THE UNIVERSITY of York

Classical Molecular Dynamics

Matt Probert
Condensed Matter Dynamics Group
Department of Physics,
University of York, U.K.
http://www-users.york.ac.uk/~mijp1

Overview of lecture

- Motivation
- Types of MD
- N_2 example
- More advanced MD

Motivation

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 MD

Types of *ab initio* MD

- Classical MD
  - 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 MD
  - 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:

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

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

  - trade-off time step vs. stability vs. accuracy
  - need accurate forces (for *ab initio*, this means converged basis set 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
- Stochastic schemes (e.g. Langevin) need to be derived differently as non-Hamiltonian!
- Recent success in merging the two – the Hoover-Langevin thermostat

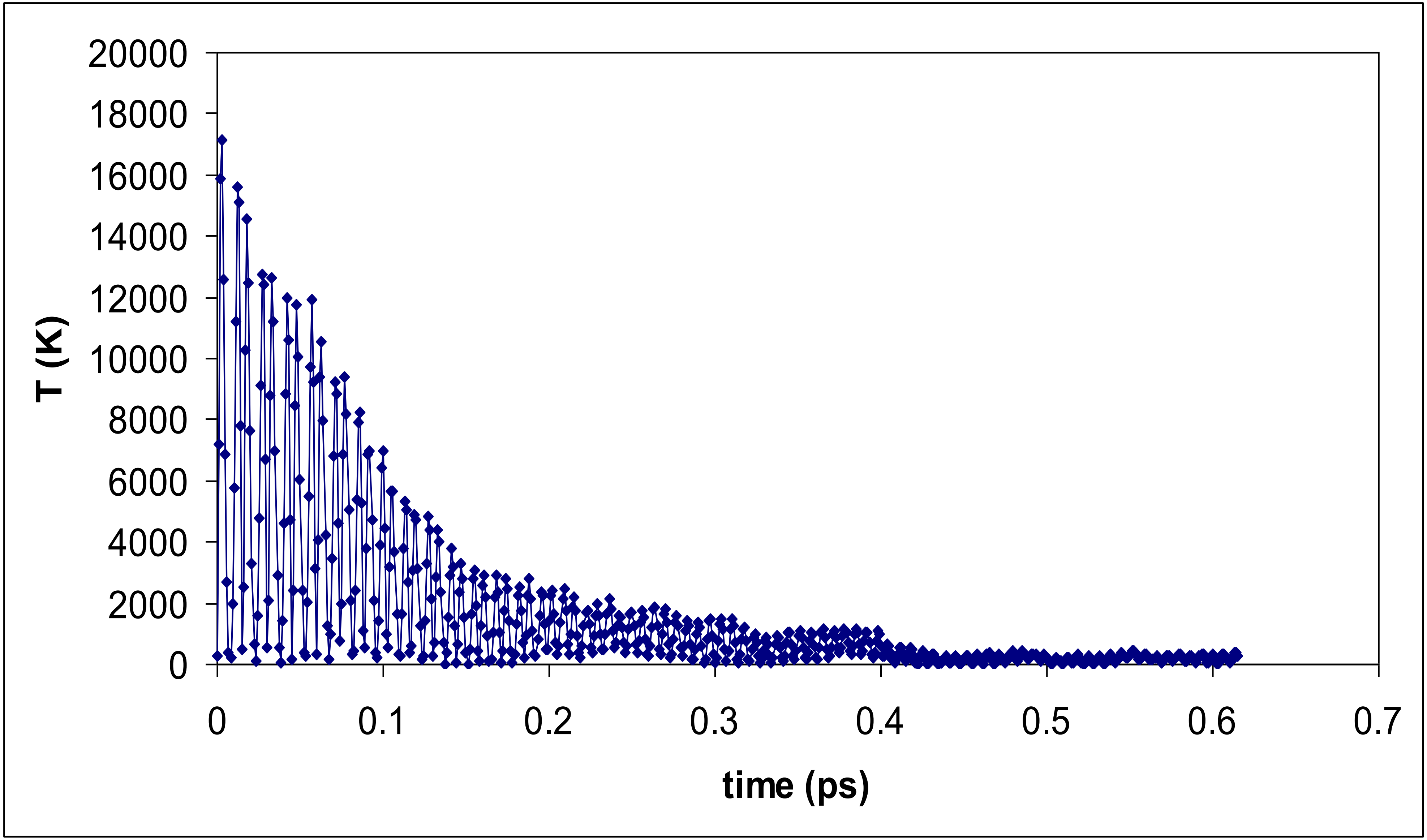
CASTEP N_2 example

Simple Example: N_2

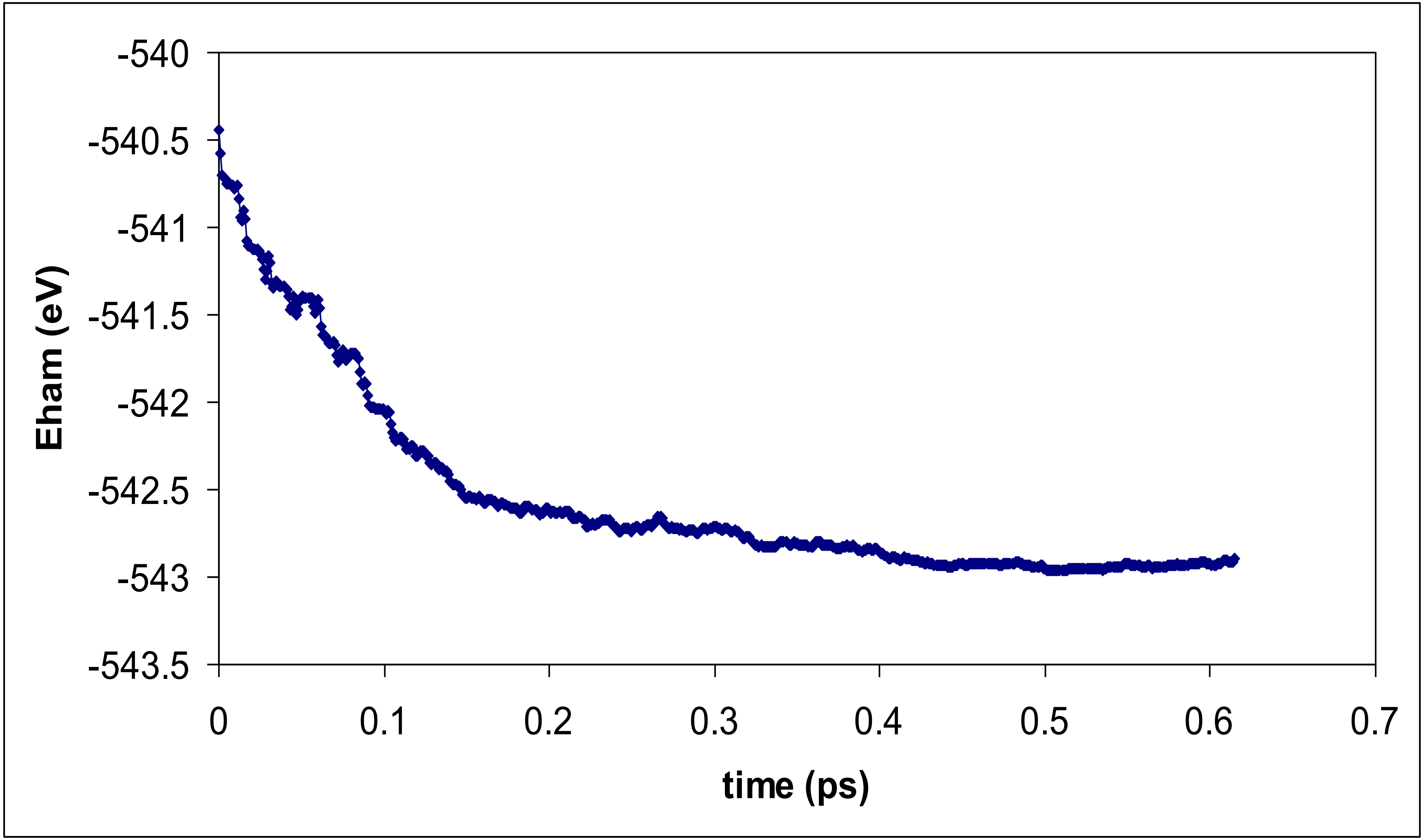
- Naïve approach:
  - put 2 N atoms in a 5 A box at (0.4,0.5,0.5) and (0.6,0.5,0.5)
  - Use Gamma point for BZ sampling (it is an isolated molecule after all ☺)
  - Use cheap settings, e.g. “medium” Ecut.
  - Run NVT dynamics at default T=273 K using Langevin thermostat with default “Langevin time” of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

Simple N_2 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 E_{Ham}?

- It depends on the ensemble
  - Essentially the closest thing to the "value of the Hamiltonian" which should be conserved:

$$NVE: \quad E_{Ham} = E_{electrons} + KE_{ions}$$

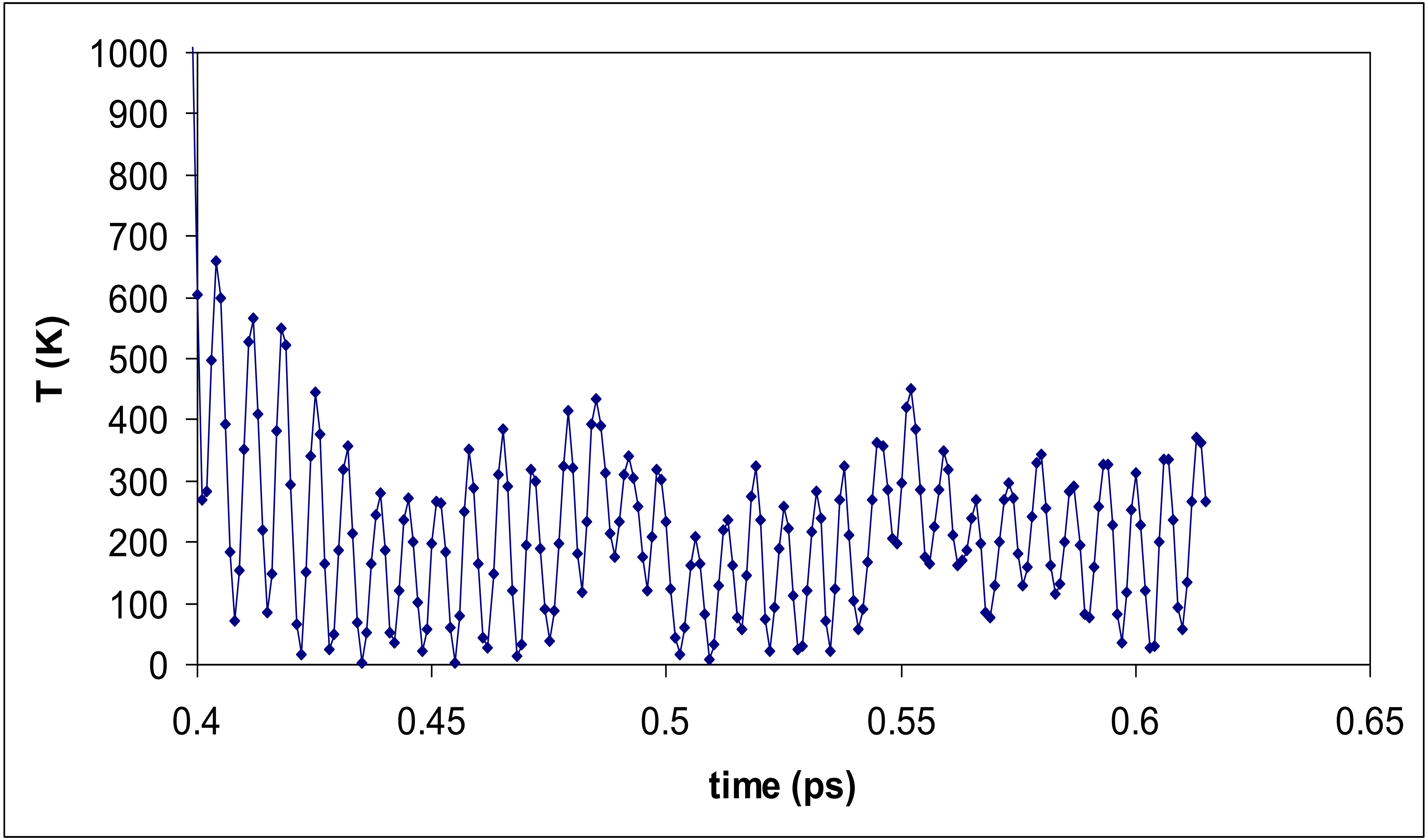
$$NVT: \quad E_{Ham} = E_{electrons} + KE_{ions} + PE_{NHC} + KE_{NHC}$$

$$NPH: \quad E_{Ham} = E_{electrons} + KE_{ions} + p_{ext}V + KE_{cell}$$

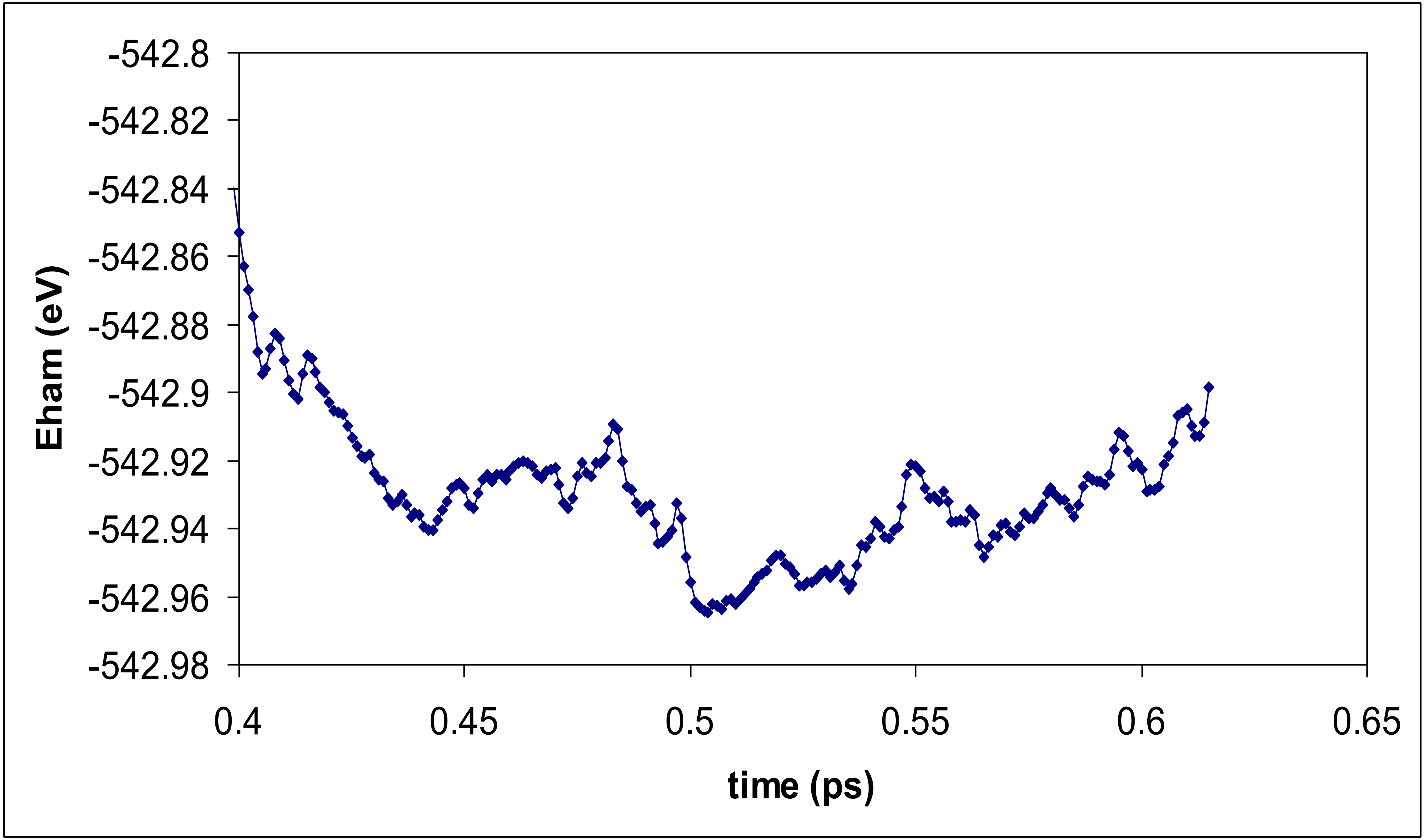
$$NPT: \quad E_{Ham} = E_{electrons} + KE_{ions} + p_{ext}V + KE_{cell} + PE_{NHC} + KE_{NHC}$$

  - It certainly does not seem very constant …
  - 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

Taking control

CASTEP MD keywords

Most set in the `.param` file, e.g.

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

should be obvious but what about things like **`md_ion_t`**? 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 = 10 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

Hoover-Langevin thermostat

- Imagine a Nosé-Hoover variable connected to your physical system, and then have a Langevin heat-bath to thermostat that
- No need for any chains!
- Physical system is deterministic, and N-H variable is ergodic – best of both worlds!
- **`md_thermostat = Hoover-Langevin`**
- **`md_ion_t`** works on the N-H variable so want it resonant with system
- MUCH LESS SENSITIVE than N-H to value

Barostat keywords

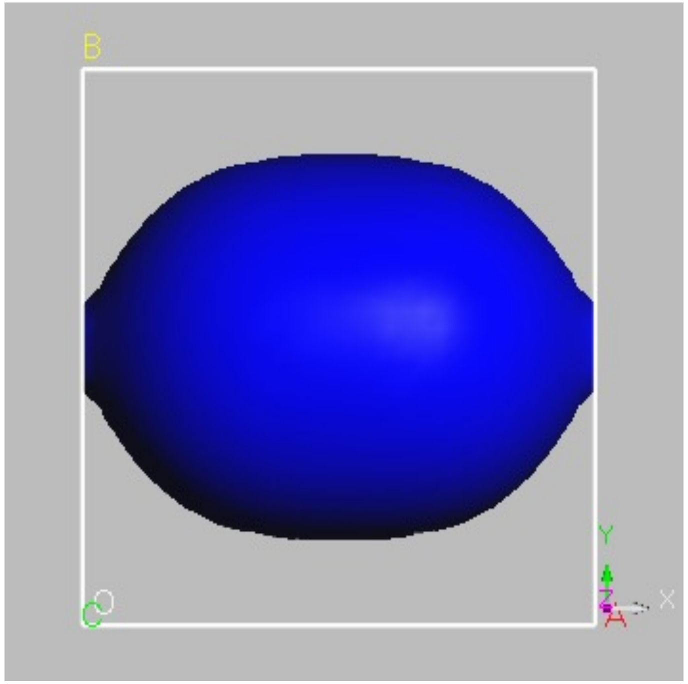
- What about the barostat? How is that controlled?
- In all MD schemes, the barostat is implemented by giving something a fictitious “mass”
  - Andersen-Hoover uses ⅓log(V/V_0) whilst Parrinello-Rahman uses the cell **h**-matrix
- In both cases, this “mass” is set by **`md_cell_t`** which sets the time scale for relaxations of the cell motion. Should be slow …

Back to N_2

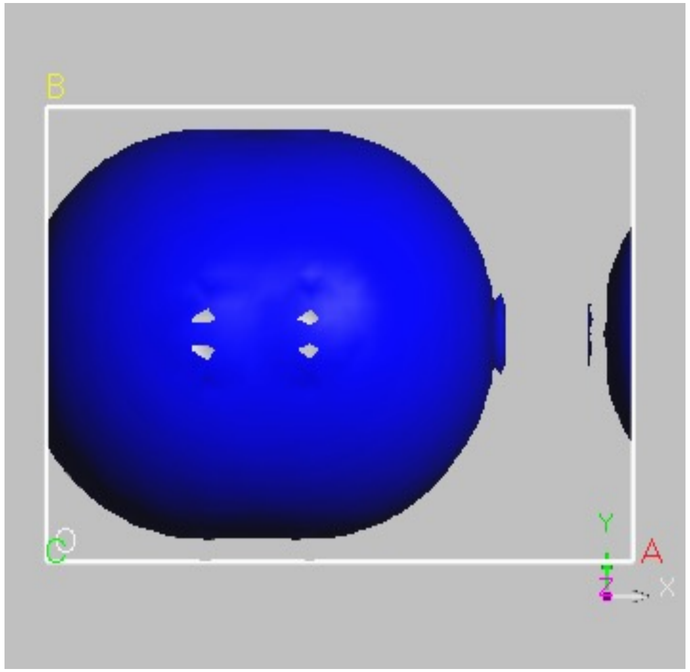
Doing N_2 “properly”

- 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



7x5x5 A, 0.001 charge isosurface



Doing N2 “properly”

- Now do geometry optimisation:

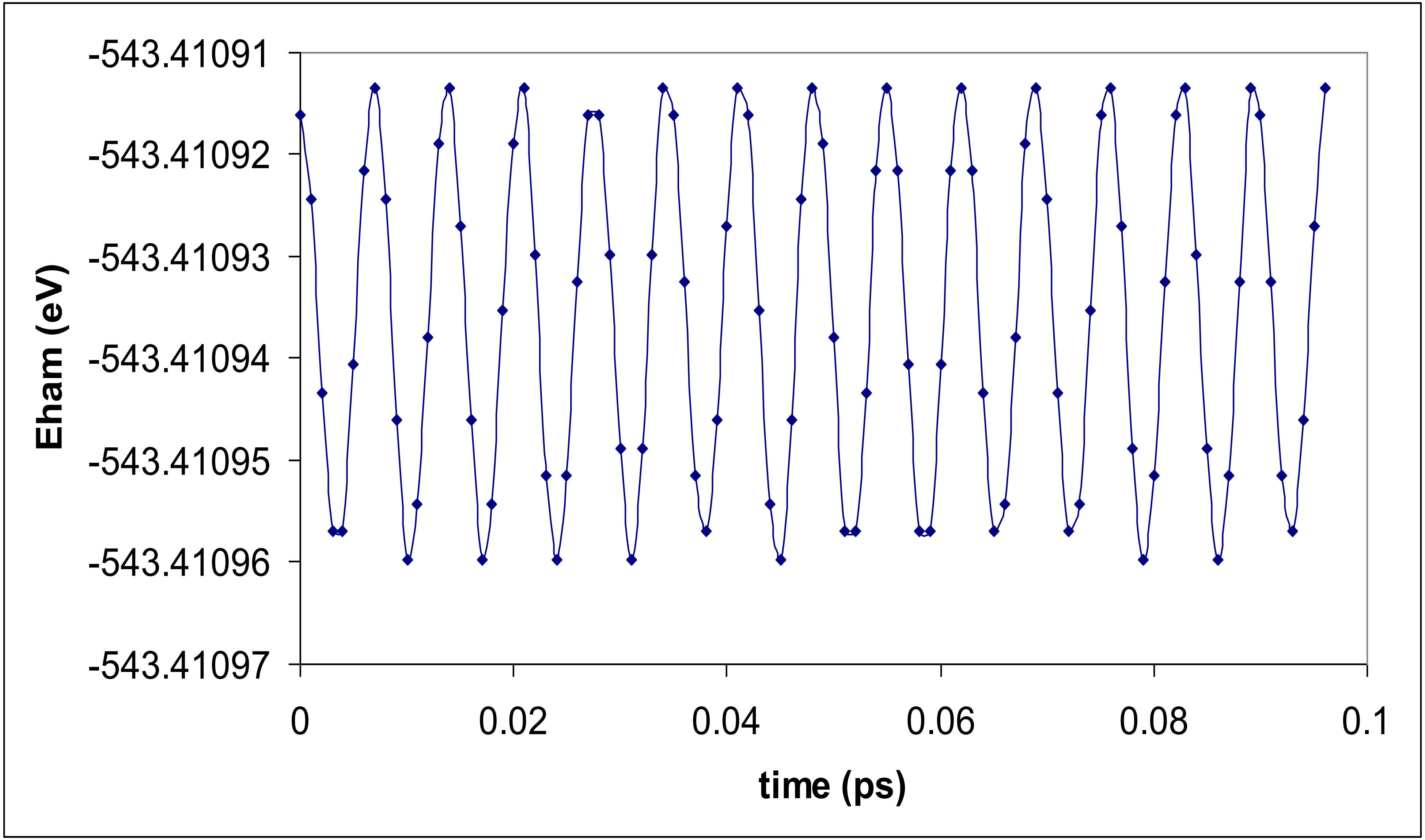
  δE ~ 0.1 meV, final freq. est. = 2387.5 cm^{-1}

  (this is automatic from BFGS analysis)

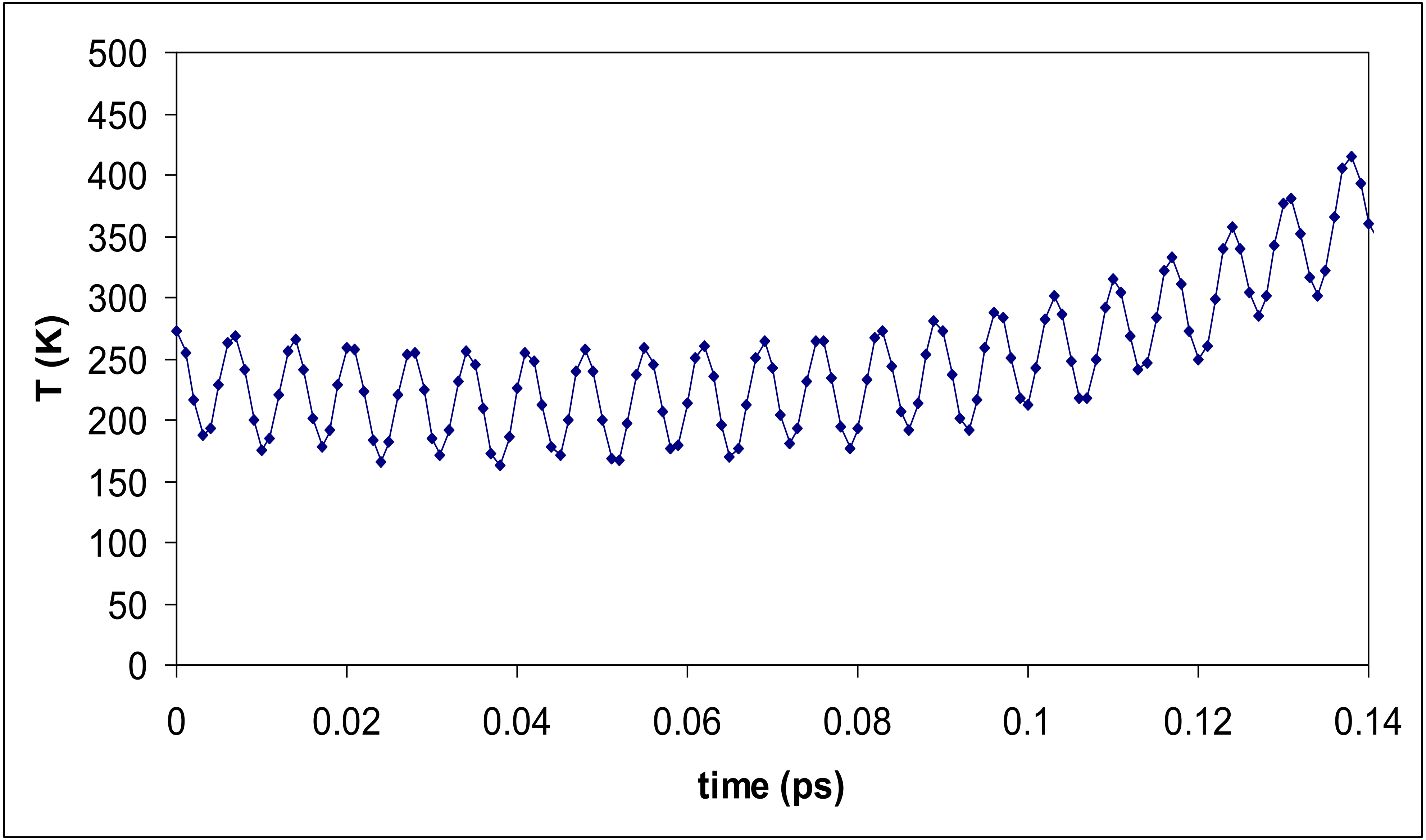
  $\rightarrow \tau = 1/(100.c.\nu) \sim 15$ fsec so δt=1 fsec OK?

- 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 N_2 “properly”



Doing N2 “properly”



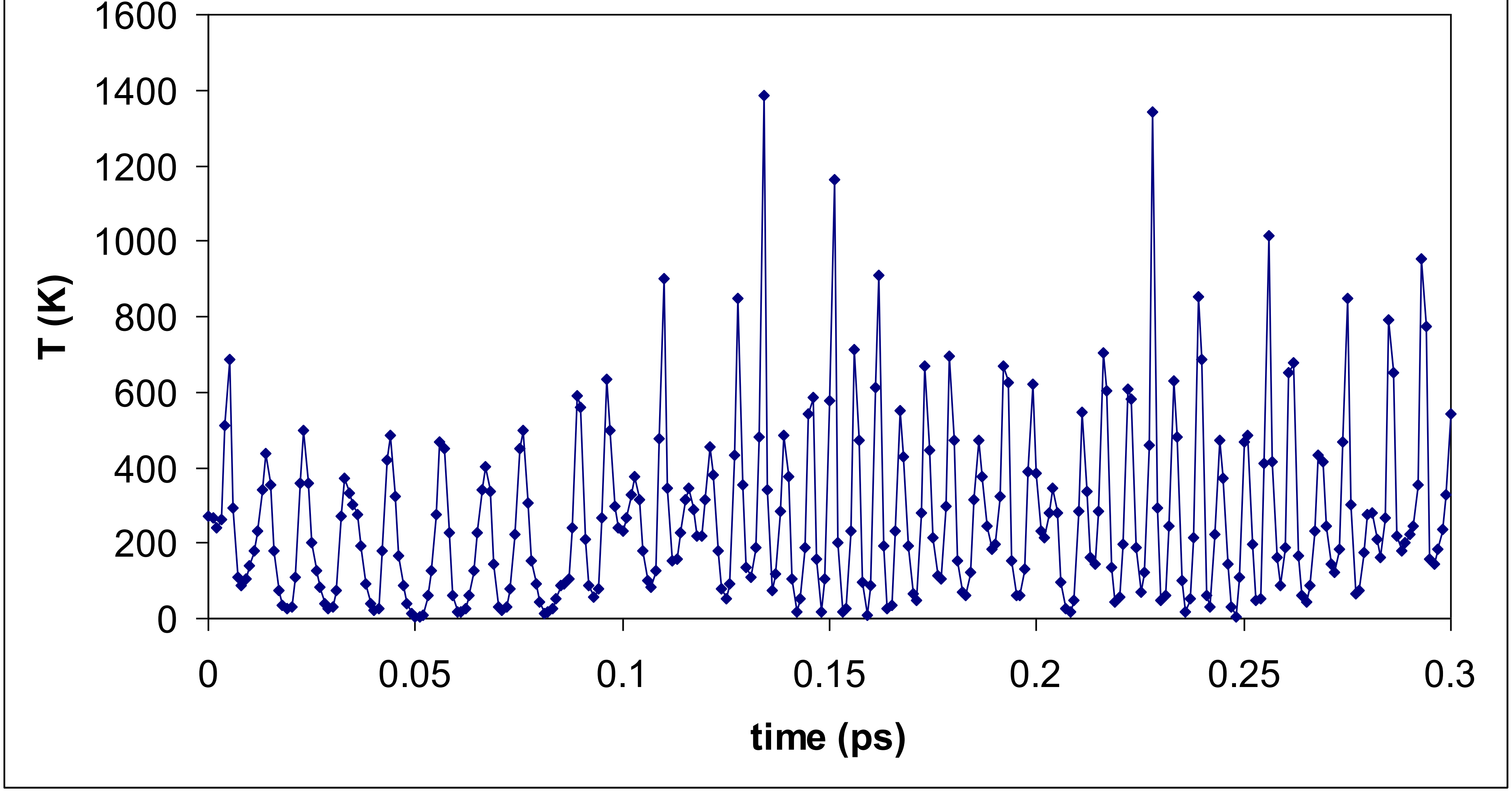
Now what?

- Problem is in the velocity initialisation:
  - assigning a temperature means a random velocity to each degree of freedom
  - this leads to motion in arbitrary directions
  - so molecule rotates
  - overlaps with small c

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 ± 256 K $\langle E_{ham} \rangle$ = -543.40 ± 0.02 eV



More advanced MD

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

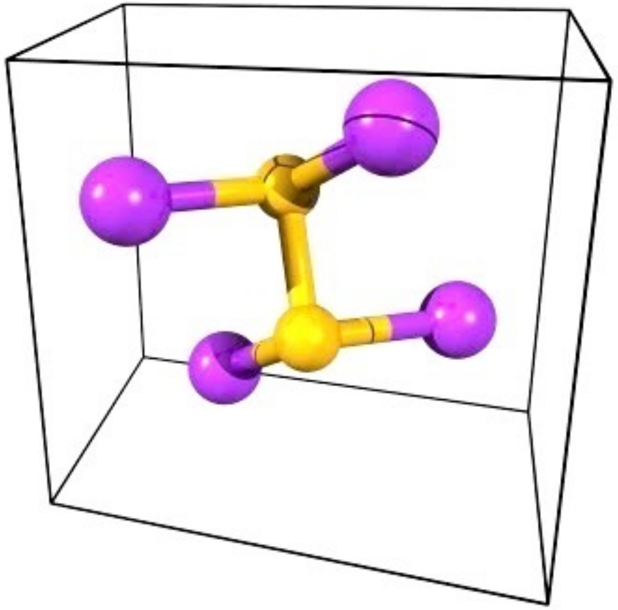
Hence can simulate high velocity shock, non-equilibrium MD, etc

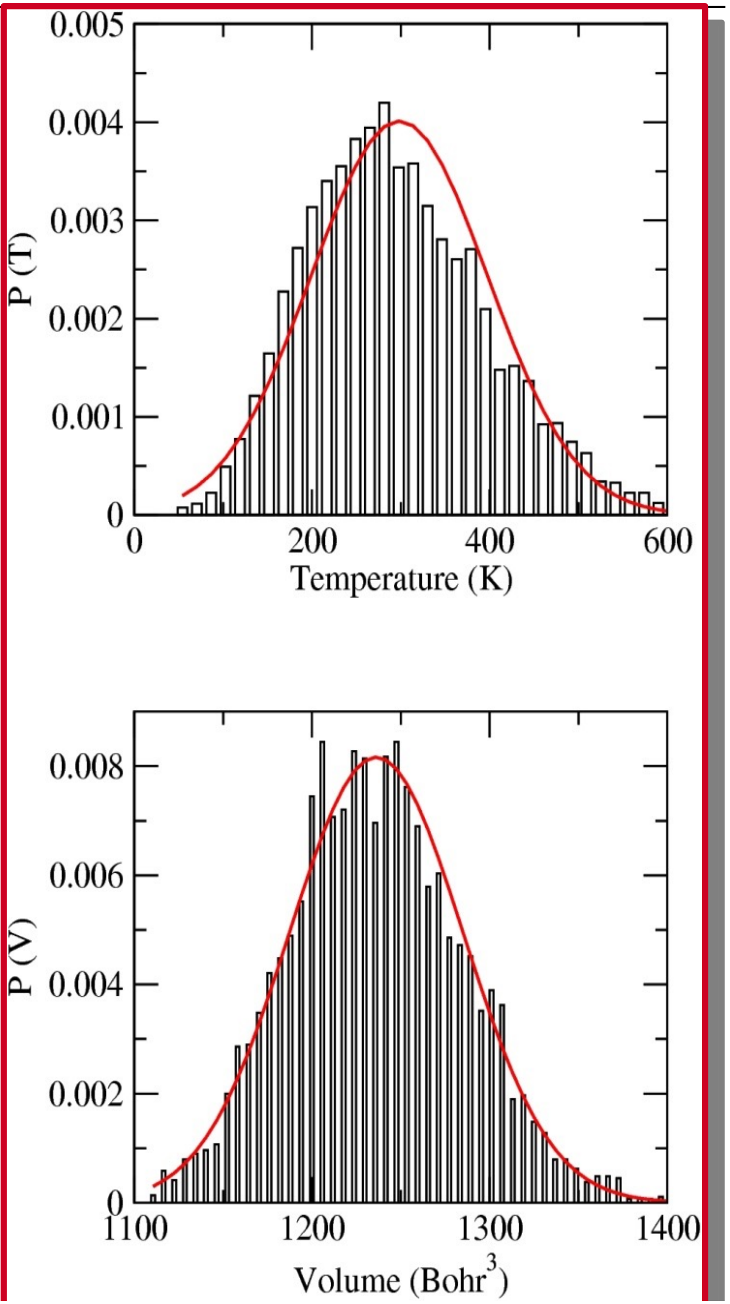
Non-Equilbrium MD

- Can set up arbitrary initial conditions and hence simulate non-equilibrium systems
  - E.g. temperature gradient to measure heat flux
  - E.g. atom hitting surface to measure sticking coefficient
  - E.g. shock wave – supersonic impact and shock propagation - useful tool for probing high Temperature and Pressure physics – and useful to simulate as experiments can be hard to interpret and perform …

NPT Statistical Mechanics

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





Analysis

- MD typically generates a LOT of data!
  - Interest in mean values + fluctuations + correlation functions + distribution functions, etc.
- Successive configurations are highly correlated so need to be careful with statistics
- Example CASTEP trajectory `.md` file:

```
                         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 – designed as an example start
- 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 and stability properties of system
  - 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`**)
  - BEWARE TRANSIENTS if sampling equilibrium properties
  - Can use fast equilibration schemes (eg Berendsen – form of velocity rescaling) to save CPU time

Constraints

- Can use to 'freeze' fast motions that are not of interest and so use larger time step
- Linear constraints, e.g. fix atom position
- Non-linear constraints using RATTLE can also fix *bond-length*

$$\sqrt{\left(r_x^1 - r_x^2\right)^2 + \ldots}$$

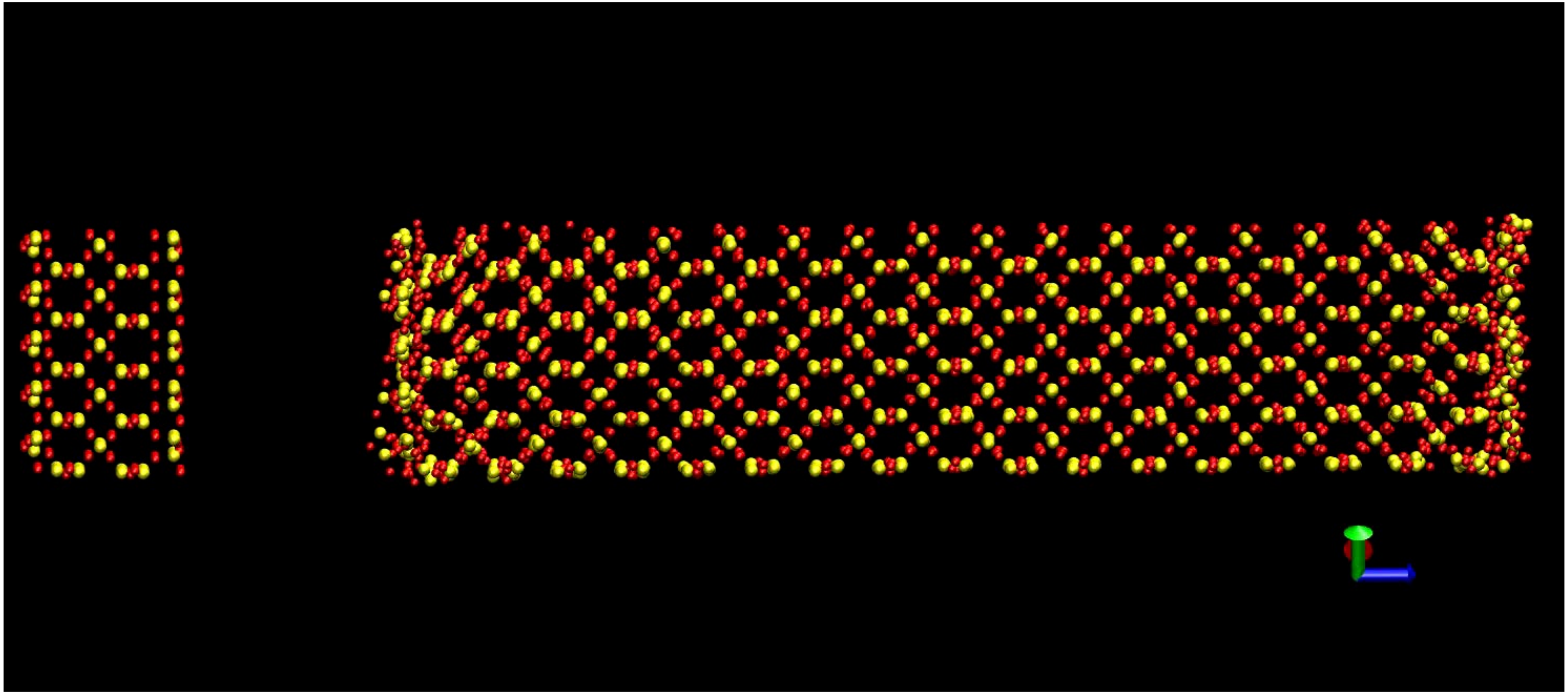
  - e.g. H_2O – can use 2-3x larger δt with 'rigid molecule' approach
  - Currently working on bond/torsion angle constraints for structure optimisation …

Sampling

- Can save time by only calculating expensive *ab initio* properties every **`md_sample_iter`** steps
- Useful as successive configurations very correlated and hence not so useful anyway!
  - E.g. Calculate pressure if **`calc_stress=true`** (useful for NVE and NVT ensembles), or
  - population analysis if **`popn_calculate=true`**
  - Or can customize code as required, e.g. to write out density to file for fancy movies, etc

Shock wave example

- 4725 atoms (672 in flyer plate) for 2 ps
- Shock = 6 km/s ~ speed of sound in quartz



Practical Tips

- MUST have good forces – watch convergence!
- 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

Summary

Summary

- 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!
- CASTEP can do all the basic MD
  - Working on some more exotic features
  - See advanced course for PIMD …

References

- "*Understanding Molecular Simulation 2nd Ed.*"
  - D. Frenkel & B. Smit (2002). Very useful.
- "*Molecular Dynamics Simulations*"
  - J.M. Haile, (1992). Beginners guide.
- "*Computer Simulation of Liquids 2nd Ed.*"
  - M.P Allen & D.J. Tildesley (2017). Updated!

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