Classical MD simulations

Miguel A. Gonzalez

Institut Laue-Langevin, Grenoble (France)

Plan

- Introduction
  - Do we need simulations? Why?
  - What kind of simulations?
- Representing intermolecular forces
  - Electronic structure methods ✉ Density functional theory **(DFT)**
  - **Force fields**
  - Coarse grained models
- Sampling phase space
  - Lattice Dynamics
  - Monte Carlo
  - **Molecular Dynamics**
- Software
- Analysis
  - What can we get from a simulation?

Introduction

Do we need computer simulations?

EXCELLENT COMPUTER SIMULATIONS ARE DONE FOR A PURPOSE. THE MOST VALID PURPOSES ARE TO EXPLORE UNCHARTED TERRITORY, TO RESOLVE A WELL-POSED SCIENTIFIC OR TECHNICAL QUESTION, OR TO MAKE A GOOD DESIGN CHOICE.

Leo P. Kadanoff: "Excellence in computer simulation", *Comp. Sci. Eng.* **6**, 57-67 (2004).

- Only in rare cases simulations have discovered new phenomena (e.g. long time tails)
- Typical role is to work with theory and experiment and serve to test models and theories in order to understand a given phenomenon
- Caution: Quite often simulations can be wrongly used to support previous conclusions

"Experiment is very good at finding unexpected behavior and describing its overall characteristics; theory often can explain what's going on. After an appropriate pause for algorithm development, simulations then can test the ideas and fill in the details."

Introduction

What kind of simulations? Compromise: Accuracy vs efficiency

From Kermode et al., in *Multiscale Simulation Methods in Molecular Sciences,*
J. Grotendorst, N. Attig, S. Blügel, D. Marx (Eds.), NIC Series, Vol. 42, pp. 215-228 (2009).

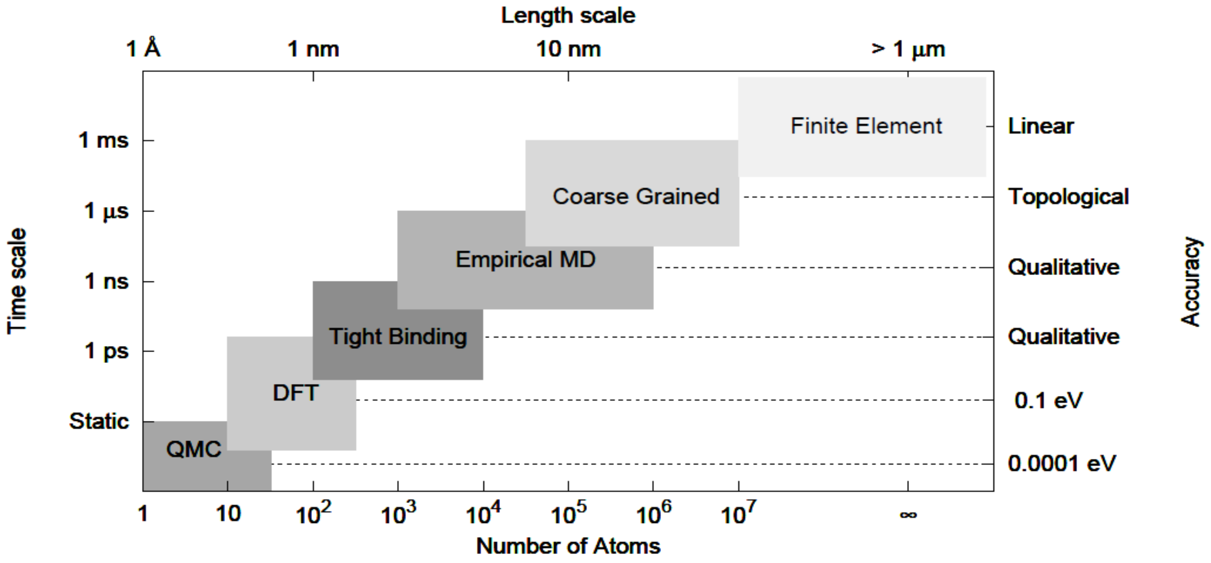


Figure 1. Schematic representation of the range of length- and time-scales accessible to a variety of modelling methods, from quantum Monte Carlo (QMC) for very accurate, very expensive static calculations through to approximate methods such as finite-element modelling.

Introduction

What kind of simulations?

Choose way of computing interatomic forces

Hartree-Fock methods

Semiempirical methods

Density Functional Theory

Tight-binding

Empirical force fields

Coarse graining

Brownian Dynamics

Langevin Dynamics

Dissipative Particle Dynamics

Choose way of sampling the phase space

Lattice Dynamics

$$D_{ij}(\mathbf{k}) = \frac{1}{\sqrt{m_i m_j}} \sum_{\ell} \Phi_{ij}(0\ell) \exp\left(i\mathbf{k} \cdot [\mathbf{r}_i(0) - \mathbf{r}_j(\ell)]\right)$$

Thermodynamics
Phonons

Monte Carlo

$$\pi_a(\Gamma \rightarrow \Gamma') = \min(1, \exp(-\beta \Delta H))$$

Thermodynamics
Structure

Molecular Dynamics

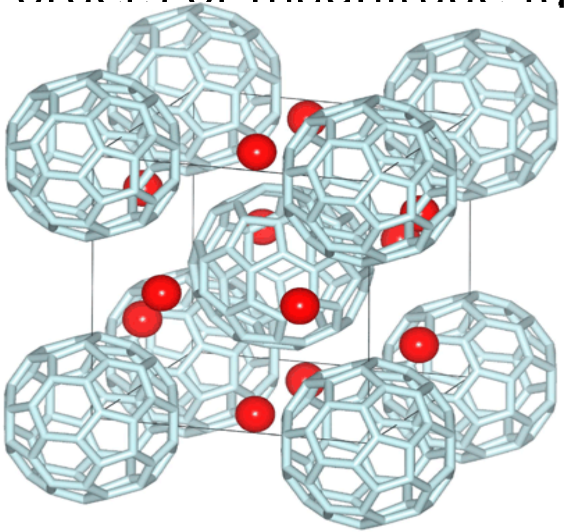
Computing energies and forces
Using empirical potentials

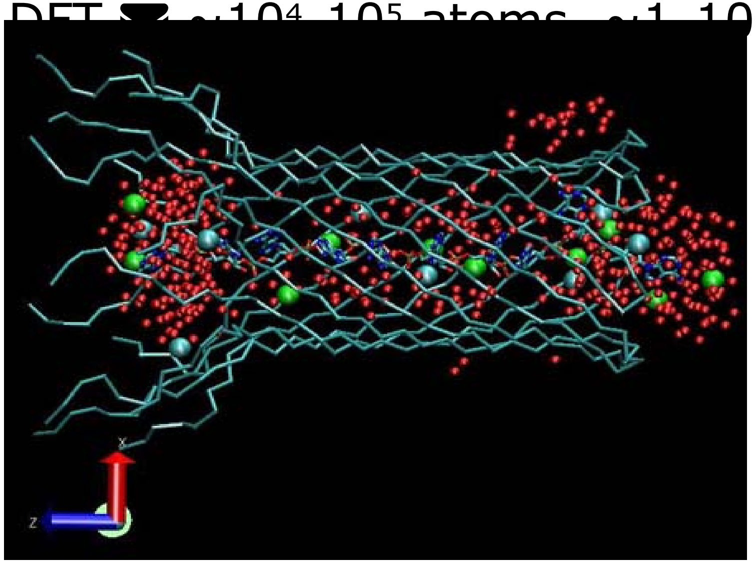
Molecular Mechanics approach

- No electrons
- Molecules represented as ‘balls’ + ‘springs’
- Empirical potential to account for intra- and intermolecular interactions

✉Force Field (FF)

- ~5-6 orders of magnitude faster than DFT ✉ • ~10^4-10^5 atoms, • ~1-100 ns

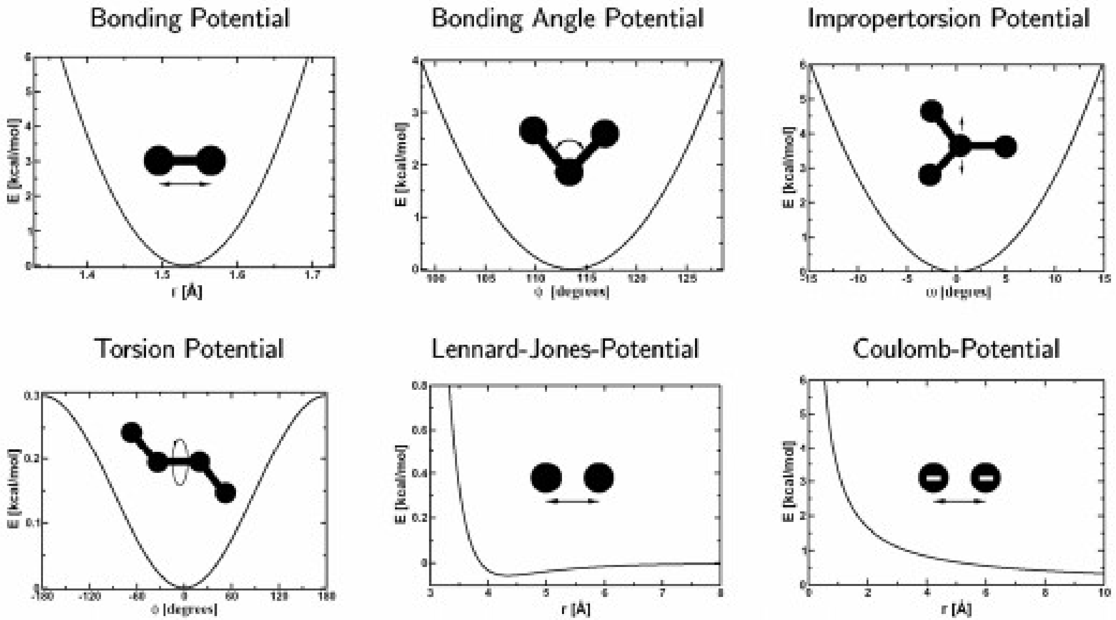




What is a Force Field?

- A mathematical expression describing the dependence of the energy of a system on the coordinates of the atoms
- Consists of:
  1. An analytical form of the interatomic potential energy, $U(r_1 \ldots r_N)$
  2. A set of parameters entering $U(r_1 \ldots r_N)$
- The parameters are obtained from *ab initio* calculations and/or fitting to experimental data
- Molecules are defined as collection of atoms held together by simple elastic (harmonic) forces

A standard FF



$$
\begin{aligned}
E \quad = \quad & \sum_b k_b(r - r_{\mathrm{b}})^2 + \sum_\theta k_\theta(\theta - \theta_{\mathrm{O}})^2 + \sum_\omega k_\omega(\omega - \omega_{\mathrm{O}})^2 \\
+ \quad & \sum_\phi k_\phi(1 - \cos(n\phi - \delta)) + \sum_{i<j} 4\ \epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right] + \sum_{i<j}\frac{q_i q_j}{r}
\end{aligned}
$$

FF parameterization

An elaborate job

- Goal: Describe in classical terms the quantum mechanical facts by partitioning the total electronic energy into well separated atom-atom contributions
- As it is impossible to fully separate the intricate electronic effects, this means applying significant approximations ✉ empirical potentials
- We will obtain different FFs depending on the set of data (experimental or QM calculations) and the procedure employed to optimize the parameters
- “**FF development is still as much a matter of art as of science**”
  (T. Halgren, *Curr. Opin. Struct. Biol.*, 1995)

FF parameterization

Components

- Functional forms of the components of the energy expression
- A list of atom types
- A list of atomic charges
- Rules for atom types
- A set of parameters for the function terms
- (Rules to generate parameters that have not been defined explicitly)

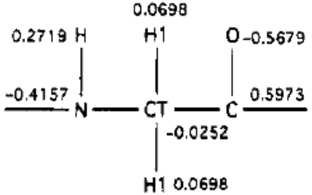
Generally developed by specialized groups and available in the literature as data tables

FF example

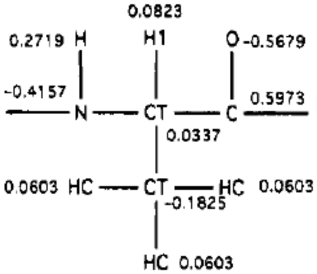
Amber

Table 1. List of Atom Types[a]

atom	type	description
carbon	CT	any sp^3 carbon
	C	any carbonyl sp^2 carbon
	CA	any aromatic sp^2 carbon and (Cϵ of Arg)
	CM	any sp^2 carbon, double bonded
	CC	sp^2 aromatic in 5-membered ring with one substituent + next to nitrogen (Cγ in His)
	CV	sp^2 aromatic in 5-membered ring next to carbon and lone pair nitrogen (e.g. Cδ in His (δ))
	CW	sp^2 aromatic in 5-membered ring next to carbon and NH (e.g. Cδ in His (ϵ) and in Trp)
	CR	sp^2 aromatic in 5-membered ring next to two nitrogens (Cγ and Cϵ in His)
	CB	sp^2 aromatic at junction of 5- and 6-membered rings (Cδ in Trp) and both junction atoms in Ade and Gua
	C*	sp^2 aromatic in 5-membered ring next to two carbons (e.g. Cγ in Trp)
	CN	sp^2 junction between 5- and 6-membered rings and bonded to CH and NH (Cϵ in Trp)
	CK	sp^2 carbon in 5-membered aromatic between N and N-R (C8 in purines)
	CQ	sp^2 carbon in 6-membered ring between lone pair nitrogens (e.g. C2 in purines)
nitrogen	N	sp^2 nitrogen in amides
	NA	sp^2 nitrogen in aromatic rings with hydrogen attached (e.g. protonated His, Gua, Trp)
	NB	sp^2 nitrogen in 5-membered ring with lone pair (e.g. N7 in purines)
	NC	sp^2 nitrogen in 6-membered ring with lone pair (e.g. N3 in purines)
	N*	sp^2 nitrogen in 5-membered ring with carbon substituent (in purine nucleosides)
	N2	sp^2 nitrogen of aromatic amines and guanidinium ions
	N3	sp^3 nitrogen



GLY



ALA

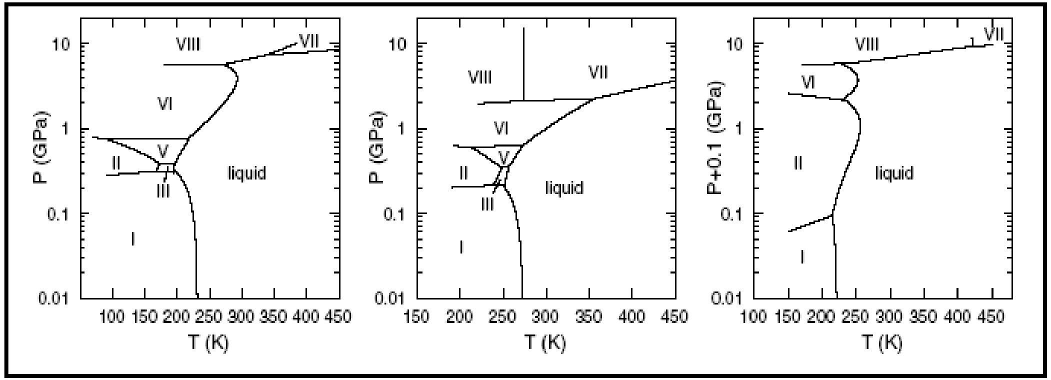
Van der Waals Parameters

atom type	R^{*j}	ϵ^k	atom type	R^{*j}	ϵ^k	atom type	R^{*j}	ϵ^k	atom type	R^{*j}	ϵ^k
C^l	1.9080	0.0860	H2	1.2870	0.0157	HS	0.6000	0.0157	O2	1.6612	0.2100
CA	1.9080	0.0860	H3	1.1870	0.0157	HW	0.0000	0.0000	OH	1.7210	0.2104
CM	1.9080	0.0860	H4	1.4090	0.0150	IP	1.8680	0.00277	OS	1.6837	0.1700
Cs	3.3950	0.0000806	H5	1.3590	0.0150	K	2.6580	0.000328	OW	1.7683	0.1520
CT	1.9080	0.1094	HA	1.4590	0.0150	Li	1.1370	0.0183	P	2.1000	0.2000
F	1.75	0.061	HC	1.4870	0.0157	N^m	1.8240	0.1700	Rb	2.9560	0.00017
H	0.6000	0.0157	HO	0.0000	0.0000	$N3^n$	1.875	0.1700	S	2.0000	0.2500
H1	1.3870	0.0157	HP	1.1000	0.0157	O	1.6612	0.2100	SH	2.0000	0.2500

Choosing a FF

Highly subjective, e.g. water case

Sanz, PRL (2004)



TIP4P **Exp** **SPC/E**

But ...

$$D_{SPC/E} = 2.49,\ D_{TIP4P} = 3.29,\ D_{exp} = 2.4\ (10^{-5}\ cm^2/s)$$

$$\varepsilon_{SPC/E} = 71,\ \varepsilon_{TIP4P} = 53,\ \varepsilon_{exp} = 78$$

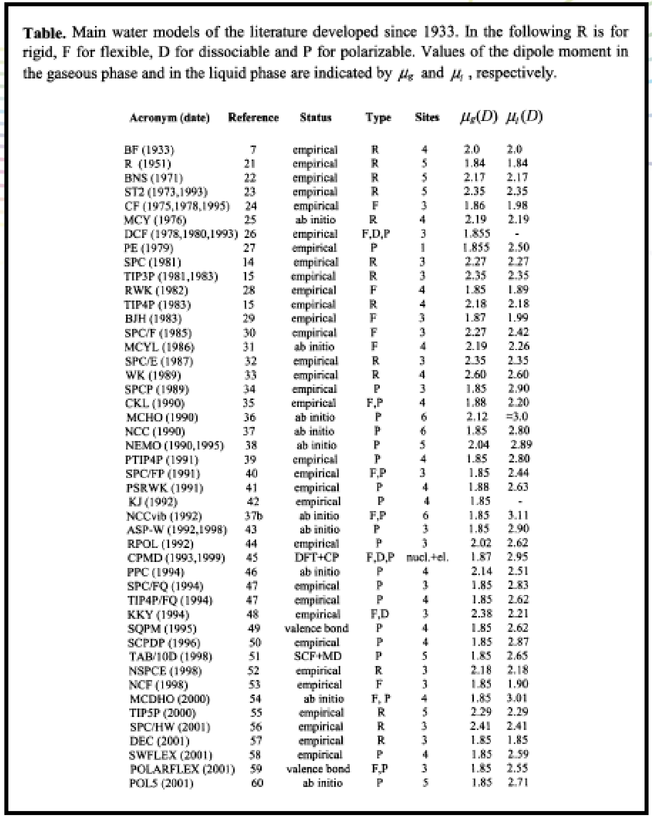


Table. Main water models of the literature developed since 1933. In the following R is for rigid, F for flexible, D for dissociable and P for polarizable. Values of the dipole moment in the gaseous phase and in the liquid phase are indicated by μ_g and μ_l , respectively.

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

Guillot, J. Mol. Liq. (2002)

Choosing a FF

Highly subjective, e.g. water case

- Compute 17 different properties and note each of them
- TIP4P/2005 is probably close to the best that can be achieved with a rigid non-polarizable model

Vega & Abascal, *Phys. Chem. Chem. Phys.* **13**, 19663-19688(2011)

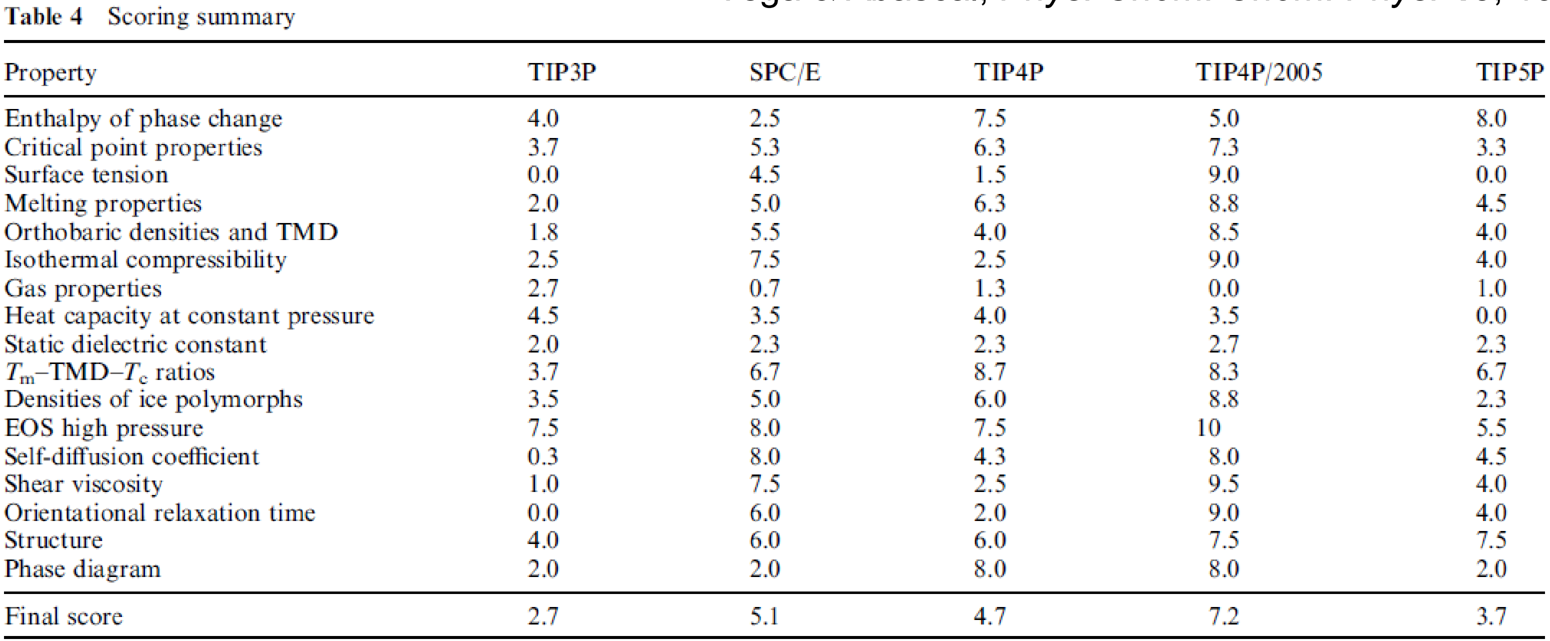


Table 4 Scoring summary

Property	TIP3P	SPC/E	TIP4P	TIP4P/2005	TIP5P
Enthalpy of phase change	4.0	2.5	7.5	5.0	8.0
Critical point properties	3.7	5.3	6.3	7.3	3.3
Surface tension	0.0	4.5	1.5	9.0	0.0
Melting properties	2.0	5.0	6.3	8.8	4.5
Orthobaric densities and TMD	1.8	5.5	4.0	8.5	4.0
Isothermal compressibility	2.5	7.5	2.5	9.0	4.0
Gas properties	2.7	0.7	1.3	0.0	1.0
Heat capacity at constant pressure	4.5	3.5	4.0	3.5	0.0
Static dielectric constant	2.0	2.3	2.3	2.7	2.3
T_m–TMD–T_c ratios	3.7	6.7	8.7	8.3	6.7
Densities of ice polymorphs	3.5	5.0	6.0	8.8	2.3
EOS high pressure	7.5	8.0	7.5	10	5.5
Self-diffusion coefficient	0.3	8.0	4.3	8.0	4.5
Shear viscosity	1.0	7.5	2.5	9.5	4.0
Orientational relaxation time	0.0	6.0	2.0	9.0	4.0
Structure	4.0	6.0	6.0	7.5	7.5
Phase diagram	2.0	2.0	8.0	8.0	2.0
Final score	2.7	5.1	4.7	7.2	3.7

Choosing a FF

Highly subjective

"All leading force fields now provide reasonable results for a wide range of properties of isolated molecules, pure liquids, and aqueous solutions", Jorgensen & Tirado-Rives, *PNAS* **102**, 6665 (2005).

But the best force field does not exist!

Choice will depend on the type of molecular system and the type of properties we are interested in!

FF vs DFT

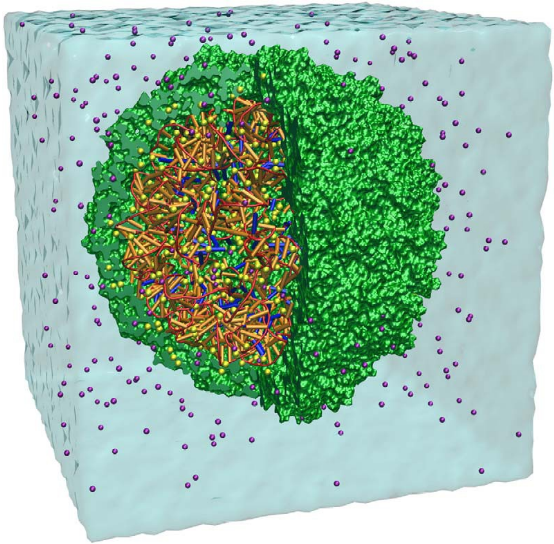
Disadvantages

- **Intrinsic limitations:**
  - No electronic structure information is available
  - Unable to handle reactions (bond breaking/forming, charge transfer, ...)
- **Limited prediction power:**
  - The accuracy depends on the parameterization
  - It can only be used for systems having the functional groups that were included in parameterization
  - Limited transferability of the FF (care needed when applying a FF under conditions that are very different from the conditions used in its parameterization, e.g. P, T)

Needs experimental validation!

Empirical force fields Advantages

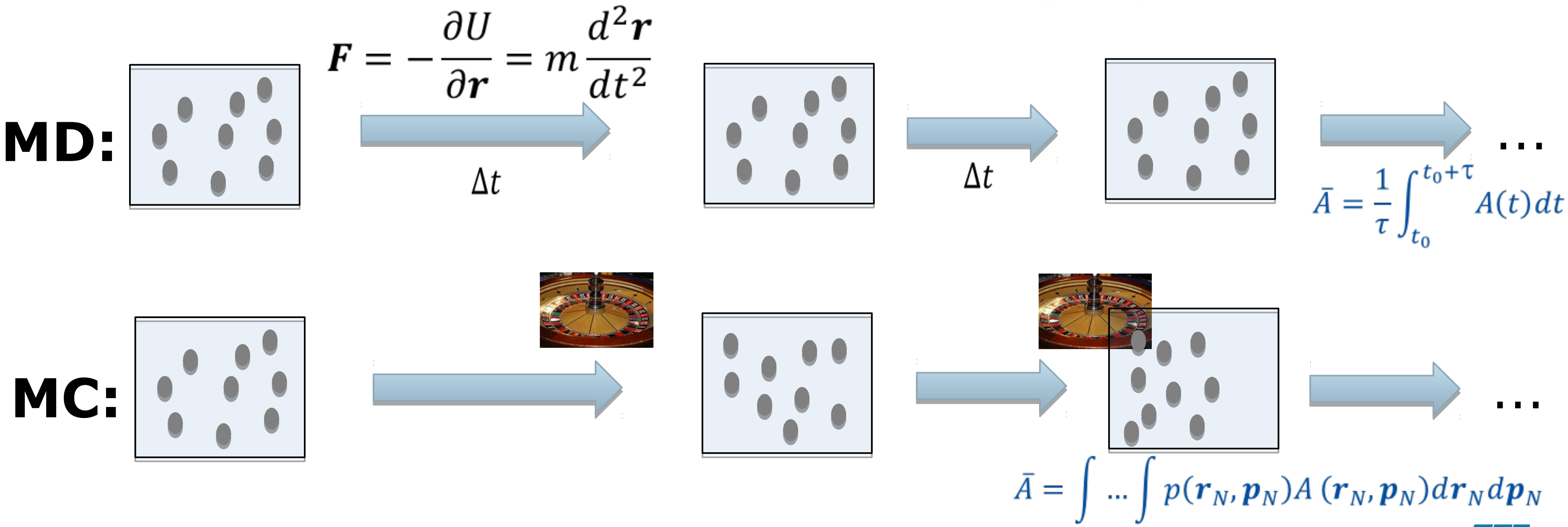
- Analysis of energy contributions can be done at the level of individual interactions or classes of interactions
- Possible to modify the energy expression to bias the calculation
- Allow to handle large systems and simulate relatively long times (several orders of magnitude faster than quantum-based calculations):
  - Small cluster (16 processors) ~ 10^4 atoms @ 1 ns/day
  - BlueGene/L (131072 processors) ~ 320 billion atoms (a cubic piece of metal of side ~1 μm) @ 10 ps/day (Kadau, *Int. J. Modern Physics C* (2006))
  - Anton (specialized machine for MD simulations) ~ 10^4 atoms @ 10 μs/day (Klepeis, *Curr. Opin. Str. Biol.* (2009))
  - NCS (256 nodes) ~ 10^6 atoms (all-atom satellite tobacco mosaic virus, NAMD) @ 1ns/day (Freddolino, *Structure* (2006))
  - Perspectives: multimillion-atom (~100 nm scale) @ 30 ns/day on a Cray XT5 (10^5 cores) (Schulz, *J. Chem. Theory Comput.* (2009))



Sampling phase space using Molecular Dynamics

Some statistical mechanics

- Impossible to explore all possible states of the system
- But it is enough to generate a set of representative states



MD:

$$\boldsymbol{F} = -\frac{\partial U}{\partial \boldsymbol{r}} = m\frac{d^2\boldsymbol{r}}{dt^2}$$

Δt

Δt

…

$$\bar{A} = \frac{1}{\tau}\int_{t_0}^{t_0+\tau} A(t)dt$$

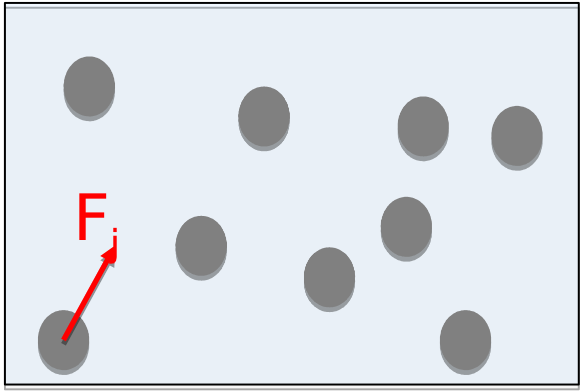
MC:

…

$$\bar{A} = \int \dots \int p(\boldsymbol{r}_N, \boldsymbol{p}_N) A(\boldsymbol{r}_N, \boldsymbol{p}_N) d\boldsymbol{r}_N d\boldsymbol{p}_N$$

Molecular Dynamics The basics

- Generate a dynamical trajectory by integrating Newton's equations of motion, with suitable initial and boundary conditions
- We need a good way to determine the forces acting on each atom (e.g. FF or DFT)
- Also need an accurate numerical method to integrate the equations of motion



$$\boldsymbol{F}_i = -\frac{\partial U(\boldsymbol{r}^N)}{\partial \boldsymbol{r}_i} = m_i \frac{d^2\boldsymbol{r}_i}{dt^2}$$

- System of N coupled ODEs
- No exact solution
- Step-by-step numerical integration

Molecular Dynamics The basics II

$$\frac{d\mathbf{r}_j}{dt} = \frac{\mathbf{p}_j}{m}$$

$$\frac{d\mathbf{p}_j}{dt} = \mathbf{F}_j$$

- **Discretize solution, using a time step Δt**

$$\boldsymbol{r}_i(t_0) \to \boldsymbol{r}_i(t_0 + \Delta t) \to \boldsymbol{r}_i(t_0 + 2\Delta t) \to \cdots \to \boldsymbol{r}_i(t_0 + n\Delta t)$$

- **Important features of our numerical integrator:**
  - minimal need to compute forces (a very expensive calculation)
  - good stability for large time steps
  - good accuracy
  - conserves energy and momentum
  - time-reversible
  - symplectic: conserves volume in phase space

- **Taylor expansion?**

$$\boldsymbol{r}_i(t_0 + \Delta t) = \boldsymbol{r}_i(t_0) + \boldsymbol{v}_i(t_0)\Delta t + \frac{1}{2}\boldsymbol{a}_i(t_0)\Delta t^2 + O(\Delta t^3)$$

$$\boldsymbol{v}_i(t_0 + \Delta t) = \boldsymbol{v}_i(t_0) + \boldsymbol{a}_i(t_0)\Delta t + O(\Delta t^2)$$

Unstable and inaccurate!

Molecular Dynamics The basics III

Verlet algorithm:

$$\boldsymbol{r}_i(t_0+\Delta t) = \boldsymbol{r}_i(t_0) + \boldsymbol{v}_i(t_0)\Delta t + \frac{1}{2}\boldsymbol{a}_i(t_0)\Delta t^2 + \frac{1}{3!}\dddot{\boldsymbol{r}}_i(t_0)\Delta t^3 + O(\Delta t^4)$$

$$\boldsymbol{r}_i(t_0-\Delta t) = \boldsymbol{r}_i(t_0) - \boldsymbol{v}_i(t_0)\Delta t + \frac{1}{2}\boldsymbol{a}_i(t_0)\Delta t^2 - \frac{1}{3!}\dddot{\boldsymbol{r}}_i(t_0)\Delta t^3 + O(\Delta t^4)$$

$$\boldsymbol{r}_i(t_0+\Delta t) + \boldsymbol{r}_i(t_0-\Delta t) = 2\boldsymbol{r}_i(t_0) + \boldsymbol{a}_i(t_0)\Delta t^2 + O(\Delta t^4)$$

- Positions at $t_0+\Delta t$ computed from actual positions and forces, and previous positions
- Error is $O(\Delta t^4)$
- Velocities not available, but easy to derive as $\boldsymbol{r}_i(t_0+\Delta t) - \boldsymbol{r}_i(t_0-\Delta t) = 2\boldsymbol{v}_i(t_0) + O(\Delta t^3)$
- Verlet and equivalent algorithms (e.g. leap-frog) are simple, efficient, stable and reasonably accurate. Also time-reversible and symplectic, giving low drifts in energy
- Used in most MD software
- Predictor-corrector algorithms are also used occasionally. They are more accurate for small Δt, but not time reversible, introducing larger drifts in E_{total}

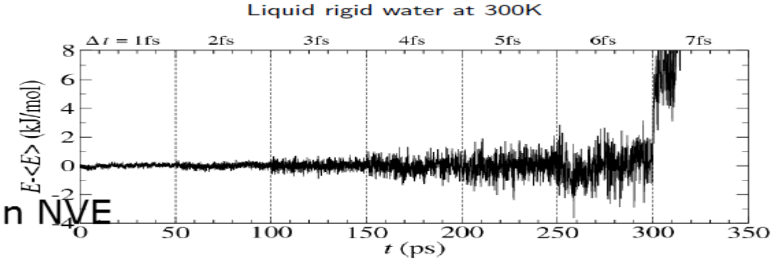
Molecular Dynamics The basics IV

Time step:

- Compromise between efficiency and accuracy of the integration
- Small Δt ✉ Inefficient phase space sampling
- Large Δt ✉ Numerical instabilities
- Hints:
    - Δt < mean time between collisions or shorter relaxation time
    - LJ atomic system ~ 10 fs
    - Flexible molecules with rigid bonds ~ 2 fs
    - Flexible bonds ~ 0.5-1 fs
    - Good test: Total energy should be conserved in NVE simulations. Acceptable

- Compromise between efficiency and accuracy of the integration
- Small Δt → Inefficient phase space sampling
- Large Δt → Numerical instabilities
- Hints:
    - Δt < mean time between collisions or shorter relaxation time
    - LJ atomic system ~ 10 fs
    - Flexible molecules with rigid bonds ~ 2 fs
    - Flexible bonds ~ 0.5-1 fs
    - Good test: Total energy should be conserved in NVE simulations. Acceptable $\sqrt{\langle \delta E^2 \rangle}/_E < 10^{-4}$

Frenkel & Smit: "Understanding molecular simulation", Academic Press (2002)

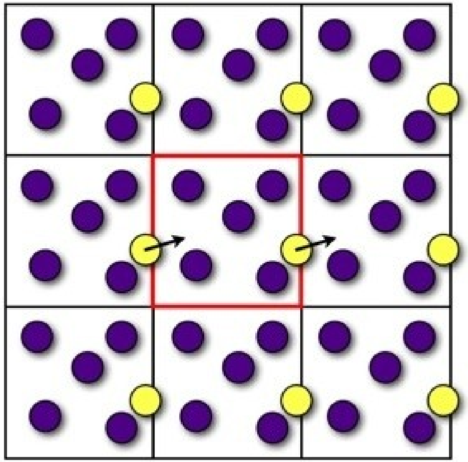


Molecular Dynamics The basics V

- **Periodic Boundary Conditions**
- **Potential truncation**
  - May need to ensure that forces (and not only energy) are 0 at R_c
- **Electrostatic forces**
- **Thermodynamic ensembles**
  - The integration of the equations of motion keeps constant N, V, and E (microcanonical ensemble)
  - Integration errors, force fluctuations and inconsistencies in the forces (e.g. generated by the cutoff) may cause slow drifts in the total energy: check E_{total} conservation!
  - E is constant in the NVE ensemble, but not K and U, so systems not in equilibrium will go to equilibrium while the temperature changes
  - Need to modify the Lagrangian or couple the system to a heat or pressure bath to keep T and/or P constant

Monte Carlo Practical aspects

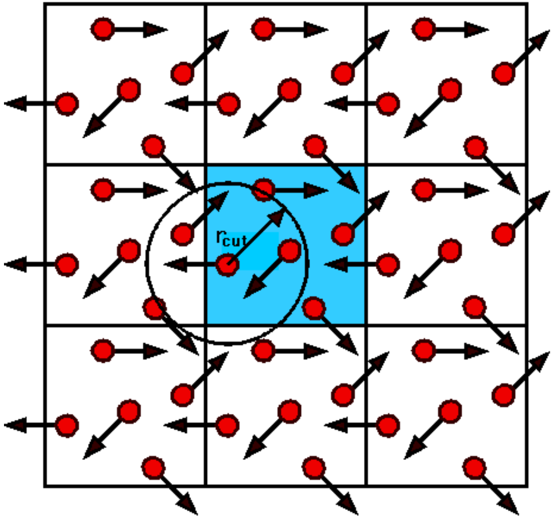
- Same as for Molecular Dynamics simulations
- Use **periodic boundary conditions (PBC)** to avoid surface effects
    - Atoms in surface ~ $N^{-1/3}$ (49% for a box containing 1000 atoms, 6% for 10^6)
    - Need PBC to get bulk properties using a limited number of atoms
    - When an atom leaves the simulation cell, it is replaced by another with the same velocity, entering from the opposite cell face (N conserved)



- Normally OK, but beware of artificial periodicity
    - Only allowed fluctuations: λ compatible with box length
    - Longest wavelength $\lambda = L$ ✉ problem if long wavelength fluctuations are important (e.g. phase transitions)
    - When simulating solids, the strain field generated by inhomogeneities will be artificially truncated and modified by the boundary
    - A macromolecule may interact with its own image if the box is too small ✉ need to add layer of water sufficiently large
    - In some cases, spurious correlations may appear when dealing with charged or strongly polar solutes

Monte Carlo Practical aspects II

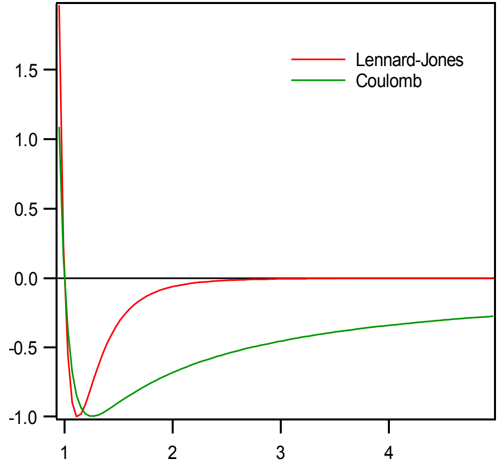
- Potential truncation of non-bonding interactions
- This is the most time-consuming part of the simulation
  - For bonding interactions ✉ O(N)
  - For non-bonding interactions ✉ O(N^2)
- Unfeasible to compute interactions with a very large number of images



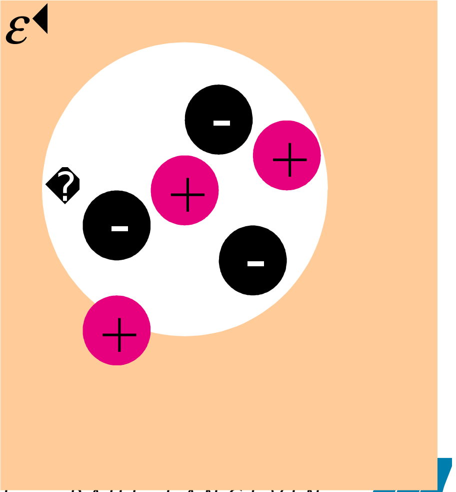
- Minimum image convention
  - Each atom only sees the closest image of the remaining N-1 atoms
  - “Truncation” surface not spherical
- Or apply an spherical cut-off
  - Only interactions with atoms at $r < R_c$ are considered
  - $R_c < L/2$. Typically
  - Introduces a discontinuity in energies and forces (for MD) at R_c. Often the potential is modified in order to bring it smoothly to 0 at R_c (switching function)

Monte Carlo Practical aspects III

- Treatment of long-range (electrostatic) interactions
- Truncation method is not appropriate, as they extend beyond the primitive cell
- **Reaction Field:**
  - Use spherical truncation and compute interactions for $r < R_c$
  - Beyond R_c, consider the medium as a dielectric continuum that will react to the dipole moment of the sphere given by $r < R_c$

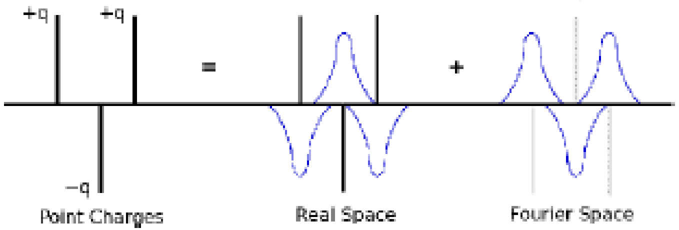


$$U_{rf} = \frac{1}{2}\sum_{i=1}^{N} \mu_i \cdot \left(\frac{2(\varepsilon' - 1)}{2\varepsilon' + 1} \frac{1}{r_c^3} \sum_{j \neq i} \mu_j \right)$$



Monte Carlo Practical aspects IV

- Treatment of long-range (electrostatic) interactions
- Compute the full lattice sum over the infinite set of images
- **Ewald sum:**
  - Reference method, often in one of its particle-mesh variants (PME, SPME, PPPE)
  - At each point charge, add a Gaussian distribution of opposite sign to screen it
  - Screened charges can be treated in real space (short-range interactions)
  - Compensating Gaussian distribution treated in reciprocal space



$$\rho = -q_i \left(\frac{\alpha}{\pi} \right)^{3/2} \exp\left(-\alpha r^2\right) \qquad U_{sr} = \frac{1}{2} \sum_{i \neq j}^{N} q_i q_j \operatorname{erfc}\left(\sqrt{\alpha}\, r_{ij}\right) / r_{ij}$$

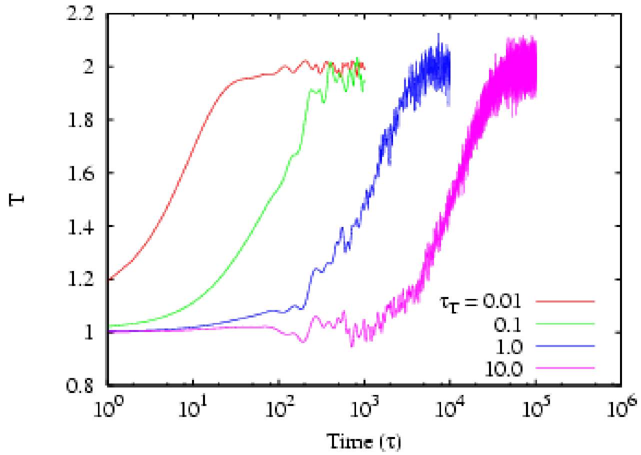
$$U_{lr} = \frac{1}{2V} \sum_{\boldsymbol{k} \neq 0} \frac{4\pi}{k^2} |\rho(\boldsymbol{k})|^2 \exp\left(-k^2/4\alpha\right) \qquad U_{\text{self}} = \left(\frac{\alpha}{\pi} \right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2$$

$$\text{with } \rho(\boldsymbol{k}) = \sum_{i=1}^{N} q_i \exp\left(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{r}_i\right)$$

Molecular Dynamics The basics VI

Thermodynamic ensembles

- Several thermostats allow to do **NVT** simulations: velocity scaling, Berendsen, Andersen, Nosé-Hoover, but not all of them sample strictly the correct thermodynamic ensemble.
- There are also different barostats to do **NPT** simulations.



Berendsen: Simple, fast response and flexible, but no canonical ensemble

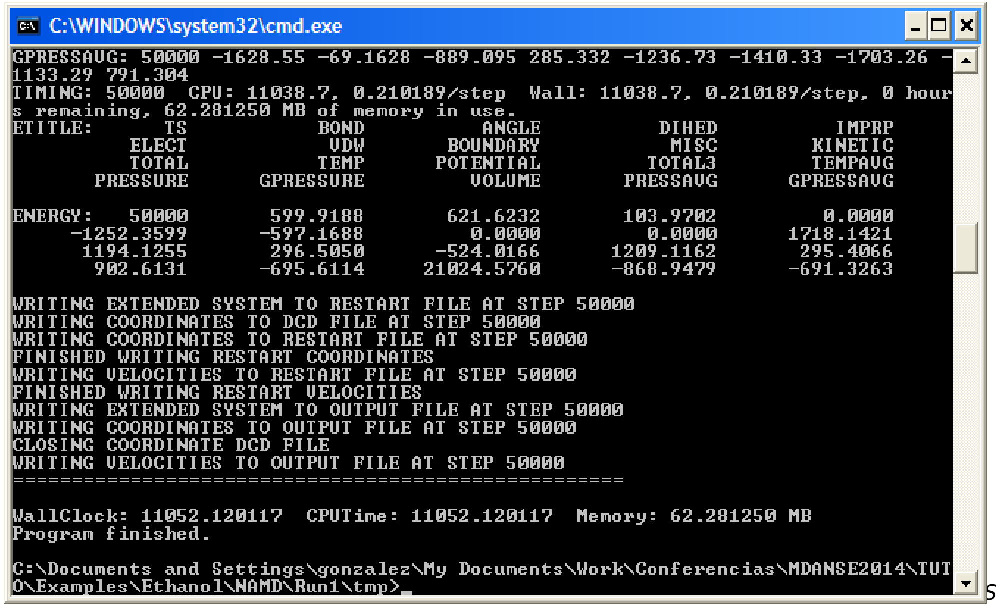
$$\boldsymbol{a} = \boldsymbol{F}/m + \frac{1}{2\tau_T}\left(\frac{T_B}{T(t)} - 1\right)\boldsymbol{v}$$

Nosé-Hoover: Extended lagrangian. Sample correct canonical ensemble, but goes to T_{ref} in an oscillatory way (poorer T control)

Analysis of MD simulations

Once our MD run is over …

- Snapshot of the final state of the system
- Trajectory file containing positions (and velocities, if requested) at selected intervals
- File containing thermodynamic information (T, P, energies, etc.)



```
GPRESSAVG: 50000 -1628.55 -69.1628 -889.095 285.332 -1236.73 -1410.33 -1703.26 -
1133.29 791.304
TIMING: 50000  CPU: 11038.7, 0.210189/step  Wall: 11038.7, 0.210189/step, 0 hour
s remaining, 62.281250 MB of memory in use.
ETITLE:      TS           BOND          ANGLE          DIHED          IMPRP
          ELECT            VDW       BOUNDARY           MISC        KINETIC
          TOTAL           TEMP      POTENTIAL         TOTAL3        TEMPAVG
       PRESSURE      GPRESSURE         VOLUME       PRESSAVG      GPRESSAVG

ENERGY:   50000       599.9188       621.6232       103.9702         0.0000
     -1252.3599      -597.1688         0.0000         0.0000      1718.1421
      1194.1255       296.5050      -524.0166      1209.1162       295.4066
       902.6131      -695.6114     21024.5760      -868.9479      -691.3263

WRITING EXTENDED SYSTEM TO RESTART FILE AT STEP 50000
WRITING COORDINATES TO DCD FILE AT STEP 50000
WRITING COORDINATES TO RESTART FILE AT STEP 50000
FINISHED WRITING RESTART COORDINATES
WRITING VELOCITIES TO RESTART FILE AT STEP 50000
FINISHED WRITING RESTART VELOCITIES
WRITING EXTENDED SYSTEM TO OUTPUT FILE AT STEP 50000
WRITING COORDINATES TO OUTPUT FILE AT STEP 50000
CLOSING COORDINATE DCD FILE
WRITING VELOCITIES TO OUTPUT FILE AT STEP 50000
====================================================

WallClock: 11052.120117  CPUTime: 11052.120117  Memory: 62.281250 MB
Program finished.

C:\Documents and Settings\gonzalez\My Documents\Work\Conferencias\MDANSE2014\TUT
O\Examples\Ethanol\NAMD\Run1\tmp>_
```

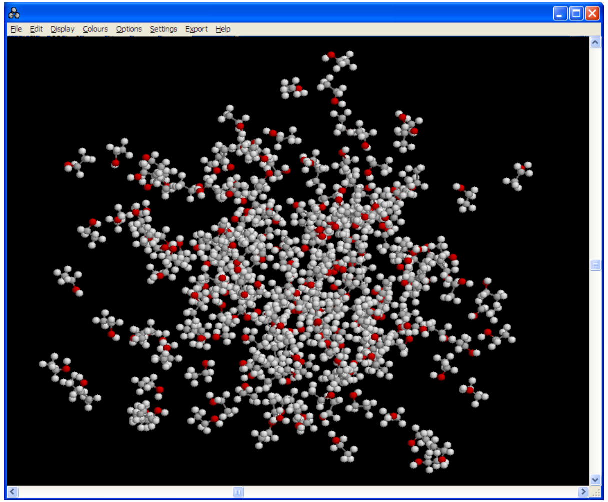
Did everything go well?

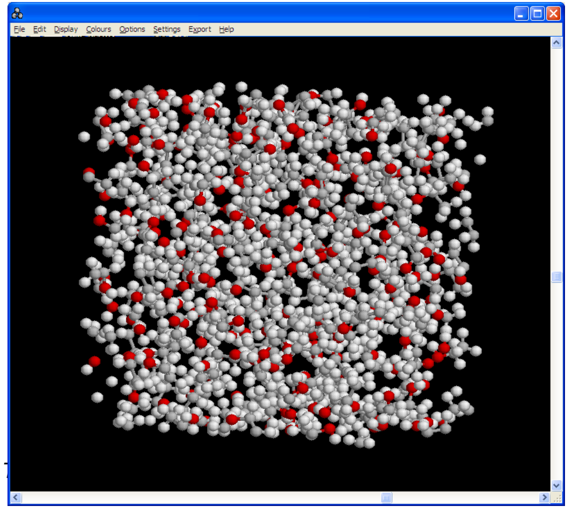
Can I get the information I'm looking for?

Once our MD run is over …

Visual checking of last configuration:

- Many visualization tools available (VMD, Rasmol, MDANSE, …)
- Don’t be puzzled by PBC’s (remember that the simulated system is really an infinite replica of images of the primary box)
- Are molecular geometries reasonable?
- Are there pairs of atoms that come ‘too close’ to each other?

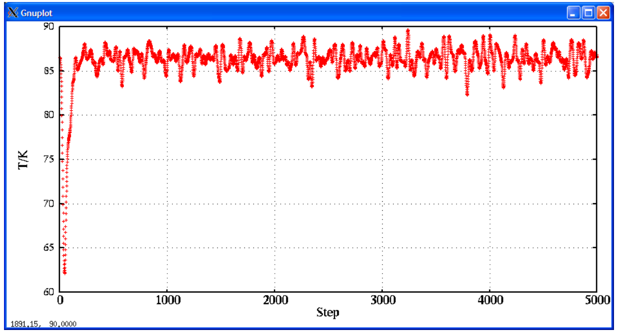


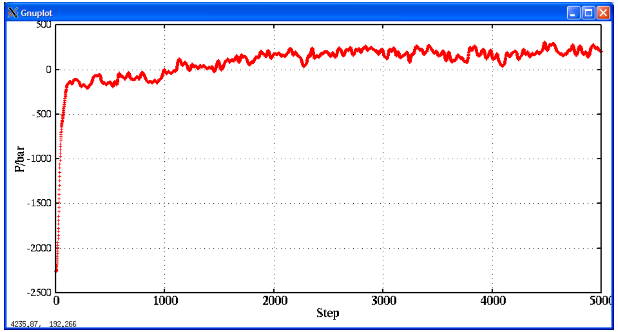


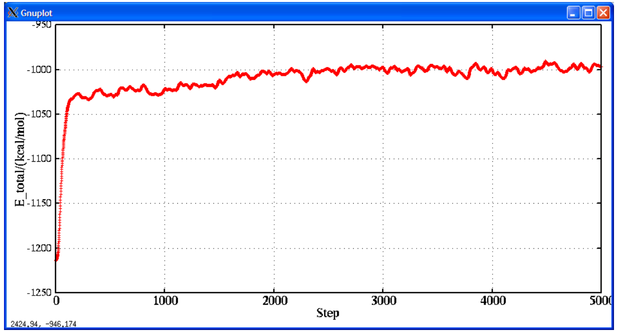
Once our MD run is over …

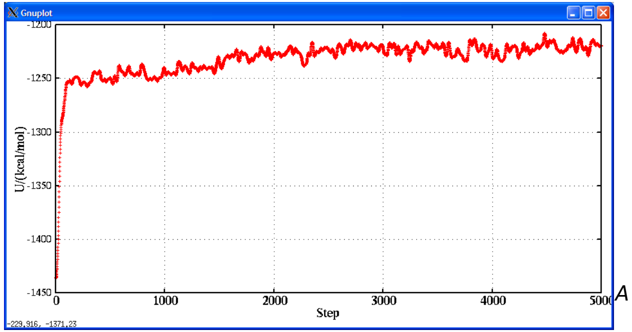
Is the system well equilibrated?

- Not all the properties converge at same speed



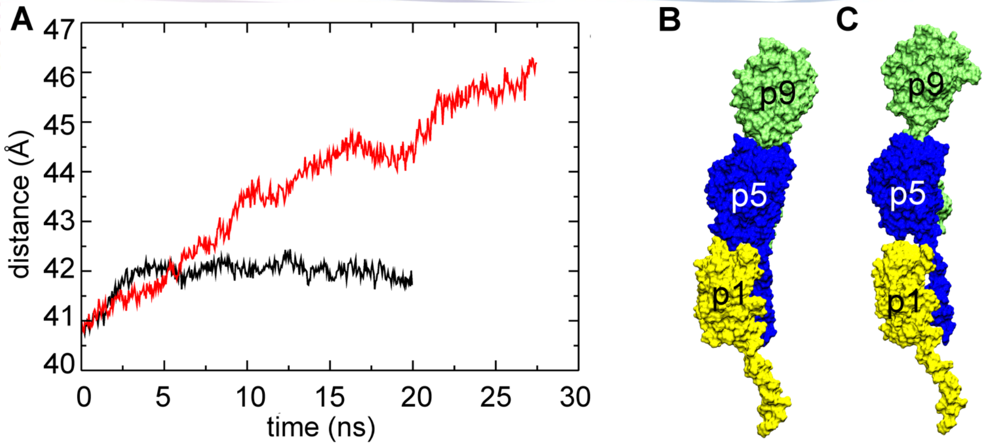






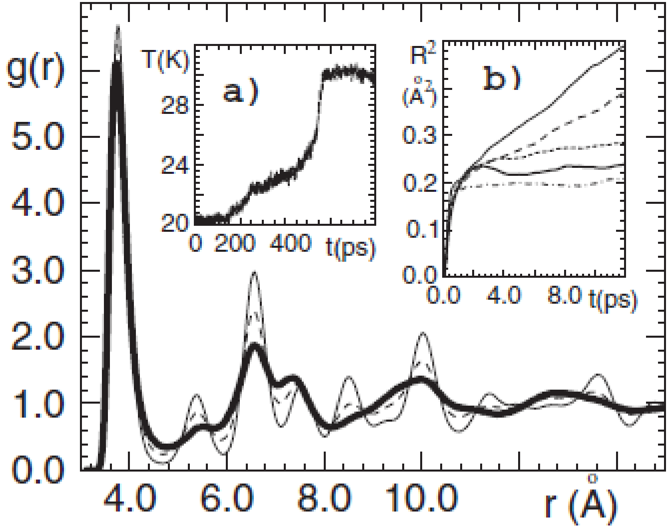
Once our MD run is over ...

Baker, PLOS Comp. Biol. (2013)

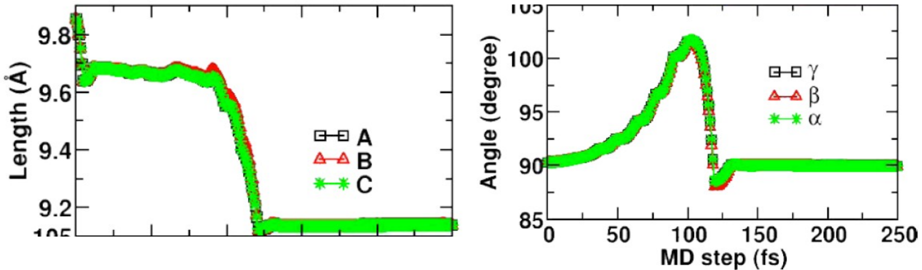


Does the system change during the simulation?

De Santis, JPCM **14** (2002)



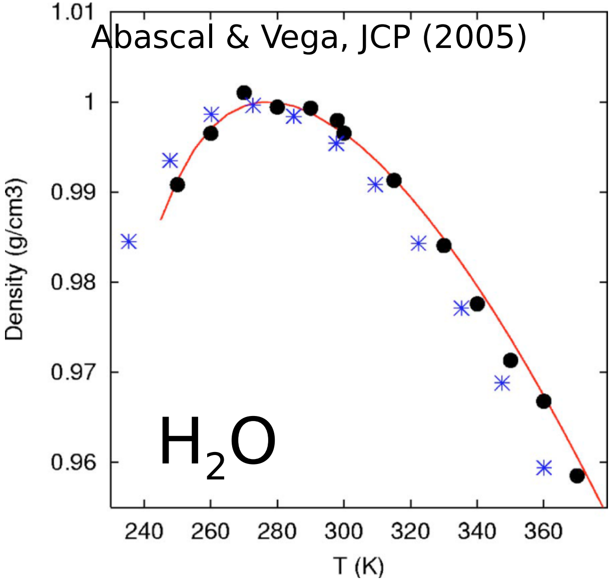
Durandurdu, JPCM **21** (2009)



Thermodynamic averages

Does our model give reasonable average values for T, P, density, …?

- T, P, V, energy contributions are usually directly given by the MD code
- Other thermodynamic properties may need to be calculated later
- If possible, compare with as much experimental data as possible
- Note that formulae based on fluctuations depend on the ensemble



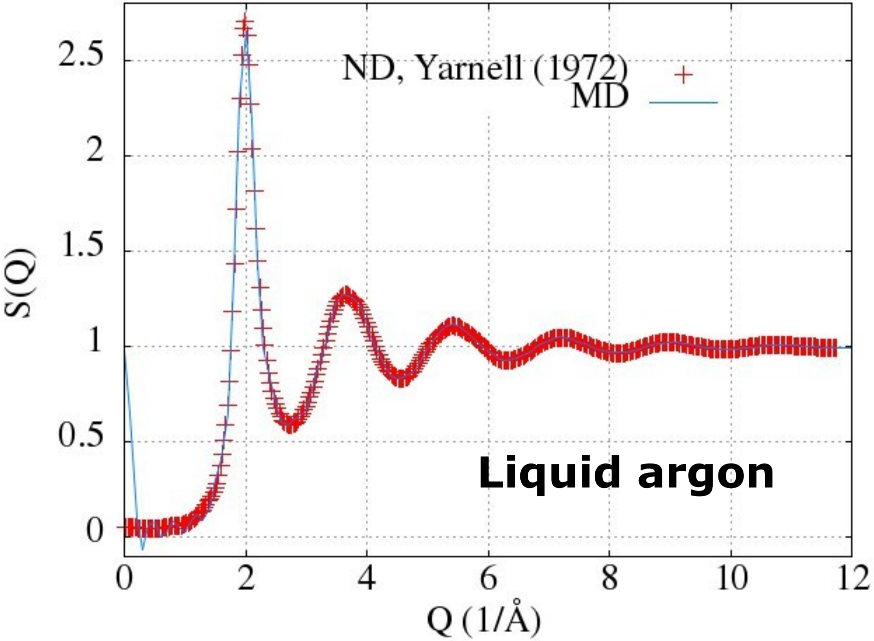
$$\Delta H_{vap} = H_{gas} - H_{liq} = (U + PV)_{gas} - (U + PV)_{liq}$$

$$C_V = \frac{\left\langle \left(H - \langle H\rangle\right)^2 \right\rangle}{k_B T^2} = \frac{\left\langle \left(U - \langle U\rangle\right)^2 \right\rangle}{k_B T^2} + \frac{3}{2} N k_B T \qquad \text{(NVT)}$$

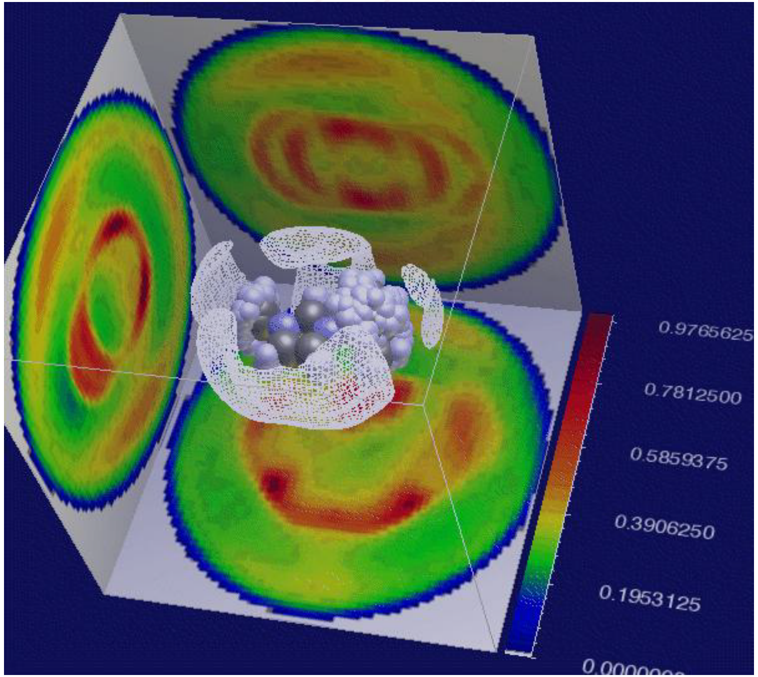
$$C_V = \frac{9 N k_B^3 \langle T\rangle^2}{6 k_B^2 \langle T\rangle^2 - 4\left\langle \left(U - \langle U\rangle\right)^2 \right\rangle} \qquad \text{(NVE)}$$

Structural properties

The static structure factor can be easily calculated and compared to neutron or x-ray diffraction data



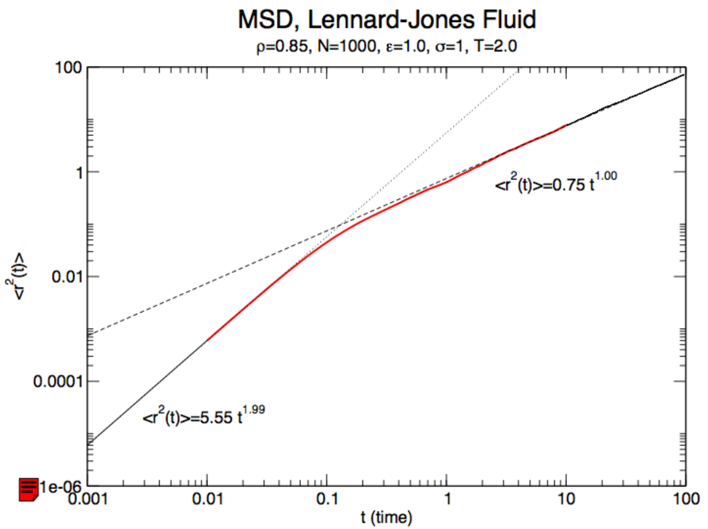
Liquid argon

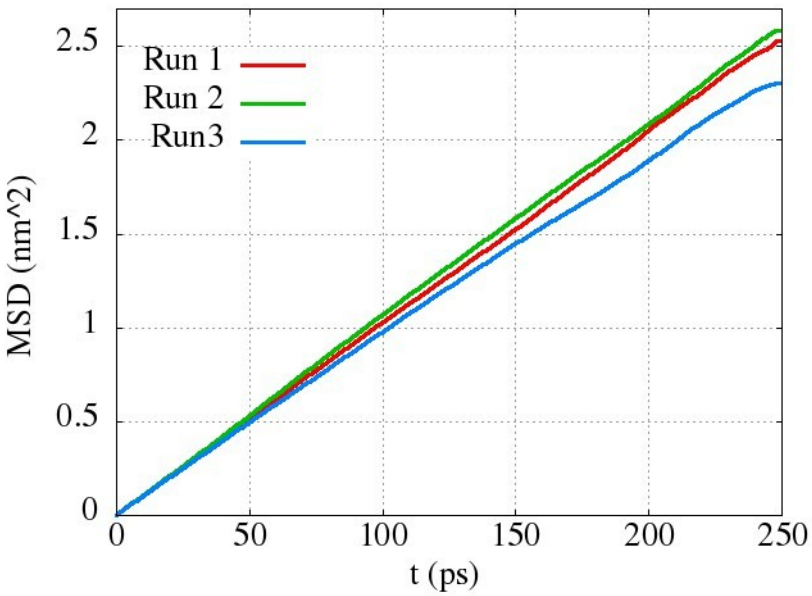


Dynamical properties

The mean square displacement (total or for a particular atom) can be easily calculated and provides a direct way to compute the self-diffusion coefficient (available e.g. from NMR, QENS or tracer diffusion)

- $u^2(t)$ can be calculated over different time origins, but the number of origins decreases with t, so error in $u^2(t)$ will increase with time

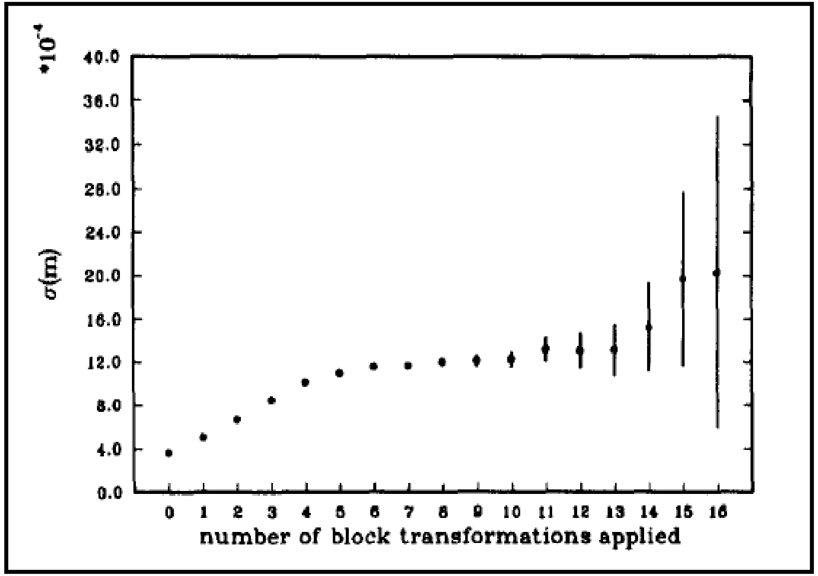




Error analysis

Error in correlated measurements: Block averaging method

- Flyvbjerg and Petersen, *J. Chem. Phys.* **91**, 461 (1989)
- Calculate averages over M_b blocks of increasing size b
- As b increases, the blocks are less correlated until they become statistically independent. Then a plateau is observed which gives us the correct σ



Error analysis

Error in time correlation functions

Error in time correlation functions

- Zwanzig and Ailawadi, *Phys. Rev.* **182**, 280 (1969)

- Zwanzig and Ailawadi, *Phys. Rev.* **182**, 280 (1969)

- For an autocorrelation function $C_X^{MD}(t) = \langle X(t_0+t)X(t_0)\rangle_{MD}$, assuming that X($t_0$) is a Gaussian random variable, the variance is

- For an autocorrelation function , assuming that X(t_0) is a Gaussian random variable, the variance is

$$\sigma^2(\langle X(t_0+t)X(t_0)\rangle_{MD}) \approx \frac{2\tau_X C_X^2(0)}{t_{MD}}$$

τ_X: Mean relaxation time for X,

t_{MD}: Total simulation time

- And for a normalized a.c.f.

- And for a normalized a.c.f.

- E.g.:

$$\sigma\left(\frac{\langle X(t_0+t)X(t_0)\rangle_{MD}}{\langle X^2\rangle_{MD}}\right) \approx \sqrt{\frac{2\tau_X}{t_{MD}}} \times (1 - C_X^{MD}(t))$$

- E.g.: $\frac{t_{MD}}{\tau_X} = 1000 \ \Rightarrow \ \sigma \approx 0.04 \text{ (for } t \gg \tau_X, \text{ so } C_X^{MD}(t) \approx 0)$

$$\frac{t_{MD}}{\tau_X} = 100 \ \Rightarrow \ \sigma \approx 0.14$$

$$\frac{t_{MD}}{\tau_X} = 10 \ \Rightarrow \ \sigma \approx 0.45$$

But things are much better for single-particle c.f.:

$$\boldsymbol{\sigma} \approx \sqrt{\frac{2\tau_X}{N t_{MD}}}$$

Comparing to neutron scattering data

G. R. Kneller et al.: "nMOLDYN: A program package for a neutron scattering oriented analysis of Molecular Dynamics simulations", *Comp. Phys. Comm.* **91**, 191 (1995)

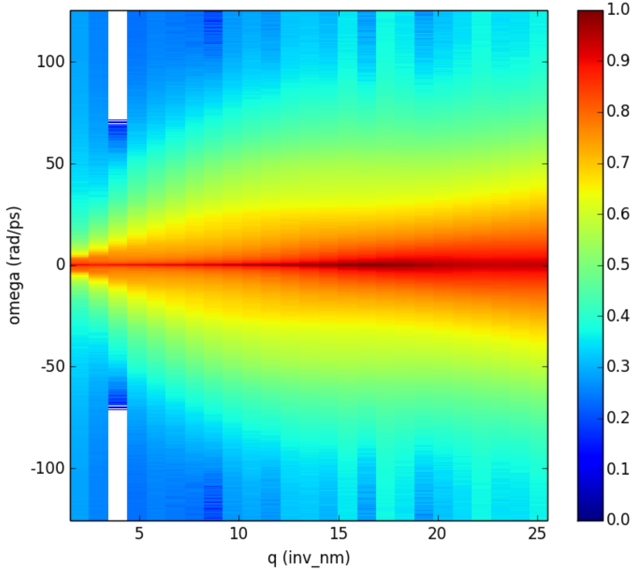
$$S(\mathbf{q}) = \frac{1}{N}\sum_{\alpha,\beta} b_{\alpha,\mathrm{coh}} b_{\beta,\mathrm{coh}} \langle \exp[i\mathbf{q}\cdot(\mathbf{R}_\beta - \mathbf{R}_\alpha)]\rangle.$$

$$\mathcal{F}_{\mathrm{coh}}(\mathbf{q},t) = \frac{1}{N}\sum_{\alpha,\beta} b_{\alpha,\mathrm{coh}} b_{\beta,\mathrm{coh}} \langle \exp[-i\mathbf{q}\cdot\mathbf{R}_\alpha(0)]\exp[i\mathbf{q}\cdot\mathbf{R}_\beta(t)]\rangle.$$

$$\mathcal{F}_{\mathrm{inc}}(\mathbf{q},t) = \frac{1}{N}\sum_{\alpha} b^2_{\alpha,\mathrm{inc}} \langle \exp[-i\mathbf{q}\cdot\mathbf{R}_\alpha(0)]\exp[i\mathbf{q}\cdot\mathbf{R}_\alpha(t)]\rangle,$$

$$EISF(\mathbf{q}) = \lim_{t\to\infty} \mathcal{F}_{\mathrm{inc}}(\mathbf{q},t),$$

$$EISF(\mathbf{q}) = \frac{1}{N}\sum_{\alpha} b^2_{\alpha,\mathrm{inc}} \langle |\exp[i\mathbf{q}\cdot\mathbf{R}_\alpha]|^2\rangle,$$



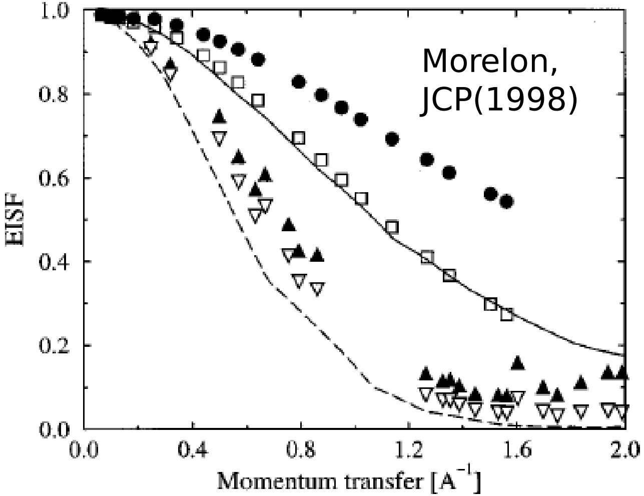
Comparing to neutron scattering data

Think carefully how measured and computed quantities are exactly determined

The EISFs were computed from the MD trajectories with the relation:

$$\mathrm{EISF}(Q)=\sum_i b_i^2 |\langle\langle e^{-i\mathbf{Q}\cdot\mathbf{R}_i(t)}\rangle_t\rangle_{\hat{Q}}|^2, \tag{19}$$

$$EISF(\mathbf{q}) = \lim_{t\to\infty} \mathcal{F}_{\mathrm{inc}}(\mathbf{q}, t),$$



Morelon, JCP(1998)

In the quasielastic energy transfer range, the dynamical structure factor can be written as:[9]

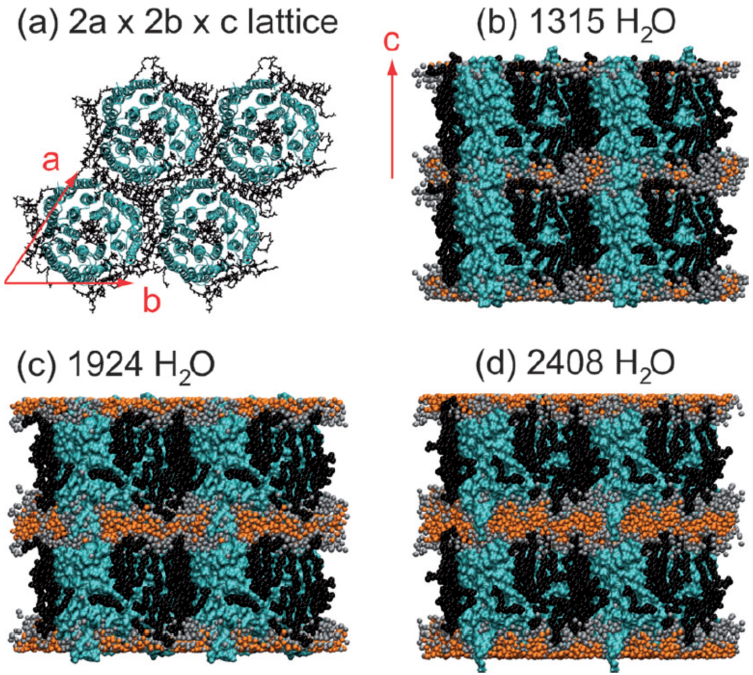
$$S(\mathbf{Q},\omega)=e^{-Q^2\langle u^2\rangle/6}\times[I^D(\mathbf{Q},+\infty)\,\delta(\omega)+\Gamma(\mathbf{Q},\omega)] + Background, \tag{14}$$

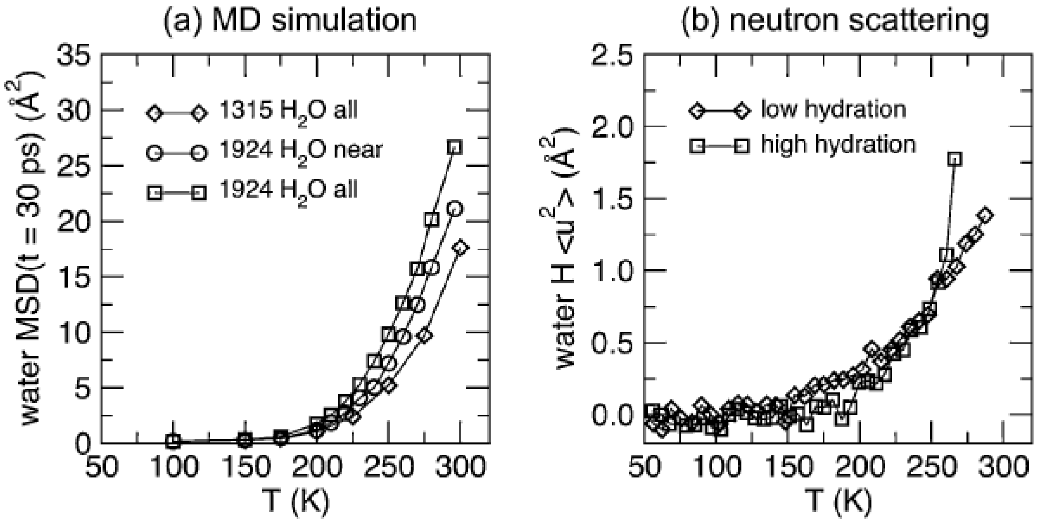
$$\mathrm{EISF}(\mathbf{Q})=I^D(\mathbf{Q},+\infty).$$

Due to the experimental EISF extraction procedure, the vibrational librational motions do not contribute to the measured EISF. In constrast, the simulation-derived EISFs obtained from Eq. (19) are computed without decomposition of the atomic trajectories into their vibrational and diffusive parts. The relation between the experimental and simulated elastic incoherent structure factors is therefore, with a spherical averaging:

$$\mathrm{EISF}_{\mathrm{sim}}(Q)=e^{-Q^2\langle u^2\rangle/6}\,\mathrm{EISF}_{\mathrm{exp}}(Q). \tag{20}$$

Comparing to neutron scattering data



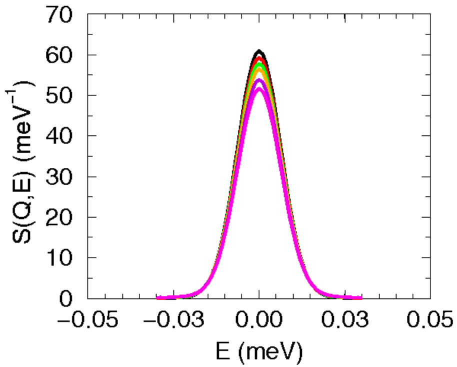


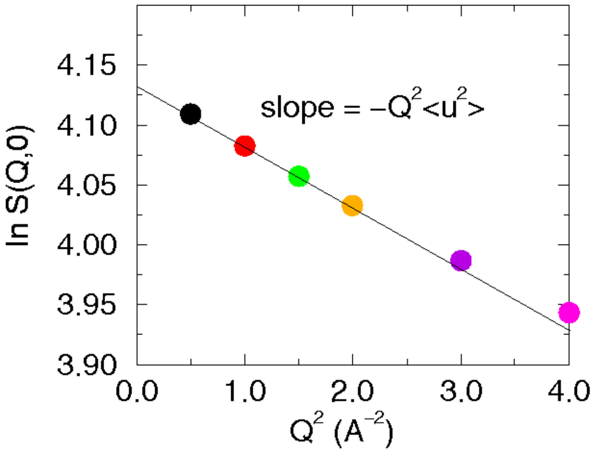
From Tobias, Faraday Discuss. (2009)

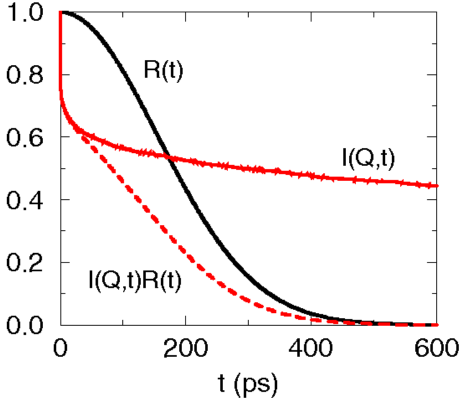
Comparing to neutron scattering data

MD: MSD computed directly as $\boldsymbol{u^2(t) = \langle [r(t_0 + t) - r(t_0)]^2 \rangle}$

Neutron scattering: MSD from elastic intensity via DW $I_{el} \approx e^{-Q^2 u(t)^2}$







Can not compare the msd directly obtained from MD.

Instead compute I(Q,t) ✉ S(Q,ω) ✉ S(Q,ω) ⊗ R(ω) and then follow same procedure as for experiment.

Comparing to neutron scattering data

Standard MD treats nuclei as classical particles (for both empirical and AIMD)

Computed correlation functions are classical correlation functions

Limits in computed S(Q,ω):

- Quantum nature of the simulated system must be negligible (otherwise need to consider methods of incorporating quantum mechanics in the simulation, e.g. using PIMD)
- $\hbar\omega \ll k_B T$, so care when interpreting high-frequency features or low temperature studies (For $T = 298\,K \Rightarrow \nu \approx 200\ \mathrm{cm}^{-1}\ (25\ \mathrm{meV}), \tau \approx 0.2\ \mathrm{ps}$)
- Recoil effect on scattering atom must be negligible (Kneller, *Mol. Phys.* **83**, 63 (1994)), so $\frac{\hbar^2 Q^2}{2M_I} \ll k_B T$
- Analysis soft use tabulated **bound** coh and inc neutron scatt lengths

The end