A Brief Introduction to *Ab Initio* Molecular Dynamics

Matt Probert
Condensed Matter Dynamics Group
Department of Physics,
University of York, UK

http://www.cmt.york.ac.uk/cmd
http://www.castep.org

THE UNIVERSITY of York

Overview of Talk

- In this talk I hope to give you some ideas as to why you might want to do MD and what it can tell you.
- I hope to pass on some practical tips and advice, and answer some of your questions, particular w.r.t. CASTEP
- I shall illustrate with examples where possible. Time is short …

Why MD?

- Atoms move!
  - We may be interested in studying time dependent phenomena, such as molecular vibrations, phonons, diffusion, etc.
  - We may be interested in studying temperature dependant phenomena, such as free energies, anharmonic effects, etc.
- Ergodic Hypothesis
  - One of the key principles behind the usefulness of MD for statistical mechanics studies
  - Iff our MD trajectory is “good enough” then a time average over the trajectory is equivalent to an ensemble average – hence MD averages are useful.

Alternatives

- Monte Carlo
  - – can do thermal averages
  - – hard to do time dependant things
- Hybrid MD/MC
  - – bad MD as good MC
  - – generate configurations using poor/cheap/fast MD but then evaluate contribution to ensemble average using MC

Types of *ab initio* MD

- Classical Motion
  - – We use classical mechanics to move the atoms
    - Born-Oppenheimer approximation decouples nucleus and electrons
  - – But using forces and stresses derived from the electronic wavefunction
  - – No quantum fluctuations, tunneling, zero point motion, etc.
- Quantum Motion
  - – Can include ZPM etc using *ab initio* Path Integral MD
- Damped MD as a geometry optimizer
  - – BFGS *ought* to be a lot better but not always – see Probert, J. Comput. Phys. 191, 130 (2003)

Choice of Ensemble

- NVE
  - Micro-canonical ensemble
  - Constant Number of atoms, Volume and Energy
  - Corresponds to Newtonian mechanics
  - Good for non-equilibrium situations, e.g. watching a bond vibrate or doing impact movies
- NVT
  - Canonical ensemble – constant Temperature
  - More physical as it allows energy exchange with a heat bath
  - Good for simulating thermal equilibrium
  - Choice of thermostating algorithms

Choice of Ensemble

- NPH
  - constant pressure P and enthalpy H
  - Choice of barostats to handle pressure:
  - Andersen can allow cell to change size isotropically (liquids) whilst Parrinello-Rahman can allow changes in size and shape (solids)
  - External pressure can be isotropic (hydrostatic) or anisotropic (shear stress etc).
- NPT
  - Most physically relevant as system is now connected to a piston and a heatbath.
  - Again, choice of thermostats and barostats
- µVT - constant chemical potential µ

How do you do it? NVE …

- Integrate classical equations of motion
  - discretize time → time step
  - different integration algorithms, e.g. Velocity Verlet:

$$r(t+\delta t)=r(t)+v(t).\delta t+\frac{f(t)}{2m}.\delta t^{2}+O\left(\delta t^{3}\right)$$

$$v(t+\delta t)=v(t)+\frac{f(t)+f(t+\delta t)}{2m}.\delta t+O\left(\delta t^{2}\right)$$

  - trade-off time step vs. stability vs. accuracy
  - need accurate forces (high cutoff energies and good k-point sampling)

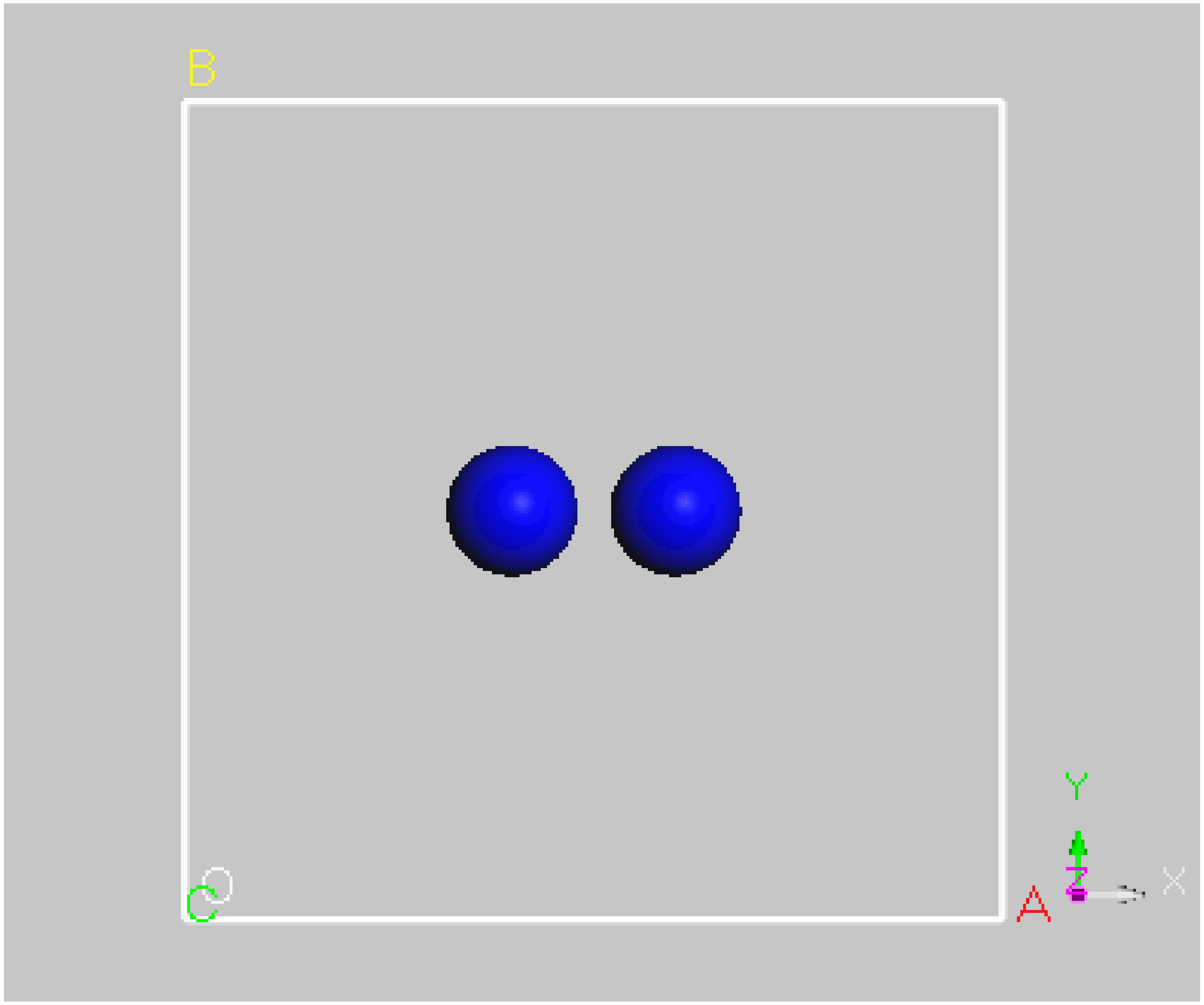
Other Ensembles

- Other ensembles can be simulated by using appropriate equations of motion
  - Usually derived from an extended Lagrangian (e.g. Nosé-Hoover, Parrinello-Rahman)
  - Recent developments in Liouvillian formulation have been very successful in deriving new symplectic integration schemes
- Langevin schemes need to be derived differently as non-Hamiltonian!
  - Need Focker-Planck & Liouville equation
  - see Quigley & Probert, J. Chem. Phys. 120, 11432 (2004) or my last talk at the FHI Berlin!

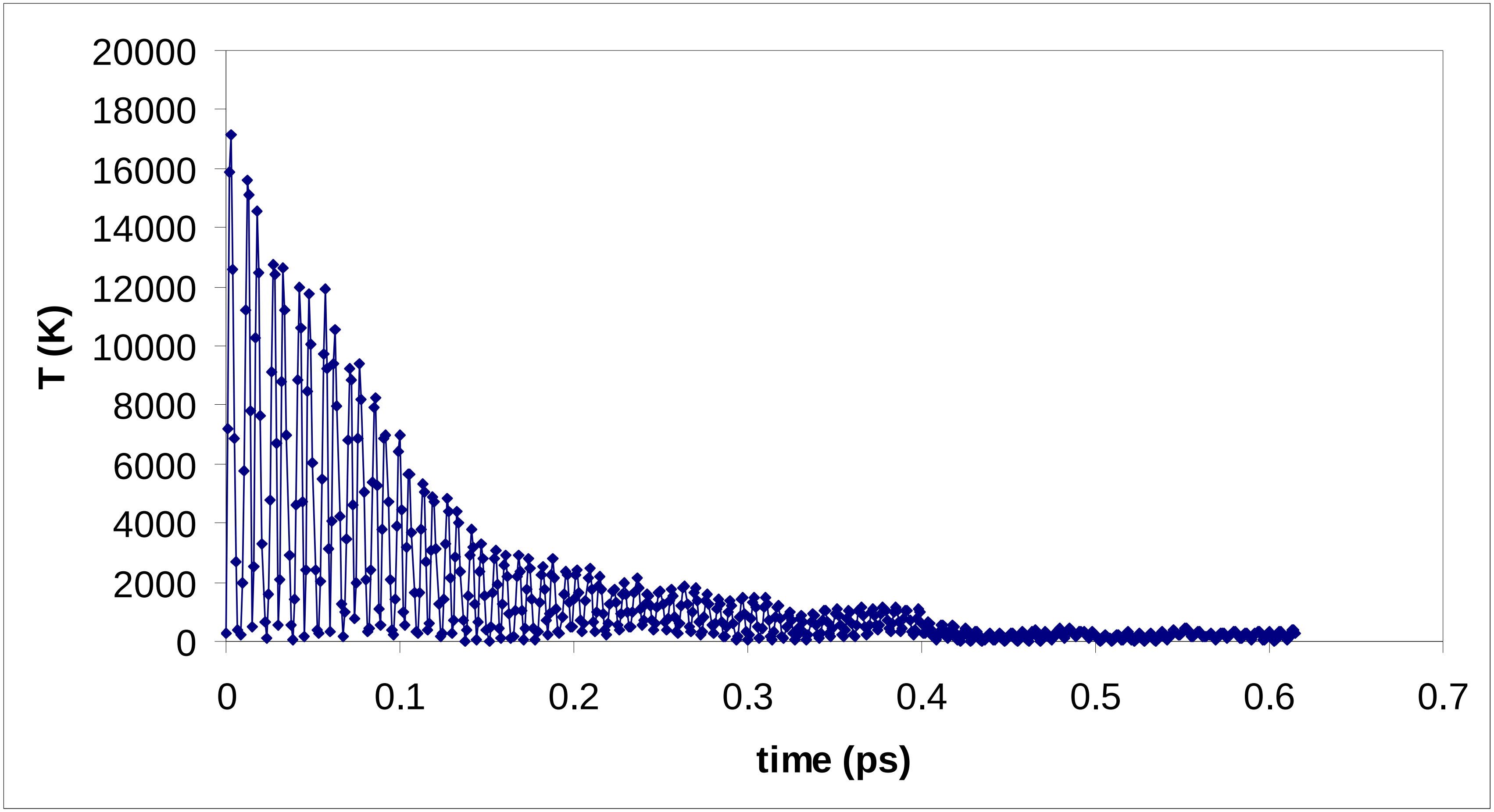
Simple Example: N2

- Naïve Materials Studio approach:
  - put 2 N atoms in a 5 A box at (0.4,0.5,0.5) and (0.6,0.5,0.5)
  - Use Gamma point for BZ sampling (it is an isolated molecule after all!)
  - Use default settings, e.g. “medium” Ecut.
  - Run NVT dynamics at default T=273 K using Langevin thermostat with default “Langevin time” of 0.1 ps and default time step of 1.0 fs
  - What do you see?

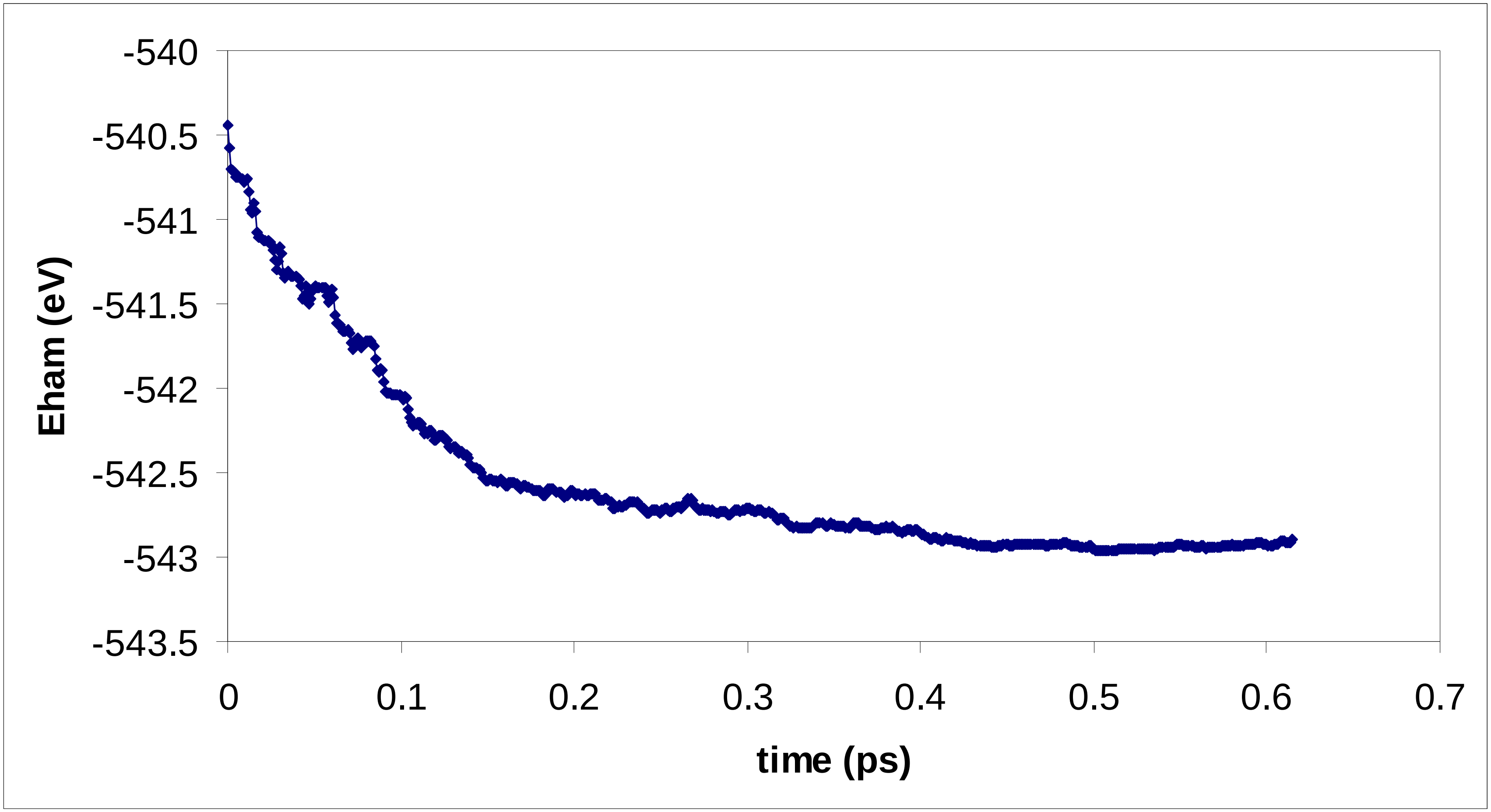
Simple N2 Movie



Temperature ???



Constant of Motion ???



What is Going On?

- Why is the temperature not constant if it is supposed to be NVT?
- The initial conditions were a long way from equilibrium. Doing a simple fixed-cell geometry optimisation relaxed > 2 eV.
- This excess PE is turned into KE by the MD – hence the huge initial temperatures before the thermostat is able to control it.
- The 2 eV excess PE shows up in the change in “constant of motion”

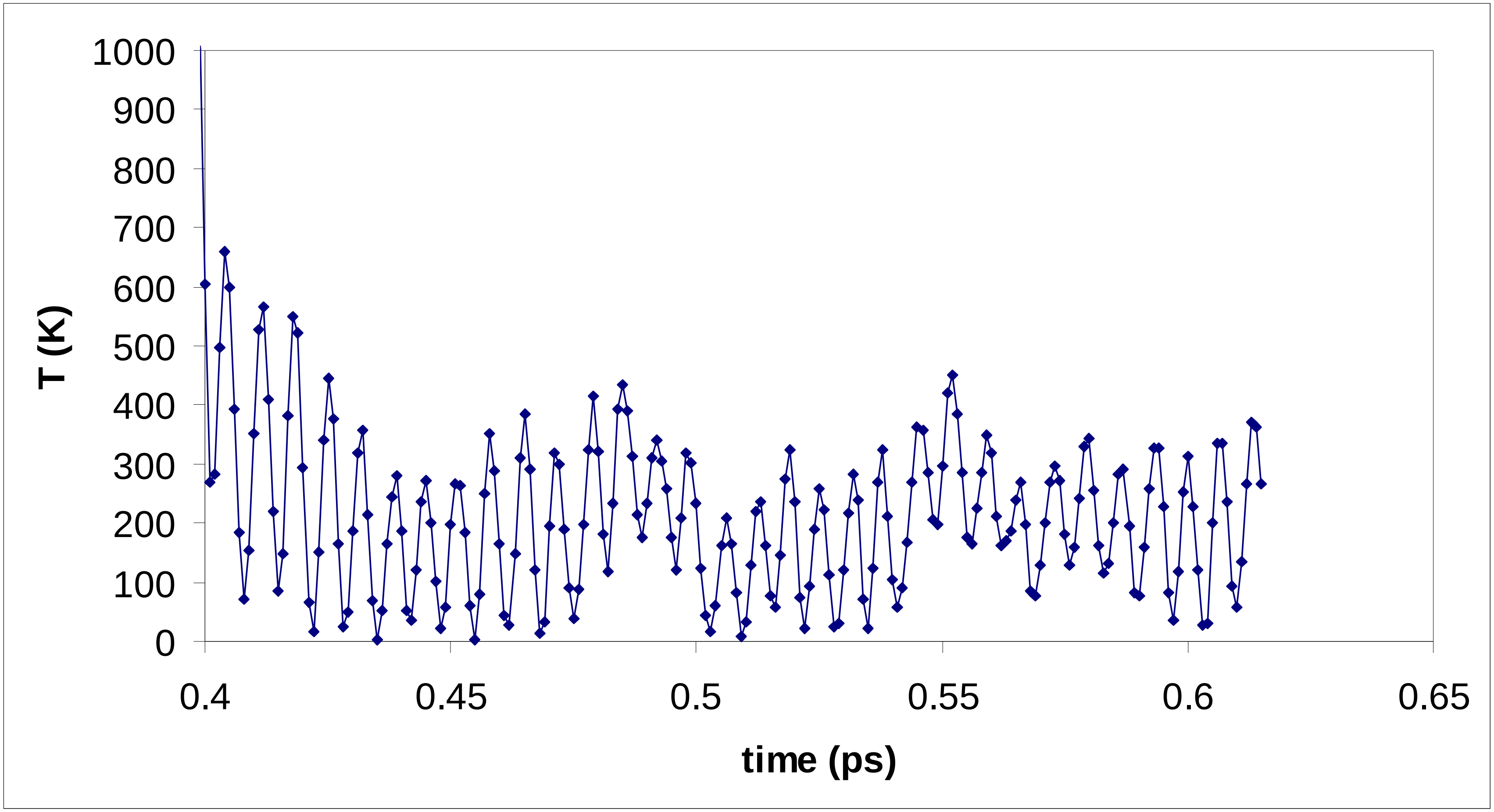
What is this "constant of motion"?

- It certainly does not seem very constant!
  - It depends on the ensemble but is essentially the closest thing to the "value of the Hamiltonian" which should be a conserved quantity:

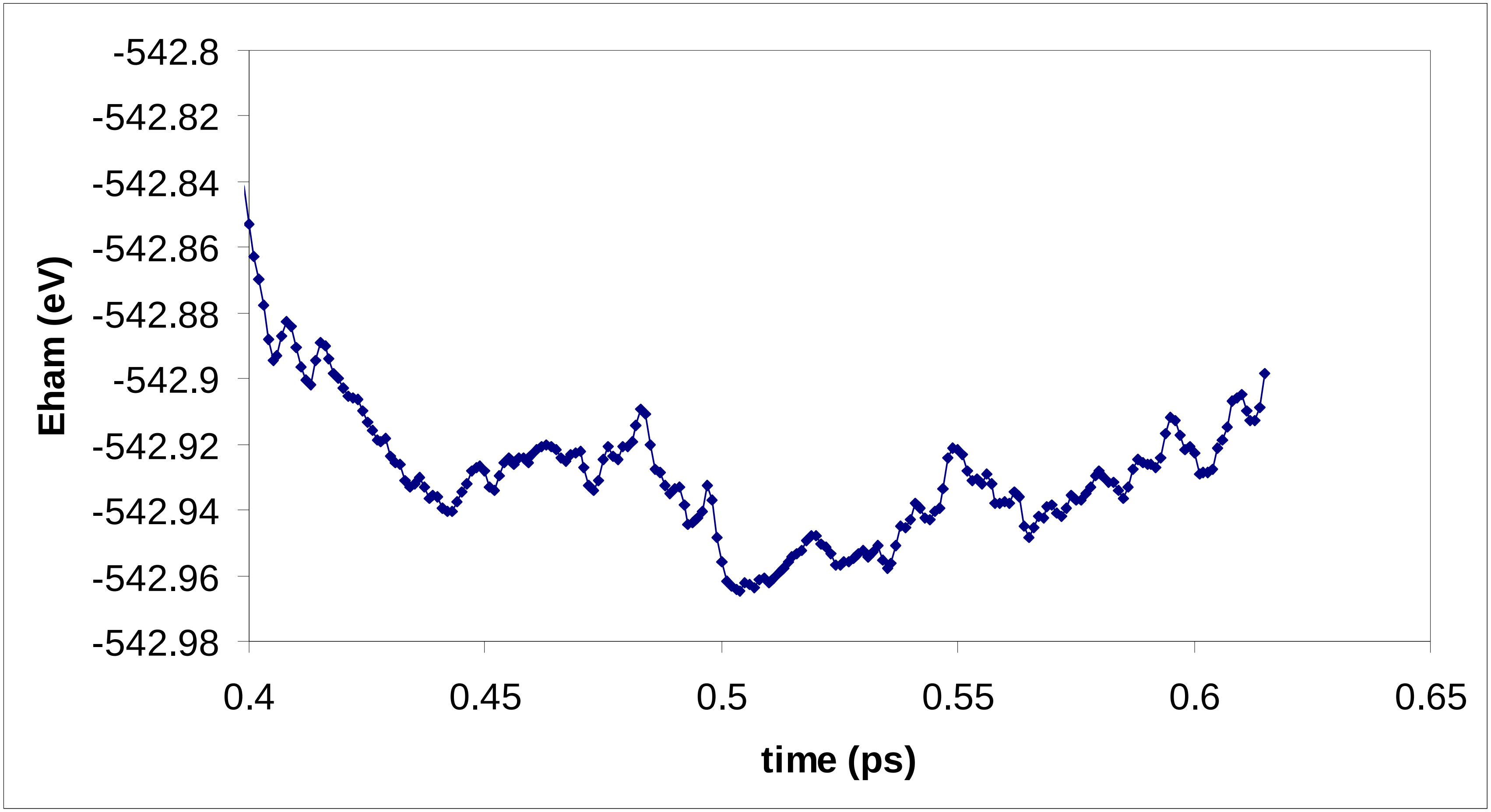
$$
\begin{aligned}
NVE: &\quad E_{Ham} = E_{electrons} + KE_{ions} \\
NVT: &\quad E_{Ham} = E_{electrons} + KE_{ions} + PE_{NHC} + KE_{NHC} \\
NPH: &\quad E_{Ham} = E_{electrons} + KE_{ions} + p_{ext}V + KE_{cell} \\
NPT: &\quad E_{Ham} = E_{electrons} + KE_{ions} + p_{ext}V + KE_{cell} + PE_{NHC} + KE_{NHC}
\end{aligned}
$$

May fluctuate on short times but no long-term drift!

Ignoring initial T transient



Ignoring initial T transient



So …

- Better but still some wobble in T – why?
- T is only strictly defined as a *macroscopic* quantity – what you are seeing is the *instantaneous* KE of a 2-particle system!
- Hence it is the *average* T that is important and should be conserved: <T>=217± 140 K
- And that will have a stat. mech. finite size variation given by

$$\frac{\delta T}{T} \approx \frac{2}{3\sqrt{N_{ions}}}$$

- T*=273 ± 129 K

CASTEP MD keywords

Most set in the param file:

- `task=Molecular Dynamics`
- `md_num_iter=10000`
- `md_delta_t=1.0 fs`
- `md_ensemble=`NVE or NVT or NPH or NPT
- `md_temperature=300 K`
- `md_thermostat=Langevin` or `Nose-Hoover`
- `md_barostat=Andersen-Hoover` or `Parrinello-Rahman`

should be obvious but what about `md_ion_t` or `md_extrap`? What do they do?

Nosé-Hoover keywords

- Nosé-Hoover chains are a standard deterministic way of thermostating system
  - Add an extra degree of freedom to the Lagrangian, to represent heat-bath with coupling depending on the instantaneous and target temperatures
  - But is not guaranteed to be ergodic
- One way to improve this is to add a thermostat to the thermostat etc … resulting in a Nosé-Hoover chain
  - `md_nhc_length=5` sets the length of this chain
  - `md_ion_t = 100 fs` sets the characteristic time for the feedback – for most efficient thermostating you want to set this time to resonate with dominant period of your system

Langevin keywords

- Langevin dynamics are an alternative and stochastic way of thermostating system
  - Implements a heat bath via Fluctuation-Dissipation theorem
  - `md_ion_t = 100 fs` sets the characteristic time for the feedback - set this to be longer than the dominant period of your system
  - Typically `5*md_ion_t` is sufficient to lose all trace of initial conditions and be in equilibrium
  - Guaranteed to be ergodic if run long enough

Barostat keywords

- What about the barostat? How is that controlled?
- In all MD schemes, the barostat is implemented by giving something a fictitious “mass”
  - Andersen-Hoover uses $\frac{1}{3}\log(V/V_0)$ whilst Parrinello-Rahman uses the cell **h**-matrix
- In both cases, this “mass” is set by `md_cell_t` which sets the time scale for relaxations of the cell motion. Should be slow …

Extrapolation Explained

Background

- With *ab initio* MD, the forces and stresses are derived from the wavefunction φ
  - Hence need a converged φ at each time step
- With CPMD, this is achieved by integrating the wavefunction and the ionic positions together
- CASTEP uses BOMD and hence must re-minimise φ each time, which is costly
- Wavefunction extrapolation is a useful speedup:
  - instead of using $\varphi(t)$ as the initial guess at the new $\varphi(t+\delta t)$ we extrapolate forwards in time using the MD integrator as a framework

Extrapolation keywords

- BUT we do not know the acceleration of φ
  - Approximate it using *known change in* φ *over previous time steps*
- If we use the current value of φ and that at the previous time step, then we have a 1st order extrapolation scheme: `md_extrap = first`
- Using the pre-previous time as well leads to a 2nd order scheme: `md_extrap = second`
- We can also switch between 1st and 2nd on alternate steps as a compromise: `md_extrap = mixed`
- The extrapolation can be done using coefficients fitted to the instantaneous behaviour of the ionic MD (`md_extrap_fit=true`) or using constant coefficients (`md_extrap_fit=false`)

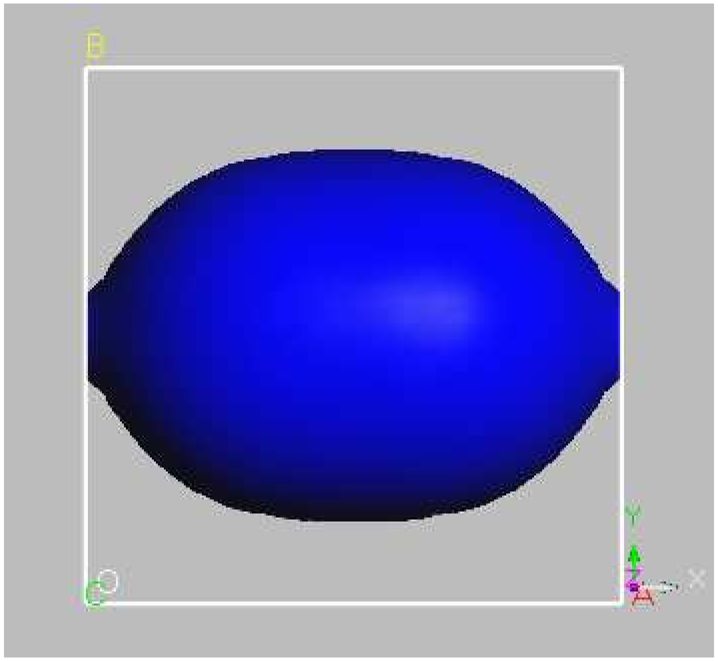
Go-faster Stripes

- CASTEP uses convergence window to determine SCF convergence
  - default is for `elec_convergence_win=3` SCF iterations to be within `elec_energy_tol` (default 10^{-5} eV/atom)
- With well-behaved MD this can be over-kill
  - The extrapolation saves many SCF cycles
  - Hence can use `md_elec_energy_tol` and `md_elec_convergence_win` to slacken tolerances if all is well.

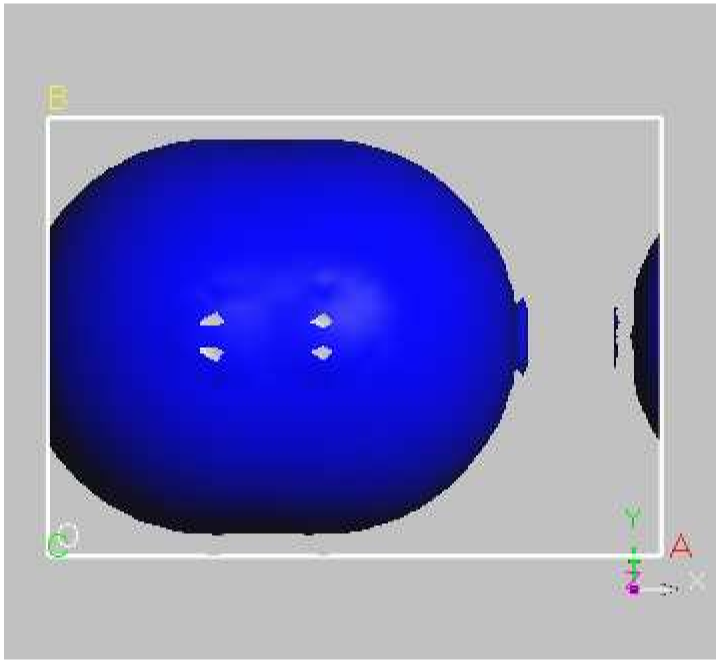
Doing N2 “properly”

- 1) Do a proper convergence test for cut-off energy at fixed k-sampling → 400 eV
- 2) Check for finite size interactions

5x5x5 A, 0.01 charge isosurface



7x5x5 A, 0.001 charge isosurface



Doing N2 “properly”

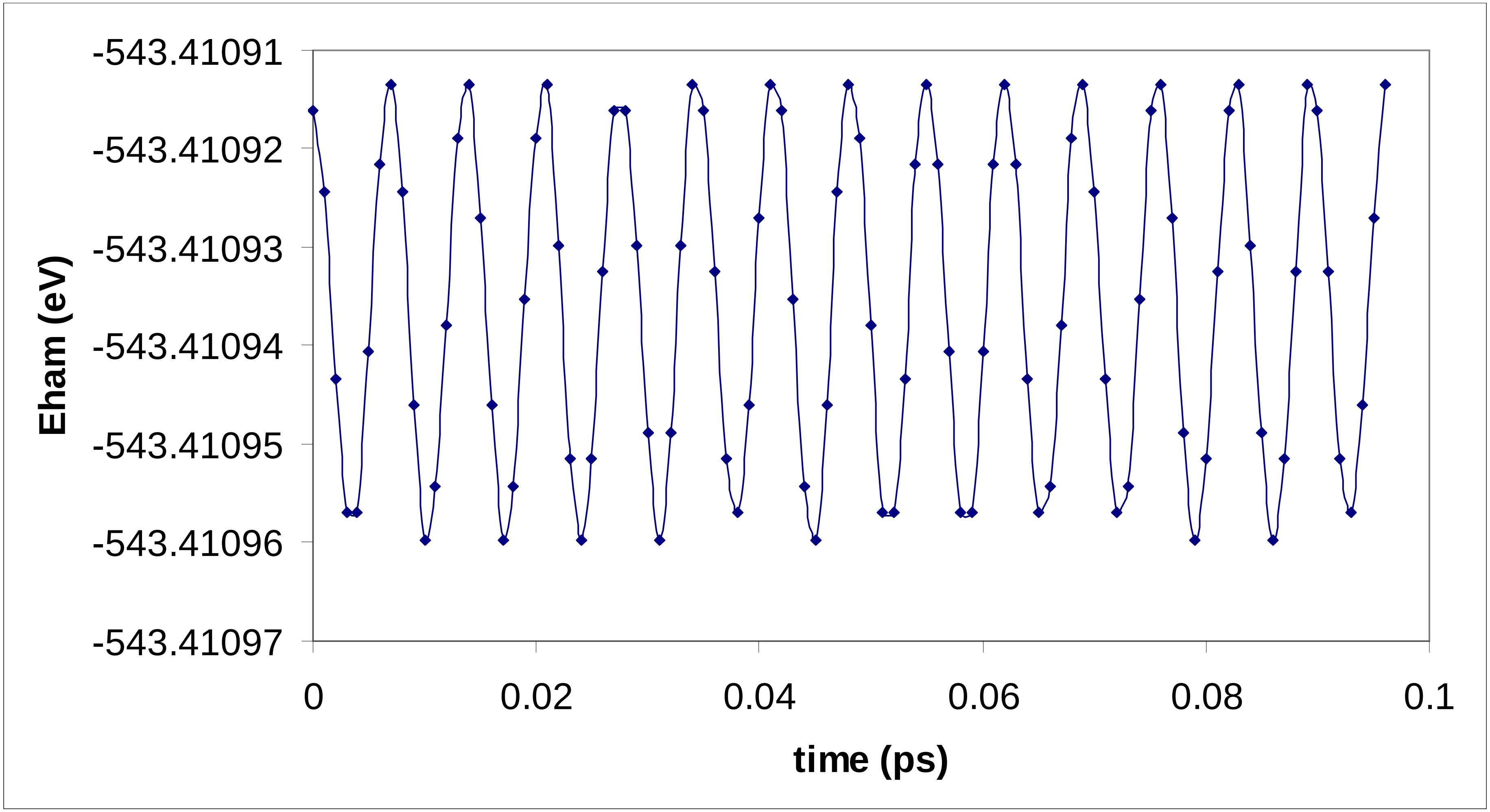
- Now do geometry optimisation:

  $\delta E \sim 0.1$ meV, final freq. est. = 2387.5 cm^{-1}

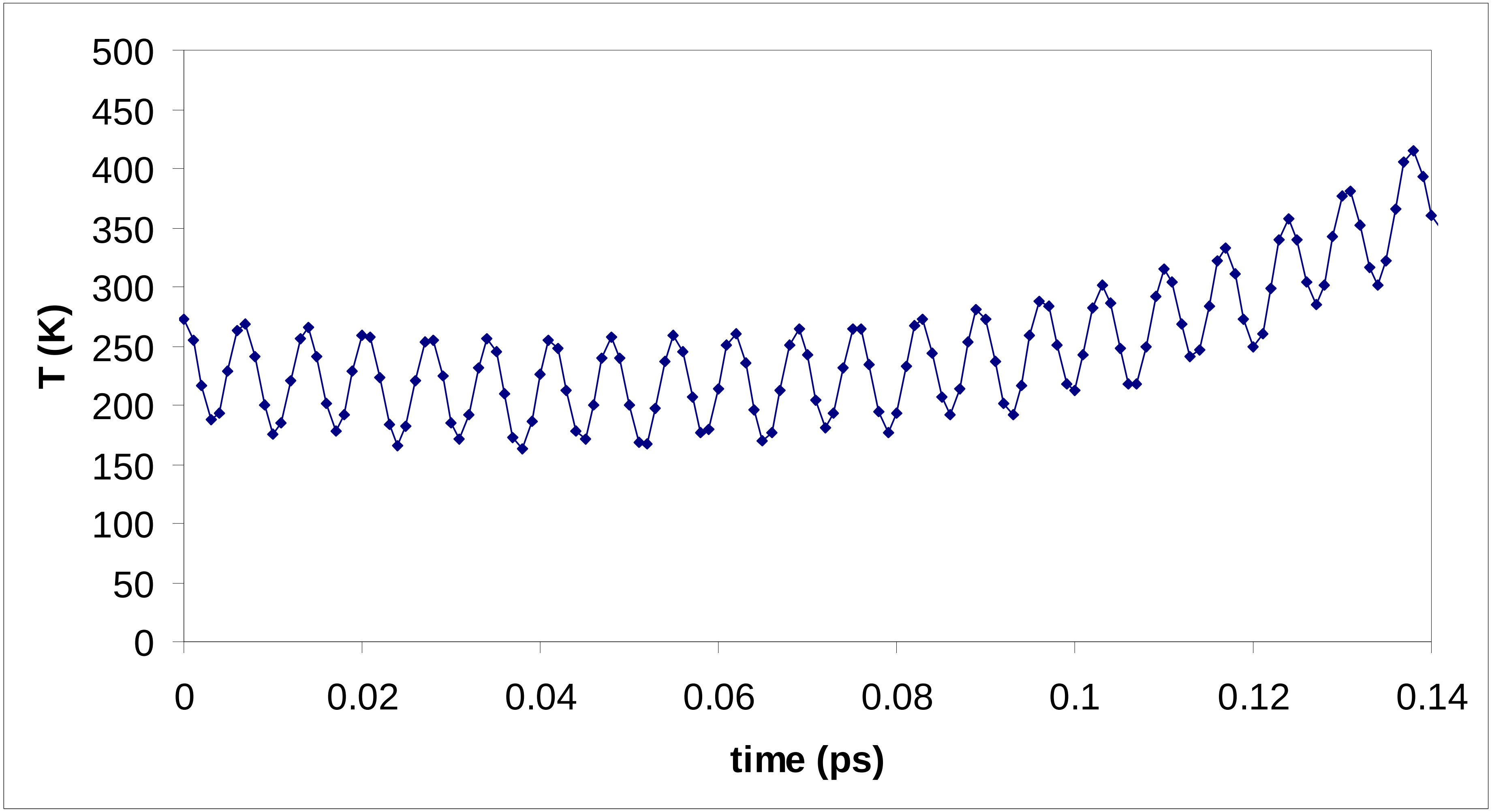
  (this is automatic from BFGS analysis)

  $\rightarrow \tau = 1/(100.c.\nu) \sim 15$ fsec so $\delta t=1$ fsec OK?
- Can change units of CASTEP input/output
  - e.g. `energy_unit = kcal/mol`
  - e.g. `frequency_unit = THz`, etc
- Now do NVE run – best for testing quality of MD – using default T=273 K:

Doing N2 “properly”

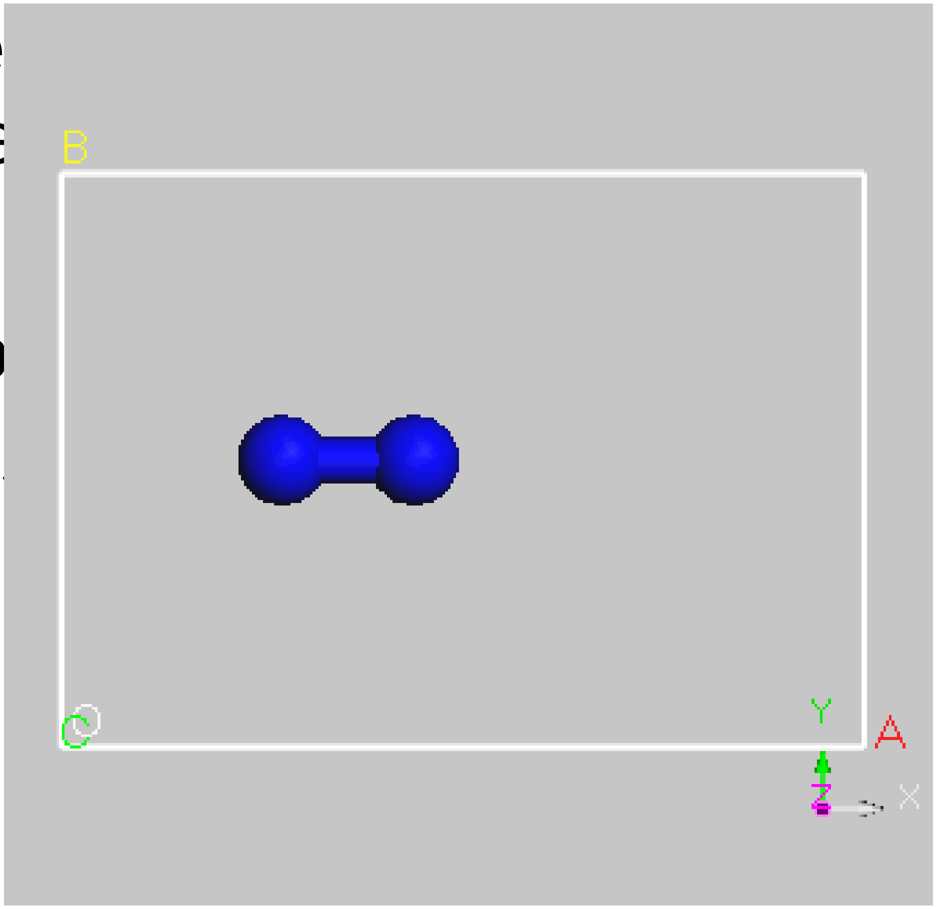


Doing N2 "properly"



Now what?

- Problem is in the velocity initialisation:
  - – assigning a tempe
    velocity to each de
  - – this leads to motio
  - – hence molecule ro
  - – hence falls foul of

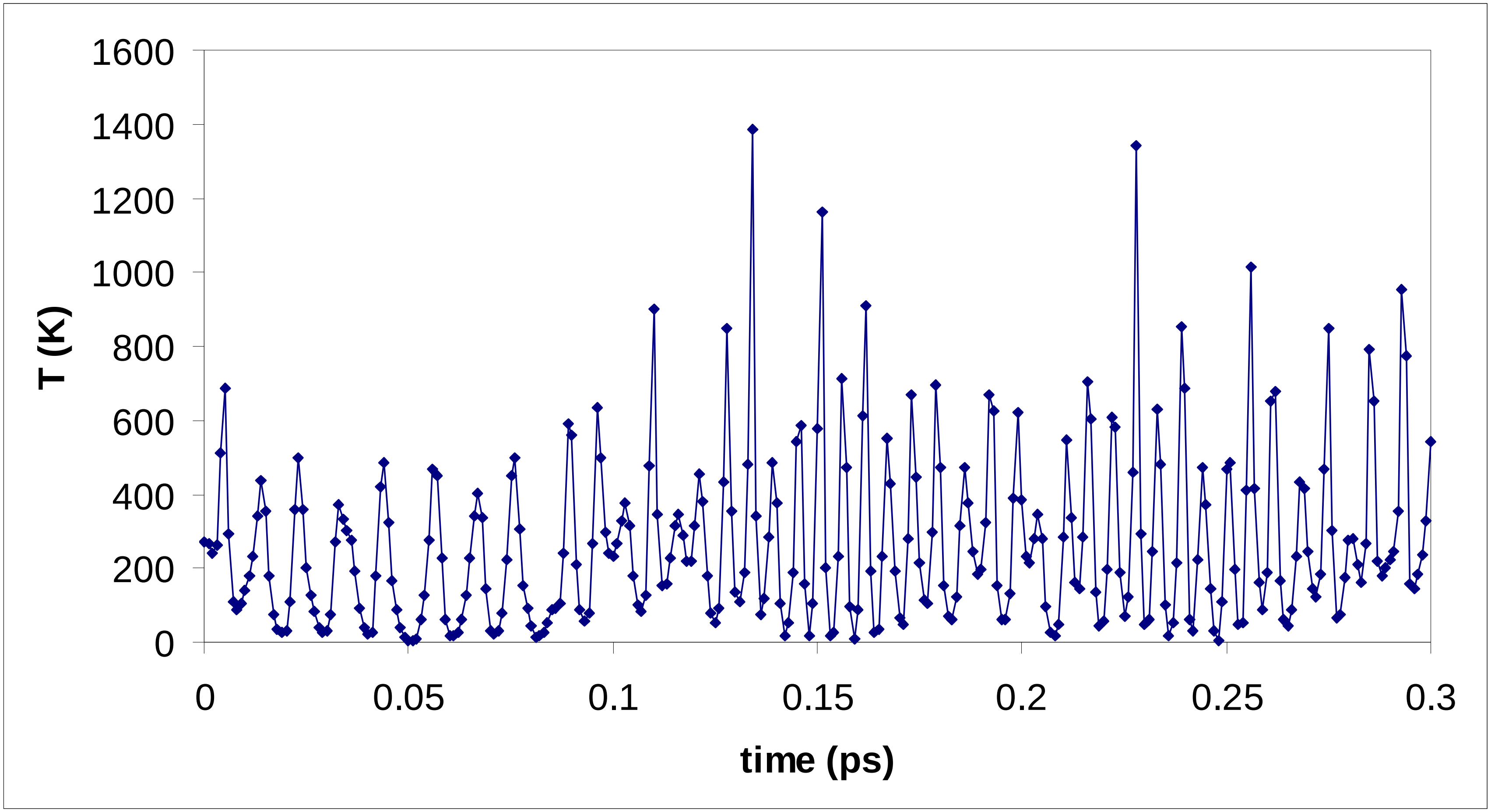


Solution is to use a 7x7x7 box or control the initial velocities

Default Nosé-Hoover in 7A^3 box

$\langle T_{150\text{-}300} \rangle = 299 \pm 256$ K

$\langle E_{ham} \rangle = -543.40 \pm 0.02$ eV



Velocity Control

- If doing NVE or NPH then can set T=0 K
  - But not if doing NVT or NPT!
  - So any initial velocity comes from the initial strain w.r.t. equilibrium, or by user input
- Can set up any condition by editing the .cell file, e.g.

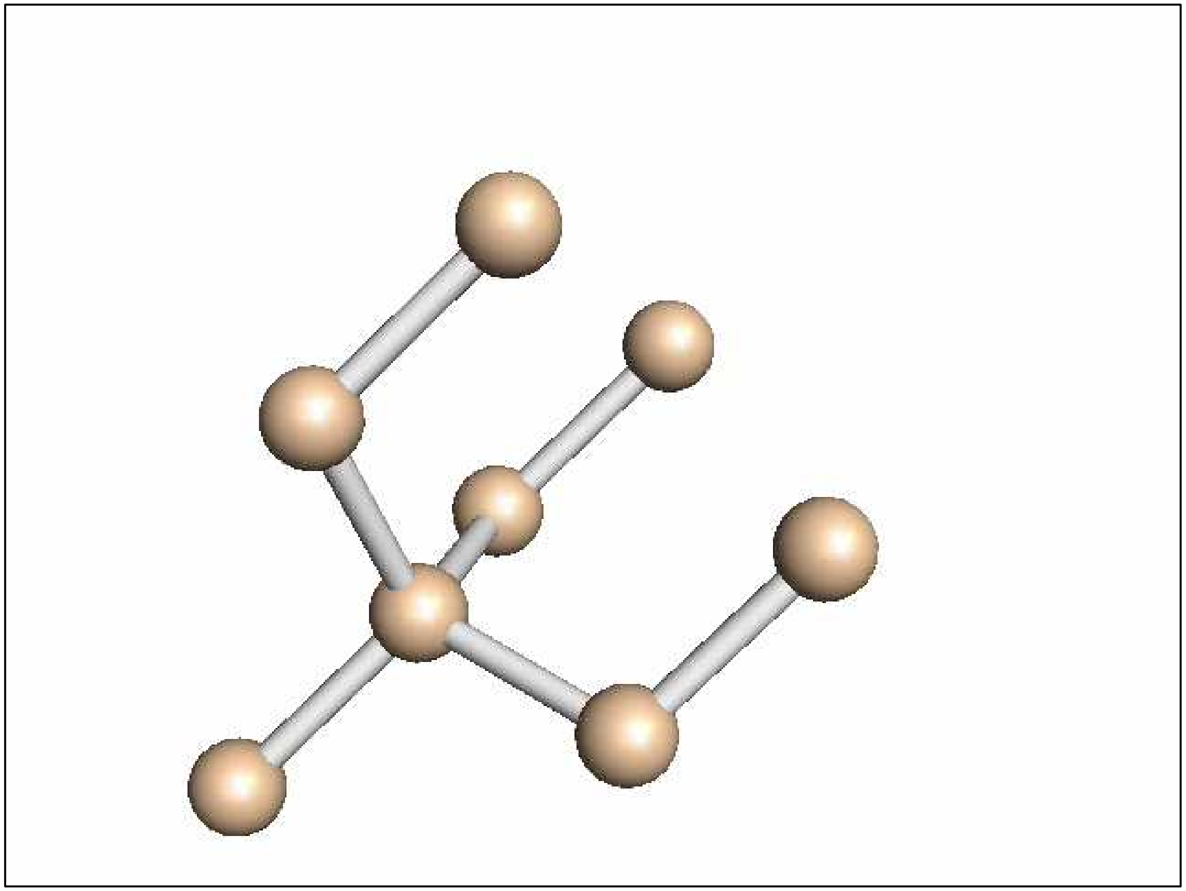
```
%block IONIC_VELOCITIES
ang/ps
  12.7 12.7 12.7
   0.0  0.0  0.0
   0.0  0.0  0.0
   0.0  0.0  0.0
<etc>
%endblock IONIC_VELOCITIES
```

NB $\sqrt{3}*12.7$ Ang/ps ~ speed of sound in silicon

Hence can simulate high velocity shock, non-equilibrium MD, etc

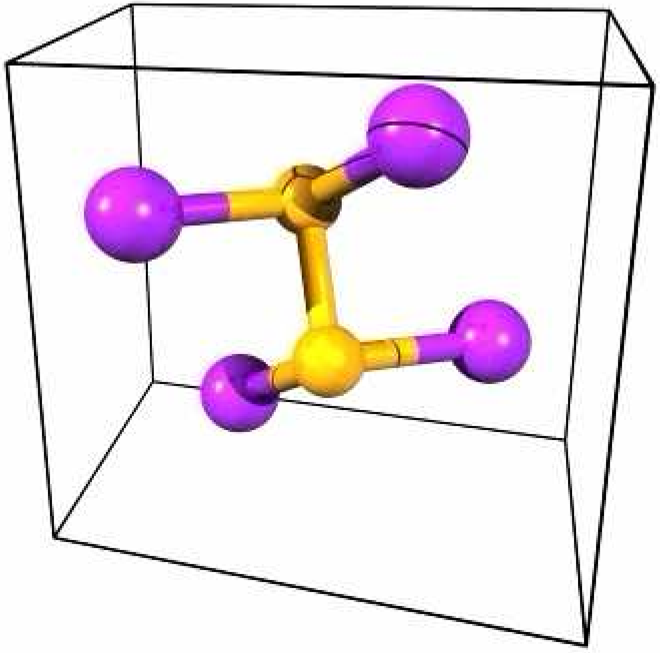
Non-Equilbrium MD

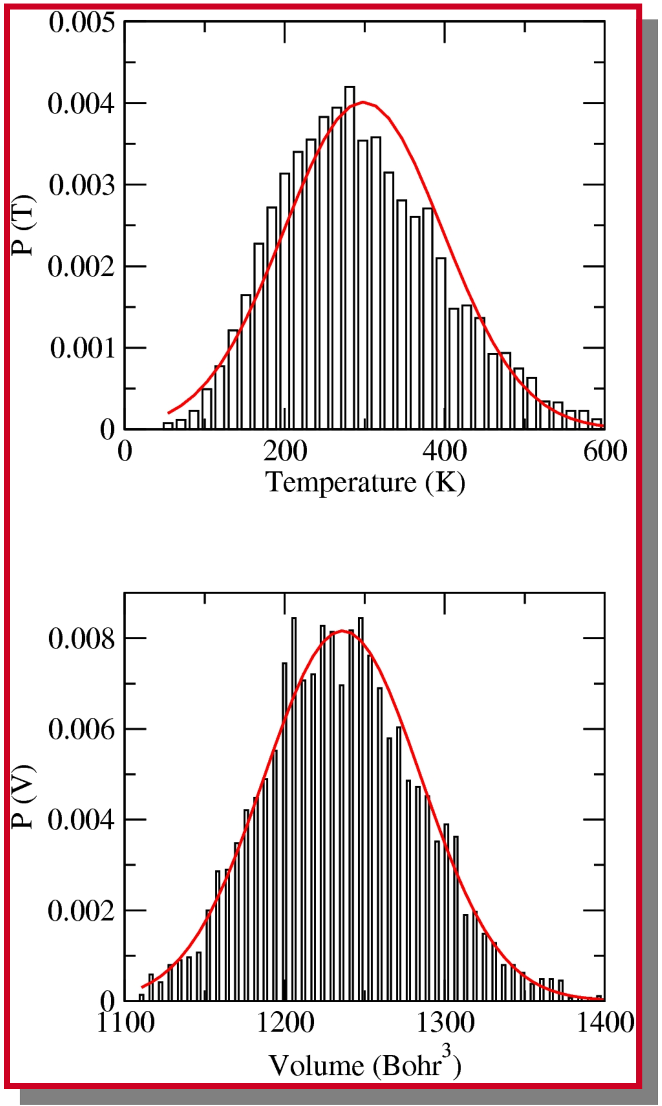
- Movie generated in FHI using PyMol and MovieMaker
- Bottom-most atom given initial velocity, others at rest …



NPT Statistical Mechanics

- *Phosphorus (II) iodide*
- soft molecular crystal with triclinic cell
- $E_{\text{cut-off}}$ = 300 eV, 3x2x2 k-points
- T=250 K, P=50 MPa
- Highly flexible cell - β_T =5.4 ± 0.1 GPa





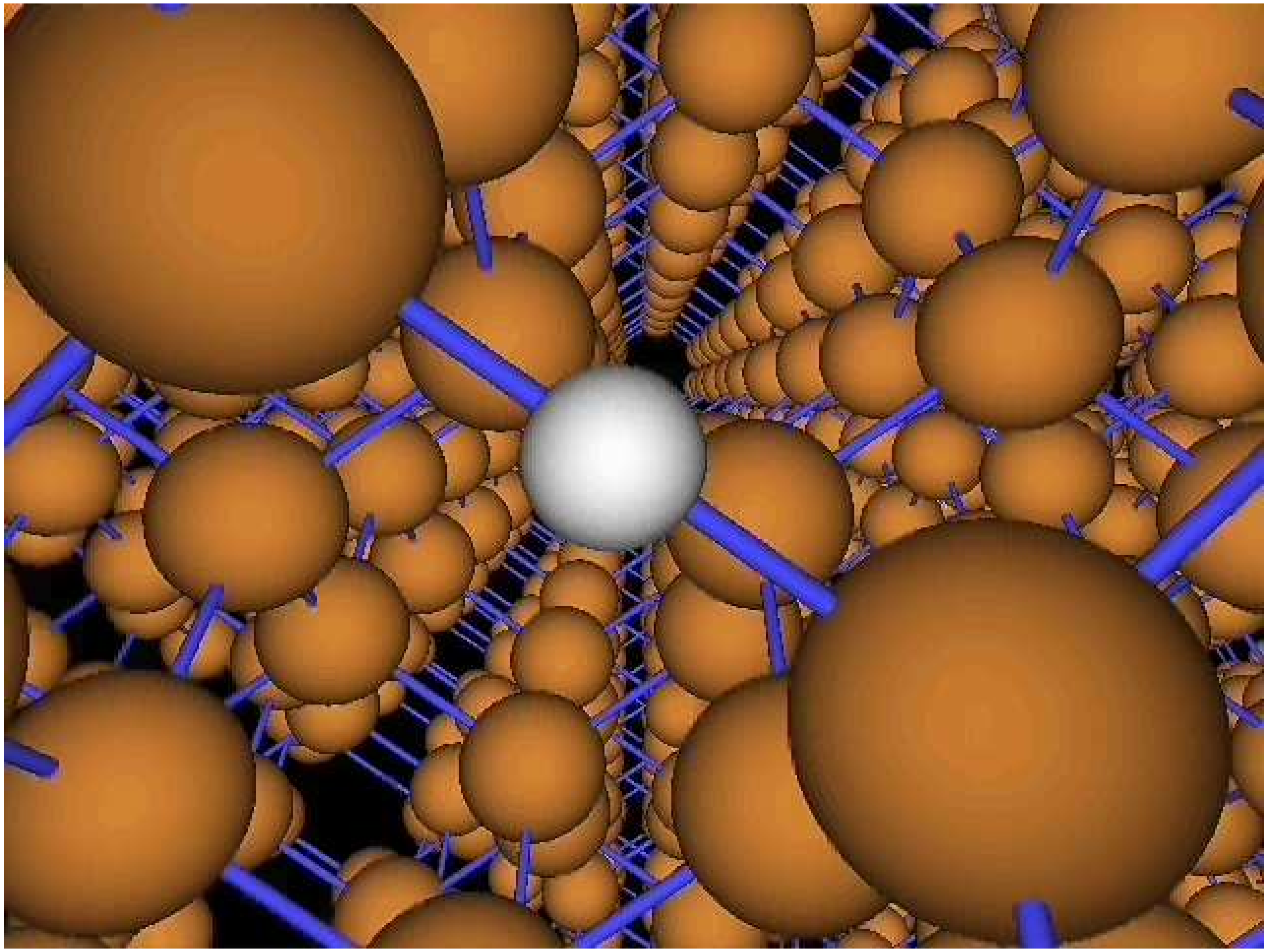
Path Integral MD

- Hydrogen defect in silicon
  - – Important defect with strong coupling of quantum ZPM to surrounding silicon lattice

```
md_use_pathint=true
md_num_beads=16
md_pathint_staging=true
num_farms=16
```

- Movie generated from .md file using PovRay to render each timestep

Path Integral MD



Computational Steering

- … is trendy
- But it has been in CASTEP for ages!
- The .param file is re-read every time step
  - – Many parameters can be changed “on-the-fly” to *steer* the calculation, e.g. `md_temperature` or `md_num_iter` or `md_delta_t`
  - – and even more parameters can be changed upon a continuation
- But not the .cell file!

Analysis

- Materials Studio will give you elementary data and analysis
- The .castep file gives a brief summary of what is happening in the user units …

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
x                                               MD Data:      x
x                                                             x
x              time :      0.001000                   ps      x
x                                                             x
x   Potential Energy:   -543.432706                   eV      x
x   Kinetic   Energy:      0.034494                   eV      x
x   Total     Energy:   -543.398212                   eV      x
x   Hamilt    Energy:   -543.397578                   eV      x
x                                                             x
x        Temperature:    266.854751                    K      x
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

Analysis

- More advanced analysis requires more data, for which we use the .md file.
- This contains a LOT of information, for each time step, always using atomic units:

```
                       1.19476569E+004
             -1.99707968E+001     -1.99692125E+001      9.64993404E-004  <-- E
               6.43328936E-04                                            <-- T
              1.32280829E+001      0.00000000E+000      0.00000000E+000  <-- h
              0.00000000E+000      1.32280829E+001      0.00000000E+000  <-- h
              0.00000000E+000      0.00000000E+000      1.32280829E+001  <-- h
N      1      4.83250673E+000      3.95868000E+000     -3.95873877E+000  <-- R
N      2      4.61612393E+000      5.48995066E+000     -5.48989189E+000  <-- R
N      1      1.15732344E-004      1.10453835E-004     -1.10452023E-004  <-- V
N      2     -1.15732344E-004     -1.10453835E-004      1.10452023E-004  <-- V
N      1     -1.83347496E-004      1.53896599E-003     -1.53886170E-003  <-- F
N      2      1.83347496E-004     -1.53896599E-003      1.53886170E-003  <-- F
```

More Analysis of MD?

- Using the .md file as input you can easily write your own analysis codes
  - e.g. MDTEP on www.castep.org
- MDTEP can calculate
  - radial distribution function, velocity autocorrelation function, mean-squared displacement, heat capacity, thermal expansion coefficient, bulk modulus, temperature and volume distributions
  - and generate .xmol and .axsf files for standard Linux visualisation programs

Miscellaneous Tips

- The choice of time step should reflect the physics not the algorithm
  - e.g. smallest phonon period/10
  - effects the conservation properties of system and long-time stability
  - Langevin: `md_ion_t` ~ 10*period
  - Nosé-Hoover: `md_ion_t` ~ period
  - NPH or NPT: `md_cell_t` ~ 100*period
  - equilibration time ~ 5*max(`md_ion_t`, `md_cell_t`)
- Can use constraints to increase time step
  - freeze motions that are not of interest

Use of Constraints

- Based upon an extended Lagrangian
- Can do any number of linear constraints
  - e.g. Fix atom, centre of mass, relative positions, plane, etc.
- Non-linear constraints requires extra coding for each different constraint
  - e.g. to fix relative separation
  - bond-length constraints written but not yet fully tested and ready for general release
- Can increase time step if freeze unimportant motions, e.g. C-H bond vibrations etc.

Choice of Electronic Minimizer?

- All-bands/EDFT
  - self consistent φ and ρ
  - Variational E~$O(h^2)$, F~$O(h)$
  - Best for high-quality MD but slow
- Density-Mixing
  - Non-variational minimisation and non-self consistent φ and ρ => need high accuracy φ
  - Harris-Foulkes functional has E~$O(h^3)$
  - Energy-based convergence criteria deceiving!

Practical Tips

- Beware Equilibration
  - – sensitivity to initial conditions
  - – depends on the quantity of interest
- Not all configurations are equal
  - – sampling and correlation
  - – statistical inefficiency
- Apply basic physics to the results
  - – conservation laws, equipartition, etc

Conclusions

- MD is a useful general-purpose tool for computer experiments
  - Widely applicable
  - e.g. to study finite temperature or time dependant or non-equilibrium phenomena
  - Much more than shown here!
- This has been a brief overview
  - see references for details

References

- “*Molecular Dynamics Simulations*”
  - J.M. Haile, (1992). Beginners guide.
- “*Computer Simulation of Liquids*”
  - M.P Allen & D.J. Tildesley (1987). Old but useful.
- “*Understanding Molecular Simulation 2nd Ed.*”
  - D. Frenkel & B. Smit (2002). Very useful.

- **www.castep.org** web site
  - Useful MD and geometry optimisation tutorials, plus FAQs, on-line keyword listing, MDTEP download, etc.