



2549-1

#### ADGLASS Winter School on Advanced Molecular Dynamics Simulations

18 - 21 February 2013

#### Fundamentals of Classical Molecular Dynamics

M P Allen
University of Warwick
Department of Physics
United Kingdom

# Fundamentals of Classical Molecular Dynamics

M P Allen

Department of Physics University of Warwick

Trieste ICTP

18 Feb 2013



## A General Bibliography

- D. Frenkel and B. Smit, Understanding Molecular Simulation: from Algorithms to Applications, 2nd edition (Academic Press, 2002).
- D. C. Rapaport, The Art of Molecular Dynamics Simulation, 2nd ed (Cambridge University Press, 2004).
- M. E. Tuckerman, Statistical Mechanics: Theory and Molecular Simulation, (Oxford University Press, 2010).
- M. P. Allen and D. J. Tildesley, Computer Simulation of Liquids (Clarendon Press, 1989).
- B. J. Leimkuhler and S. Reich, Simulating Hamiltonian Dynamics (Cambridge University Press, 2004).

- Molecular Interactions
- Molecular Dynamics Algorithms
- Periodic Boundaries: Short and Long-Ranged Forces
- 4 Thermostats

Molecular dynamics: numerical, step-by-step, solution of the classical equations of motion.

#### Newton's or Hamilton's Equations

$$m_i \ddot{\boldsymbol{r}}_i = \boldsymbol{f}_i$$
 or  $\begin{cases} \dot{\boldsymbol{r}}_i &= \boldsymbol{p}_i / m_i \\ \dot{\boldsymbol{p}}_i &= \boldsymbol{f}_i \end{cases}$  where  $\boldsymbol{f}_i = -\frac{\partial}{\partial \boldsymbol{r}_i} U = -\nabla_i U$ 

System of coupled ordinary differential equations.

- Need to be able to calculate the forces  $f_i$
- usually derived from a potential energy  $U(\mathbf{r})$
- $r = r_1, r_2, \dots r_N = \{r_i\}$  are atomic coordinates

The non-bonded potential energy  $U_{nb}$  is traditionally split into 1-body, 2-body, 3-body . . . terms:

#### Non-bonded Potential

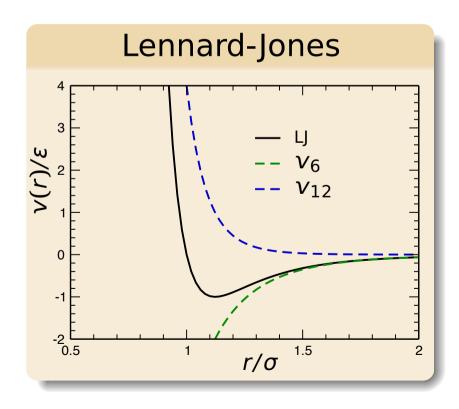
$$U_{\text{nb}}(\mathbf{r}) = \sum_{i} u(\mathbf{r}_{i}) + \sum_{i,j>i} v(\mathbf{r}_{i},\mathbf{r}_{j}) + \sum_{ijk} w(\mathbf{r}_{i},\mathbf{r}_{j},\mathbf{r}_{k}) + \dots$$

- The external field or container walls
  - usually dropped for simulations of bulk systems
- The interatomic pair potential
- Usually neglect higher order interactions.

There is an extensive literature on the experimental determination of these potentials.

# Lennard-Jones Potential

- Sometimes sufficient to use the simplest models that faithfully represent the essential physics.
- The Lennard-Jones potential is the most commonly used form, developed for studies of inert gases.



#### Lennard-Jones

$$v^{LJ}(r) = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right]$$
$$= v_{12}^{LJ}(r) + v_{6}^{LJ}(r)$$

- $\sigma$  = diameter
- $\varepsilon$  = well depth

Electrostatic charges interact via long-ranged potentials

#### Coulomb Potential

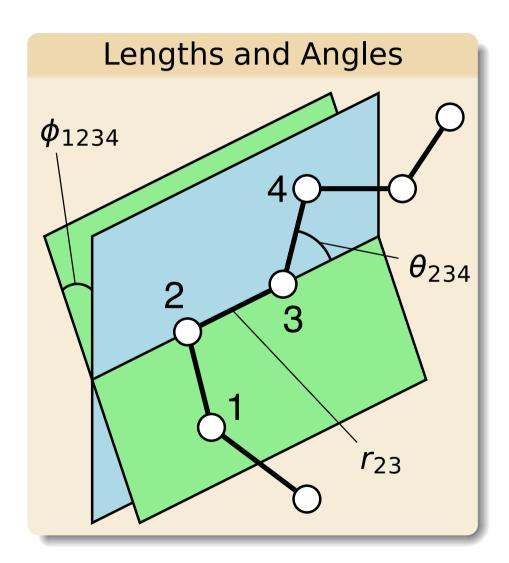
$$v^{qq}(r) = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

- $\bullet$   $q_1$ ,  $q_2$  are the charges
- $\bullet$   $\epsilon_0$  is the permittivity of free space.
- The correct handling of long-range forces provokes much discussion in the literature.
- Interactions involving dipole moments and higher-order multipoles expressed in similar way.

## Atoms to Molecules

- For molecular systems, we simply build the molecules out of Lennard-Jones site-site potentials, or similar.
- Typically, a single-molecule quantum-chemical calculation may be used to estimate the electron density throughout the molecule.
- This may then be modelled by a distribution of partial charges,
- or more accurately by a distribution of electrostatic multipoles.

## **Bonding Potentials**



For molecules we must also consider the intramolecular bonding interactions. Consider this geometry of an alkyl chain (just showing the carbons).

- interatomic distance  $r_{23}$
- bend angle  $\theta_{234}$
- torsion angle  $\phi_{1234}$

# **Bonding Potentials**

A very simple example:

### Intramolecular Bonding Potentials

$$U_{\text{int}} = \frac{1}{2} \sum_{\text{bonds}} k_{ij}^{r} (r_{ij} - r_{\text{eq}})^{2}$$

$$+ \frac{1}{2} \sum_{\substack{\text{bend} \\ \text{angles}}} k_{ijk}^{\theta} (\theta_{ijk} - \theta_{\text{eq}})^{2}$$

$$+ \frac{1}{2} \sum_{\substack{\text{torsion} \\ \text{angles}}} \sum_{m} k_{ijkl}^{\phi, m} [1 + \cos(m\phi_{ijkl} - \gamma_{m})]$$

Packaged force fields specify  $k_{ij}^r$ ,  $k_{ijk}^{\theta}$ ,  $k_{ijkl}^{\phi,m}$ ,  $r_{eq}$ ,  $\theta_{eq}$ ,  $\gamma_m$  or similar parameters.

#### Harmonic Potential

$$\frac{1}{2}k_{ij}^r(r_{ij}-r_{eq})^2$$

- The "bonds" involve separations  $r_{ij} = |\mathbf{r}_i \mathbf{r}_j|$  between atoms in a molecular framework.
- We assume a harmonic form with specified equilibrium separation - not the only possibility.
- Deriving forces from this is straightforward.
- Vibration frequencies relatively high
  - E.g. for C—H bonds, period ~ 10fs
  - in a step-by-step solution of the equations of motion, need timestep  $\Delta t \approx 5 \text{fs}$ .

## Quadratic Approximation or Trigonometric Form

$$\frac{1}{2}k_{ijk}^{\theta}(\theta_{ijk}-\theta_{eq})^2$$
 or  $\frac{1}{2}k_{ijk}^{\theta}(1-\cos^2(\theta_{ijk}-\theta_{eq}))$ 

- The "bend angles"  $\theta_{ijk}$  are between successive bond vectors such as  $\mathbf{r}_i \mathbf{r}_j$  and  $\mathbf{r}_j \mathbf{r}_k$ .
- Therefore, they involve three atom coordinates:

#### **Bend Angle Definition**

$$\cos\theta_{ijk} = \hat{\boldsymbol{r}}_{ij} \cdot \hat{\boldsymbol{r}}_{jk} = (\boldsymbol{r}_{ij} \cdot \boldsymbol{r}_{ij})^{-1/2} (\boldsymbol{r}_{jk} \cdot \boldsymbol{r}_{jk})^{-1/2} (\boldsymbol{r}_{ij} \cdot \boldsymbol{r}_{jk})$$

- Derived forces affect all three atoms.
- Calculated using the chain rule.
- Angle-bend timescales, e.g. in  $H_2O$ , are  $\sim 20 fs$ .

- Molecular Interactions
- Molecular Dynamics Algorithms
- Periodic Boundaries: Short and Long-Ranged Forces
- 4 Thermostats

## The MD Algorithm

Calculating forces is expensive, typically a pairwise sum over atoms, so we need to perform this as infrequently as possible.

- Wish to make the timestep as large as possible.
  - Hence, simulation algorithms tend to be low order (i.e. do not use high derivatives of r);
  - this allows the time step to be increased as much as possible without jeopardizing energy conservation.
- Cannot accurately follow true trajectory for very long times  $t_{run}$ .
  - Classical trajectories are 'ergodic' and 'mixing';
  - trajectories diverge from each other exponentially;
  - however long-term energy conservation is possible.

# The Verlet Algorithm

There are various, essentially equivalent, versions of the Verlet algorithm, including this one:

### Velocity Verlet Equations

$$p(t + \frac{1}{2}\Delta t) = p(t) + \frac{1}{2}\Delta t f(t)$$

$$r(t + \Delta t) = r(t) + \Delta t p(t + \frac{1}{2}\Delta t)/m$$

$$p(t + \Delta t) = p(t + \frac{1}{2}\Delta t) + \frac{1}{2}\Delta t f(t + \Delta t)$$

 $r = \{r_i\}$  (all coordinates),  $p = \{p_i\}$  (all momenta) and  $f = \{f_i\}$  (all forces).

After the middle step, a force evaluation is carried out, to give  $\mathbf{f}(t + \Delta t)$  for the last step. This scheme advances the coordinates and momenta over a timestep  $\Delta t$ .

# The Verlet Algorithm

A piece of pseudo-code illustrates how this works:

### Velocity Verlet Algorithm

```
do step = 1, nstep
    p = p + 0.5*dt*f
    r = r + dt*p/m
    f = force(r)
    p = p + 0.5*dt*f
enddo
```

- The force routine carries out the time-consuming calculation of all the forces, and potential energy U.
- The kinetic energy K can be calculated after the second momentum update.
- At this point the total energy is U + K.

## The Verlet Algorithm

#### Important features of the Verlet algorithm are:

- 1 it is exactly time reversible;
- it is symplectic (to be discussed later);
- it is low order in time, permitting long timesteps;
- it requires just one force evaluation per step;
- it is easy to program.

Formally, for  $A = \mathbf{r}$ ,  $\mathbf{p}$ , or any  $A(\mathbf{r}, \mathbf{p})$ :

- $\dot{A} = i\mathcal{L}A$  (Liouville operator);
- $A(t + \Delta t) = e^{i\mathcal{L}\Delta t}A(t)$  (propagator).

Useful approximations arise from splitting  $i\mathcal{L}$  in two:

#### **Split Propagator**

$$i\mathcal{L}_{p} = \boxed{ \boldsymbol{f} \cdot \frac{\partial}{\partial \boldsymbol{p}} }$$

$$i\mathcal{L}_r = m^{-1} \boldsymbol{p} \cdot \frac{\partial}{\partial \boldsymbol{r}}$$

$$i\mathcal{L} = i\mathcal{L}_p + i\mathcal{L}_r$$

$$e^{i\mathcal{L}_{p}\Delta t}\boldsymbol{p}=\boldsymbol{p}+\boldsymbol{f}\Delta t$$
 kick

$$e^{i\mathcal{L}_r\Delta t}\mathbf{r} = \mathbf{r} + m^{-1}\mathbf{p}\Delta t$$
 drift

The following approximation is asymptotically exact in the limit  $\Delta t \rightarrow 0$ .

#### Symmetric Splitting

$$e^{i\mathcal{L}\Delta t} = e^{(i\mathcal{L}_p + i\mathcal{L}_r)\Delta t} \approx e^{i\mathcal{L}_p\Delta t/2} e^{i\mathcal{L}_r\Delta t} e^{i\mathcal{L}_p\Delta t/2}$$

- For nonzero  $\Delta t$  this is an approximation to  $e^{i\mathcal{L}\Delta t}$ because in general  $i\mathcal{L}_p$  and  $i\mathcal{L}_r$  do not commute,
- but it is still exactly time reversible and symplectic.
  - Symplectic (roughly) implies conserving phase space volume  $d\mathbf{r}(t + \Delta t)d\mathbf{p}(t + \Delta t) = d\mathbf{r}(t)d\mathbf{p}(t)$ .

It is then easy to see that the three successive steps embodied in the above equation, with the above choice of operators, generate the velocity Verlet algorithm.

## Propagators and the Verlet Algorithm

- The trajectories generated by the above scheme are approximate, and will not conserve the true energy H.
- Nonetheless, they do exactly conserve a "pseudo-hamiltonian" or "shadow hamiltonian" H<sup>‡</sup>
- H and  $H^{\ddagger}$  differ from each other by a small amount,  $H^{\ddagger} = H + \mathcal{O}(\Delta t^2)$ .
- This means that the system will remain on a hypersurface in phase space which is "close" to the true constant-energy hypersurface.

Such a stability property is extremely useful in MD, since we wish to sample constant-energy states.

# Example

Consider a simple harmonic oscillator, of natural frequency  $\omega$ , representing perhaps an interatomic bond in a diatomic molecule. The equations of motion and conserved hamiltonian are

#### Harmonic Oscillator Equations

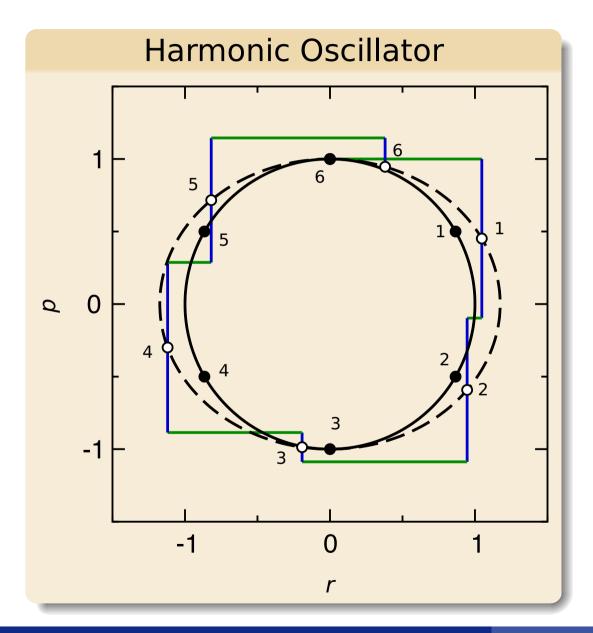
$$\dot{r} = p/m$$
,  $\dot{p} = -m\omega^2 r$ ,  $H(r, p) = p^2/2m + \frac{1}{2}m\omega^2 r^2$ 

For this system, velocity Verlet exactly conserves:

#### The Shadow Hamiltonian

$$H^{\ddagger}(r,p) = p^2/2m + \frac{1}{2}m\omega^2r^2(1 - (\omega\Delta t/2)^2)$$

## Example



In a "phase portrait", the simulated system remains on the constant- $H^{\ddagger}$  ellipse (dashed line) which differs only slightly (for small  $\omega \Delta t$ ) from the true constant-H ellipse (full line), e.g. here for  $\omega \Delta t = \pi/3$ .

One approach to handling the fast bond vibrations is to use a shorter timestep for them.

- Use Liouville operator formalism to generate time-reversible Verlet-like multiple-timestep algorithm.
- Suppose there are "slow" F, and "fast" f, forces.
- Momentum satisfies  $\dot{\boldsymbol{p}} = \boldsymbol{F} + \boldsymbol{f}$ .
- Break up Liouville operator  $i\mathcal{L} = i\mathcal{L}_p + i\mathcal{L}_p + i\mathcal{L}_r$ :

### Multiple Timestep Liouville Operator

$$i\mathcal{L}_{p} = \boxed{\mathbf{F} \cdot \frac{\partial}{\partial \mathbf{p}}}, \quad i\boldsymbol{\ell}_{p} = \boxed{\mathbf{f} \cdot \frac{\partial}{\partial \mathbf{p}}}, \quad i\mathcal{L}_{r} = \boxed{m^{-1}\mathbf{p} \cdot \frac{\partial}{\partial \mathbf{r}}}$$

## Multiple Timesteps

The propagator approximately factorizes

#### Long Timestep Splitting

$$e^{i\mathcal{L}\Delta t} \approx e^{i\mathcal{L}_p\Delta t/2} e^{(i\ell_p+i\mathcal{L}_r)\Delta t} e^{i\mathcal{L}_p\Delta t/2}$$

where  $\Delta t$  represents a long time step. The middle part is then split again, using the conventional separation as usual, iterating over small time steps  $\delta t = \Delta t/n$ :

#### Short Timestep Splitting

$$e^{(i\ell_p + i\mathcal{L}_r)\Delta t} \approx \left(e^{i\ell_p \delta t/2} e^{i\mathcal{L}_r \delta t} e^{i\ell_p \delta t/2}\right)^n$$

- Fast-varying forces computed at short intervals.
- Slow forces computed once per long timestep.

## Multiple Timesteps

#### Multiple Timestep Algorithm

```
do STEP = 1, NSTEP
   p = p + (DT/2)*F
   do step = 1, n
      p = p + (dt/2)*f
      r = r + dt*p/m
      f = force(r)
      p = p + (dt/2)*f
   enddo
   F = FORCE(r)
   p = p + (DT/2)*F
enddo
```

Some pseudo-code illustrates how simple this is.

The simulation run consists of NSTEP long steps, of length DT, each consisting of n sub-steps of length dt, where DT = n\*dt.

# Multiple Timesteps

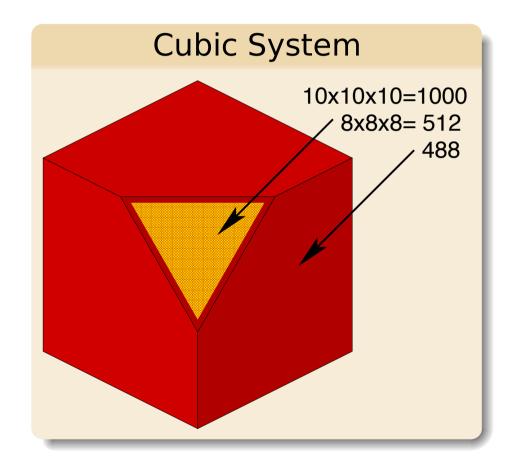
Non-bonded interactions may be calculated more efficiently this way too. A typical approach is to split the interatomic force law into a succession of components covering different ranges:

- the short-range forces change rapidly with time and require a short time step,
- the long-range forces vary more slowly, so we use a longer time step and less frequent evaluation.

Multiple-time-step algorithms are still under active study, and there is some concern that resonances may occur between the natural frequencies of the system and the various timesteps used in schemes of this kind. The area remains one of active research.

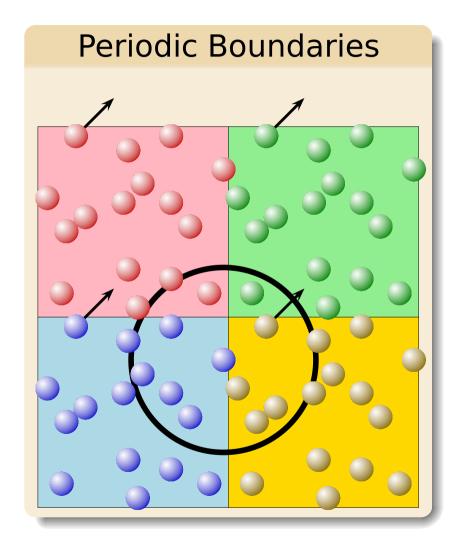
- Molecular Interactions
- Molecular Dynamics Algorithms
- Periodic Boundaries: Short and Long-Ranged Forces
- 4 Thermostats

## Periodic Boundary Conditions



- Consider  $N = 10^3$  atoms arranged in a cube.
- Nearly half the atoms are on the outer faces,
- will have a large effect on the measured properties.
- Even for
   N = 100<sup>3</sup> = 10<sup>6</sup> atoms,
   6% of atoms are on surface.

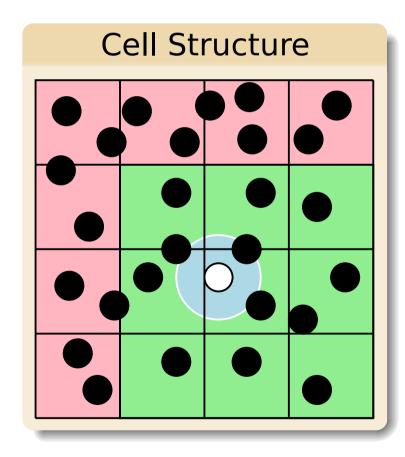
## Periodic Boundary Conditions



- Surround the cube with replicas
- For short-range potentials, adopt the minimum image convention: each atom interacts with the nearest atom or image in the periodic array.
- If an atom leaves the basic simulation box, attention can be switched to the incoming image.

## **Neighbour Lists**

For short-range potentials v(r) = 0,  $r > r_{\rm cut}$ , speed up search for those interactions which are in range.



- Divide  $L \times L \times L$  simulation box into  $n \times n \times n$  sub-cells
- Side of the cell

$$\ell = L/n \ge r_{\rm cut}$$

 In searching for atoms within range, examine the atom's own cell, and nearest neighbour cells.

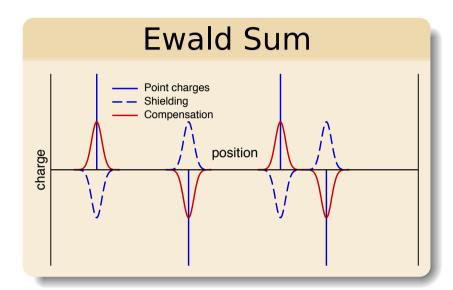
## Long-range forces

In periodic systems, the evaluation of the Coulomb interactions is non-trivial.

$$U = \sum_{i=1}^{N} \sum_{j>i} \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}}$$

- i and j vary over all ions in all cells: no cutoff
- Formally, this sum is only conditionally convergent
  - the result depends on the ordering of the terms
- There are some subtleties associated with the dielectric medium assumed to be "outside" the infinitely periodic system

## **Ewald Sum**



A trick allows us to evaluate the sum. Add to the system of point charges a set of positive and negative Gaussian distributions.

#### Real Space and Reciprocal Space Terms

$$U_{\text{r-space}} = \sum_{i=1}^{\infty} \sum_{j>i} \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \operatorname{erfc} \alpha r_{ij}$$

$$U_{\text{k-space}} = \frac{1}{2V\varepsilon_0} \sum_{\mathbf{k}} \frac{e^{-k^2/4\alpha^2}}{k^2} \left| \sum_{j} q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \right|^2$$

## **Ewald Sum**

#### The Coulomb energy is thus transformed into

- a real-space sum, involving the much shorter-ranged screened Coulomb interaction
- a reciprocal-space sum, i.e. a sum over wave-vectors k, involving the interaction between Gaussian charge clouds
- some correction terms handling the self-interaction between the added Gaussians

#### Smoothed Particle-Mesh Ewald

Mapping onto a regular grid by interpolation.

- The primary mathematical operation may now be performed by Fast Fourier Transform
- Greatly speeds up biological simulations

33

## Outline



- Molecular Interactions
- Molecular Dynamics Algorithms
- Periodic Boundaries: Short and Long-Ranged Forces
- 4 Thermostats

How to simulate systems at given T by MD?

#### Andersen Thermostat

- Periodically reselect atomic velocities at random from the Maxwell-Boltzmann distribution.
- Like occasional random coupling with thermal bath.
- The resampling may be done to individual atoms, or to the entire system.
- Simple to implement and reliable.
- Proven to sample the canonical ensemble
  - if MD is accurate!



An alternative, deterministic, approach:

#### Nosé-Hoover Equations

$$\dot{\mathbf{r}} = \mathbf{v} = \mathbf{p}/m \qquad \dot{\mathbf{p}} = \mathbf{f} - \xi \mathbf{v}$$

$$\dot{\xi} = M^{-1} \left[ \sum v^2 - gk_B T/m \right] \qquad \sum v^2 \equiv \sum_{i\alpha} v_{i\alpha}^2$$

- ξ: friction coefficient, allowed to vary in time;
- M is a thermal inertia parameter, determining a relaxation rate for thermal fluctuations;
- $g \approx 3N$  is the number of degrees of freedom.

If the system is too hot (cold), then  $\xi$  will tend to increase (decrease) tending to cool (heat) the system.

## **Hoover Derivation**



WG Hoover, Molecular dynamics, Springer Berlin (1987).

#### Assumed Equations of Motion

$$\dot{r} = p/m$$
,  $\dot{p} = f(r) - \xi v$ ,  $\dot{\xi} = G(r, p)$ 

- $\bullet$  G(r, p) is the object of the derivation.
- Ansatz: G(r, p) depends only on r, p, not  $\xi$ .

#### Generalized Liouville Equation

$$\sum \frac{\partial}{\partial \mathbf{r}} \cdot (\rho \dot{\mathbf{r}}) + \sum \frac{\partial}{\partial \mathbf{p}} \cdot (\rho \dot{\mathbf{p}}) + \frac{\partial}{\partial \xi} (\rho \dot{\xi}) = 0$$

Follows from continuity  $d\varrho/dt = 0$ , and stationarity  $\partial \rho/\partial t = 0$  of phase space distribution function  $\rho(\mathbf{r}, \mathbf{p}, \xi)$ .

Try: 
$$\rho(\mathbf{r}, \mathbf{p}, \xi) \propto \exp\{-H(\mathbf{r}, \mathbf{p})/k_{\rm B}T\} \exp\{-\frac{1}{2}M\xi^2/k_{\rm B}T\}$$

#### Required Form of G

$$G(\mathbf{r}, \mathbf{p}) = M^{-1} \sum_{i} \left( \frac{\mathbf{p}_{i}}{m_{i}} \cdot \mathbf{v}_{i} - k_{B}T \frac{\partial}{\partial \mathbf{p}_{i}} \cdot \mathbf{v}_{i} \right)$$
$$= M^{-1} \sum_{i} \left[ \mathbf{v}_{i}^{2} - 3k_{B}T/m \right]$$

- M arbitary constant (thermal inertia)
- Term in square brackets vanishes if averaged over canonical momentum distribution.



LD Landau and EM Lifshitz, *Statistical Physics* (1958). Based on hypervirial theorem in canonical ensemble.

#### Configurational and Kinetic Temperature

$$k_{\rm B}T_{\rm c} = rac{\left<\left(\partial H/\partial r_{ilpha}
ight)^2\right>}{\left<\partial^2 H/\partial r_{ilpha}^2\right>} \,, \quad i=1\dots N, \; lpha=x,y,z$$
  $k_{\rm B}T_{\rm k} = \left< p_{ilpha}^2/m \right>$ 

Both sides may be averaged over i and  $\alpha$ .

#### The configurational temperature is useful:

- to define T in the microcanonical ensemble
  - HH Rugh, *Phys. Rev. Lett.*, **78**, 772 (1997).
- as a test for simulation nonequilibrium
  - BD Butler, et al, J. Chem. Phys., **109**, 6519 (1998).
- in experiments on colloidal suspensions
  - YL Han, DG Grier, *Phys. Rev. Lett.*, **92**, 148301 (2004);
- As the basis of a thermostat in MD.
  - © C Braga and KP Travis, *J. Chem. Phys.*, **123**, 134101 (2005).

## Configurational Nosé-Hoover Thermostat

WARWICK

© C Braga and KP Travis, *J. Chem. Phys.*, **123**, 134101 (2005).

#### **Equations of Motion**

$$\dot{\mathbf{r}} = \mathbf{p}/m + \mu \mathbf{f} , \quad \dot{\mathbf{p}} = \mathbf{f}$$

$$\dot{\mu} = M^{-1} \left( \sum_{j} \left| \frac{\partial H}{\partial \mathbf{r}_{j}} \right|^{2} - k_{\mathrm{B}} T \frac{\partial}{\partial \mathbf{r}_{j}} \cdot \frac{\partial H}{\partial \mathbf{r}_{j}} \right)$$

- ullet  $\mu$  is a dynamical "mobility" coefficient.
- The driving force is related to  $T_c$ .

#### We have discussed the fundamentals of classical MD:

- specifying the molecular model;
- a good algorithm to advance the system in time;
- some techniques to improve efficiency;
- modifications for different physical conditions.

#### We have not discussed:

- how to analyse the results
  - structural and dynamical properties
- how to efficiently use different hardware
  - parallel computers or GPUs
- the relation between MD and Monte Carlo
  - Hybrid Monte Carlo, Brownian/Langevin Dynamics