Molecular Dynamics

Lecture 1: Basic Ideas, Newtonian motion, trajectories, sampling.

David J. Willock

Cardiff University, U.K.

Books

"Molecular Modelling, Principles & applications", A.R.Leach

"Computer Simulation of Liquids", M.P.Allen and D.J.Tildesley, Oxford Science Pubs., (1993), ISBN 0-19-855645-4

"Understanding Molecular Simulation", Daan Frenkel and Berend Smit, Academic Press, (1996), ISBN 0-12-267370-0

"Theory of Simple Liquids", J.-P. Hansen and I.R. McDonald, Academic Press (1986).

"The Art of Molecular Dynamics Simulation", D.C. Rapaport, Camb. Univ. Press (2004)

"Classical Mechanics", H. Goldstein, Addison-Wesley.

What is Molecular Dynamics?

- MD is the solution of the <u>classical</u> equations of motion for atoms and molecules to obtain the <u>time evolution</u> of the system**.**

- Applied to many-particle systems - a general analytical solution not possible. Must resort to numerical methods and computers.

- Classical mechanics only - fully fledged many-particle time-dependent quantum method not yet available.

- Gives time evolution of systems: trajectories. These allow time dependant phenomena, such as diffusion, to be followed.

- Maxwell-Boltzmann averaging process for thermodynamic properties (time averaging).

Lennard-Jones Spheres (Argon).

Pair Potential: $V(r_{ij}) = 4\varepsilon\left\{\left(\dfrac{\sigma}{r_{ij}}\right)^{12} - \left(\dfrac{\sigma}{r_{ij}}\right)^{6}\right\}$
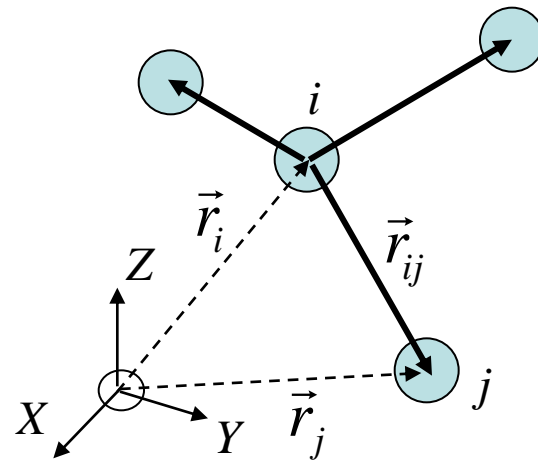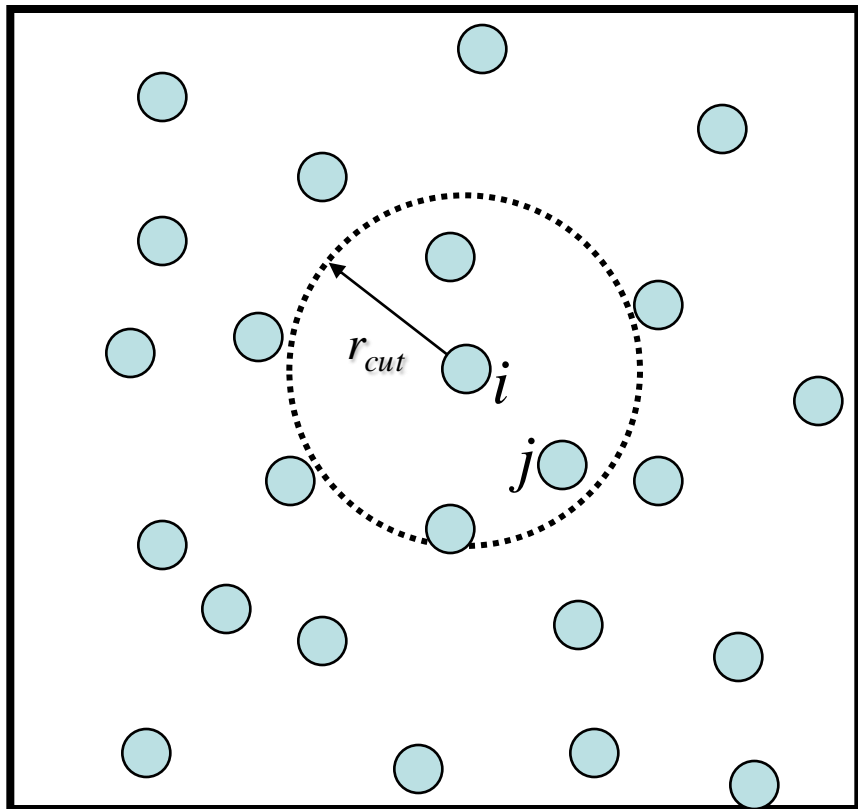
Pair force:

$$f_{ix} = -\frac{\partial V(r_{ij})}{\partial x_i} = -\frac{\partial V(r_{ij})}{\partial r_{ij}}\frac{\partial r_{ij}}{\partial x_i}$$
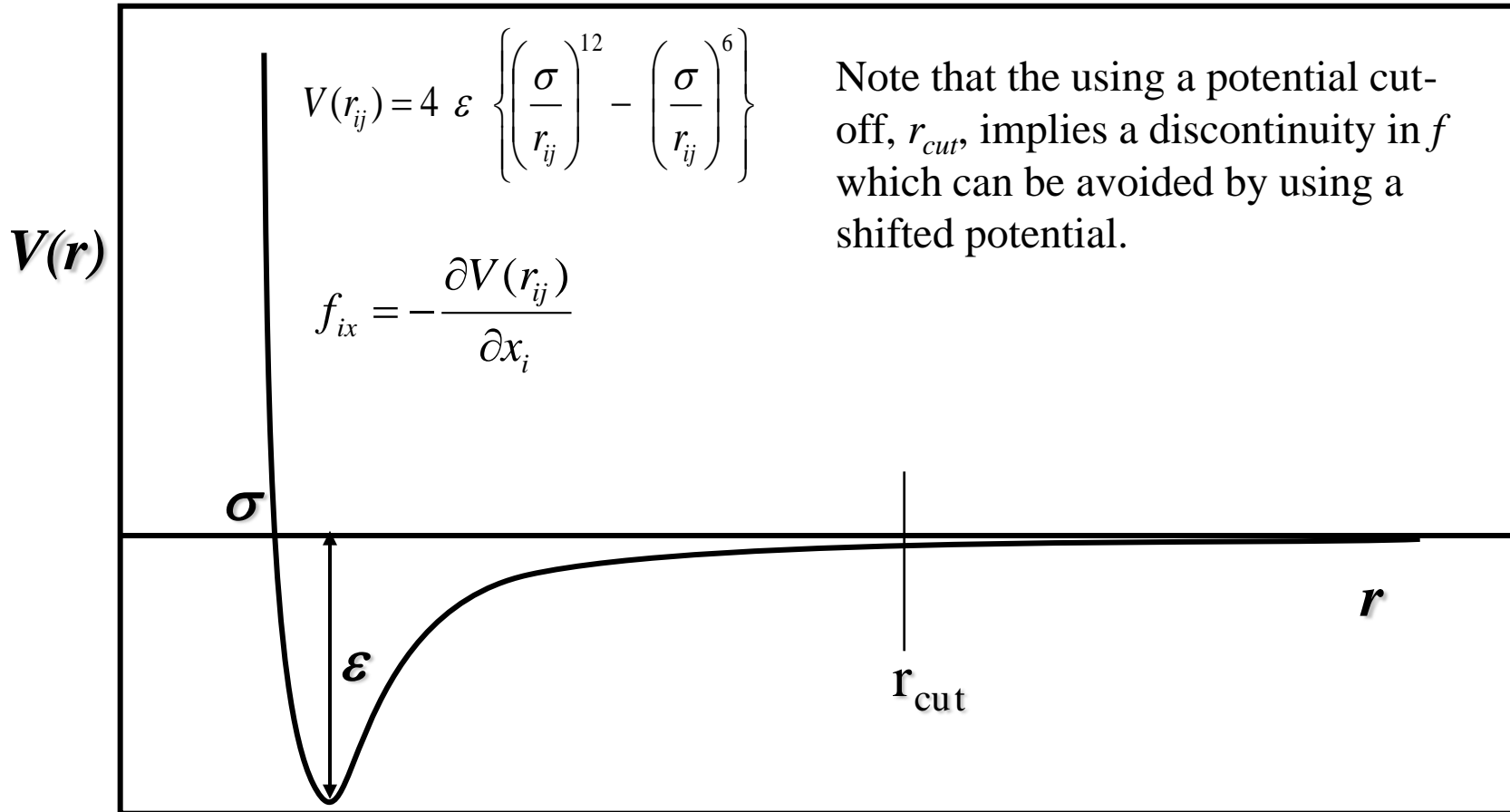
$$\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$$

$$r_{ij} = \left((x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2\right)^{\frac{1}{2}}$$

$$\frac{\partial r_{ij}}{\partial x_i} = -\frac{(x_j - x_i)}{r_{ij}} = -\frac{r_{ijx}}{r_{ij}}$$

Note this leads to equal and opposite forces on the two particles.

r_{cut}

4

Lennard -Jones Potential

$$V(r_{ij}) = 4\,\varepsilon \left\{ \left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right\}$$

Note that the using a potential cut-off, r_{cut}, implies a discontinuity in f which can be avoided by using a shifted potential.

$$f_{ix} = - \frac{\partial V(r_{ij})}{\partial x_i}$$

V(r)

σ

ε

r

r_{cut}

Molecular Dynamics: Physical properties, *e.g.* water

Molecular dynamics can be used to simulate the collective behaviour of large systems and so provide ensemble averages as estimates of marcoscopic properties.
Water has been studied extensively: Important and accurate experimental data.

Table 2 Calculated physical properties of the presented water models and their respective experimental values (all the data are at 298 K and 1 atm, except * at 300 K and [#] at 293 K)

Model	SPC	SPC/E	TIP3P	TIP4P	Experimental
Density, ρ (kg/m^3)	970*[a]	**998*[a]**	1002[i]	1001[i]	997[i]
Dipole moment, μ (D)	2.27[b]	**2.39[b]**	2.35[b]	2.18[b]	2.95[l]
Dielectric constant, ε (adimensional)	65[c]	**68[c]**	97[c]	53[#j] 29*[k]	78.39[m]
Self-diffusion constant, D (10^{-5} cm^2/s)	3.85[d]	**2.49[d]**	5.19[d]	3.31[d]	2.30[n]
Average configurational energy, U (kJ mol^{-1})	−37.7[a]	**−41.4[a]**	−41.1[i]	−42.1[i]	−41.5[o]
Density maximum, T_{TMD} (K)	228[e]	235[g]	182[e]	**253[e]**	277.13[p]
Expansion coefficient, α_p (10^{-4} °C^{-1})	7.3*[f]	5.14[h]	9.2[i]	**4.4[i]**	2.53[q]

[a] Berendsen et al. (1987), [b] van der Spoel et al. (1998), [c] Höchtl et al. (1998), [d] Mahoney and Jorgensen (2001), [e] Vega and Abascal (2005), [f] Yu and van Gunsteren (2004), [g] Báez and Clancy (1994), [h] Wu et al. (2006), [i] Mahoney and Jorgensen (2000), [j] Neumann (1986), [k] Strauch and Cummings (1989), [l] Gubskaya and Kusalik (2002), [m] Vidulich et al. (1967), [n] Krynicki et al. (1978), [o] Postma (1985), [p] Franks (2000), [q] Kell (1975)
Physical properties that show most similar values to experimental determinations are indicated in bold
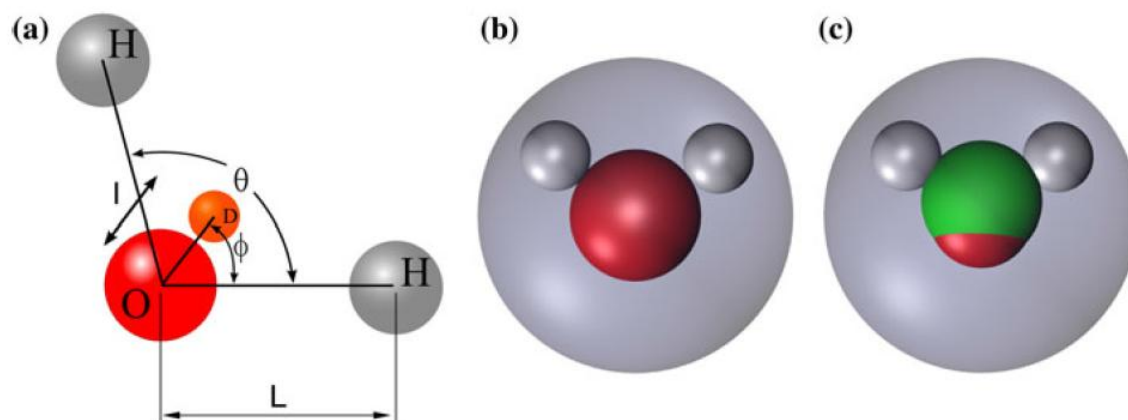
Taken from Ozu *et al.*, Eur. Biophys. J. (2013) **42**: 223

Table 1 Parameters of the described models

Model	L-J parameters		L (10^{-10} m)	l (10^{-10} m)	q_H (e^-)	q_O (e^-)	$\theta°$	$\phi°$
	σ (10^{-10} m)	ε (kJ mol^{-1})						
SPC	3.166	0.650	1.0000	–	+0.410	−0.8200	109.47	–
SPC/E	3.166	0.650	1.0000	–	+0.4238	−0.8476	109.47	–
TIP3P	3.15061	0.6364	0.9572	–	+0.4170	−0.8340	104.52	–
TIP4P	3.15365	0.6480	0.9572	0.15	+0.5200	−1.0400	104.52	52.26

σ: Lennard-Jones collision diameter; ε: depth of the potential well; L: distance between the centres of mass (COM) of the oxygen and hydrogen atoms. qH: positive charge; qO: negative charge; is called $\theta°$: HOH angle; $\varphi°$: angle between the OH segment and the OD segment (i.e. $\theta° = 2\varphi°$); l: distance between the dummy position (D in Fig. 1) where the negative charge is placed (near the oxygen along the bisector of $\theta°$) and the oxygen centre of mass

Taken from Ozu *et al.*, Eur. Biophys. J. (2013) **42**: 223

Water models: TIP4P and simple three site models
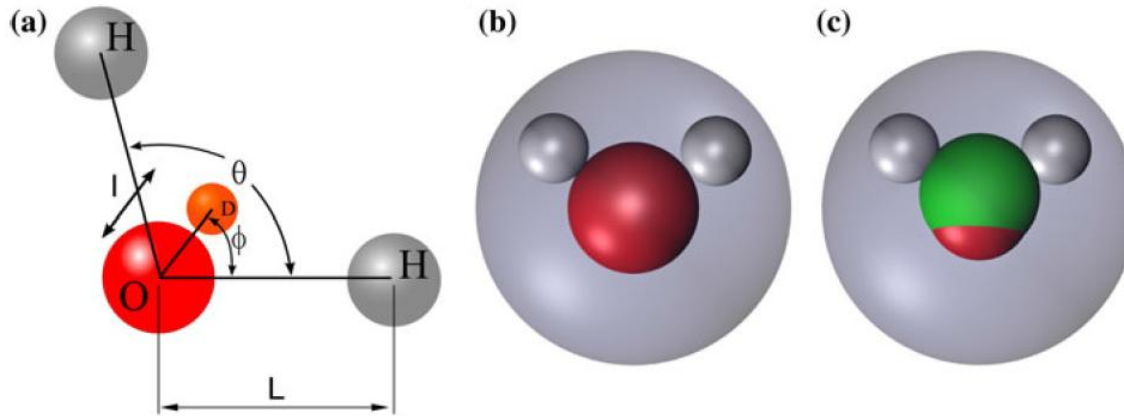
Water TIP4P/2005
UNITS kcal
ATOMS 4
```
   ow      15.9996      0.0000    1
   hw       1.00080     0.5564    1
   hw       1.00080     0.5564    1
   M        0.10000    -1.1128    1
rigid 1
4  1 2 3 4
FINISH
VDW   1
   ow     ow  lj   0.18519   3.1589
close
```

Part of a DLPOLY FIELD file for TIP4P water.

ATOMS lines read as "mass" "charge" "number"
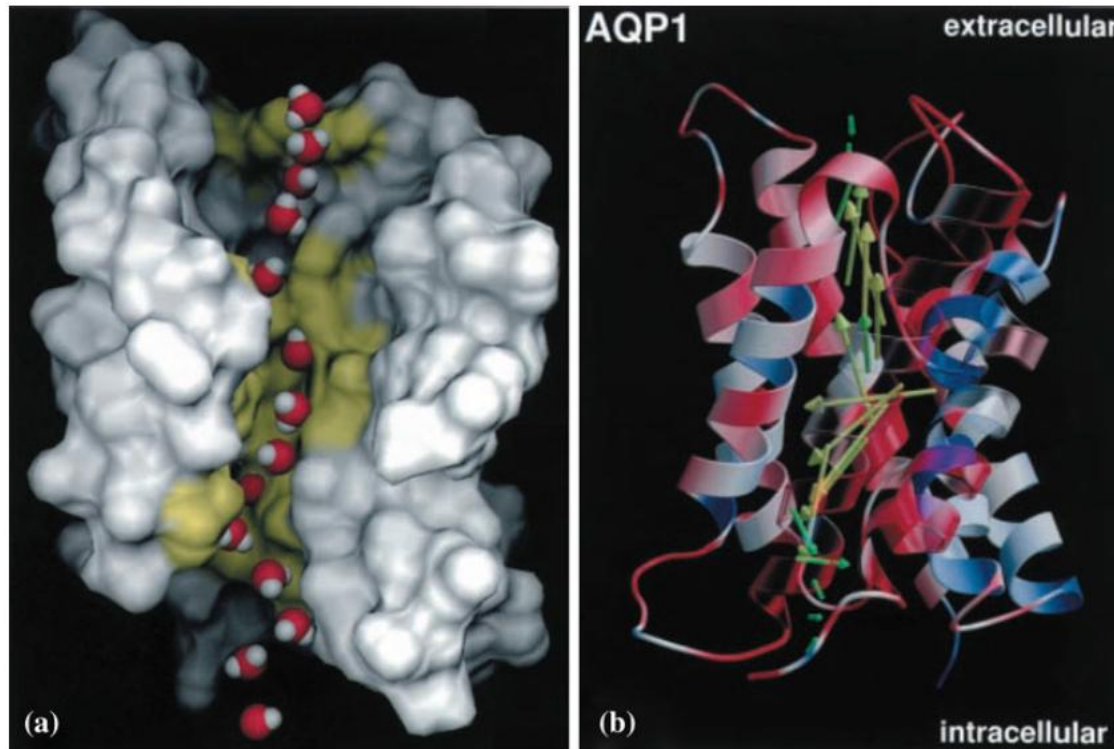
Additional site marked as "M"

Molecule treated as a rigid body.

8

Fig. 4 Schematic representation of AQP1. **a** Snapshot showing the single file of water molecules inside the channel. **b** Scheme of AQP1 showing the water dipole inversion (*green arrows*) during translocation through the channel. Adapted from de Groot and Grubmüller (2001), ISSN 0036-8075 (print), ISSN 1095-9203 (online). Reprinted with permission from AAAS

Taken from Ozu *et al.*, Eur. Biophys. J. (2013) **42**: 223

Equations of Motion, the Lagrangian

The Lagrangian is the difference between the system kinetic and potential energy:

$$L = K(\dot{q}) - U(q)$$

Here q is a general co-ordinate and a dot over a letter means a time derivative (Newton). In forcefield dynamics U depends only on position, q, and K only on velocity $\dot{q}$.

$$\frac{\partial}{\partial t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) = \frac{\partial L}{\partial q_i}$$

General property of the Lagrangian: the time derivative of the momentum is equal to the derivative of L w.r.t. position. Allows a relationship between the co-ordinate and its time derivatives to be established.

For our gas of Lennard-Jones Argon atoms:

$$L(\vec{r}_i, \vec{v}_i) = \frac{1}{2}\sum_i^N m_i v_i^2 - \sum_i \sum_{j>i}^{N-1} V(r_{ij})$$

q is the atom co-ordinates and we have
$$\dot{q}_i = \frac{\partial r_i}{\partial t} = v_i$$

For particle *i*:

LHS of gen. prop.
$$\frac{\partial L}{\partial v_i} = m_i v_i$$
Momentum, p_i

so

$$\frac{\partial}{\partial t}\left(\frac{\partial L}{\partial v_i}\right) = m_i \frac{\partial v_i}{\partial t}$$

RHS of gen. prop.
$$\frac{\partial L}{\partial r_i} = -\frac{\partial V}{\partial r_i}$$

giving
$$-\frac{\partial V}{\partial r_i} = m_i \frac{\partial v_i}{\partial t}$$

Which says force is rate of change of momentum, the familiar *f = ma*, for each particle.

10

$$f_i = m_i \frac{\partial v_i}{\partial t}$$

The force on atom *i* for any given configuration can be calculated from the forcefield. This equation relates force and acceleration.

$$\frac{\partial v_i}{\partial t} = \frac{f_i}{m_i}$$

So for any given configuration we know the acceleration of each particle.

$$v_i(\tau) = v_i(0) + \int_0^\tau \frac{dv_i}{dt} dt$$

If a particle has an initial velocity, $v_i(0)$, and moves under the action of this force for a time, τ, its velocity after the time, τ, will be given by integration.

$$r_i(\tau) = r_i(0) + \int_0^\tau v_i(t) dt$$

Similarly the position of the particle after time, τ, is given by an integral of the velocity.

Unfortunately once the particles move the distances governing the potential change and so the forces are altered.

Molecular dynamics is about integrating these equations of motion such that the continuous trajectories are obtained numerically.

Integration: Essential Idea

Constant v over τ would follow this path

$\underline{r}'(t+\tau)$

$\underline{v}(t)\,\tau$

Net displacement

$\underline{r}(t+\tau)$

$\underline{r}(t)$

Correction assuming acceleration is constant over τ

$\underline{f}(t)\tau^2/m$

$$r_i(\tau) = r_i(0) + \int_0^\tau v_i(t)\,dt$$

$$v_i(\tau) = v_i(0) + \int_0^\tau \frac{dv_i}{dt}\,dt$$

$[\underline{r}(t),\ \underline{v}(t),\ \underline{f}(t)] \longrightarrow [\underline{r}(t+\tau),\ \underline{v}(t+\tau),\ \underline{f}(t+\tau)]$
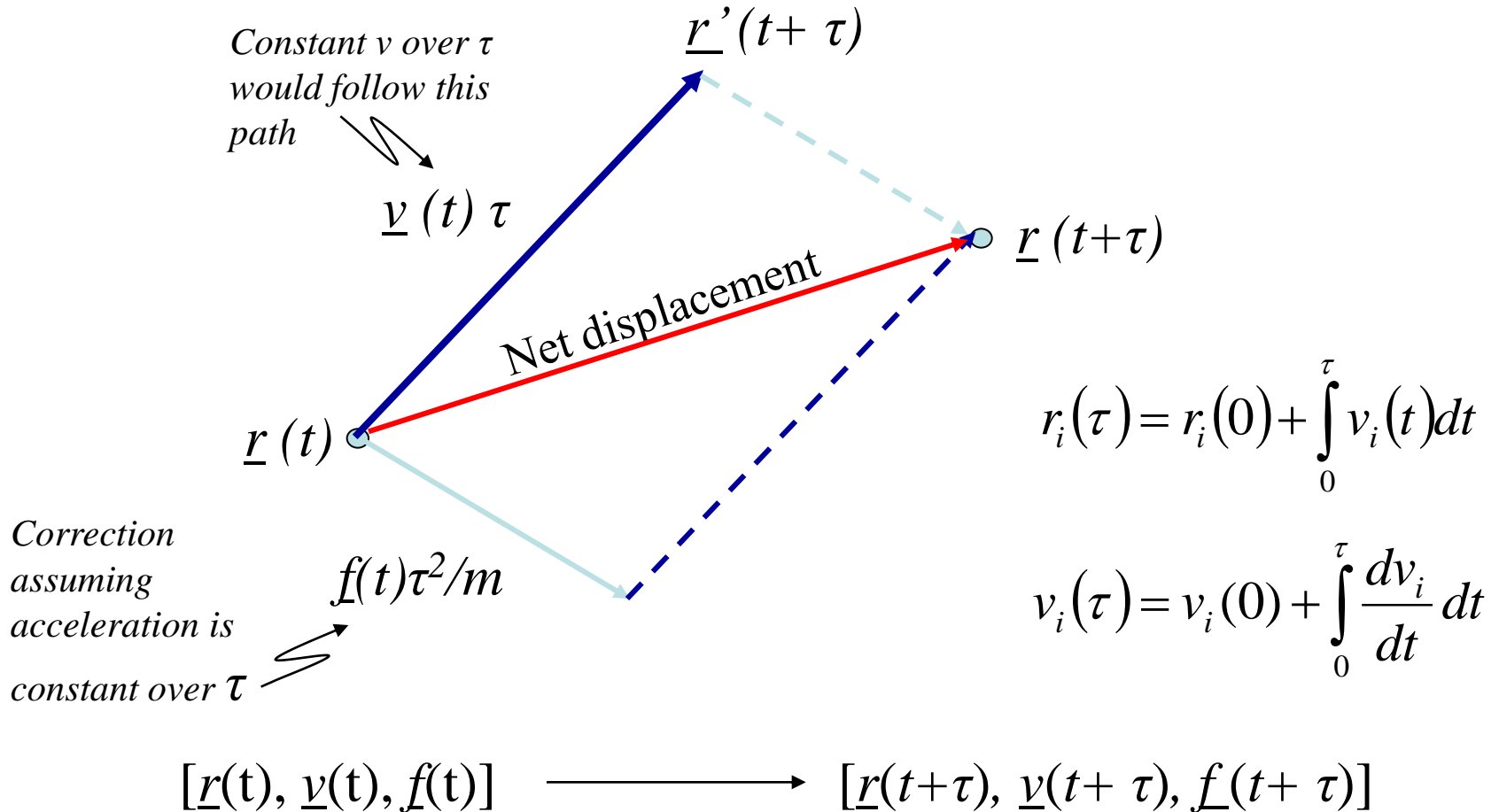
The Verlet algorithm (1967).

In a molecular dynamics simulations the configuration at time is split into a series of steps of length, τ, the time step. Information from previous steps can be used to generate future steps.

We have looked at Newton's laws for the position of any atom at time $t+\tau$. This is the first part of a Taylor expansion:

$$r_i(t+\tau) = r_i(t) + v_i(t)\tau + \frac{1}{2}\frac{f_i(t)}{m_i}\tau^2 + \frac{1}{6}\dddot{r}_i(t)\tau^3 ......$$

We could equally well back track to time t-τ by using a negative time step, this would give:

$$r_i(t-\tau) = r_i(t) - v_i(t)\tau + \frac{1}{2}\frac{f_i(t)}{m_i}\tau^2 - \frac{1}{6}\dddot{r}_i(t)\tau^3 ......$$

Note: Using further extensions of the Taylor expansion is the basis of the Gear Predictor-Corrector integrator methods (see Allen & Tildesley).

By adding these equations we can remove the need to calculate the velocity at any timestep since:

$$r_i(t+\tau) + r_i(t-\tau) = 2r_i(t) + \frac{f_i(t)}{m_i}\tau^2$$

So the next position for the atom in the trajectory is:

$$r_i(t+\tau) = 2r_i(t) - r_i(t-\tau) + \frac{f_i(t)}{m_i}\tau^2$$

Note that in this scheme the cubic terms in the timestep have been cancelled out and so the error in the interpolation to the next point is of order τ^4.

This means that the Verlet algorithm will follow the "real" trajectory of the system more accurately than a simple scheme based on the velocity and force equation for r_i.

Time line for a Verlet simulation

Time: $t\text{-}2\tau$ $t\text{-}\tau$ t $t\text{+}\tau$ $t\text{+}2\tau$

Position :

Force:

$$r_i(t+\tau) = 2r_i(t) - r_i(t-\tau) + \frac{f_i(t)}{m_i}\tau^2$$

15
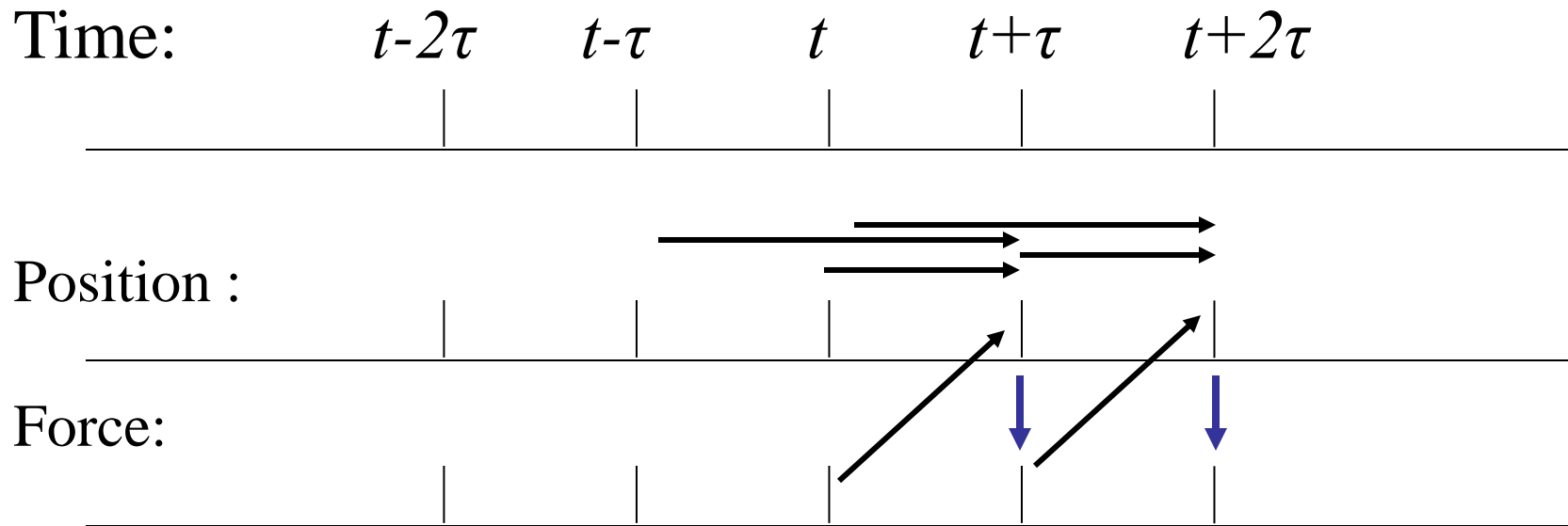
The velocity is not required for the trajectory in the Verlet algorithm but it is usually required for calculating the kinetic energy of the system.

The velocity can be calculated from any two positions using the finite difference equation:

$$v_i(t) = \frac{r_i(t + \tau) - r_i(t - \tau)}{2\tau}$$

This allows us to calculate the kinetic energy at any time in the simulation just after making a step.

This gives an error in the velocity of order τ^2.

So although the trajectory error is small, accurate velocities requires a small timestep.

Boundary Conditions

- None - biopolymer simulations
- Stochastic boundaries - biopolymers
- Hard wall boundaries - pores, capillaries
- **Periodic boundaries - most MD simulations**
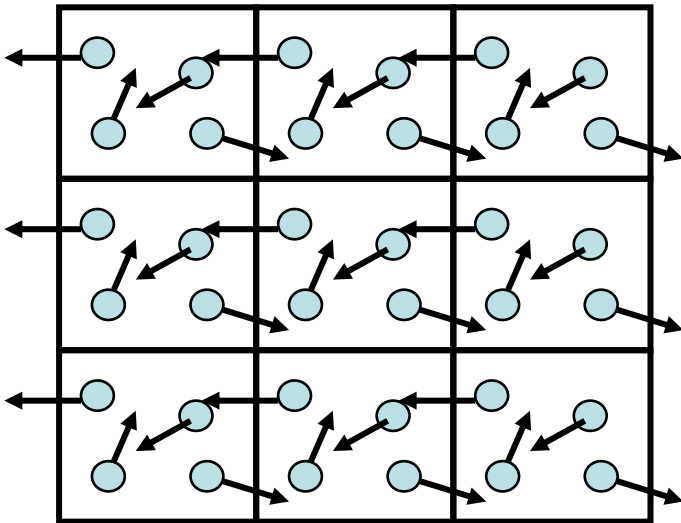
Periodic Boundaries - Consequences

- System no longer has a *surface*.
- System becomes *pseudo-periodic* (used to advantage for Ewald sums).
- Correlations in space beyond *half cell width* (L/2) are artificial. For this reason, the cut-off r_{cut} is usually no greater than L/2.
- Correlations in time beyond t=L/c are (in principle) subject to *recurrence*. In practice this does not seem to be the case.
- Use with *Minimum Image* convention.

Minimum Image Convention

Use r_{ij}, not r_{ij} to find $r_{ij'}$ work in fractional co-ordinates:

$$f_{aij} = a^* \bullet r_{ij} \quad f_{bij} = b^* \bullet r_{ij} \quad f_{aij} = c^* \bullet r_{ij}$$

(* indicates reciprocal space vectors)
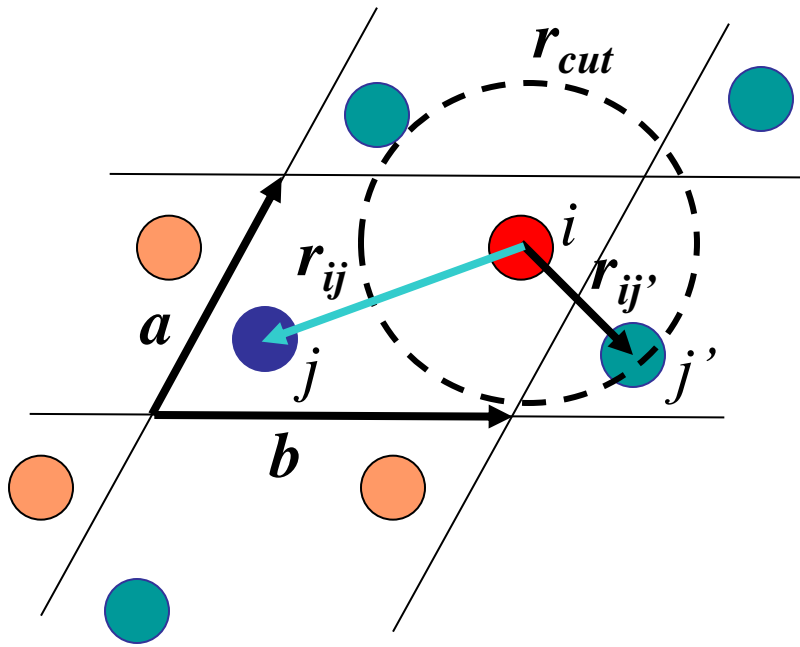
These need to be in the range $-\tfrac{1}{2} < f \le \tfrac{1}{2}$

$$f_{ij'} = f_{ij} - \text{nint}\left(f_{ij}\right)$$

(nint(f)=nearest integer to f)

Then convert back to Cartesian:

$$\begin{pmatrix} r_{ij'x} \\ r_{ij'y} \\ r_{ij'z} \end{pmatrix} = \begin{pmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{pmatrix} \begin{pmatrix} f_{aij'} \\ f_{bij'} \\ f_{cij'} \end{pmatrix}$$

For van der Waals interactions use only the nearest images of atoms. The minimum image convention limits the cut off used in the potential sum to half the shortest lattice parameter.

18

Considerations for MD

In the potentials we are using the force generated between a pair of atoms is equal and opposite. This means that the total momentum of the system cannot change, *i.e.*:

$$\sum_i m_i \underline{v}_i = 0$$

Similarly the centre of mass position should be fixed:

$$\underline{r}_{cm} = \frac{\sum_i m_i \underline{r}_i}{\sum_i m_i}$$

since

$$\left(\dot{\underline{r}}_{cm} = \frac{\sum_i m_i \dot{\underline{r}}_i}{\sum_i m_i} \right)$$

Observations such as this allow checks that coding/integration are working!

In MD we have two contributions to the energy.

1) Kinetic energy, which sets the temperature:

$$KE(t) = \tfrac{1}{2} \sum_i m_i \left| \underline{v}_i(t) \right|^2$$

2) Potential energy, which is dependant on the atomic configuration according to the forcefield:

$$PE(t) = \sum_i \sum_{j>i} V_{ij}\left(r_{ij}(t) \right)$$

The energy of the system is continually interchanging between potential and kinetic contributions.

Key Stages in MD Simulation

Initialise

Forces

Motion

Properties

Summarise
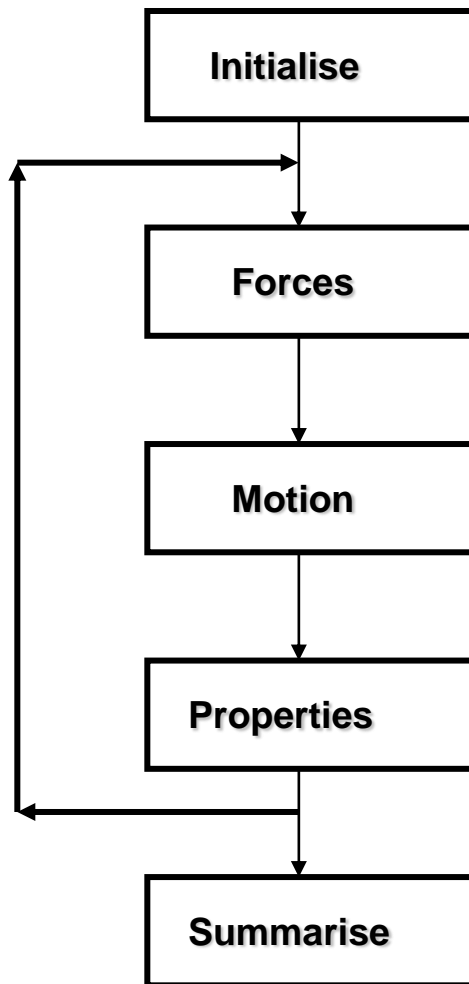
- *Set up initial system*

  positions of atoms and initial velocities (Boltzmann with random directions)

- *Calculate atomic forces*

  based on potential model.

- *Calculate atomic motion*

  via the integration algorithm, e.g. Verlet.

- *Calculate physical properties*
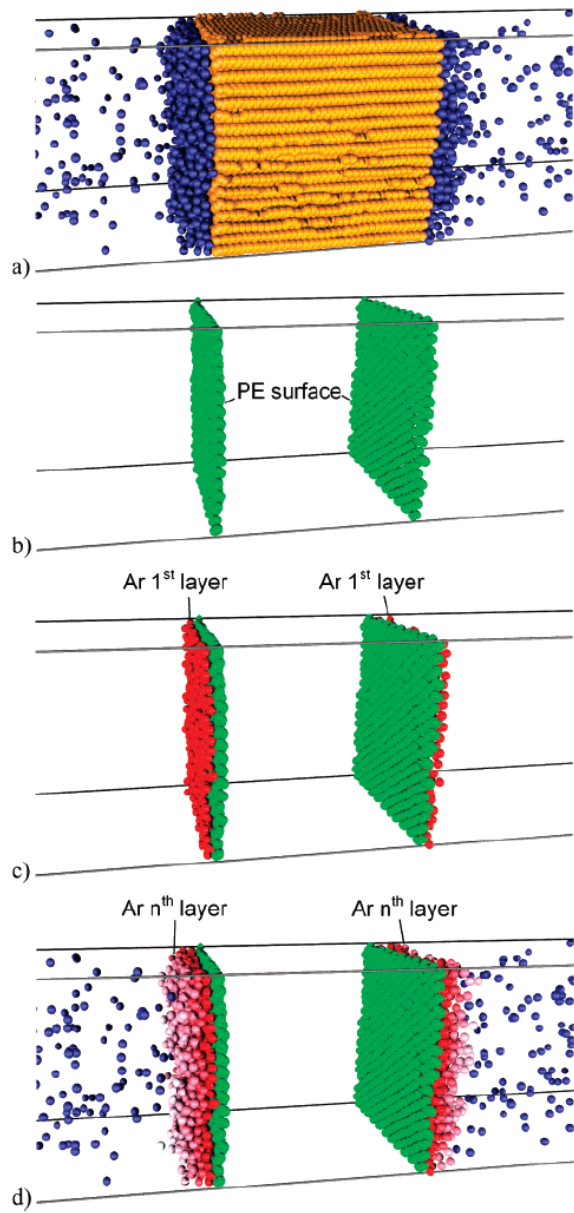
- *Repeat !*

- *Produce final summary*

a)

b)   PE surface

      Ar 1st layer   Ar 1st layer

c)   Ar nth layer   Ar nth layer

d)

Figure 2. Snapshots of a condensation simulation at $T_{PE} = 80$ K and $\rho = 50$ g/dm^3 after 5 ns. (a) Complete polymer film with condensed argon and a part of the argon vapor phase. (b) Surface layers of polyethylene. (c) Polyethylene surface layer and first argon layer. (d) Polyethylene surface layer with first and all subsequent argon layers. Each layer has a different grayscale (red shade).

Lennard-Jones Spheres (Argon).

Provides a relatively simple system in which the ideas of using molecular dynamics to study kinetics and thermodynamics of physical processes can be tested.

Here a simulation of Ar condensation at the surface of a polyethylene film is simulated.

System size:
Film consists of 374 chains, 70 sites each.
Vapor phase: 5000 argon atoms.

Run times: 1ns equilibration of components.
            12 ns run times.

R. Rozas & T. Kraska, *J. Phys. Chem. C,* **111**, 15784, (2007).
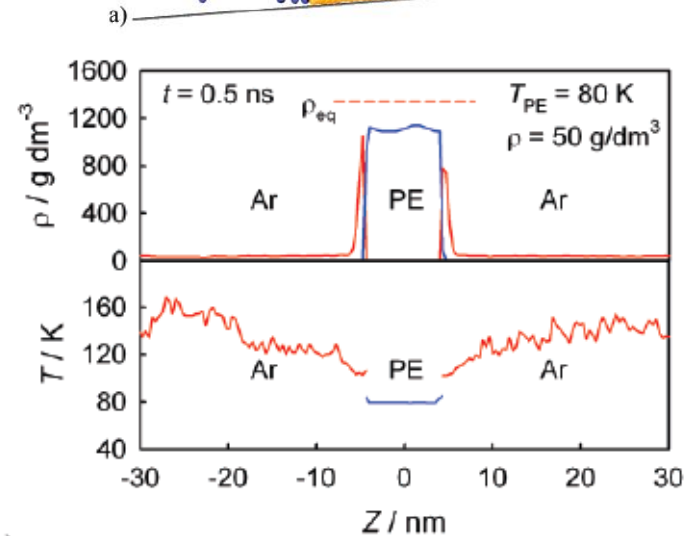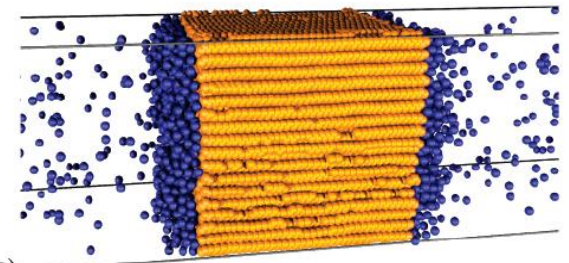
21

System Properties (1)

Thermodynamic Properties
MD Kinetic Energy:

$$\langle K.E. \rangle = \left\langle \frac{1}{2} \sum_i^N m_i v_i^2 \right\rangle$$

The average <...> kinetic energy per particle is directly related to the statistical mechanics definition of temperature:

$$T = \frac{2}{3Nk_B} \langle K.E. \rangle$$

Data from: R. Rozas & T. Kraska, *J. Phys. Chem. C,* **111**, 15784, (2007).
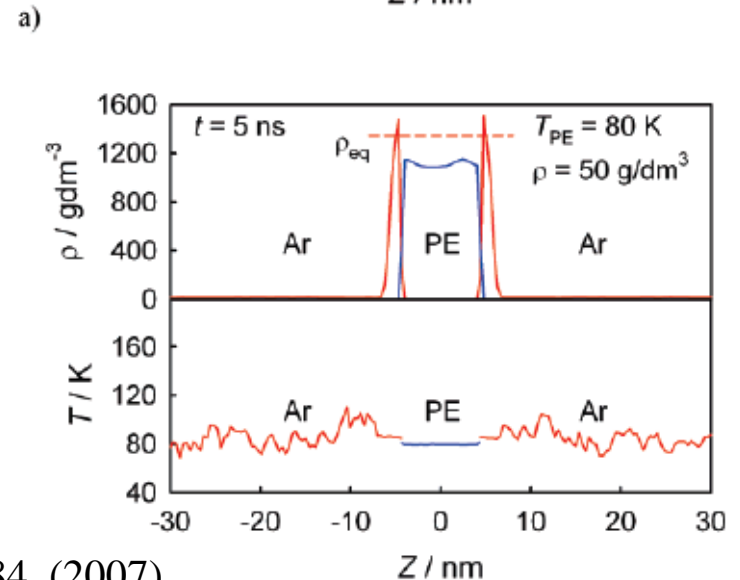
$$T = \frac{2}{3Nk_B}\langle K.E.\rangle$$

Velocity scaling

To get around the problem of temperature drift we can simply remove or inject kinetic energy by scaling the velocities in the initial stages of the calculation:

$$KE(t) = \tfrac{1}{2}\sum_i m_i\left|s.\underline{v}_i(t)\right|^2 \qquad s = \left(\frac{T(set)}{T(t)}\right)^{1/2}$$

Where *T(set)* is the target temperature for the simulation. This technique is referred to as velocity scaling and is used during the equilibration period of the simulation.
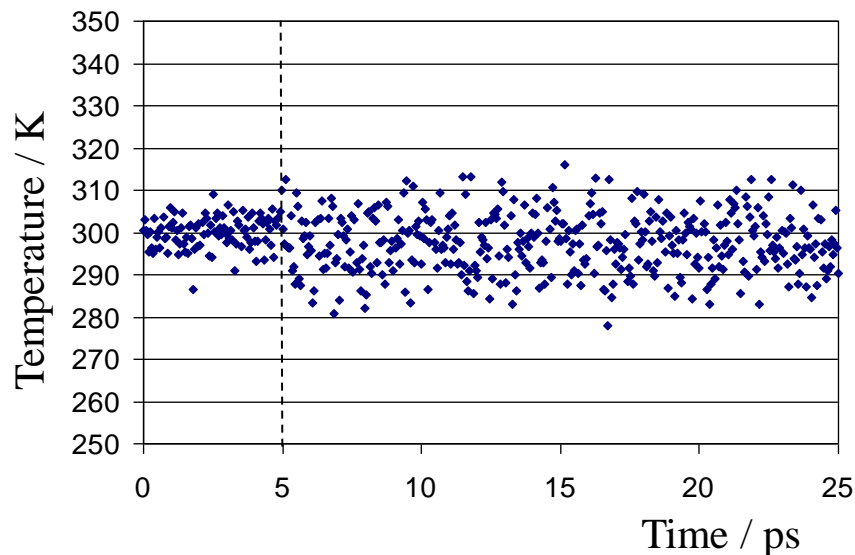
Example MD run with T set at 300K, velocity scaling used up to 5ps.

Thereafter fluctuations in T increase but no drift away from average.

23

Radial Distribution Function, RDF

*The radial distribution function is the probability of finding a pair of atoms a distance **r** apart relative to the probability for a completely uniform distribution.*

Edge of simulation cell

δr

r
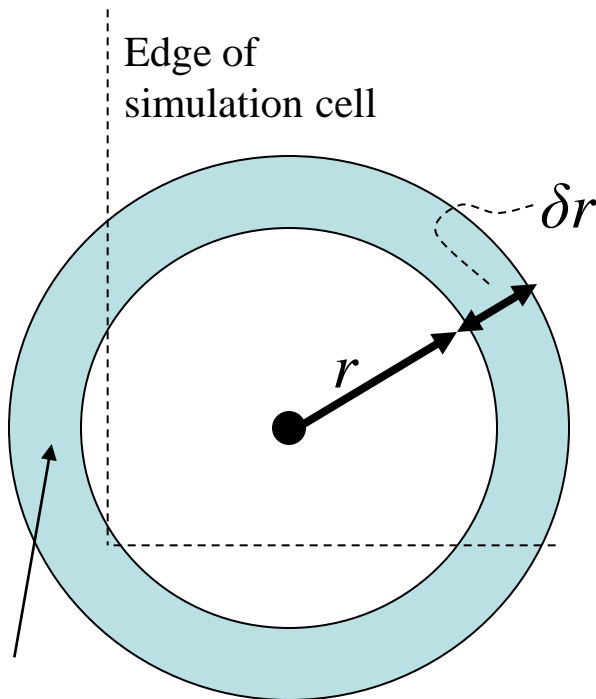
To calculate distances in this region requires min. image

Volume of the sampling shell:

$$\delta V = \frac{4}{3}\pi(r+\delta r)^3 - \frac{4}{3}\pi\, r^3 = 4\pi\, r^2 \delta r$$

If this were a perfectly uniform distribution the number of particle pairs at a particular r will depend on the concentration and volume of space sampled:

$$d_{uni}(r) = \frac{N(N-1)}{V} 4\pi\, r^2 \delta r$$

Num pairs per unit volume

Volume sampled at r.

24

Radial distribution function

The radial distribution function is calculated for each frame and then averaged:

$$g(r) = \frac{\sum\limits_{frames} d(r)}{N_f d_{uni}(r)} = \frac{V}{4\pi \; N_f N(N-1)r^2 \delta r} \sum_{frames} \sum_i \sum_{j \neq i} \delta(r - r_{ij})$$
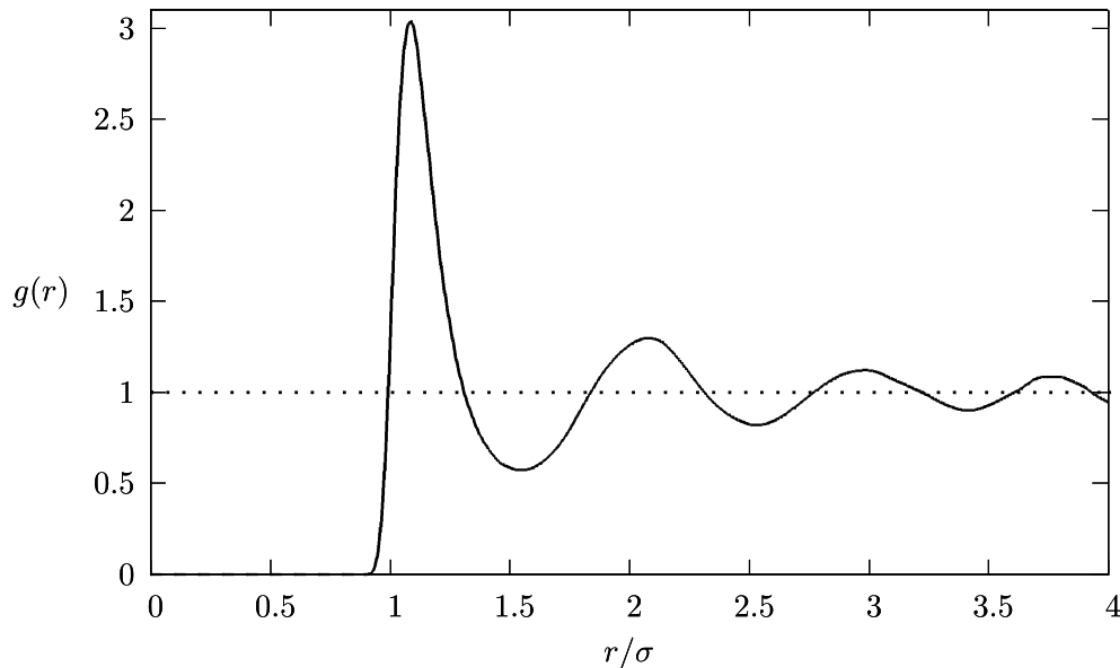
The function indicates if there is local structure in the fluid.

A uniform distribution of particles would have $g(r) = 1$ for all r.
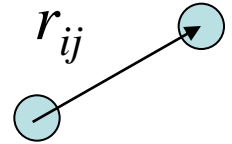
So accumulation of particles shows up as $g(r) > 1$ and depletion as $g(r) < 1$.

25

Radial Distribution Function, RDF

Edge of
simulation cell

δr

The distribution of particles at a particular instant (*i.e.* for a particular frame) can be calculated from the atom positions:

r_{ij}

r

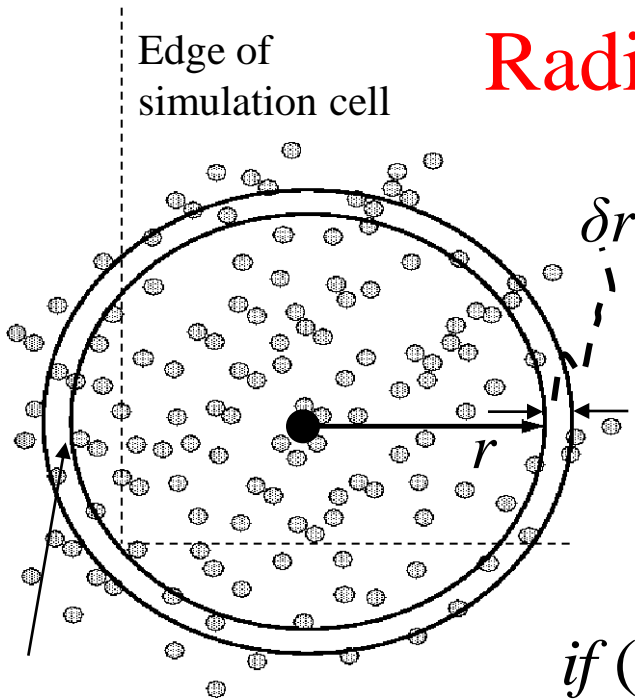$$d(r) = \sum_i \sum_{j \neq i} \delta(r - r_{ij})$$

Where we apply a binning criterion:

$$if\,((r - \delta r/2) < r_{ij} > (r + \delta r/2)) \quad \delta(r - r_{ij}) = 1$$

To calculate distances in this region requires min. image

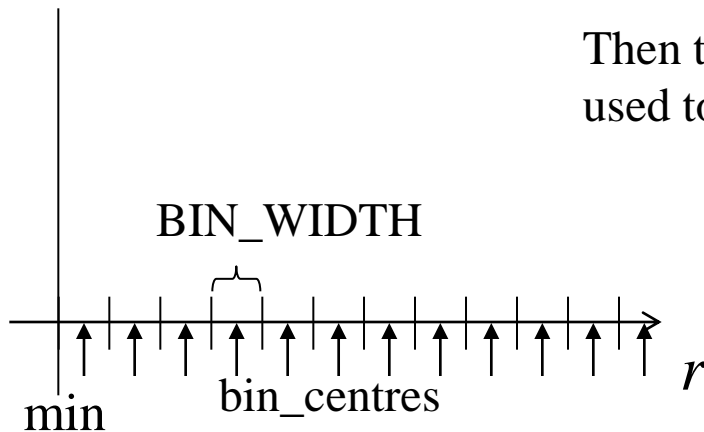To do this an array structure of "bins" is set up.

Then the relationship between r_{ij} and the bin indices is used to increment the corresponding bin counter.

BIN_WIDTH

$$index = \text{int}\left(\frac{(r_{ij} - \min)}{bin\_width}\right) + 1$$

For Fortran programmers ..

r
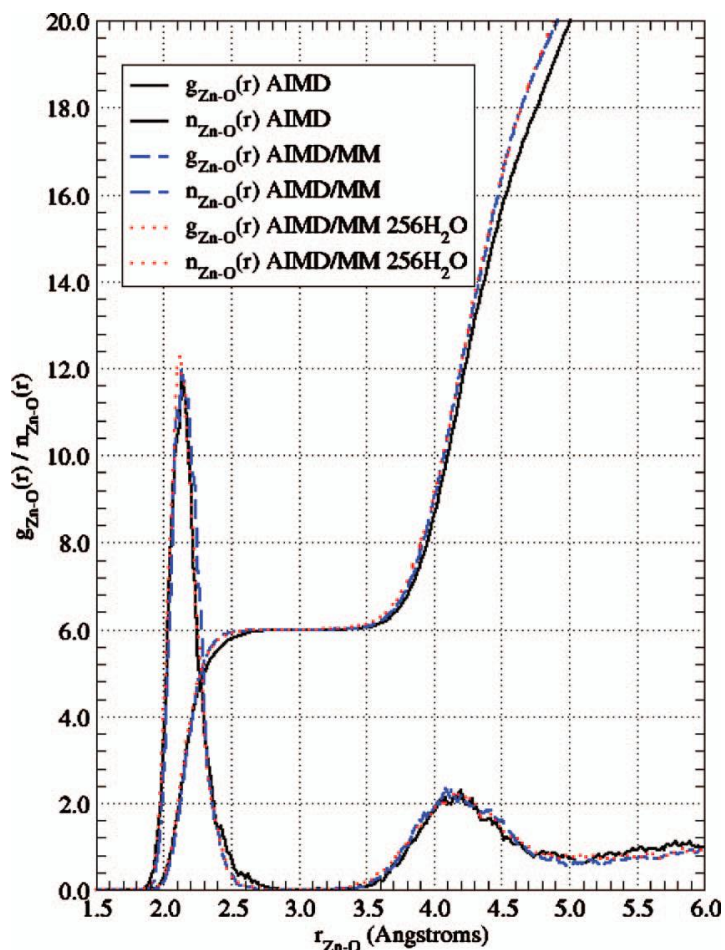
min

bin_centres

Information from Radial Distribution Functions.

Example DFT-MD/ QM-MM calculation of the distribution of H_2O around a Zn^{2+} cation.

Note:

First $g_{Zn-O}(r)$ peak integrates to 6. (co-ordination N^o)

QM/MM combination captures $g_{Zn-O}(r)$.

Orientation of water in first solvation shell leads to depletion in $g_{Zn-O}(r)$ between 2.5 and 3.5 Å.

E. Cauët *et al.*, *J. Chem. Phys.* **132**, 194502, (2010)

27