http://tinyurl.com/CP2KSchool2018
#CP2KSummerSchool

CP2K:
MOVING ATOMS
CP2K Summer School, 19-22 June 2018

Iain Bethune

iain.bethune@stfc.ac.uk
@iainbethune

(material from Jürg Hutter, Matt Watkins, Konstanze Hahn)
Outline

• Geometry & cell optimisation
  • Local Minimisation
  • Global optimisation

• 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: $E_{\text{tot}}[n(r)]$ and $n(r) \leftrightarrow V(R)$ (Hohenberg-Kohn)
  • In molecular mechanics there is a forcefield:

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

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
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 $3N$ where $N$ is number of atoms being optimized

• L-BFGS

• Conjugate gradients
  • Only uses gradients rather than approximation to curvature, should be more robust when far from minima
Geometry optimisation

- **RUN_TYPE GEO_OPT** in **GLOBAL** section
- **GEO_OPT%OPTIMIZER** in **MOTION** section
  - CG, use with poor initial guesses, noisy forces, rough optimization
  - (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 be defined 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`)

```fortran
&CELL
    ABC 9.167 9.167 11.808
    SYMMETRY ORTHORHOMBIC
    MULTIPLE_UNIT_CELL 2 2 2
&END CELL

...```

```fortran
&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 \texttt{CELL\_OPT\_TYPE}
  - \texttt{GEO\_OPT}: Original implementation.
    - 1. Inner cycle optimize atomic positions
    - 2. Outer cycle optimize cell vectors
  - \texttt{DIRECT\_CELL\_OPT} (default): New implementation from version 2.4 onwards
    - Cell parameters (stresses) go into the optimizer along with atomic coordinates
  - \texttt{MD}: Optimize at finite temperature.
    - Uses MD, so only of use if you have a cheap Hamiltonian
  - \texttt{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/Å.
- Good enough for most purposes.
  - May need tighter e.g. for subsequent vibrational analysis.
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 (\texttt{MOTION\%MD\%ANNEALING})
  • Monte Carlo
  • Basin Hopping (\texttt{GLOBAL\%SWARM\%GLOBAL\_OPT\%METHOD})

• Details of methods and implementation in Ole Shütt’s Masters Thesis
  • Linked from \url{https://www.cp2k.org/docs}
Optimisation Exercises

• Geometry Optimisation of a water molecule
  • https://www.cp2k.org/howto:geometry_optimisation

• NaCl clusters (classical) and NaCl cell opt (DFT)
  • https://www.cp2k.org/exercises:2016_summer_school:geometry_and_cell_optimization
Molecular Dynamics

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

\[ m_i \ddot{r}_i = F_i \quad F_i = -\frac{dU(R)}{dr_i} \]

• 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 $t_0$

\[
\begin{align*}
\mathbf{R}(t_0) & \rightarrow \mathbf{R}(t_0 + \partial t) \rightarrow \mathbf{R}(t_0 + 2\partial t) \\
\mathbf{V}(t_0) & \rightarrow \mathbf{V}(t_0 + \partial t) \rightarrow \mathbf{V}(t_0 + 2\partial t)
\end{align*}
\]

- 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

\[
\begin{align*}
  r_i(t + \partial t) &\rightarrow r_i(t) + \partial t \cdot v_i(t) + \frac{\partial t^2}{2m_i} f_i(t) \\
  v_i(t + \partial t) &\rightarrow v_i(t) + \frac{\partial t}{2m_i} [f_i(t) + f_i(t + \partial t)]
\end{align*}
\]

- **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 Approximation:

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

\[ m_i \ddot{r}_i = F_i \]

**Kohn-Sham BO potential:**

\[ F_i = -\frac{dU(R)}{dr_i} \]

\[ U(R) = \min_{\phi} [E_{KS}(\{\phi(r)\}, R)] \]

\[ F_{KS}(R) = \frac{\partial E_{KS}}{\partial R} + \sum_i \frac{\partial E_{KS}}{\partial \phi_i} \frac{\partial \phi_i}{\partial R} \]
BO-MD in CP2K

- Benchmark system setup:
  - 64 water molecules
  - density $1 \text{gcm}^{-3}$
  - Temperature $\approx 330K$
  - Timestep 0.5fs

- DFT Settings:
  - GPW, TZV2P basis (2560 basis functions), PBE functional
  - $\text{CUTOFF 280 Rydberg, } \epsilon_{\text{default}} = 10^{-12}$
  - OT-DIIS, Preconditioner FULL_SINGLE_INVERSE
  - Reference trajectory (1ps), $\epsilon_{\text{SCF}} = 10^{-10}$
BO-MD in CP2K

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

<table>
<thead>
<tr>
<th>$\epsilon_{SCF}$</th>
<th>MAE $E_{KS}$ (Hartree)</th>
<th>MAE f (Hartree/Bohr)</th>
<th>Drift (Kelvin/ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-08}$</td>
<td>$1.2 \cdot 10^{-11}$</td>
<td>$5.1 \cdot 10^{-09}$</td>
<td>0.0</td>
</tr>
<tr>
<td>$10^{-07}$</td>
<td>$9.5 \cdot 10^{-10}$</td>
<td>$5.6 \cdot 10^{-08}$</td>
<td>0.1</td>
</tr>
<tr>
<td>$10^{-06}$</td>
<td>$6.9 \cdot 10^{-08}$</td>
<td>$4.8 \cdot 10^{-07}$</td>
<td>0.4</td>
</tr>
<tr>
<td>$10^{-05}$</td>
<td>$7.4 \cdot 10^{-06}$</td>
<td>$5.6 \cdot 10^{-06}$</td>
<td>2.3</td>
</tr>
<tr>
<td>$10^{-04}$</td>
<td>$3.3 \cdot 10^{-04}$</td>
<td>$5.9 \cdot 10^{-05}$</td>
<td>50</td>
</tr>
</tbody>
</table>
BO-MD in CP2K

4th order Gear predictor (PS extrapolation in CP2K)

<table>
<thead>
<tr>
<th>Method</th>
<th>$\epsilon_{\text{SCF}}$</th>
<th>Iterations</th>
<th>Drift (Kelvin/ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guess</td>
<td>$10^{-06}$</td>
<td>14.38</td>
<td>0.4</td>
</tr>
<tr>
<td>Gear(4)</td>
<td>$10^{-07}$</td>
<td>6.47</td>
<td>5.7</td>
</tr>
<tr>
<td>Gear(4)</td>
<td>$10^{-06}$</td>
<td>5.22</td>
<td>11.8</td>
</tr>
<tr>
<td>Gear(4)</td>
<td>$10^{-05}$</td>
<td>4.60</td>
<td>86.8</td>
</tr>
</tbody>
</table>

What is the problem?

Time reversibility has been broken!
### BO-MD in CP2K

**DFT%QS%EXTRAPOLATION ASPC**
**DFT%QS%EXTRAPOLATION_ORDER 3**

<table>
<thead>
<tr>
<th>Method</th>
<th>$\epsilon_{\text{SCF}}$</th>
<th>Iterations</th>
<th>Drift (Kelvin/ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guess</td>
<td>$10^{-06}$</td>
<td>14.38</td>
<td>0.4</td>
</tr>
<tr>
<td>ASPC(3)</td>
<td>$10^{-06}$</td>
<td>5.01</td>
<td>0.2</td>
</tr>
<tr>
<td>ASPC(3)</td>
<td>$10^{-05}$</td>
<td>3.02</td>
<td>4.5</td>
</tr>
<tr>
<td>Gear(4)</td>
<td>$10^{-07}$</td>
<td>6.47</td>
<td>5.7</td>
</tr>
<tr>
<td>Gear(4)</td>
<td>$10^{-06}$</td>
<td>5.22</td>
<td>11.8</td>
</tr>
<tr>
<td>Gear(4)</td>
<td>$10^{-05}$</td>
<td>4.60</td>
<td>86.8</td>
</tr>
</tbody>
</table>

Kolafa, JCC (2004)
VandeVondele *et al.*, CPC (2005)
## BO-MD in CP2K

DFT\%QS\%EXTRAPOLATION ASPC
DFT\%QS\%EXTRAPOLATION\_ORDER 4...

<table>
<thead>
<tr>
<th>Method</th>
<th>$\epsilon_{\text{SCF}}$</th>
<th>Iterations</th>
<th>Drift (Kelvin/ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASPC(4)</td>
<td>$10^{-04}$</td>
<td>1.62</td>
<td>1742.4</td>
</tr>
<tr>
<td>ASPC(5)</td>
<td>$10^{-04}$</td>
<td>1.63</td>
<td>1094.0</td>
</tr>
<tr>
<td>ASPC(6)</td>
<td>$10^{-04}$</td>
<td>1.79</td>
<td>397.4</td>
</tr>
<tr>
<td>ASPC(7)</td>
<td>$10^{-04}$</td>
<td>1.97</td>
<td>445.8</td>
</tr>
<tr>
<td>ASPC(8)</td>
<td>$10^{-04}$</td>
<td>2.06</td>
<td>24.1</td>
</tr>
</tbody>
</table>
BO-MD in CP2K: Summary

- Defaults settings are ASPC(3)
- SCF tolerance for ‘acceptable’ drift is system-dependent but $\text{EPS\_SCF} \approx 1E^{-5}$ or $1E^{-6}$ is a good guess
- Use OT and appropriate preconditioner to speed up SCF

Further reading:
- 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 $\mu$ (not implemented in CP2K)

  - NVE – microcanonical
  - NVT – canonical
  - NPT – isothermal–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

- Velocity rescaling

MD step

compute instantaneous kinetic energy

rescale velocity by \( \lambda \)

\[
\lambda = \sqrt{\frac{T_0}{T(t)}}
\]

\( T < 290 \text{ K or } T > 310 \text{ K} \)
Thermostats in CP2K

&MOTION
  &MD
  ENSsemble 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 equilibration
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

• Langevin Dynamics:
  • 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 equilibration

- Default is usually OK for production MD

- Check the `PROJECT.ener` file that the constant of motion is indeed conserved

- Check for large fluctuations in the temperature

- Almost all of the same options apply for barostats
  - `MOTION%MD%BAROSTAT`
MD Exercises

• Acetic acid binding to anatase surface
  • https://www.cp2k.org/exercises:2016_summer_school:gga

• Bulk liquid water
  • https://www.cp2k.org/exercises:2016_summer_school:aimd