CP2K: MOVING ATOMS

CP2K Summer School, 23-26 Aug 2016

Iain Bethune
(material from Jürg Hutter, Matt Watkins, Konstanze Hahn)

http://tinyurl.com/CP2K2016
#CP2KSummerSchool

Outline

- Geometry & cell optimisation
  - Local Minimisation
- Molecular Dynamics
  - Born-Oppenheimer MD
  - Accuracy and stability
- Ensembles
  - Thermostats

Geometry & Cell optimisation

- What do we mean by optimisation?
  - Minimising the total energy
  - aka. relaxation

- In atomistic simulations, the total energy is a function of atomic positions:
  - In DFT: $E_{tot}[n(r)]$ and $n(r) \Leftrightarrow V(\mathbf{R})$ (Hohenberg-Kohn)
  - In molecular mechanics there is a forcefield:

$$U(\mathbf{R}) = \sum_{bonds:i,j} V_{bond}(R_i, R_j) + \sum_{angles:i,j,k} V_{angle}(R_i, R_j, R_k) + \ldots$$

Geometry & Cell Optimisation

- We can think of the potential energy as a surface in a 3N-dimensional space (N = number of atoms)
  - + 9 more if we include lattice vectors for a periodic system!

- Minimas may be local or global!

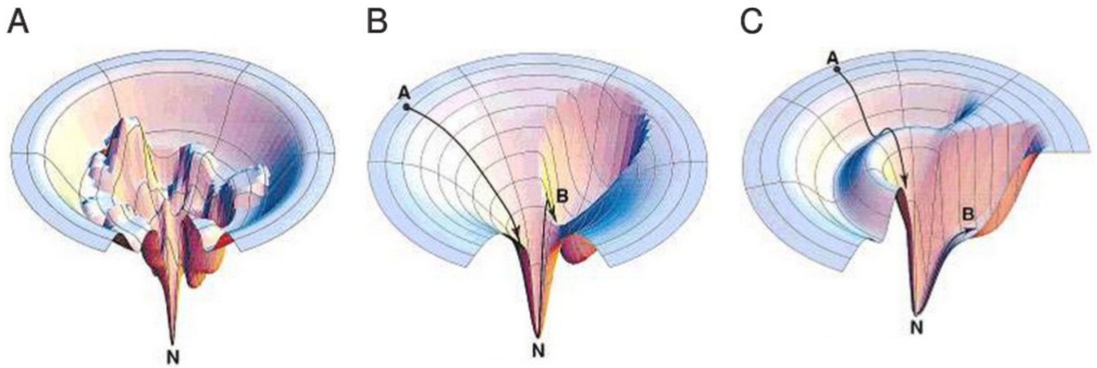


Fig. 12. Different folding scenarios. The vertical axis is internal free energy. Each conformation is represented as a point on the landscape. The two horizontal axes represent the many chain degrees of freedom. **a:** A rugged landscape with hills and traps, folding kinetics is likely multiple-exponential (from Ref. 8). **b:** A landscape in which folding is faster than unfolding. A is a through-way folding path, whereas unfolding chains (path B) must surmount a barrier to reach the most stable denatured conformations. **c:** A landscape in which folding is slower than unfolding. Most folding paths (path A) pass through a kinetic trap, whereas some low-lying denatured conformations are readily accessible from the native state during unfolding (path B).

Pic: Chan & Dill, *Proteins* (1998)

Local minimisation

- BFGS (Broyden-Fletcher-Goldfarb-Shanno)
  - most efficient for small–medium size systems with a reasonable guess at the geometry
  - requires inversion/diagonalization of approximate Hessian matrix – Hessian matrix has dimension 3*N* where *N* is number of atoms being optimized
- L-BFGS
  - A linear-scaling version of BFGS (Byrd, et al *SIAM Journal on Scientific Computing* (1995)")
- Conjugate gradients
  - Only uses gradients rather than approximation to curvature, should be more robust when far from minima

Local minimisation

- What can CP2K minimise with respect to?
  - `MOTION%GEO_OPT` – vary atomic coordinates only
  - `MOTION%CELL_OPT` – both atomic coordinates and lattice vectors
  - Some values may be constrained e.g. cell angles, certain atomic positions
  - Collective variables (distances, angles) can be constrained

Geometry optimisation

- `RUN_TYPE GEO_OPT` in `GLOBAL` section
- `GEO_OPT%OPTIMIZER` in `MOTION` section
  - CG, use with poor intial guesses, noisy forces, rough optmization
  - (L)BFGS, for most QS calculations – consider switching to LBFGS above ~1000 atoms. Look for diagonalization routine timings at end of run to see relative cost
- `MAX_ITER` number of optimization steps
- Constraints may bedefined in `MOTION%CONSTRAINT` section:

```
&FIXED_ATOMS
  COMPONENTS_TO_FIX X
  LIST 1
&END
&FIXED_ATOMS
  COMPONENTS_TO_FIX Y
  LIST 2
&END
```

Cell optimisation

- CP2K can respect cell symmetry (only for `CELL_OPT`)

```
&CELL
  ABC 9.167 9.167 11.808
  SYMMETRY ORTHORHOMBIC
  MULTIPLE_UNIT_CELL 2 2 2
&END CELL
...
&CELL_OPT
  KEEP_SYMMETRY TRUE
&END CELL_OPT
```

- Also `KEEP_ANGLES` (e.g. allows cubic symmetry to break)

Cell optimisation

- Three algorithms in CP2K controlled by `CELL_OPT%TYPE`
  - `GEO_OPT`: Original implementation.
    - 1. Inner cycle optimize atomic positions
    - 2. Outer cycle optimize cell vectors
  - `DIRECT_CELL_OPT` (default): New implementation from version 2.4 onwards
    - Cell parameters (stresses) go into the optimizer along with atomic coordinates
  - `MD`: Optimize at finite temperature.
    - Uses MD, so only of use if you have a cheap Hamiltonian
- `DIRECT_CELL_OPT` should be much more efficient – try for yourself
- Generally best to enforce symmetry / fix angles to start with to minimize number of degrees of freedom.

Output

- Grep for "`Max. grad`" in output file to see the progress of the optimization
  - this gives maximum energy gradient on atoms being optimized
- Below "`Convergence check :`" there is a summary of the progress
  - convergence requires Max and RMS step size and Max and RMS gradients to be converged.
  - Pressure extra criteria for CELL_OPT
- The convergence criteria can be set in the `MOTION%[CELL|GEO]_OPT` section
- Default Max. grad is equal to 0.025 eV/A
- Good enough for most purposes
  - May need tighter e.g. for subsequent vibrational analysis

```
ENERGY| Total FORCE_EVAL ( QS ) energy (a.u.):            -2314.583506508490700

 *******************************************************************************
 ***                          2PNT LINE SEARCH INFO                          ***
 ***                                                                         ***
 ***   DX (EVALUATED)=    0.025697          DX (THRESHOLD)=    0.250000      ***
 ***   DX (FITTED   )=    0.252697          DX (ACCEPTED )=    0.250000      ***
 *******************************************************************************

 --------  Informations at step =     4 ------------
  Optimization Method        =                   SD
  Total Energy               =     -2314.5828781392
  Internal Pressure [bar]    =     35379.3762511153
  Real energy change         =        -0.0099708467
  Decrease in energy         =                  YES
  Used time                  =              465.309

  Convergence check :
  Max. step size             =         0.1443379699
  Conv. limit for step size  =         0.0030000000
  Convergence in step size   =                   NO
  RMS step size              =         0.0456435465
  Conv. limit for RMS step   =         0.0015000000
  Convergence in RMS step    =                   NO
  Max. gradient              =         0.0148361109
  Conv. limit for gradients  =         0.0004500000
  Conv. for gradients        =                   NO
  RMS gradient               =         0.0046915771
  Conv. limit for RMS grad.  =         0.0003000000
  Conv. for gradients        =                   NO
  Pressure Deviation [bar]   =     35279.3762511153
  Pressure Tolerance [bar]   =       100.0000000000
  Conv. for  PRESSURE        =                   NO
 ---------------------------------------------------

 CELL| Volume [angstrom^3]:                                           180.108
 CELL| Vector a [angstrom]:       5.647     0.000     0.000    |a| =     5.647
 CELL| Vector b [angstrom]:       0.000     5.647     0.000    |b| =     5.647
 CELL| Vector c [angstrom]:       0.000     0.000     5.647    |c| =     5.647
 CELL| Angle (b,c), alpha [degree]:                                    90.000
```

```
 --------  Informations at step =     5 ------------
  Optimization Method        =                   CG
  Total Energy               =     -2314.5861062993
  Internal Pressure [bar]    =        44.8922780589
  Real energy change         =        -0.0032281601
  Decrease in energy         =                  YES
  Used time                  =              484.509

  Convergence check :
  Max. step size             =         0.0000008113
  Conv. limit for step size  =         0.0030000000
  Convergence in step size   =                  YES
  RMS step size              =         0.0000004766
  Conv. limit for RMS step   =         0.0015000000
  Convergence in RMS step    =                  YES
  Max. gradient              =         0.0000003312
  Conv. limit for gradients  =         0.0004500000
  Conv. in gradients         =                  YES
  RMS gradient               =         0.0000001945
  Conv. limit for RMS grad.  =         0.0003000000
  Conv. in RMS gradients     =                  YES
  Pressure Deviation [bar]   =       -55.1077219411
  Pressure Tolerance [bar]   =       100.0000000000
  Conv. for  PRESSURE        =                  YES
 ---------------------------------------------------

 *******************************************************************************
 ***                    GEOMETRY OPTIMIZATION COMPLETED                      ***
 *******************************************************************************

                    Reevaluating energy at the minimum
```

Global optimisation

- Brute force approach:
  - Generate a grid of points (size m) in each of 3M dimensions
  - m^{3N} energy evaluations – exponential in system size X

- Practical methods exploit shape of PES
  - Genetic algorithms
  - Simulated annealing (MOTION%MD%ANNEALING)
  - Monte Carlo
  - Basin Hopping (GLOBAL%SWARM%GLOBAL_OPT%METHOD)

- Details of methods and implementation in Ole Shütt's Masters Thesis
  - Linked from https://www.cp2k.org/docs

Molecular Dynamics

- Classical Molecular Dynamics, particles obey Newton's 2nd Law and move subject to a position-dependent interaction potential:

$$m_i \ddot{r}_i = F_i \qquad F_i = -\frac{dU(\mathbf{R})}{dr_i}$$

- For a fixed number of particles N in a volume V these equations of motion generate the microcanonical (NVE) ensemble.
- The total energy U + the kinetic energy is **conserved**

Molecular Dynamics

- We solve the equations of motion by discretisation in time, given positions $\mathbf{R}$ and velocities $\mathbf{V}$ at time t_0

$$\mathbf{R}(t_0) \rightarrow \mathbf{R}(t_0 + \partial t) \rightarrow \mathbf{R}(t_0 + 2\partial t)\ldots$$

$$\mathbf{V}(t_0) \rightarrow \mathbf{V}(t_0 + \partial t) \rightarrow \mathbf{V}(t_0 + 2\partial t)\ldots$$

- Want a scheme which is:
  - Efficient: minimal number of force evaluations, stored data
  - Stable: minimal drift in conserved quantity
  - Accurate: minimal distance to exact trajectory

Velocity Verlet Integrator

$$r_i(t+\partial t) \rightarrow r_i(t) + \partial t \cdot v_i(t) + \frac{\partial t^2}{2m_i} f_i(t)$$

$$v_i(t+\partial t) \rightarrow v_i(t) + \frac{\partial t}{2m_i}[f_i(t) + f_i(t+\partial t)]$$

- Efficient: 1 force evaluation, 3 stored quantities
- Stable: time reversible
- Accurate: symplectic, integration error $O(\partial t^2)$

+ extensions for constraints (SHAKE, RATTLE, ROLL)
+ multiple timesteps (r-RESPA) and thermostats

Born-Oppenheimer MD

- Born-Oppenheimer Approximation:
  - Ionic mass >> electron mass so equations of motion for (classical) nuclei and (quantum) electrons are separable

$$m_i \ddot{r}_i = F_i$$

$$F_i = -\frac{dU(\mathbf{R})}{dr_i}$$

- Kohn-Sham BO potential: $U(\mathbf{R}) = \min_{\phi}[E_{KS}(\{\phi(\mathbf{r})\}, \mathbf{R})]$

$$F_{KS}(\mathbf{R}) = \frac{\partial E_{KS}}{\partial \mathbf{R}} + \sum_i \frac{\partial E_{KS}}{\partial \phi_i} \frac{\partial \phi_i}{\partial \mathbf{R}}$$

BO-MD in CP2K

- Benchmark system setup:
  - 64 water molecules
  - density $1gcm^{-3}$
  - Temperature ≈ $330K$
  - Timestep $0.5fs$
- DFT Settings:
- GPW, TZV2P basis (2560 basis functions), PBE functional
- `CUTOFF` 280 Rydberg, $\varepsilon_{default} = 10^{-12}$
- OT-DIIS, Preconditioner `FULL_SINGLE_INVERSE`
- Reference trajectory ($1ps$), $\varepsilon_{SCF} = 10^{-10}$

BO-MD in CP2K

Unbiased initial guess; $\Phi(t) = \Phi_0(\boldsymbol{R}(t))$

ϵ_{SCF}	MAE E_{KS} Hartree	MAE f Hartree/Bohr	Drift Kelvin/ns
10^{-08}	$1.2 \cdot 10^{-11}$	$5.1 \cdot 10^{-09}$	0.0
10^{-07}	$9.5 \cdot 10^{-10}$	$5.6 \cdot 10^{-08}$	0.1
10^{-06}	$6.9 \cdot 10^{-08}$	$4.8 \cdot 10^{-07}$	0.4
10^{-05}	$7.4 \cdot 10^{-06}$	$5.6 \cdot 10^{-06}$	2.3
10^{-04}	$3.3 \cdot 10^{-04}$	$5.9 \cdot 10^{-05}$	50

BO-MD in CP2K

Unbiased initial guess; $\Phi(t) = \Phi_0(\boldsymbol{R}(t))$

ϵ_{SCF}	MAE E_{KS} Hartree	MAE f Hartree/Bohr	Drift Kelvin/ns
10^{-08}	$1.2 \cdot 10^{-11}$	$5.1 \cdot 10^{-09}$	0.0
10^{-07}	$9.5 \cdot 10^{-10}$	$5.6 \cdot 10^{-08}$	0.1
10^{-06}	$6.9 \cdot 10^{-08}$	$4.8 \cdot 10^{-07}$	0.4
10^{-05}	$7.4 \cdot 10^{-06}$	$5.6 \cdot 10^{-06}$	2.3
10^{-04}	$3.3 \cdot 10^{-04}$	$5.9 \cdot 10^{-05}$	50

BO-MD in CP2K

```
DFT%QS%EXTRAPOLATION PS
DFT%QS%EXTRAPOLATION_ORDER 4
```

4th order Gear predictor (PS extrapolation in CP2K)

Method	ϵ_{SCF}	Iterations	Drift (Kelvin/ns)
Guess	10^{-06}	14.38	0.4
Gear(4)	10^{-07}	6.47	5.7
Gear(4)	10^{-06}	5.22	11.8
Gear(4)	10^{-05}	4.60	86.8

What is the problem?

Time reversibility has been broken!

BO-MD in CP2K

```
DFT%QS%EXTRAPOLATION ASPC
DFT%QS%EXTRAPOLATION_ORDER 3
```

Method	ϵ_{SCF}	Iterations	Drift (Kelvin/ns)
Guess	10^{-06}	14.38	0.4
ASPC(3)	10^{-06}	5.01	0.2
ASPC(3)	10^{-05}	3.02	4.5
Gear(4)	10^{-07}	6.47	5.7
Gear(4)	10^{-06}	5.22	11.8
Gear(4)	10^{-05}	4.60	86.8

Kolafa, JCC (2004)
VandeVondele *et al.*, CPC (2205)

BO-MD in CP2K

```
DFT%QS%EXTRAPOLATION ASPC
DFT%QS%EXTRAPOLATION_ORDER 4…
```

Method	ϵ_{SCF}	Iterations	Drift (Kelvin/ns)
ASPC(4)	10^{-04}	1.62	1742.4
ASPC(5)	10^{-04}	1.63	1094.0
ASPC(6)	10^{-04}	1.79	397.4
ASPC(7)	10^{-04}	1.97	445.8
ASPC(8)	10^{-04}	2.06	24.1

BO-MD in CP2K : Summary

- Defaults settings are ASPC(3)
- SCF tolerance for ‘acceptable’ drift is system-dependent but `EPS_SCF 1E-5` or `1E-6` is a good guess
- Use OT and appropriate preconditioner to speed up SCF

- Further reading:
  - “Car-Parrinello molecular dynamics”, Jürg Hutter, *WIREs Comput Mol Sci*, 2: 604-612, 2012
  - Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods, Dominik Marx & Jürg Hutter

Ensembles

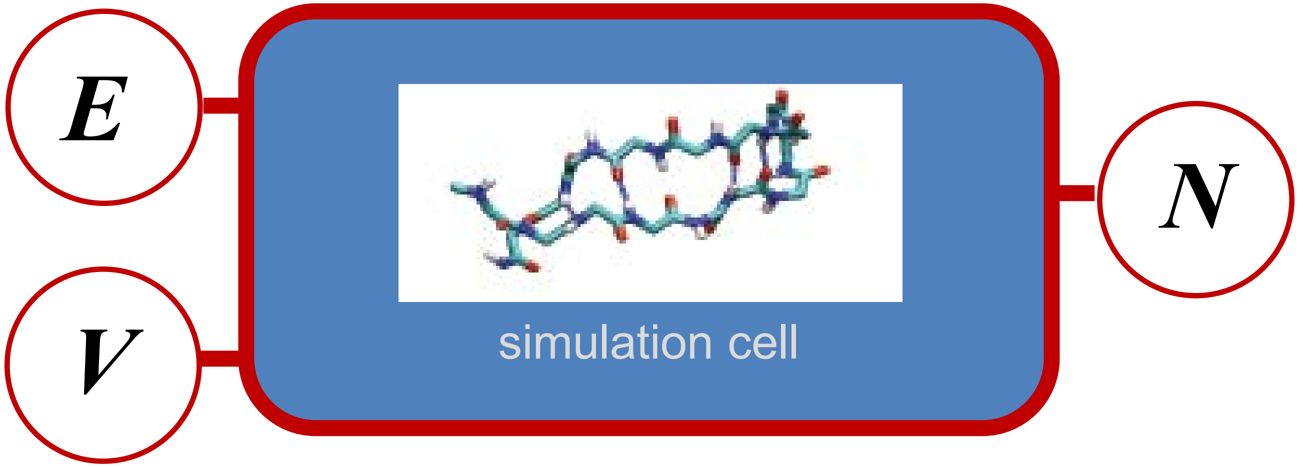
- Ensemble: set of all microstates $\{r_i, \dot{r}_i\}$ accessible to the simulation, each microstate occurring with a particular probability

- Various possibilities for quantities that may be conserved or fixed in the simulations:
  - Number of particles N
  - Volume V
  - Energy E
  - Temperature T
  - Pressure P
  - Chemical Potential μ (not implemented in CP2K)

NVE – microcanonical
NVT – canonical
NPT – isotherhmal-isobaric

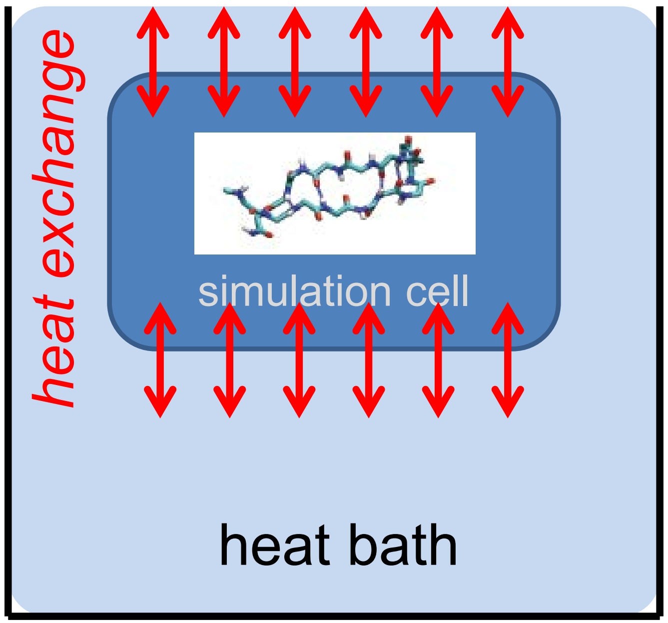
Ensembles

- Newton's second law applied to a set of N particles in a fixed box of volume V produces the microcanonical (NVE) ensemble
- Total Energy is conserved as the system is isolated



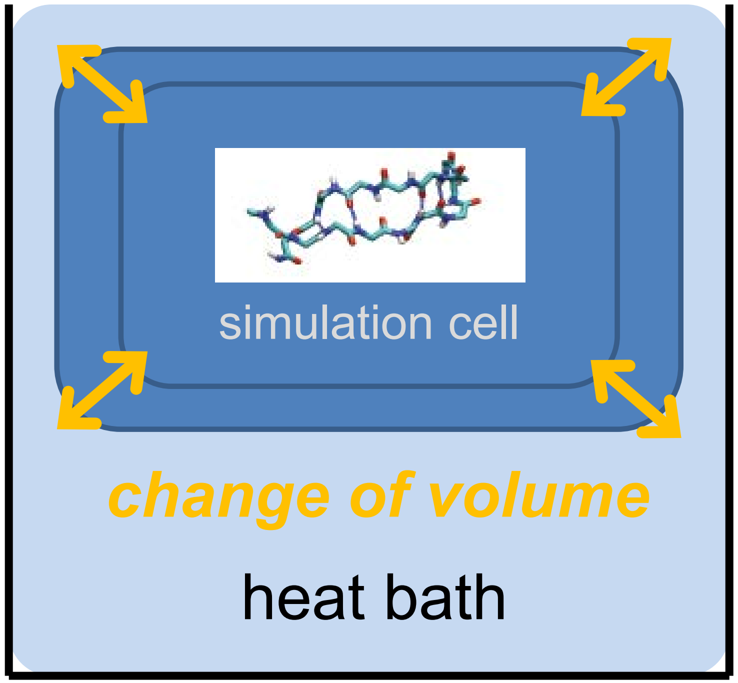
Ensembles

- If the system is in thermal contact with a heat bath at temperature T (canonical / NVT ensemble) the total energy of the system is no longer conserved
  - It may gain or lose energy from/to the heat bath
  - Instead the constant of the motion is the energy of the system + the energy of the bath



Ensembles

- If the box size/shape is allowed to change in response to internal stress and external pressure (isobaric-isothermal / NPT ensemble) then energy is exchanged with the environment via $dW = PdV$



- Cons. Quantity =
  - Energy of the system +
  - Energy of the 'thermostat'
  - Energy of the 'barostat'

Ensembles

```
&MOTION
  &MD
    ENSEMBLE NVE
    STEPS 1000
    TIMESTEP 0.5
    TEMPERATURE 300
  &END MD
&END MOTION
```

Possible choices

- microcanonical: `NVE`
- canonical: `NVT`
- canonical using Langevin dynamics: `LANGEVIN`
- isobaric-isothermal: `NPT_F` / `NPT_I`
- Constant pressure: `NPE_F` / `NPE_I`
- Also: `ISOKIN`, `HYDROSTATICSHOCK`, `MSST`, `MSST_DAMPED`, `NVT_ADIABATIC`

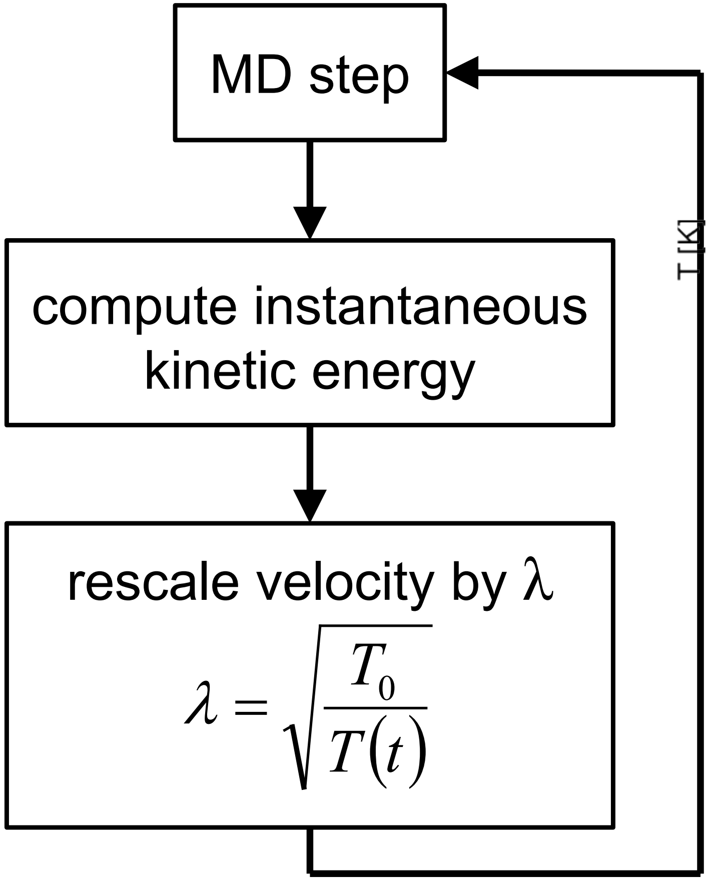
Thermostats in CP2K

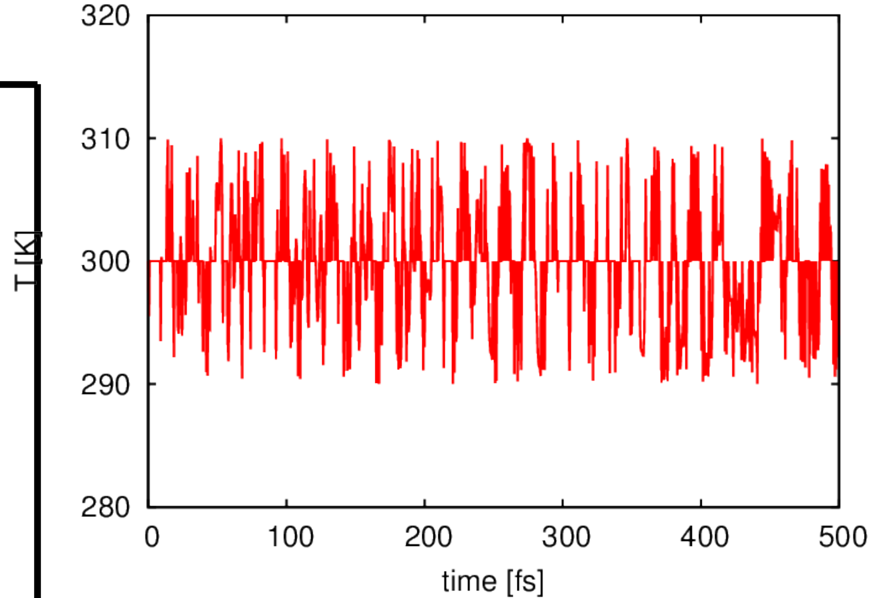
```
&MOTION
   &MD
      ENSEMBLE NVE
      STEPS 1000
      TIMESTEP 0.5
      TEMPERATURE 300
      TEMP_TOL 10
   &END MD
&END MOTION
```

- Rescales velocities when T < 290K or T > 310K
- Does not produce the canonical ensemble
- Use only for equilibriation

Thermostats in CP2K

- Velocity rescaling





T < 290 K or T > 310 K

Thermostats in CP2K

- Langevin Dynamics – adds a dissipative (frictional) force and a stochastic force

$$m_i \ddot{r}_i = -\frac{\partial U(r)}{\partial r_i} - m\Gamma \dot{r}_i + W_i(t)$$

- Magnitude of the perturbation depends on the instantaneous temperature

- Surprisingly useful in practice!

Thermostats in CP2K

- Produces canonical ensemble (NVT)
- Local thermostat
- Ergodic
- Stable at large timesteps

but

- does not conserve momentum (due to drag force)
- only useful for sampling, not dynamical properties (e.g. diffusion)

Thermostats in CP2K

- Nosé-Hoover (chains)

- Define an extended system with a (set of) thermal reservoirs with effective ‘position’ and ‘momenta’
  - So associated potential and kinetic energies

- Thermostat couples to the particle momenta through modified equations of motion

- Integrate these variables alongside the particle positions, momenta

Thermostats in CP2K

- Produces canonical ensemble (NVT)
- Local thermostat
- Ergodic (N-H chain only)
- Second order – temperature may oscillate towards target

Thermostats in CP2K

```
&MOTION
  &MD
    …
    &THERMOSTAT
      TYPE NOSE
      &NOSE
        LENGTH 3
        TIMECON 1000 [fs]
      &END NOSE
    &END THERMOSTAT
  &END MD
&END MOTION
```

- Defaults to 3 (1 recovers original Nosé thermostat
- 1000fs is the target relaxation time

Thermostats in CP2K

- Use a small `TIMECON` for rapid equilibibriation
- Default is usually OK for production MD
- Check the `PROJECT.ener` file that the constant of motion is indeed conserved
- Check for large fluctations in the temperature
- Almost all of the same options apply for barostats
  - `MOTION%MD%BAROSTAT`