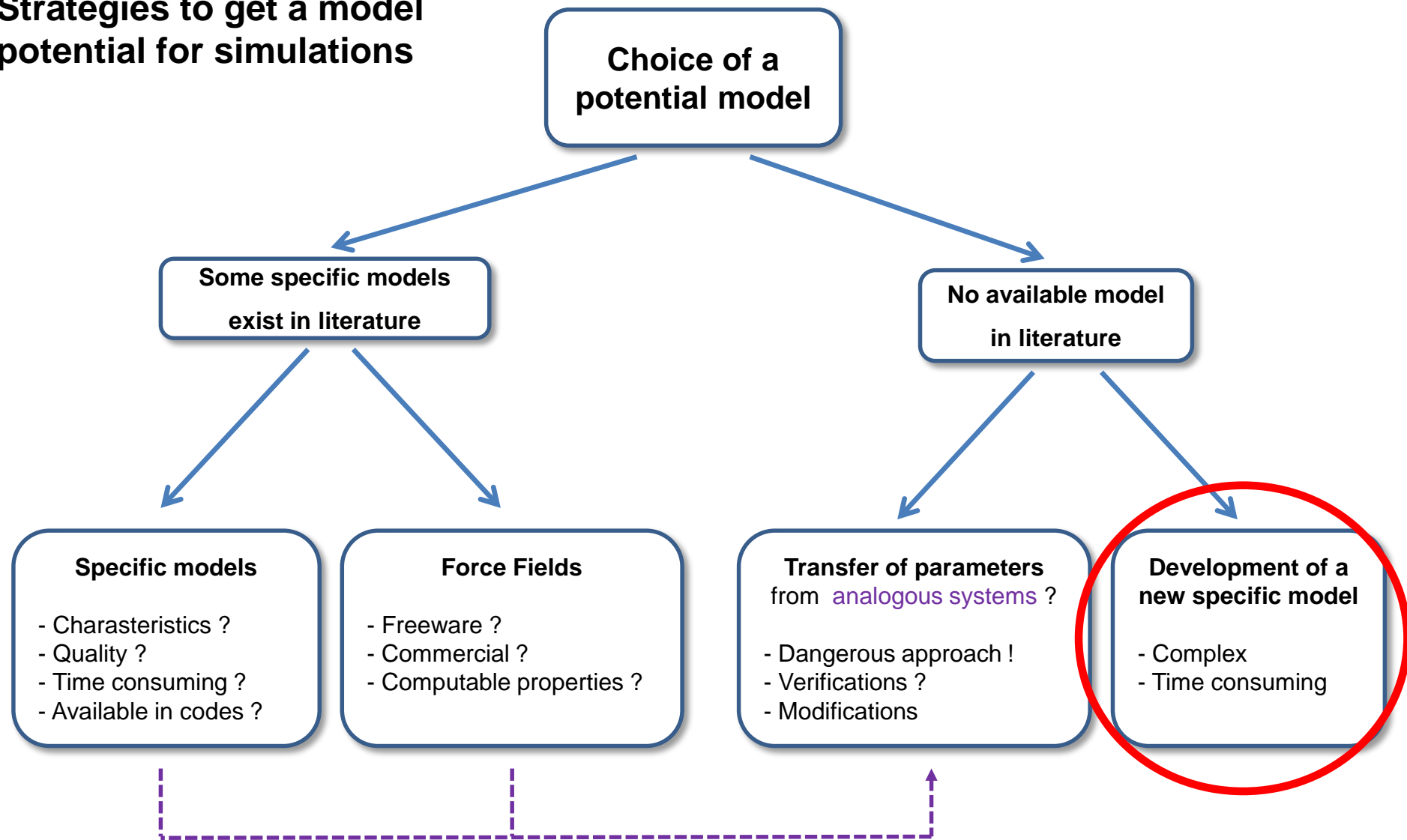
ab initio models : fitted from quantum chemistry

H₂O

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.+rel.	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)

Strategies to get a model potential for simulations



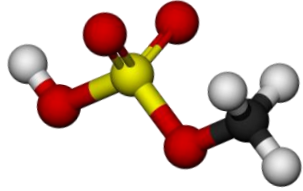
You want to study a system by the use of molecular simulations but :

- *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: *the development of models is a science in itself.*

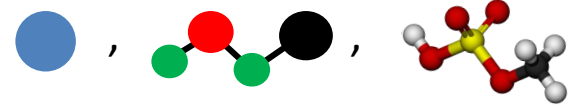


The development of models is a science in itself !

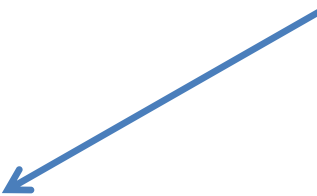
Development of a new potential model



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..."

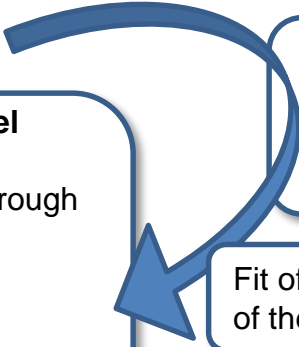


Empirical model
Fit the parameters through simulations (MD or MC) of some experimental properties



Ab initio model
Fit the parameters through QM calculations

- Geometry
- Frequencies
- PES of dimers, trimers...

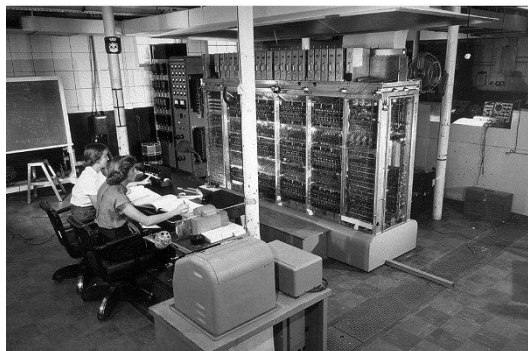


Electrostatic model
-Model derived from electrostatic properties of single molecule.

Fit of the rest (non electrostatic) of the potential energy

Historical perspectives on molecular simulations...

60 years ago : first Monte Carlo simulation (1953)



MANIAC
*Mathematical Analyzer,
Numerical Integrator,
and Computer*

*Mars 1952
Los Alamos*

1 KOPS

THE JOURNAL OF CHEMICAL PHYSICS

VOLUME 21, NUMBER 6

JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

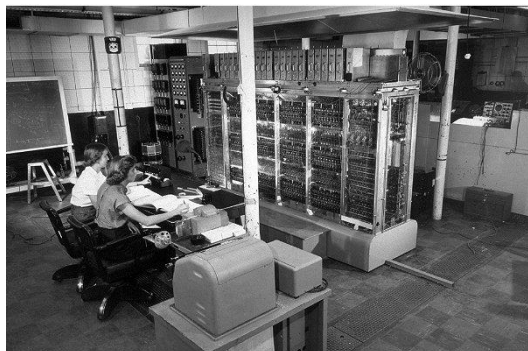
EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Historical perspectives on molecular simulations...

First MC simulation (1953)



MANIAC

Mathematical Analyzer,
Numerical Integrator,
and Computer

Mars 1952
Los Alamos

1 KOPS

If we know the positions of the N particles in the square, we can easily calculate, for example, the potential energy of the system,

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N V(d_{ij}). \quad (1)$$

(Here V is the potential between molecules, and d_{ij} is the minimum distance between particles i and j as defined above.)

In order to calculate the properties of our system we use the canonical ensemble. So, to calculate the equilibrium value of any quantity of interest F ,

$$\bar{F} = \frac{\left[\int F \exp(-E/kT) d^{2N}p d^{2N}q \right]}{\left[\int \exp(-E/kT) d^{2N}p d^{2N}q \right]}, \quad (2)$$

where $(d^{2N}p d^{2N}q)$ is a volume element in the $4N$ -dimensional phase space. Moreover, since forces between

This we do as follows: We place the N particles in any configuration, for example, in a regular lattice. Then we move each of the particles in succession according to the following prescription:

$$\begin{aligned} X &\rightarrow X + \alpha \xi_1 \\ Y &\rightarrow Y + \alpha \xi_2, \end{aligned} \quad (3)$$

where α is the maximum allowed displacement, which for the sake of this argument is arbitrary, and ξ_1 and ξ_2 are random numbers§ between (-1) and 1 . Then, after we move a particle, it is equally likely to be anywhere within a square of side 2α centered about its original position. (In accord with the periodicity assumption, if the indicated move would put the particle outside the square, this only means that it re-enters the square from the opposite side.)

We then calculate the change in energy of the system ΔE , which is caused by the move. If $\Delta E < 0$, i.e., if the move would bring the system to a state of lower energy, we allow the move and put the particle in its new position. If $\Delta E > 0$, we allow the move with probability $\exp(-\Delta E/kT)$; i.e., we take a random number ξ_3 between 0 and 1 , and if $\xi_3 < \exp(-\Delta E/kT)$, we move the particle to its new position. If $\xi_3 > \exp(-\Delta E/kT)$, we return it to its old position. Then, whether the move has been allowed or not, i.e., whether we are in a different configuration or in the original configuration, we consider that we are in a new configuration for the purpose of taking our averages. So

$$\bar{F} = (1/M) \sum_{j=1}^M F_j, \quad (4)$$

where F_j is the value of the property F of the system

Historical perspectives on molecular simulations...

First Molecular Dynamics simulation (1959)



IBM-704

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 31, NUMBER 2 AUGUST, 1959

Studies in Molecular Dynamics. I. General Method*

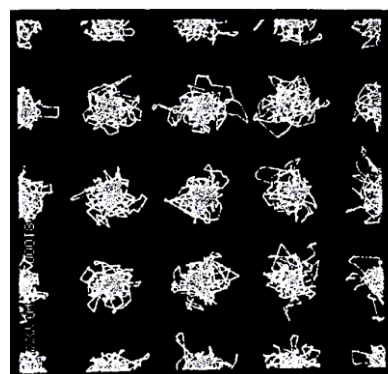
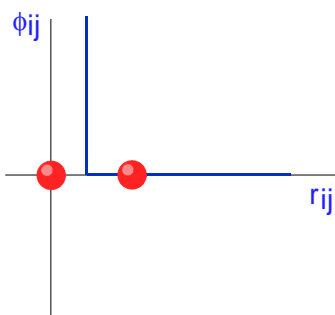
B. J. ALDER AND T. E. WAINWRIGHT

Lawrence Radiation Laboratory, University of California, Livermore, California

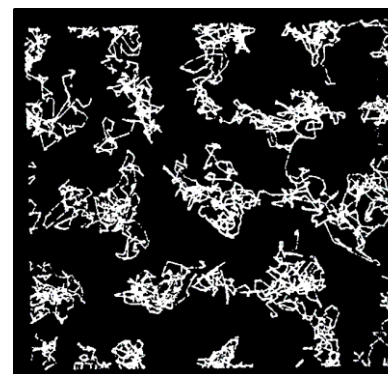
(Received February 19, 1959)

A method is outlined by which it is possible to calculate exactly the behavior of several hundred interacting classical particles. The study of this many-body problem is carried out by an electronic computer which solves numerically the simultaneous equations of motion. The limitations of this numerical scheme are enumerated and the important steps in making the program efficient on the computers are indicated. The applicability of this method to the solution of many problems in both equilibrium and nonequilibrium statistical mechanics is discussed.

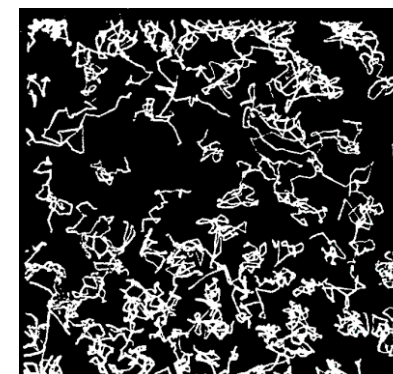
Hard disks and spheres



solid phase



liquid phase



liquid-vapour-phase

32 particules 7000 collisions / hcpu
500 particules 500 collisions / hcpu

Production run ~20000 steps

{	N=32 → 6.5 x 10 ⁵ coll. → 4 days
	N=500 → 10 ⁷ coll. → 2.3 years

Historical perspectives on molecular simulations...

First MD simulation with continuous potential (1964)



CDC 3600

PHYSICAL REVIEW

VOLUME 136, NUMBER 2A

19 OCTOBER 1964

Correlations in the Motion of Atoms in Liquid Argon*

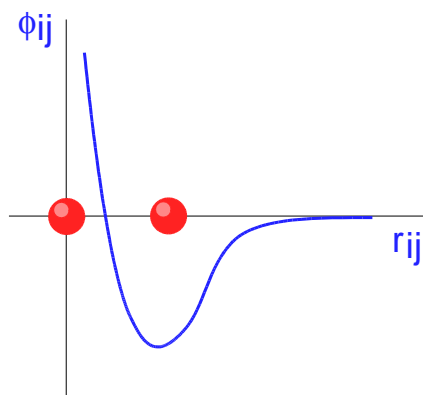
A. RAHMAN

Argonne National Laboratory, Argonne, Illinois

(Received 6 May 1964)

A system of 864 particles interacting with a Lennard-Jones potential and obeying classical equations of motion has been studied on a digital computer (CDC 3600) to simulate molecular dynamics in liquid argon at 94.4°K and a density of 1.374 g cm⁻³. The pair-correlation function and the constant of self-diffusion are found to agree well with experiment; the latter is 15% lower than the experimental value. The spectrum of the velocity autocorrelation function shows a broad maximum in the frequency region $\omega = 0.25 (k_B T / \hbar)$. The shape of the Van Hove function $G_s(r, t)$ attains a maximum departure from a Gaussian at about $t = 3.0 \times 10^{-12}$ sec and becomes a Gaussian again at about 10^{-11} sec. The Van Hove function $G_d(r, t)$ has been compared with the convolution approximation of Vineyard, showing that this approximation gives a too rapid decay of $G_d(r, t)$ with time. A delayed-convolution approximation has been suggested which gives a better fit with $G_d(r, t)$; this delayed convolution makes $G_d(r, t)$ decay as t^4 at short times and as t at long times.

Lennard-Jones potential



864 particles
Time step ~ 10 fs

Thermodynamics

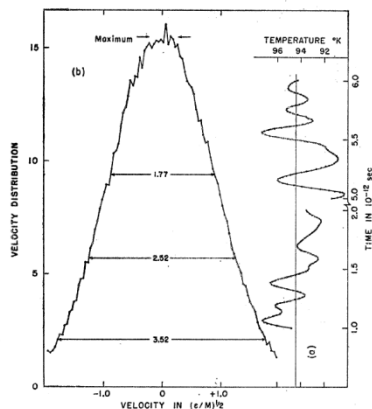


FIG. 1. Fluctuations of temperature with time in two sample regions (curve a); distribution of velocities is shown as curve b; widths of the distribution are shown at $e^{-1/2}$, e^{-1} , and e^{-2} of maximum.

Structure

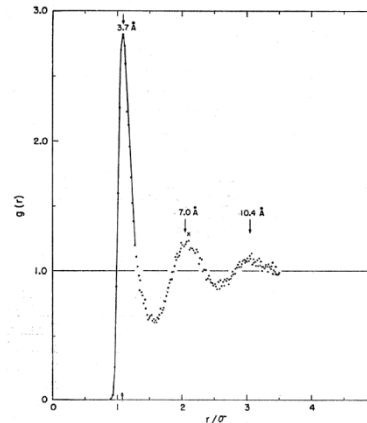


FIG. 2. Pair-correlation function obtained in this calculation at 94.4°K and 1.374 g cm⁻³. The Fourier transform of this function has peaks at $k\sigma = 6.8, 12.5, 18.5, 24.8$.

Diffusion

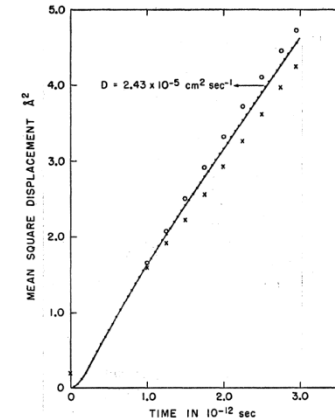
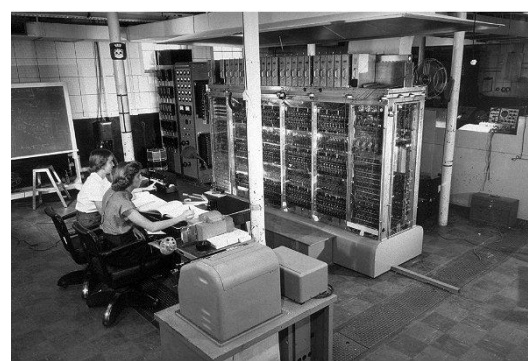


FIG. 3. Mean-square displacement of particles. The continuous curve is the mean of a set of 64 curves; the two members of the set which have maximum departures from the mean are shown as circles and as crosses. The asymptotic form of the continuous curve is $6Dt + C$, with D as shown on the figure and $C = 0.2 \text{ Å}^2$.

~20000 MD steps, 200 ps (45 s / step) \cong 10 days
(standard PC today : few minutes...)

I. Overview of Statistical Mechanics & Molecular Simulations

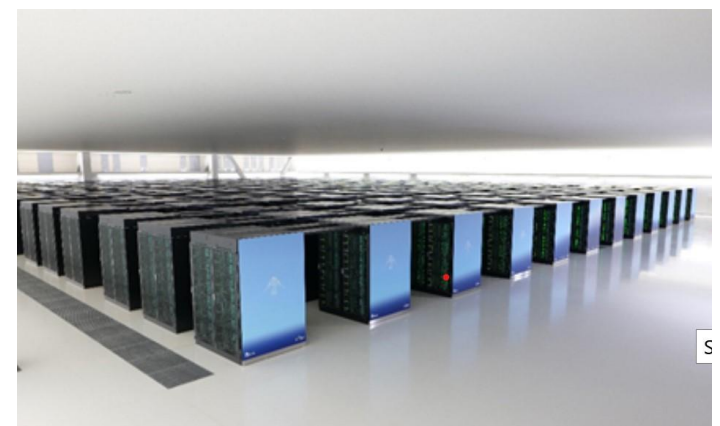


Past :

MANIAC
*Mathematical Analyzer,
 Numerical Integrator,
 and Computer*

Mars 1952
 Los Alamos

1000 OPS

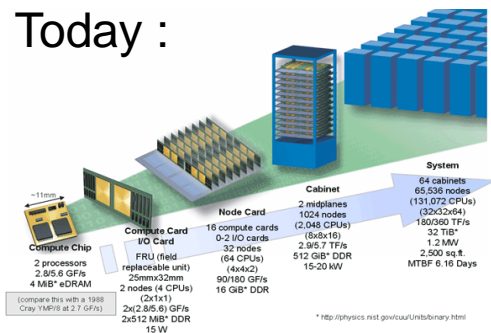


Supercomputer Fugaku (Japan)

Cores : 7,630,848
 Memory : 5,087,232 GB
 537,212.0 Tflop/s = **5,37 10^{17} Flop/s**

See: <https://www.top500.org/lists/top500/>

Today :



MCIA – Bordeaux (2015)



Electrolyte Effects on Protein Stability in Solution

Eckhard Spohr, Faculty for Chemistry, University Duisburg-Essen, Germany

S. Brudar, B. Hribar Lee, University of Ljubljana, Slovenia

Understanding Solvation in Molecular and Ionic Fluids:
Towards a Sustainable Future

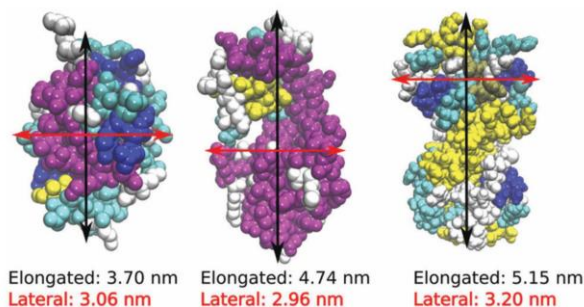
EMLG-JMLG Congress, September 4-7, 2023, Bordeaux, France



Introduction

- HEWL, T4WT, γ -D crystallin are small, almost globular proteins
- 100-200 amino acids
- Solubility measurements available (SB, BH)
- Solutions can undergo liquid-liquid phase separation
- γ -D crystallin involved in cataract formation
- Phase diagrams of HEWL, γ -D crystallin are known

HEWL = Hen egg white lysozyme

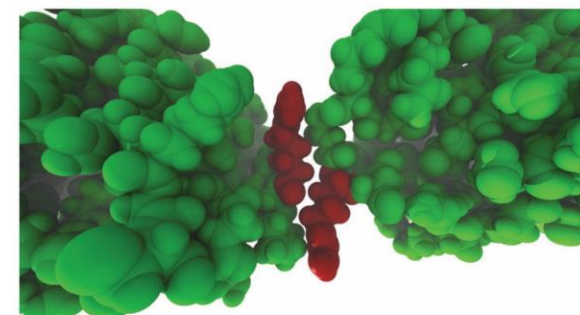
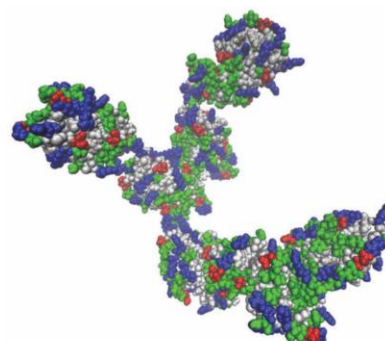


*Reproduced with kind permission
of Prof. Dr. Eckhard Spohr*

Pure Water (OPLS): Summary of Simulations

Protein	γ (mg mL ⁻¹)	T (K)	N_{water}	$N_{\text{proteins+ions}}$	L (nm)
HEWL	93	267	124 615	16 + 128	16
HEWL	93	300	124 615	16 + 128	16
HEWL	42	267	142 004	8 + 64	16.5
HEWL	42	300	142 004	8 + 64	16.5
T4 WT*	90	267	124 850	12 + 96	16
T4 WT*	90	300	124 850	12 + 96	16
T4 WT*	45	267	129 713	6 + 48	16
T4 WT*	45	300	129 713	6 + 48	16
γ -D crystallin	100	300	123 966	12 + 0	16
γ -D crystallin	100	320	123 966	12 + 0	16

Arginine (OPLS): Probability of AA Pair to Initialise Contact



- Experiments show that the addition of arginine increases the stability of the HEWL solutions.
- Computer simulations in pure water indicate that arginine residues are instrumental for self association of proteins

Arginine Solution (OPLS)

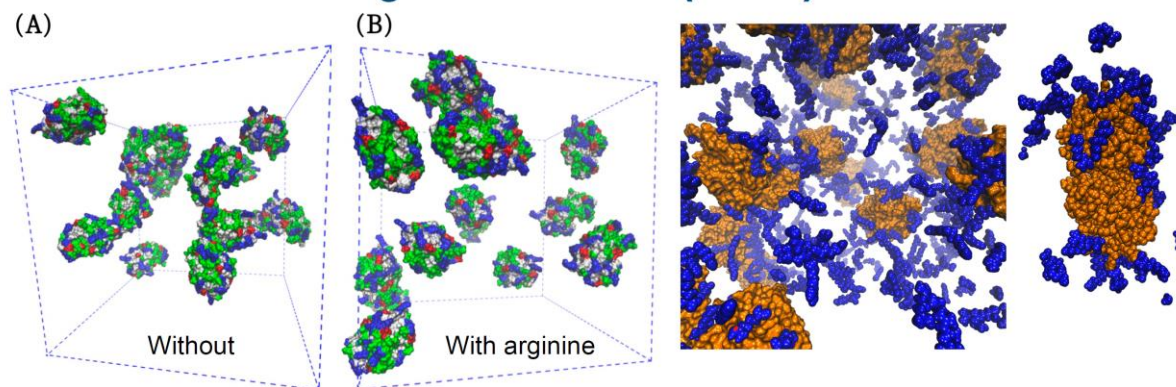


Figure 4. Snapshots from the computer simulations showing just HEWL molecules (A) without the presence of free arginine and (B) with added 0.3 M arginine.

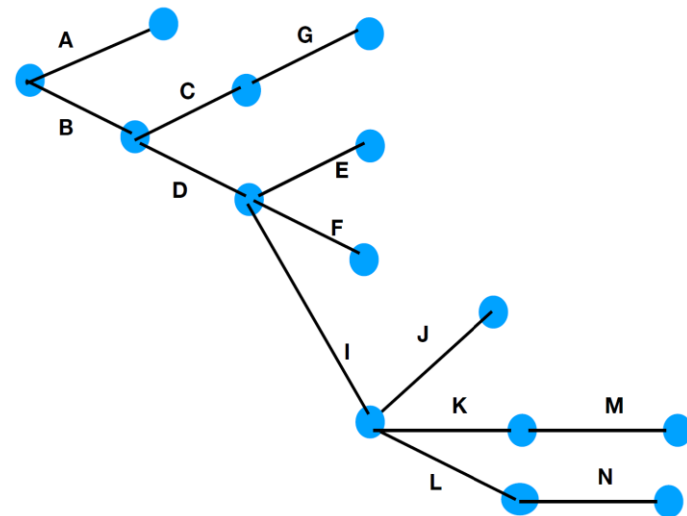
- Experiments show that the addition of arginine increases the stability of the HEWL solutions.
- The computer simulation results indicate that arginine molecules tend to self-associate
- If arginine residues are located on the protein surface, the free arginine molecules stay in their vicinity and prevent protein molecules from “connecting” through them to form clusters.

Reproduced with kind permission
of Prof. Dr. Eckhard Spohr

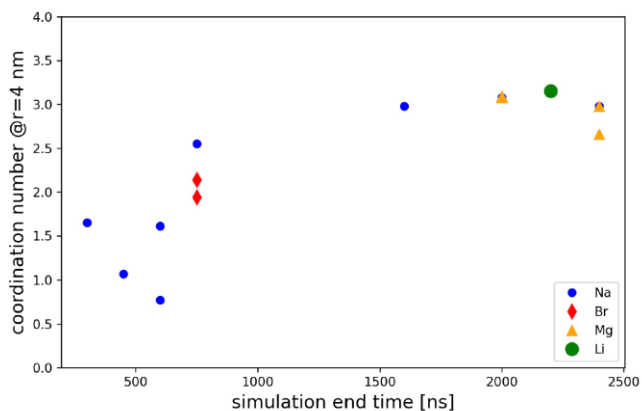
Electrolyte Solutions

T(ns)	N(4nm)	#	System	FF
450	1.07	A	0.5m NaCl 267K	OPLS
300	1.65	B	0.5m NaCl 300 K	OPLS
300	0.77	C	0.25m NaCl 267K	OPLS
300	1.61	D	0.25m NaCl 300K	OPLS
150	2.14	E	0.5 m NaBr 300K	OPLS
150	1.94	F	0.25m NaBr 300K	OPLS
150	2.55	G	0.25m* NaCl 300K	AMBER
1000	2.98	I	0.25m*NaCl 300K	AMBER
600	3.15	J	0.5m LiCl 300K	AMBER
400	3.09	K	0.25m MgCl ₂ 300K	AMBER
400	3.08	L	0.125m MgCl ₂ + 0.25m NaCl 300K	AMBER
300	2.66	M	0.125m MgCl ₂	AMBER
300	2.98	N	0.063m MgCl ₂ + 0.125m NaCl 300K	AMBER

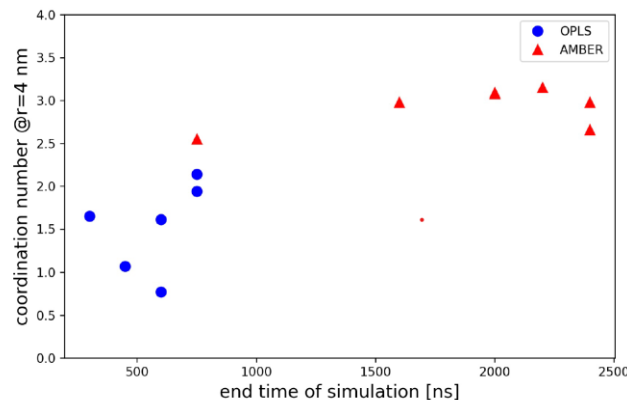
- no dynamic equilibrium (disconnection of protein pairs)
- coordination number at 4 nm (1st peak) taken as descriptor
- no convergence of RDF after 1 μ s!
- bromide ion leads to (slightly) stronger association
- LiCl shows stronger association than NaCl
- MgCl₂ increases association for all studied concentrations,
- NaCl/MgCl₂ is in between



Ion Dependence



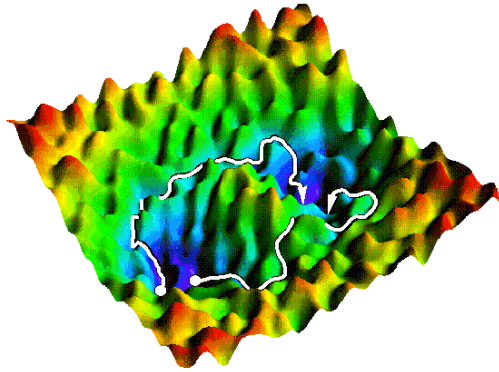
Dependence on Force Field



Nonconvergence issues after more than 2 μ s of simulation (18 months of calculation on *supercomputer magnitude*)



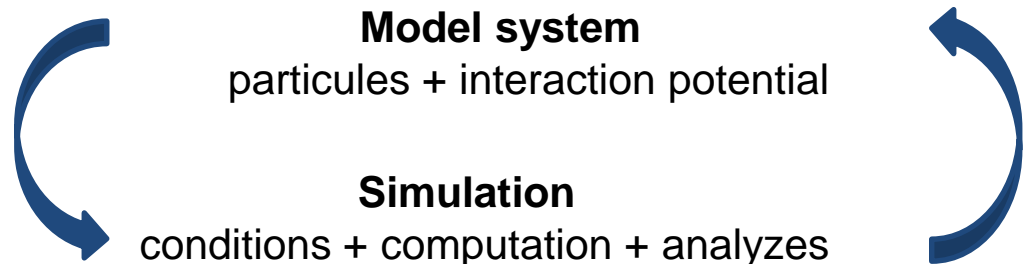
Reproduced with kind permission of Prof. Dr. Eckhard Spohr



Simulations...?

- Real condensed phase : phases space \rightarrow infinity
- It is impossible to know exhaustively the PES (Potential Energy Surface)
- Solution = exploration of the « interesting part » of the PES during the simulation
- By the way... collection and average over the generated states : $\langle \mathcal{A}(\mathcal{P}_i) \rangle = \frac{1}{N} \sum_{i=1}^N \mathcal{A}(\mathcal{P}_i)$

In practical



« **Statistical Mechanics & Simulations** »

I. Overview of Statistical Mechanics & Molecular Simulations

II. Molecular Dynamics Simulations

III. Monte Carlo methods

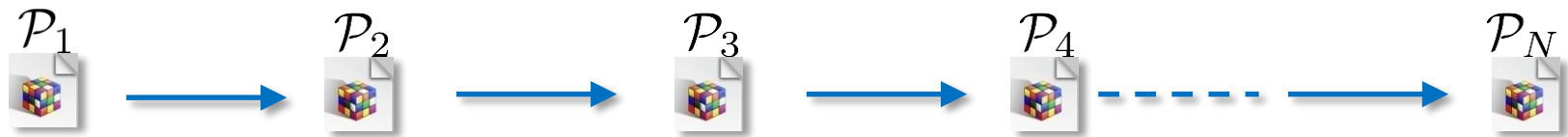
IV. « Outputs » : extracting properties from simulations

V. Initiation to statistical thermodynamics

Molecular Dynamics

The goal

Making the sample : model the motion of some group of particles



The result is a trajectory : collection of phase point $\mathcal{P}(t)$ which contain positions and velocities of all particles of the system.

The method

- Let us assume that we know one phase space point \mathcal{P} compatible with the macrostate we are interested in.
- Let us assume that our particles are heavy enough and/or the temperature high enough, so that the particles De Broglie wavelength is small compared with the distances between particles.
- We can replace the time dependant Schrodinger equation for the particle motions by Newton's equation.
- We know the Hamiltonian for the whole things, it contains the kinetic energy and a potential energy based on a chosen model.

- Let us consider a system of N particles restricted in a volume V
i.e. $P(t_0)$ which correspond to some macrostate

Question : can we characterize this macrostate ?

Answer : yes we can !

we know how to compute the total energy E of $P(t_0)$: $E = E^{total} = T + V$

$P(t_0)$ contains all velocities so we can compute the kinetic energy T

$P(t_0)$ contains all positions so we can compute the potential energy V

$P(t_0)$ is thus characterized by the values N, V and E

The N Newton's equations for all particles in the configuration $P(t_0)$ are :

$$m_i \cdot \ddot{\vec{R}}_i = -\text{grad}_i V = \vec{F}_i \quad i = 1, 2 \dots N$$

If we can solve this N 2d order differential equations, then we could get after some time δt a new configuration $P(t_0 + \delta t)$.

So we have now :

$P(t_0)$ characterized by
N
V
E

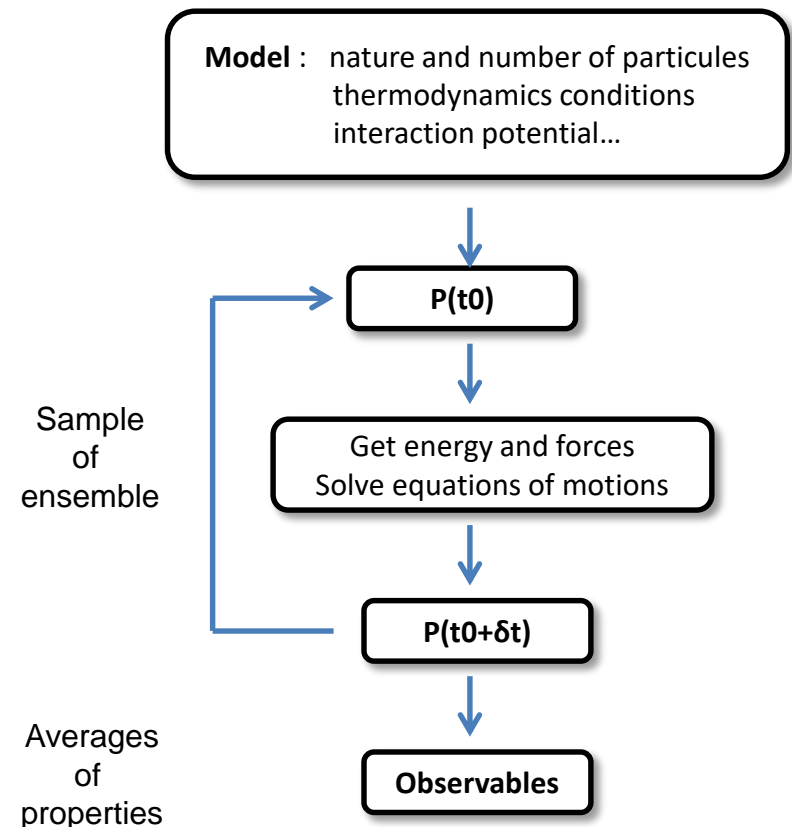
$P(t_0 + \delta t)$ characterized by
N, because Newton's equation do not change the particle numbers
V, because the particles are restricted to this volume
E, because the total energy is constant for an isolated system

Thus :

If **$P(t_0)$** is a microstate belonging to a certain macrostate

$P(t_0 + \delta t)$ is also a microstate belonging to the same macrostate

Work flow of an MD simulation



Model : nature and number of particles
thermodynamics conditions
interaction potential...

$P(t_0)$

Get energy and forces
Solve equations of motions

$P(t_0 + \delta t)$

Observables

Many questions arise :

(a) How to deal with volume restriction

(b) How to deal with the integration of Newton's equation

(c) How to choose the time step δt

(d) How to find $P(t_0)$

(e) How to control a thermodynamic ensemble

etc.

II. Molecular Dynamics Simulations

Question (a) : how to make sure that the particules are always restricted to the volume

Answer : use of Periodic Boundary Conditions (PBC)
combined with the Minimum Distance Convention (MDC)

Some general comments :

Our spatial scales are limited !

Solutions :

No boundaries

ex. droplet of liquid, molecules in vacuum

Fixed boundaries

*make use of a sacrificial region
not really physical, not really efficient*

Periodic boundaries

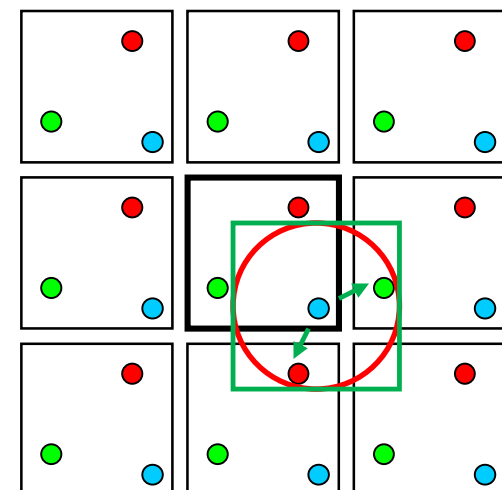
This is the most widely used solution

Basic box

Periodic boxes = mirror of basic box

Minimum Distance Convention

Cut-off sphere



II. Molecular Dynamics Simulations

Question (a) : how to make sure that the particles are always restricted to the volume

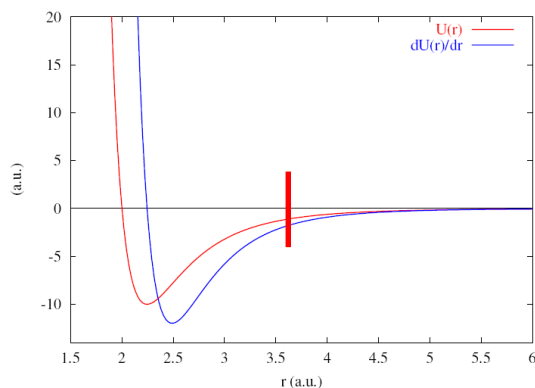
Answer : use of Periodic Boundary Conditions (PBC)
combined with the Minimum Distance Convention (MDC)

It allows to simulate an « infinite » system (or « continuous » System) with a finite number of particles.

There is no surface at the limit of the basic box.

Particles crossing a boundary of the simulation box emerge back from the opposite side.
There will always be N particles in the basic box.

Cut-off : no direct interactions beyond a certain distance



Unphysical.
Discontinuities.
Artefacts...

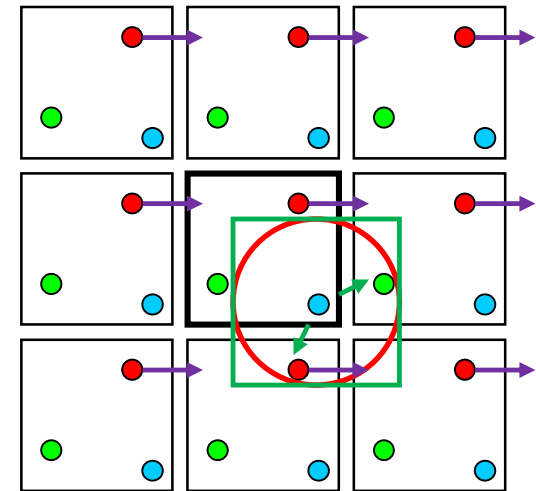
Some methods exists to avoid this problem.

Basic box

Periodic boxes = mirror of basic box

Minimum Distance Convention

Cut-off sphere



Question (b): how to deal with the integration of Newton's equation

Recall :

$$m_i \cdot \ddot{\vec{R}}_i = -\text{grad}_i V = \vec{F}_i \quad i = 1, 2 \dots N$$

But these N 2d order differential equations are coupled :

$$m_i \cdot \ddot{\vec{R}}_i = -\text{grad}_i V(\vec{R}_1, \vec{R}_2, \vec{R}_3, \dots, \vec{R}_N)$$

There is no analytic solution for this problem.

Answer : do a numerical solution, i.e. a numerical integration.

Example of the Verlet algorithm :

$$\vec{R}_i(t + \delta t) = \vec{R}_i(t) + \vec{V}_i(t)\delta t + \frac{1}{2}\delta t^2\gamma_i$$

$$\vec{R}_i(t - \delta t) = \vec{R}_i(t) - \vec{V}_i(t)\delta t + \frac{1}{2}\delta t^2\gamma_i$$

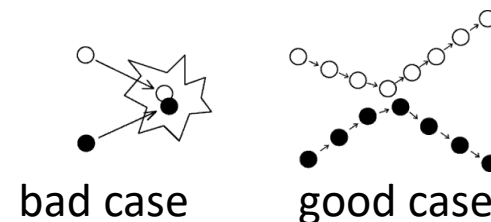
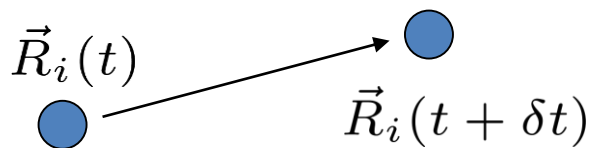
$$\vec{R}_i(t + \delta t) = 2\vec{R}_i(t) - \vec{R}_i(t - \delta t) + \delta t^2\gamma_i$$

$$\vec{V}_i(t) = \frac{1}{2\delta t}(\vec{R}_i(t + \delta t) - \vec{R}_i(t - \delta t))$$

Question (c) : how to chose the time step δt

Inside the Verlet algorithm, we see : $\vec{R}_i(t + \delta t) = 2\vec{R}_i(t) - \vec{R}_i(t - \delta t) + \delta t^2 \gamma_i$

Such a numerical integration consider that the forces acting on particules are constant during the move :



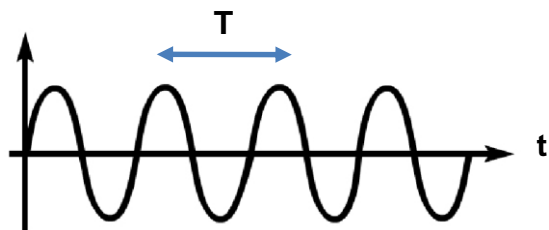
Answer : this condition is only valid for a small step in space with respect to the intermolecular distances.

What about the step in time : δt ?

It depends on the time scale of the motions in the system of interest !

\Rightarrow mass, temperature, interaction model...

Application : determination of the time step δt in a system containing vibrating OH group



$$\bar{\nu}_{OH} = 4000 \text{ cm}^{-1}$$

$$T = ???$$

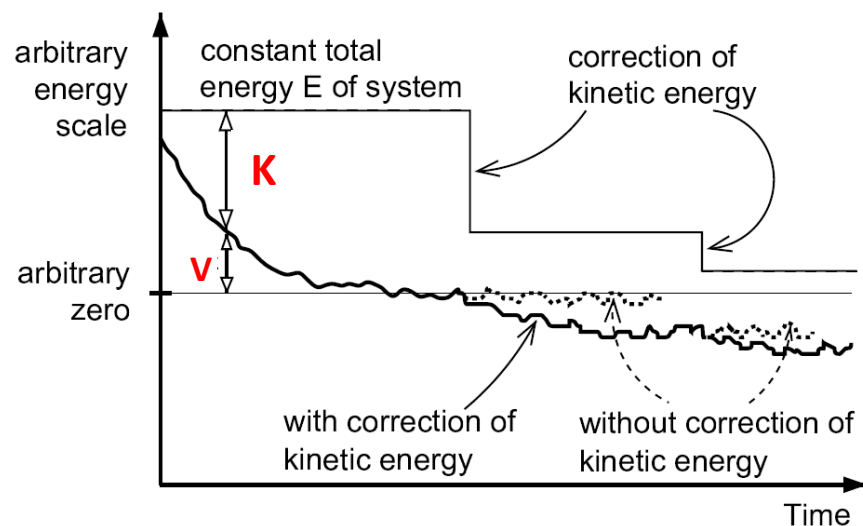
$$\delta t = T / 30 \approx ???$$

Question (d): how to find $P(t_0)$?

- our macrostates are characterized by N, V, E being constant.
- however, we usually do not know E for the system we want to study.

Answer: this figure shows how to proceed to find a $P(t_0)$

$$E = E^{total} = K + V$$



- 1) We "guess" a preliminary $\mathcal{P}(t_0)$ (educated guesses preferred!).
- 2) We start the simulation (numeric integration).
- 3) The total energy E will be constant if we integrate correctly.
- 4) However, almost always, our guessed $\mathcal{P}(t_0)$ will have too much potential energy (interactions) compared to its kinetic energy.
- 5) The kinetic energy (\propto temperature) will thus increase, the potential energy decrease, first quickly, then more and more slowly.
- 6) If our system is *well-behaved*, it will approach a *steady state*. (only fluctuations between kinetic and potential energy, no drift.)
- 7) **Trouble!** The MACROSTATE that we have is not at all the one we wanted to study (usually much too much kinetic energy!).
- 8) We now **remove kinetic energy** from the system (which is easy: scale the velocities) and use the \mathcal{P} that we have now reached as a new, **better guess for $\mathcal{P}(t_0)$** .
- 9) We restart the simulation from the new $\mathcal{P}(t_0)$ (GOTO 2) until this $\mathcal{P}(t_0)$ is really the one that we want
In this case, we forget this whole "equilibration procedure" and start the **real simulation**.

Question (e): How to control a thermodynamic ensemble

Thermodynamic ensemble

Microcanonical: NVE

$$\mathcal{W}^{NVE}(\mathbf{r}, \mathbf{p}) = \delta(\mathcal{H}(\mathbf{r}, \mathbf{p}) - E)$$

Canonical NVT

$$\mathcal{W}^{NVT}(\mathbf{r}, \mathbf{p}) = e^{-\mathcal{H}(\mathbf{r}, \mathbf{p})/k_B T}$$

Berendsen, *J. Chem. Phys.* **81**, 3684 (1984)

Nosé-Hoover, *Phys. Rev. A* **31**, 1695 (1985)

$$\ddot{\mathbf{r}}_i(t) = \frac{\mathbf{f}_i(t)}{m_i} - \xi(t)\dot{\mathbf{r}}_i(t)$$

$$\dot{\xi}(t) = \nu_T^2 \left(\frac{2E_c}{f k_B T_{ext}} - 1 \right)$$

Isotherm-isobar : NPT

$$\mathcal{W}^{NPT}(\mathbf{r}, \mathbf{p}) = e^{-(\mathcal{H}(\mathbf{r}, \mathbf{p}) + PV)/k_B T}$$

Nosé-Andersen, *Mol. Phys.* **52**, 255(1984)

$$L = \sum_i \frac{m_i}{2} s^2 V^{2/3} \dot{\mathbf{x}}_i^2 - E(V^{1/3} \mathbf{x}) + \frac{Q}{2} \dot{s}^2 + \frac{W}{2} \dot{V}^2 - (f+1)k_B T_{ext} \ln(s) - P_{ext} V$$

$$\ddot{\mathbf{r}}_i(t) = \frac{\mathbf{f}_i(t)}{m_i} - \frac{\mathbf{p}_i(t)}{m_i} \left(\frac{\dot{s}}{s} \right) + \frac{\mathbf{r}_i(t)}{3V} \left(\ddot{V} - \frac{2}{3} \frac{\dot{V}^2}{V} \right)$$

$$\ddot{s} = s \nu_T^2 \left(\frac{T(t)}{T_{ext}} - 1 \right) + \frac{\dot{s}^2}{s}$$

$$\ddot{V} = s^2 \left(\frac{P(t) - P_{ext}}{W} \right) + \frac{\dot{s}}{s} \dot{V}$$

$$\mathbf{p}_i(t) = m_i \left(\dot{\mathbf{r}}_i(t) - \frac{\dot{V}}{3V} \mathbf{r}_i(t) \right)$$

« ***Statistical Mechanics & Simulations*** »

- I. Overview of Statistical Mechanics & Molecular Simulations
- II. Molecular Dynamics Simulations
- III. Monte Carlo methods**
- IV. « Outputs » : extracting properties from simulations
- V. Initiation to statistical thermodynamics

Monte Carlo

The goal

Making the sample : create a collection of configurations of some group of particules



The result is a randomly ordered collection of configurations, which contain only positions of all particules of the system.

The method

- We put in the computer N particules in a volume V.
- Let us assume that we know one configuration compatible with the macrostate we are interested in.
- Let us assume that we know a model.
- From the interactions, we can compute the total energy of the system, and thus the energy difference between two configurations (old and new).
- The Metropolis algorithm tell us if the 'new' configuration is acceptable.

First MC simulation (1953)

Some details...

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER, Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER, * Department of Physics, University of Chicago, Chicago, Illinois

(Received March 6, 1953)

This we do as follows: We place the N particles in any configuration, for example, in a regular lattice. Then we move each of the particles in succession according to the following prescription:

X -> X + alpha xi_1
Y -> Y + alpha xi_2 (3)

where alpha is the maximum allowed displacement, which for the sake of this argument is arbitrary, and xi_1 and xi_2 are random numbers between (-1) and 1. Then, after we move a particle, it is equally likely to be anywhere within a square of side 2alpha centered about its original position. (In accord with the periodicity assumption, if the indicated move would put the particle outside the square, this only means that it re-enters the square from the opposite side.)

We then calculate the change in energy of the system delta E, which is caused by the move. If delta E < 0, i.e., if the move would bring the system to a state of lower energy, we allow the move and put the particle in its new position. If delta E > 0, we allow the move with probability exp(-delta E/kT); i.e., we take a random number xi_3 between 0 and 1, and if xi_3 < exp(-delta E/kT), we move the particle to its new position. If xi_3 > exp(-delta E/kT), we return it to its old position. Then, whether the move has been allowed or not, i.e., whether we are in a different configuration or in the original configuration, we consider that we are in a new configuration for the purpose of taking our averages. So

F_bar = (1/M) sum_{j=1}^M F_j (4)

where F_j is the value of the property F of the system

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

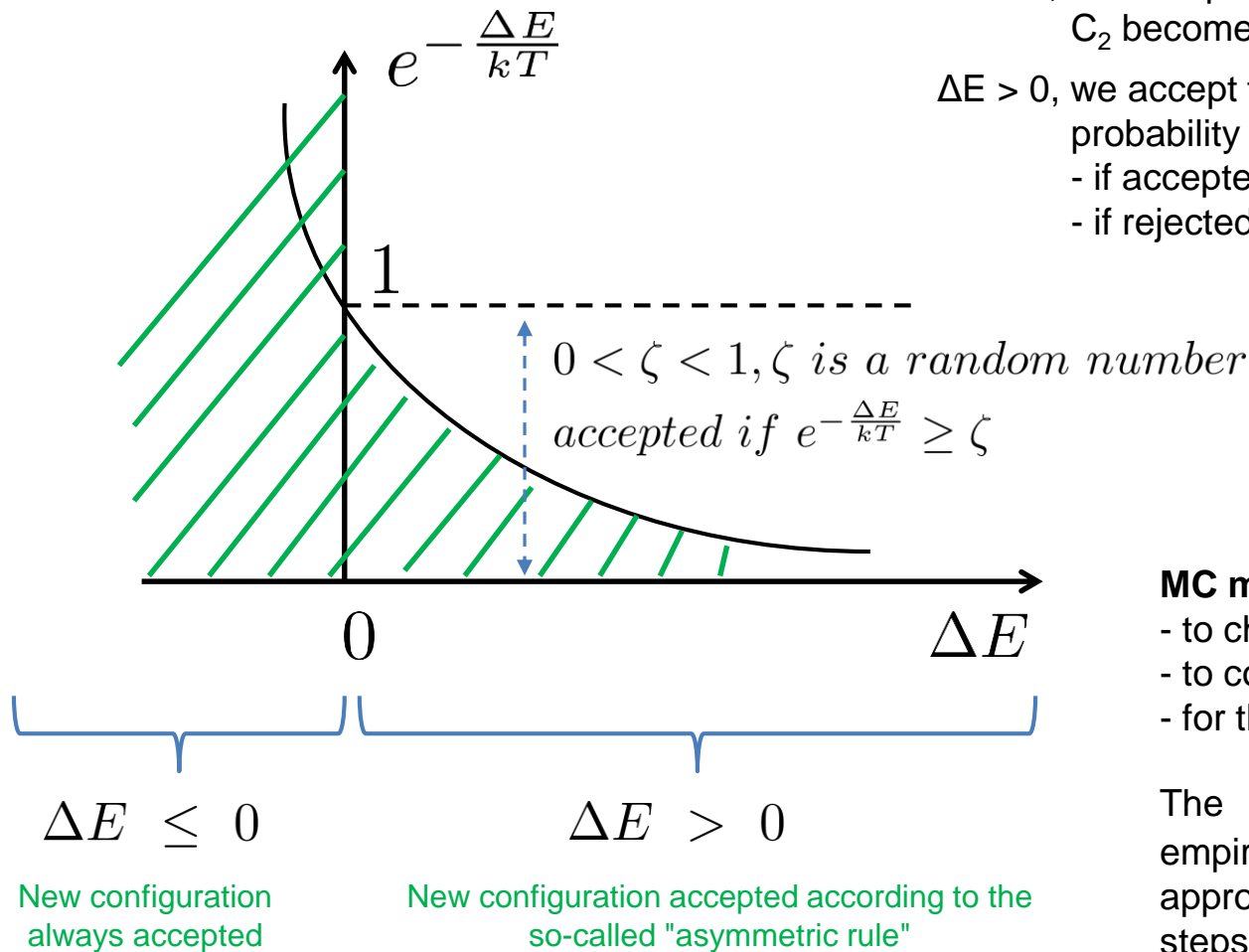
- 1) We know one configuration, call C_1, its energy E_1.
2) We create randomly a new configuration, call C_2.
3) We compute the energy E_2 of C_2 and delta E = E_2 - E_1.
4) We test the sign of delta E
delta E <= 0, we accept the new configuration: C_2 become C_1, go back to 2)
delta E > 0, we accept the new configuration according to a probability based on Boltzmann factor:
- if accepted, C_2 become C_1, go back to step 2)
- if rejected, C_2 = C_1, go back to step 2)

This kind of 'random walk' through the Cs (Markov chain) generate a representation of the canonic ensemble.

- 1) We know one configuration, call C_1 , its energy E_1 .
- 2) We create randomly a new configuration, call C_2 .
- 3) We compute the energy E_2 of C_2 and $\Delta E = E_2 - E_1$.
- 4) We test the sign of ΔE

$\Delta E \leq 0$, we accept the new configuration :
 C_2 become C_1 , go back to 2)

$\Delta E > 0$, we accept the new configuration according to a probability based on **Boltzmann factor** :
 - if accepted, C_2 become C_1 , go back to 2)
 - if rejected, $C_2 = C_1$, go back to 2)



MC makes use of random numbers

- to choose a particule to move
- to control the displacement
- for the « asymmetric rule »

The parameter α should be adjusted empirically in such a way that in step 4 approximately one out of two attempted steps is accepted.

Additional explanations :

The MC algorithm generate a stationary **Markovian stochastic** sequence which converge to the equilibrium configuration.

Stochastic means random process between two configurations $i \rightarrow j$

Markovian means that the probability of the transition $i \rightarrow j$ depend only of the configuration i

Evolution equation :

If $P(i,s)$ is the probability for the system to be in configuration i at « step » s , then :

$$P(i, s + 1) = P(i, s) + \sum_j (W(j \rightarrow i)P(j, s) - W(i \rightarrow j)P(i, s))$$

W being the conditional probability of a transition.

This equation of evolution will lead to an equilibrium if $\sum_j W(j \rightarrow i)N_j = N_i \sum_j W(i \rightarrow j)$

A solution, call *micro reversibility*, correspond to $W(j \rightarrow i)N_j = W(i \rightarrow j)N_i$

$$W(j \rightarrow i)N_j = W(i \rightarrow j)N_i$$

$$\frac{W(i \rightarrow j)}{W(j \rightarrow i)} = \frac{N_j}{N_i} = \exp(-\beta(U_j - U_i))$$

as
$$\frac{N_i}{N} = \frac{\exp(-\beta U_i)}{Z}$$

Consequence : the unknown parameters W do not depend of Z but only of the Boltzmann factor !

Proposition of Metropolis et al :

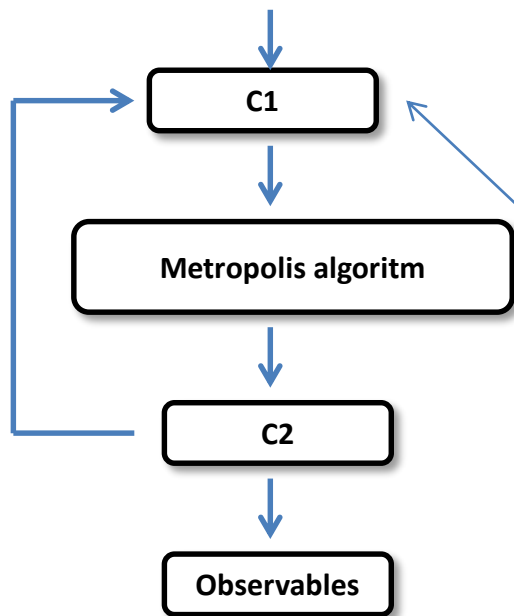
$$W(i \rightarrow j) = \alpha(i \rightarrow j)\pi(i \rightarrow j)$$

- from a configuration i we choose randomly a configuration j with a probability $\alpha(i \rightarrow j)$
- this new configuration j is accepted with a probability $\pi(i \rightarrow j)$

$$\frac{W(i \rightarrow j)}{W(j \rightarrow i)} = \frac{\pi(i \rightarrow j)}{\pi(j \rightarrow i)} = \exp(-\beta(U_j - U_i))$$

$$\begin{array}{l} \text{Solution of Metropolis} \\ = \\ \text{Metropolis algorithm} \end{array} \left\{ \begin{array}{l} \pi(i \rightarrow j) = \exp(-\beta(U_j - U_i)) \text{ if } U_j > U_i \\ \pi(i \rightarrow j) = 1 \text{ if } U_j < U_i \end{array} \right.$$

Model : nature and number of particules
 thermodynamics conditions
 interaction potential...



Not so many questions arise :

(a) How to deal with volume restriction
 ⇒ Same as MD : PBC, MIC...

~~(b) How to deal with the integration of Newton's equation~~

~~(c) How to chose the time step δt~~

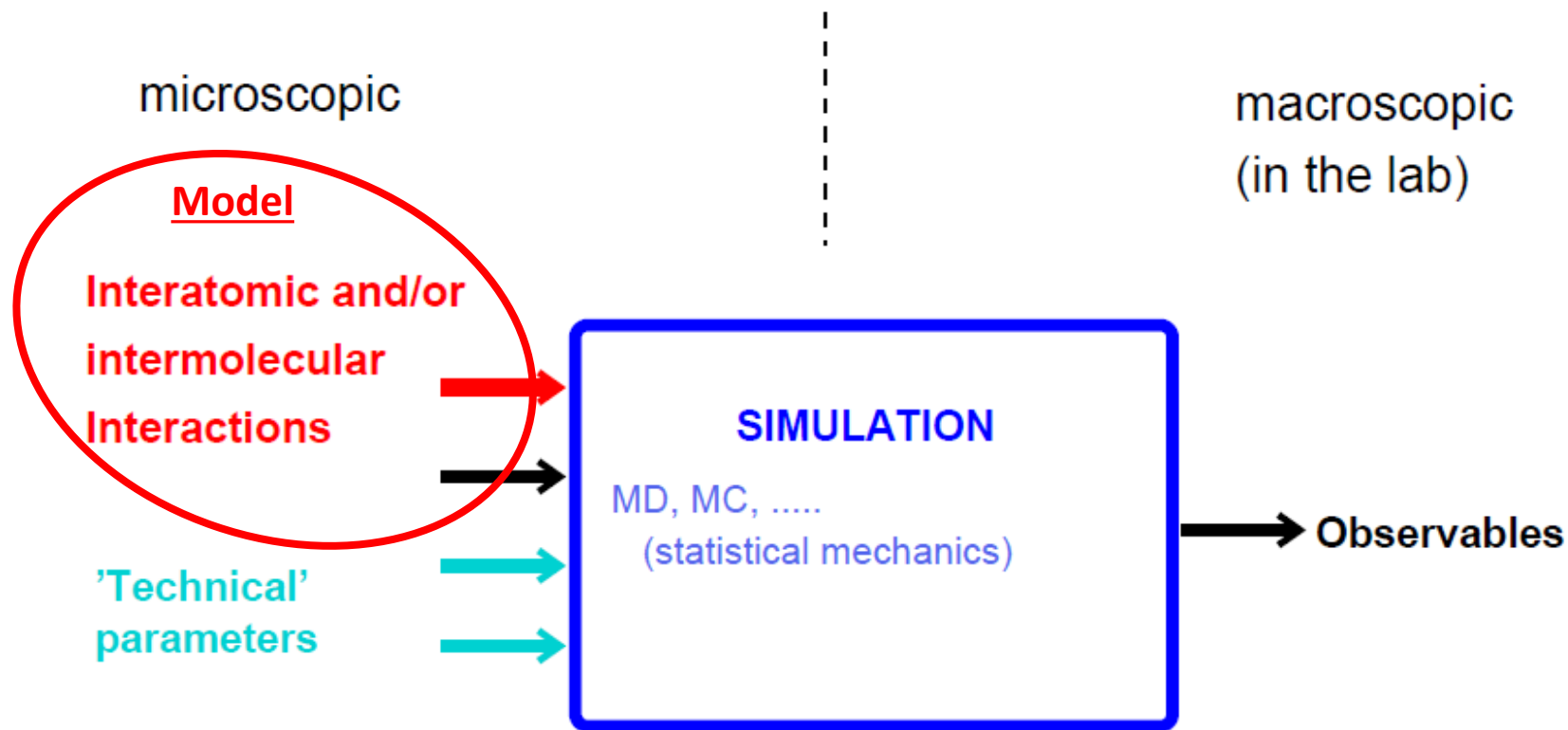
(d) How to find C1

(e) How to control a thermodynamic ensemble

} ⇒ Nearly the same as MD

etc.

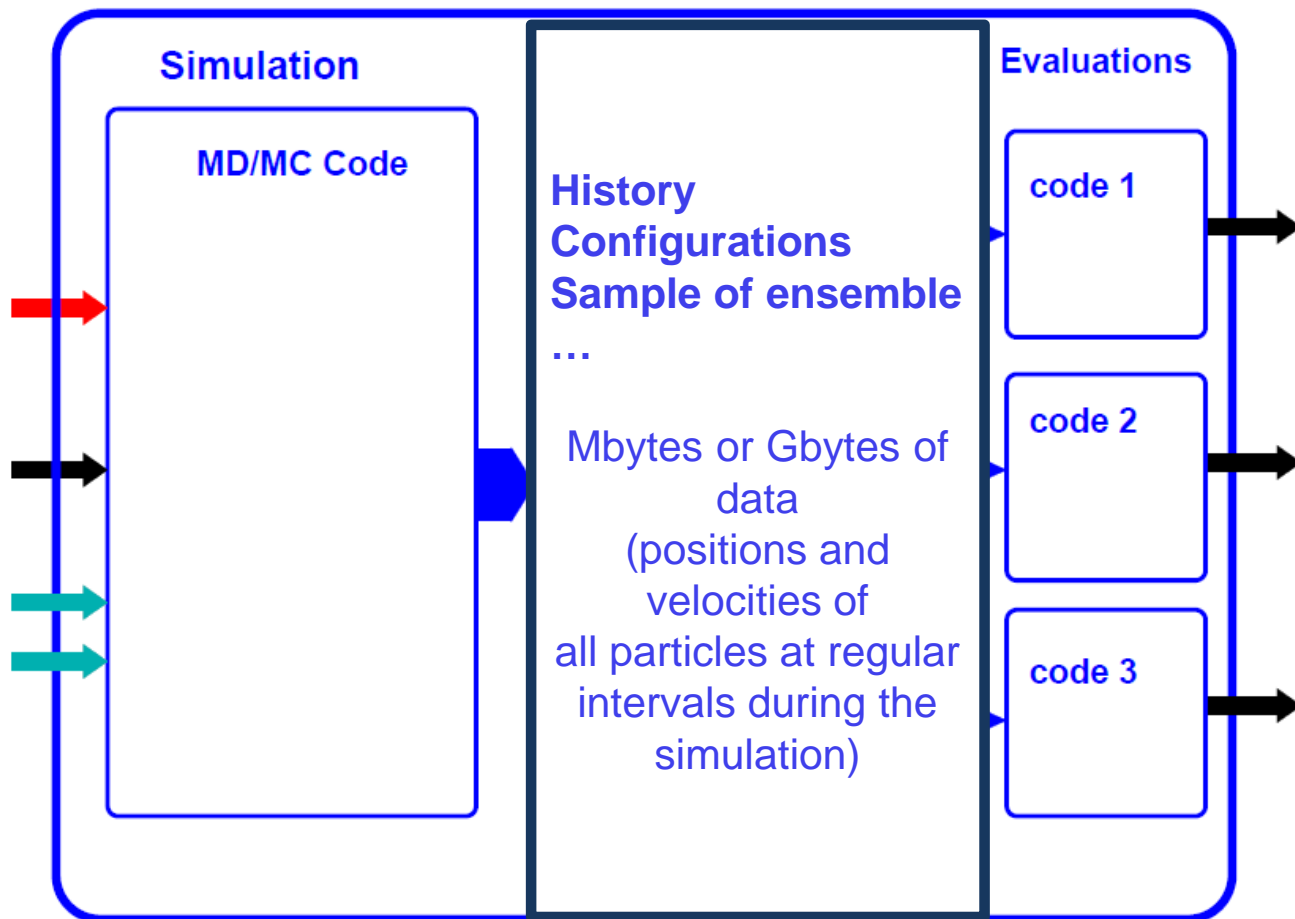
Principle of a simulation (MD or MC) :



In a « good » simulation : the **Model** should be the only input that matters.

i.e. not the technical details; number of particles (size of the box), how to compute the (long range) interactions, how to integrate the equations of motion, etc.

Principle of a simulation (MD or MC) :



« **Statistical Mechanics & Simulations** »

- I. Overview of Statistical Mechanics & Molecular Simulations
- II. Molecular Dynamics Simulations
- III. Monte Carlo methods
- IV. « Outputs » : extracting properties from simulations**
- V. Initiation to statistical thermodynamics

Evaluation of simulations experiments

We now have our sample : \mathcal{P}_i (*MD case*) or \mathcal{C}_i (*MC case*)

$$i = 1, M$$

M usually equal 100000 up to several millions...



For every quantity A than can be computed from positions (and velocities in MD) of the particules, we can calculate the thermodynamic average as :

$$\langle \mathcal{A}(\mathcal{P}_i) \rangle = \frac{1}{M} \sum_{i=1}^M \mathcal{A}(\mathcal{P}_i) \quad \text{over the } M \text{ microstates}$$

This is the most important formula of this whole simulation business !

A can be a basic thermodynamic observables (such as temperature, internal energy, pressure), a property describing the structure of the system,... **or any other property you might imagine .**

Some examples of quantities A :

Internal energy U:

$$U = \langle \mathcal{H}(\mathbf{r}, \mathbf{p}) \rangle = K + V = \langle \mathcal{K}(\mathbf{p}) \rangle + \langle \mathcal{V}(\mathbf{r}) \rangle$$

$$\text{Kinetic energy: } \langle \mathcal{K}(\mathbf{p}) \rangle = \left\langle \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{2m_i} \right\rangle$$

$$\text{Potential energy: } \langle \mathcal{V}(\mathbf{r}) \rangle$$

Temperature :

$$T = \langle \mathcal{T}(\mathbf{p}) \rangle = \frac{2}{fk_B} \langle \mathcal{K}(\mathbf{p}) \rangle = \frac{2}{fk_B} \left\langle \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{2m_i} \right\rangle$$

f : number of degrees of freedom of the system

Pressure :

$$P = \frac{Nk_B T}{V} - \frac{1}{3V} \langle \mathcal{W}(\mathbf{p}) \rangle$$

Ideal part Non ideal part ↔ interactions

$$\text{Virial: } \langle \mathcal{W}(\mathbf{p}) \rangle = - \left\langle \sum_{i=1}^{N-1} \sum_{j>i}^N \mathbf{R}_{ij} \mathbf{F}_{ij} \right\rangle$$

Structure : Pair correlation functions, the way to characterize the structure of a disordered systems.

An average of the quantity

$$a(r; \mathcal{P}_M) = \sum_i \delta(r_i - r)$$

represents the probability density of some particle being situated near r position in each configuration. The average is thus the mean fluid density at r position :

$$\langle a(r) \rangle = \rho(r) = \left\langle \sum_i \delta(r_i - r) \right\rangle$$

As the density is a constant in a fluid, we can write the « pair correlation function » (PCF) :

$$g(r) = \frac{V}{4\pi r^2 N^2} \left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle$$

It is the conditional probability density of finding a particle at r , given that there is a particle at the coordinate origin.

In a system containing different type of atoms, there are as many PCF as there are types of pairs :

$$g_{\alpha\beta}(r) = \frac{\langle N_{\beta}(r, r+dr) \rangle}{4\pi \rho_{\beta} r^2 dr}$$

where N is the number of atoms of type β at a distance between (r) and $(r+dr)$ from an atom of type α and ρ_{β} the number density of atom β

Example of pair correlation function :

Correlations in the Motion of Atoms in Liquid Argon*

A. RAHMAN

Argonne National Laboratory, Argonne, Illinois

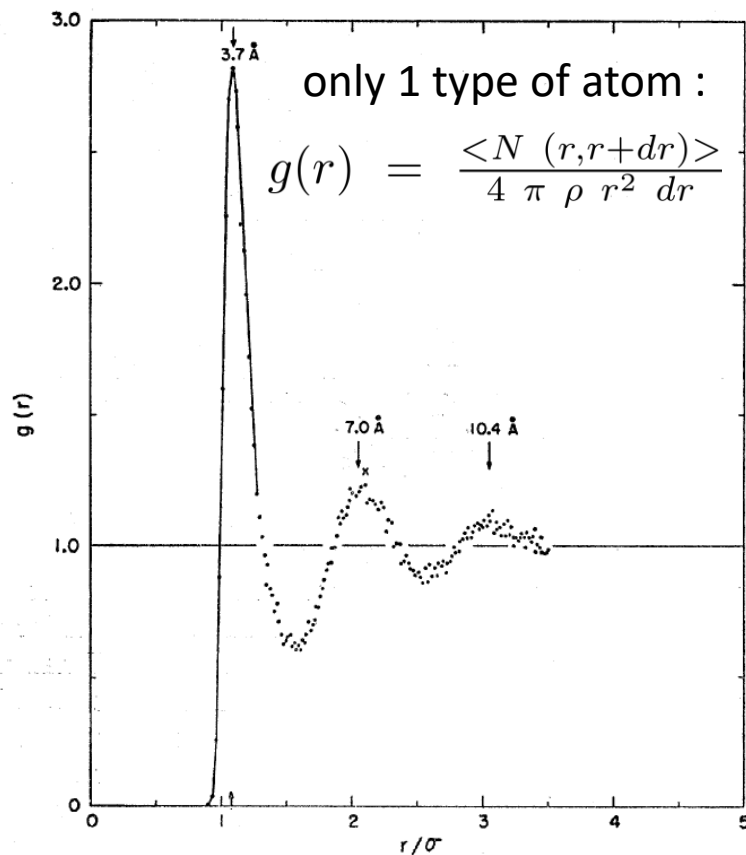
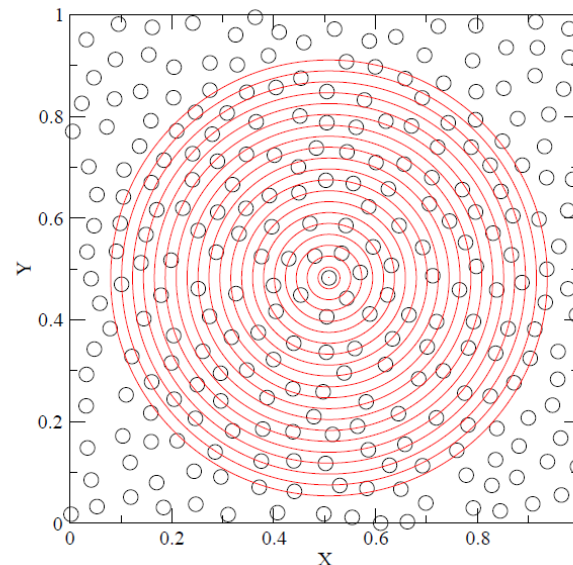


FIG. 2. Pair-correlation function obtained in this calculation at 94.4°K and 1.374 gcm⁻³. The Fourier transform of this function has peaks at $\kappa\sigma = 6.8, 12.5, 18.5, 24.8$.

Calculation principle :

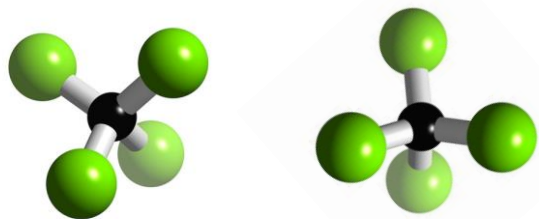


Questions :

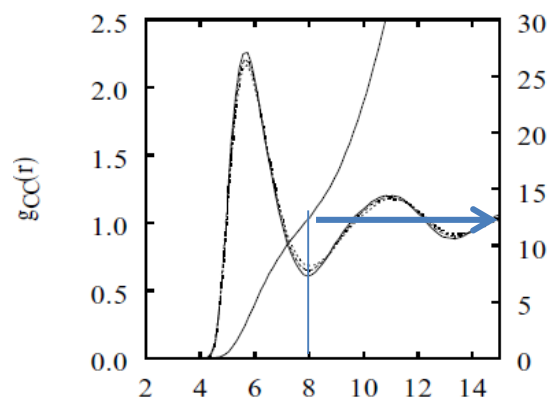
- (a) What mean these oscillations ?
 ⇒ *a deviation from the mean density*
- (b) What is the responsible of these oscillations ?
 ⇒ *the interactions between particules*
- (c) Why do it converge to 1 at high value of r ?
 ⇒ *the normalization (ρ) in the formulae*

Example of pair correlation functions :

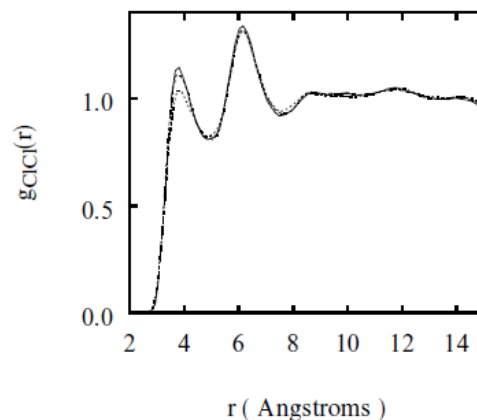
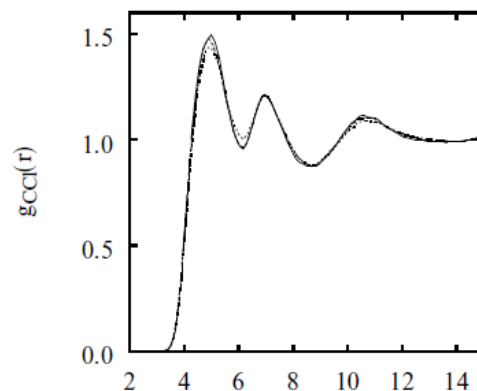
Liquid carbon tetrachloride : CCl_4



for a dimer :
 1 $g_{\text{CC}}(r)$
 8 $g_{\text{CCl}}(r)$
 16 $g_{\text{ClCl}}(r)$

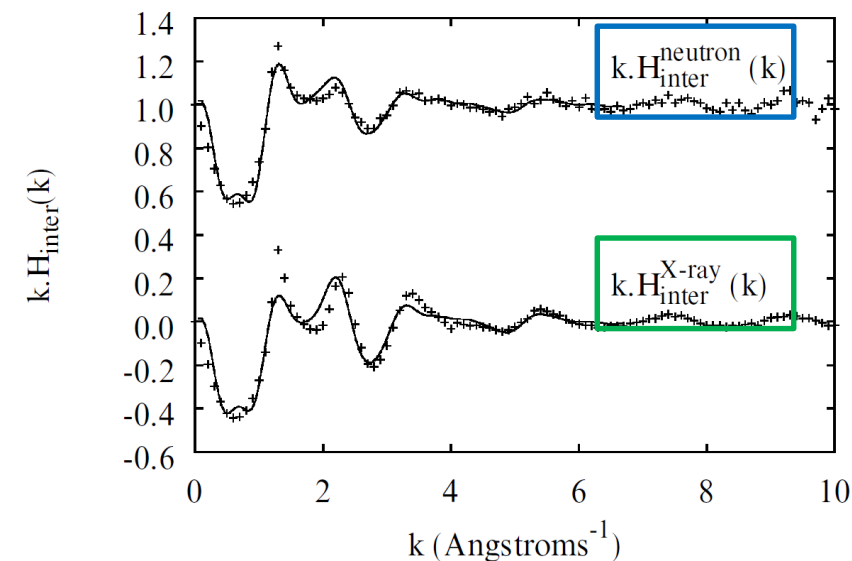
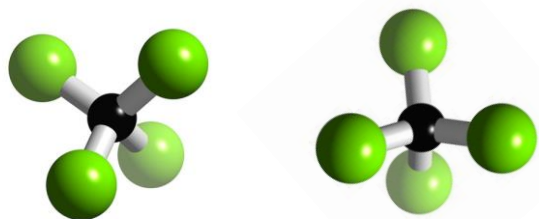


$\approx n(r)$
 ≈ 13 molecules in the first shell.



Running coordination number :

$$N_{\beta}(R) = \int_0^R 4 \pi \rho_{\beta} g_{\alpha\beta}(r) r^2 dr$$

Comparison with experiments : neutron and X-ray diffraction experiments
Liquid carbon tetrachloride : CCl_4


The intermolecular total structure factor $H_{\text{inter}}(k)$ is obtained from

$$H_{\text{inter}}(k) = 4\pi\rho \int_0^{\infty} r^2 \frac{\sin(kr)}{kr} (G_{\text{inter}}(r) - 1) dr, \quad (8)$$

where ρ is the molecular number density and $G_{\text{inter}}(r)$ is a weighted sum of the partial intermolecular radial distribution functions according to

$$G_{\text{inter}}(r) = \frac{\sum_{\alpha,\beta} b_{\alpha} b_{\beta} g_{\alpha\beta}(r)}{\left(\sum_{\alpha} b_{\alpha}\right)^2} \quad (9)$$

and b_{α} are the coherent scattering lengths of atoms ($b_{\text{C}} = 6.65$ fm, $b_{\text{Cl}} = 9.58$ fm for neutron diffraction). This leads to

$$G_{\text{inter}}^{\text{neutron}}(r) = 0.02g_{\text{CC}}(r) + 0.25g_{\text{CCl}}(r) + 0.73g_{\text{ClCl}}(r), \quad (10)$$

$$G_{\text{inter}}^{\text{X-ray}}(r) = 0.01g_{\text{CC}}(r) + 0.15g_{\text{CCl}}(r) + 0.84g_{\text{ClCl}}(r), \quad (11)$$

Question : when we use the formula

$$\langle \mathcal{A}(\mathcal{P}_i) \rangle = \frac{1}{M} \sum_{i=1}^M \mathcal{A}(\mathcal{P}_i) \quad \text{over the } M \text{ microstates obtained in a MD simulation}$$

or

$$\langle \mathcal{A}(\mathcal{C}_i) \rangle = \frac{1}{M} \sum_{i=1}^M \mathcal{A}(\mathcal{C}_i) \quad \text{over the } M \text{ microstates obtained in a MC simulation}$$

What did we miss !?!?

Answer : the fact that in MD, the \mathcal{P} s are ordered,
so the fact that they contain the time evolution of the system !

We can thus define temporal correlations functions to study the time evolution of any properties of the system we are interested in.

Of course, this concept is only valid for MD simulations !

Time auto-correlation function

$$c_{AA}(t) = \frac{1}{\text{Norm}} \cdot \langle A(\mathcal{P}(\tau)) \odot A(\mathcal{P}(\tau + t)) \rangle_{\tau}$$

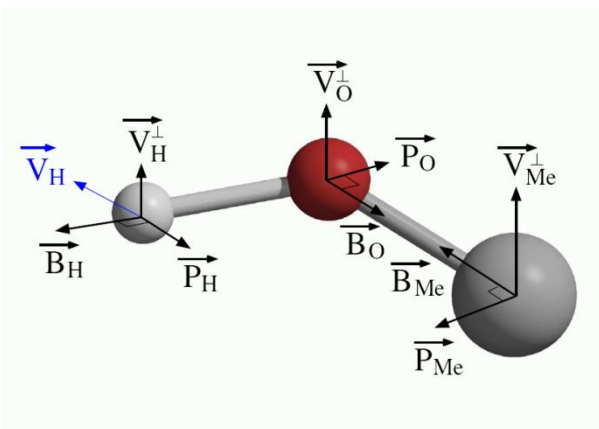
Method :

- we compute, like before, the quantity we are interested in, A, for each P.
- we relate with each other through an operator (scalar product, +, ...) the quantities A at two times (τ) and ($t + \tau$).
- we look at the result as a function of the time delay (t) between both.
- in the equilibrium ensemble, we average over all “initial” times (τ) and get the auto-correlation function.

With this concept, we are here leaving the domain of statistical thermodynamics and enter the more general domain of statistical mechanics.

These functions are important objects in statistical mechanics (see e.g. D. McQuarrie's book). In particular, one shows that:

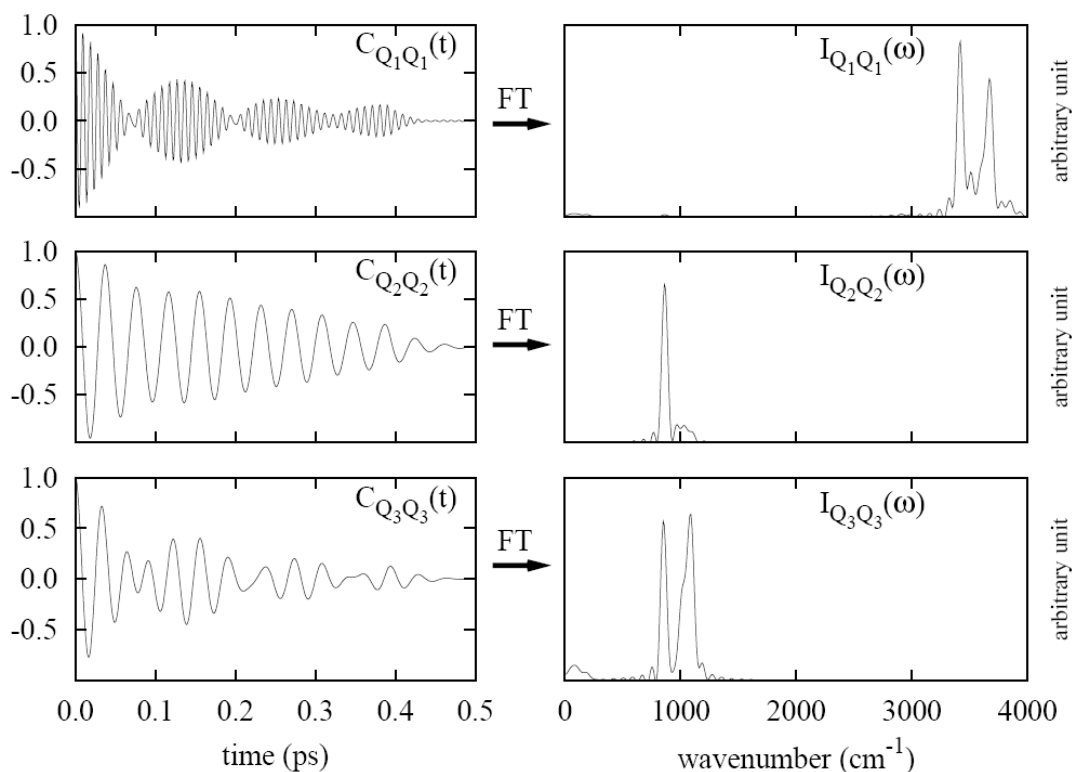
- 1) the integrals over these functions are related to the so-called “transport coefficients”.
The transport coefficient related to the velocity autocorrelation function is the self-diffusion coefficient D_s .
- 2) the Fourier transforms of these functions are called “spectra”. They are not only called spectra, they are spectra like the ones measured in Infrared, Raman, neutron inelastic scattering, spectroscopies.
The Fourier transform of the velocity autocorrelation function is related to neutron scattering.

Example : atomic velocity autocorrelation functions in methanol


MD simulation of a methanol dimer in a solvent (CCl4).

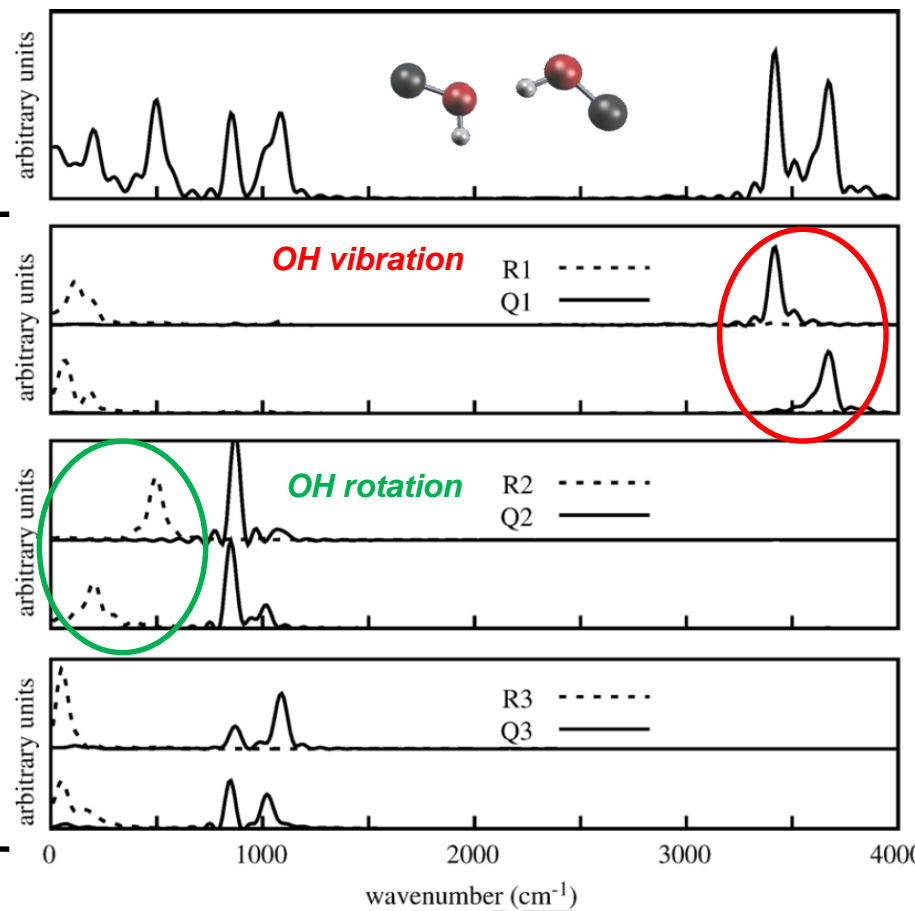
Normal mode like combinations

$Q_1 = B_H$	O–H stretching
$Q_2 = B_{Me} + B_O$	O–Me stretching
$Q_3 = P_H + P_{Me}$	H–O–Me bending
$R_1 = P_O + P_{Me}$	Rotation in plane
$R_2 = V_H^\perp$	Rotation of H roughly around the O–Me bond
$R_3 = V_O^\perp - V_{Me}^\perp$	Rotation roughly around an axis perpendicular to the O–Me bond



Example : atomic velocity autocorrelation functions in methanol (dimer in CCl_4)

Spectral density from the total velocities of both H atoms



Donor molecule

Acceptor molecule

Fourier transforms of the autocorrelation functions Qs and Rs

Generalisation : Time cross-correlation function

$$c_{AB}(t) = \frac{1}{N_{\text{orm}}} \cdot \langle A(\mathcal{P}(\tau)) \odot B(\mathcal{P}(\tau + t)) \rangle_{\tau}$$

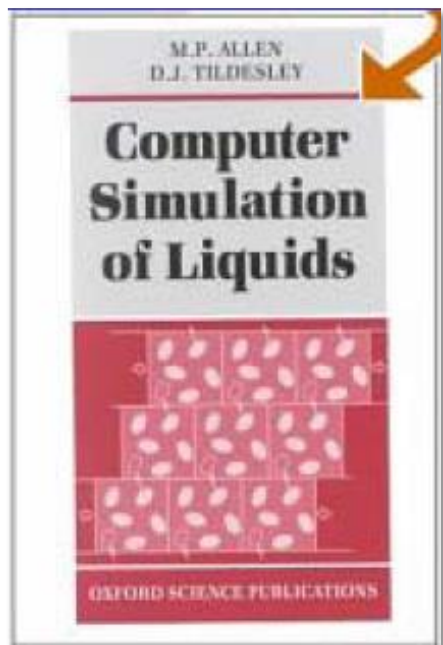
Conclusions :

We have two types of averaging procedures over our sample:

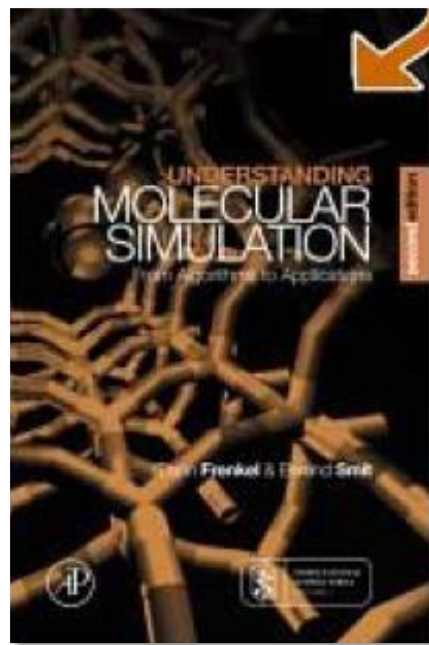
- (I) We look at one configuration $\mathcal{P}(t)$ at a time and average over all configurations.
These averages led to the often called “static properties”
- (II) We look at two configurations $\mathcal{P}(t_1)$ and $\mathcal{P}(t_2)$ and average over all pairs of configurations which have the same spacing in time
These averages led to the often called “dynamic properties”

« **Statistical Mechanics & Simulations** »

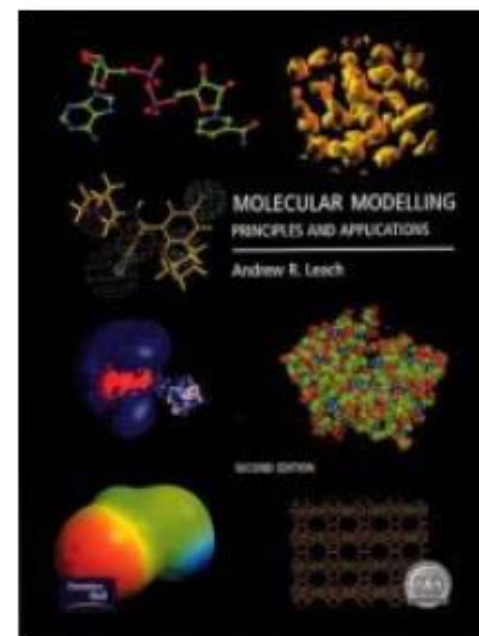
- I. Overview of Statistical Mechanics & Molecular Simulations
- II. Molecular Dynamics Simulations
- III. Monte Carlo methods
- IV. « Outputs » : extracting properties from simulations
- V. Initiation to statistical thermodynamics**



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/>

