CASTEP Workshop 2006

Finding the Kohn-Sham Groundstate

University of York

16th-21st September 2007

Overview

- The DFT Schrödinger equation
- Minimising the KS energy
- Metals and finite temperatures
- Density mixing and Ensemble DFT

The DFT Hamiltonian

- The object is to find the single-particle solutions to the Schrödinger equation

$$\hat{H}_{ks}\Psi_i=\epsilon_i\Psi_i$$

where *i* labels the different eigenstates and the Hamiltonian is given by

$$\hat{H}_{ks}=-\frac{1}{2}\nabla^2+\hat{V}_H+\hat{V}_{xc}+\hat{V}_{ps}$$

- We are using a plane-wave basis and sampling the Brillouin zone at a finite number of k-points,

$$\Psi=\sum_{\boldsymbol{G}} c_{\boldsymbol{k},\boldsymbol{G}}\, e^{i(\boldsymbol{k}+\boldsymbol{G}).\boldsymbol{r}}$$

where the sum is over all reciprocal lattice vectors below some cut-off wavevector.

- The wavefunctions at different **k**-points only interact via the density, so each **k**-point yields a separate Kohn-Sham equation.

- For simplicity we will consider only a calculation at a single **k**-point, and omit the subscript **k**.
- Each eigenstate is normalised to contain only one particle.

$$\left\langle \Psi_i \middle| \Psi_i \right\rangle = 1$$

- Each eigenstate is orthogonal to all other eigenstates (Pauli exclusion principle).

$$\left\langle \Psi_i \middle| \Psi_j \right\rangle = 0$$

Self-Consistency

- ❒ The Hartree potential represents the Coulomb repulsion between the different particles and is given by:

$$V_H(r) = \int \frac{n(r')}{|r - r'|} d^3 r'$$

where *n* is the charge density

$$n(r) = \sum_i f_i |\Psi_i(r)|^2$$

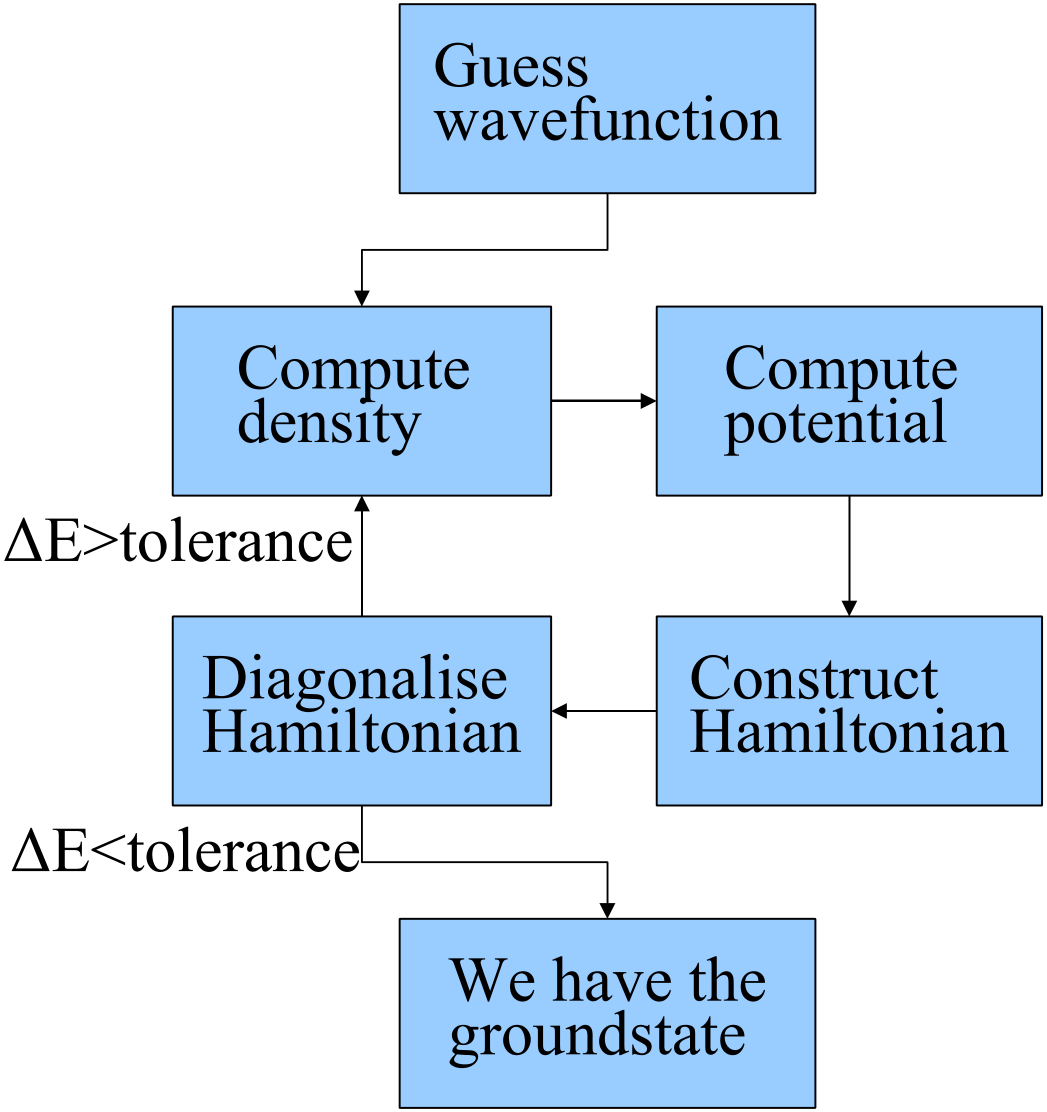
- Notice that the Hartree potential depends on the density, and the density depends on the wavefunctions, we no longer have a fixed Hamiltonian to solve.
- The exchange-correlation potential also depends on the density – the Hamiltonian changes as the wavefunctions change!
- We need to find a set of wavefunctions that lead to a Hamiltonian whose solutions are the wavefunctions we started with.
- This is called **self-consistency**.

Putting it on a computer

- The wavefunction is represented as a vector formed from the complex coefficients c.
- Operators such as the Hamiltonian are linear transformations of wavefunctions, and are represented by square Hermitian matrices.
- In principle we can now solve the Kohn-Sham equations by constructing and then diagonalising this Hamiltonian matrix to get all of the eigenstates and eigenvalues.

First Attempt - Diagonalisation

- Start with an initial set of wavefunctions, and construct the density.
- Construct the local potential from the density, and hence the Hamiltonian.
- Diagonalise to get eigenstates.
- Construct new density from eigenstates, and thus a new Hamiltonian.
- Repeat until the new Hamiltonian is the same as the old one – self-consistent field (SCF).



Guess wavefunction

Compute density

Compute potential

Construct Hamiltonian

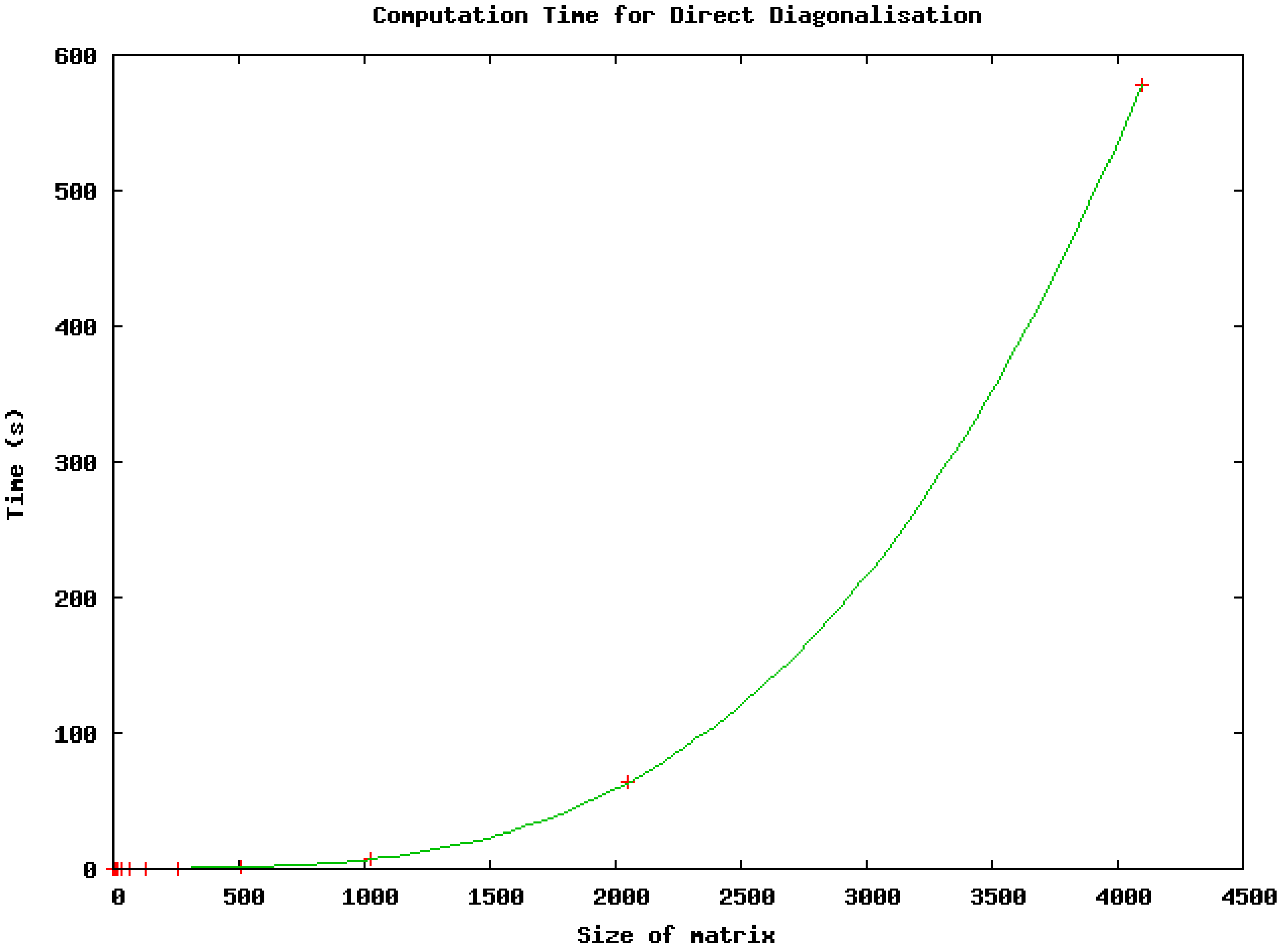
Diagonalise Hamiltonian

ΔE>tolerance

ΔE<tolerance

We have the groundstate

- Unfortunately the number of plane-waves required is large, often 10,000-100,000.
- Suppose we require 30,000 plane-waves for a calculation. The Hamiltonian is a matrix of size 30,000 x 30,000 = 9 x 10^8 elements.
- The computational time required to diagonalise a matrix scales as the cube of its size...


Computation Time for Direct Diagonalisation
Time (s)
600
500
400
300
200
100
0
0
500
1000
1500
2000
2500
3000
3500
4000
4500
Size of matrix

- Estimated time to diagonalise our Hamiltonian on a 2.8GHz Pentium IV is 74 hours (per **k**-point, per SCF cycle).
- This procedure gives us all 30,000 eigenstates, but typically fewer than 1% of these are occupied; the vast majority do not contribute to the density and have little physical meaning.

Minimising the KS energy

- An alternative method is to improve the wavefunctions iteratively. The energy of a wavefunction can be calculated as:

$$E = \sum_i \langle \Psi_i | \hat{H}_{ks} | \Psi_i \rangle$$

- We can differentiate this to find the gradient of the energy w.r.t. the wavefunction.

$$\frac{\delta E}{\delta \langle \Psi_i |} = \hat{H}_{ks} | \Psi_i \rangle$$

Steepest descent

- ❐ Using Lagrange multipliers we can include the constraint of normalisation and show that the constrained gradient is

$$\hat{G}\left|\Psi_i\right\rangle = \hat{H}\left|\Psi_i\right\rangle - \epsilon_i\left|\Psi_i\right\rangle$$

- ❐ If we include the orthogonalisation constraint as well we get:

$$\hat{G}\left|\Psi_i\right\rangle = \hat{H}\left|\Psi_i\right\rangle - \sum_j \left\langle\Psi_j\right|\hat{H}\left|\Psi_i\right\rangle\left|\Psi_j\right\rangle$$

- ❒ We can now take any wavefunction and calculate the energy gradient direction.
- ❒ This is the direction in which the energy *increases most rapidly*. Since we are looking for the lowest energy we search in the opposite direction, the *direction of steepest descent*.
- ❒ The *constrained gradient* is used rather than the ordinary gradient to prevent all the eigenstates converging to the lowest one—only one particle is allowed in each eigenstate.

The Hamiltonian

- In principle the Hamiltonian is still of size number of plane-waves squared, so difficult to store in RAM and expensive to apply.
- Fortunately the kinetic energy operator is diagonal in reciprocal space, and the local (Hartree+exchange-correlation) potential is diagonal in real-space, so we only actually need to store these diagonal components.

- Now start with initial wavefunction and construct the steepest descent direction.

$$|\Phi\rangle = -\hat{G}|\Psi\rangle$$

- Search along this direction for lowest energy.

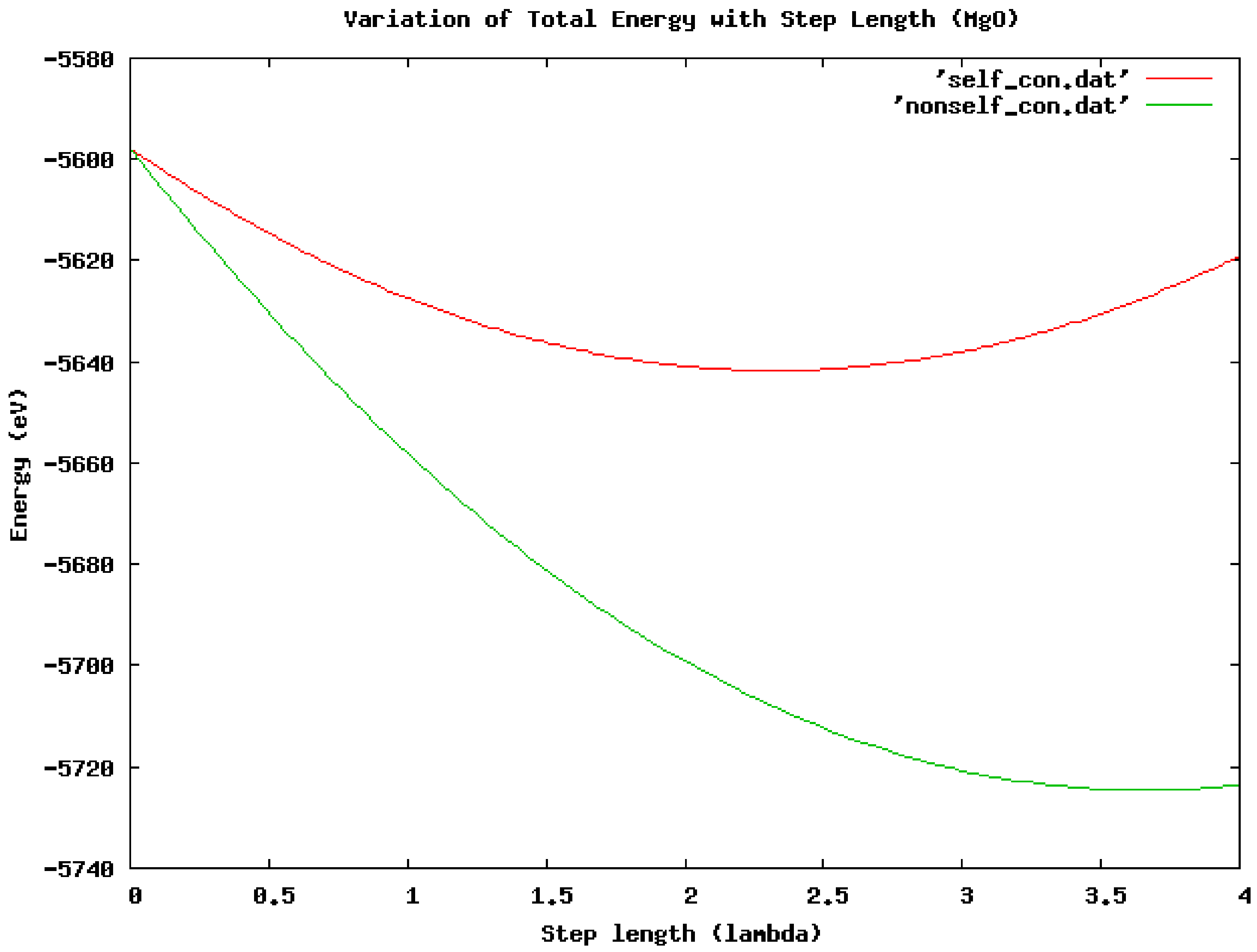
$$|\Psi'\rangle = |\Psi\rangle + \lambda|\Phi\rangle$$

- Calculate new search direction and repeat. After the first step better schemes may be used, such as conjugate gradients.

- Iterative diagonalisation is a bit slower per eigenstate than direct diagonalisation, but allows us to calculate only the ~1% of eigenstates that actually have particles in them.
- For N electrons we only need N eigenstates, so we iteratively diagonalise the Hamiltonian to obtain the N lowest eigenstates.
- E.g. 8-atom magnesium oxide cell at 330eV has 1021 plane-waves and 56 electrons. Direct diagonalisation time is 7.93s per **k**-point per SCF, iterative diagonalisation is 0.48s.

Self-consistency revisited

- ❒ We're now updating the wavefunctions iteratively. At what point do we recalculate the density?
- ❒ At each point along λ? This is called a self-consistent method. In general these are slow, but stable.
- ❒ After a certain number of wavefunction updates? This is called a non-self-consistent method. In general these are fast, but unstable.


Variation of Total Energy with Step Length (MgO)
'self_con.dat'
'nonself_con.dat'
Energy (eV)
-5580
-5600
-5620
-5640
-5660
-5680
-5700
-5720
-5740
0
0.5
1
1.5
2
2.5
3
3.5
4
Step length (lambda)

- The non-self-consistent minimum is usually found at larger step sizes than the self-consistent one, so non-self-consistent methods often overshoot the minimum.
- Consistent overshooting causes a phenomenon known as charge sloshing (see later).
- Charge sloshing is worse for large cells — consider the Hartree potential again but this time in reciprocal space:

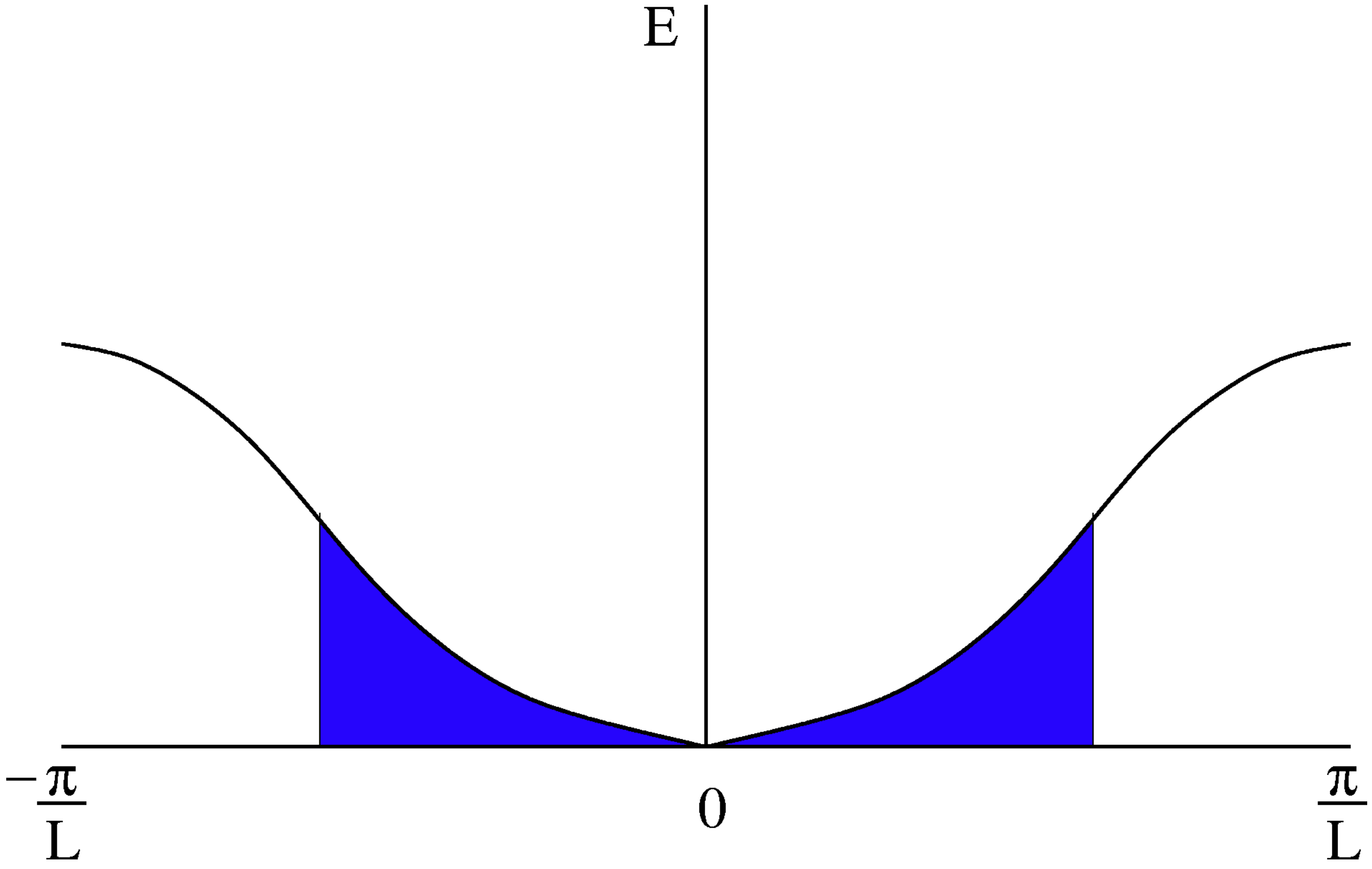
$$V_H(\boldsymbol{G})=\frac{4\pi n(\boldsymbol{G})}{\Omega_{cell}|\boldsymbol{G}|^2}$$

Large cell means small G-vectors

Metallic systems

- ❒ When we construct the density, we do so from the lowest N eigenstates.
- ❒ As the band-gap decreases, it becomes increasingly difficult to identify the lowest eigenstates (i.e. to find the Fermi surface).
- ❒ Bands below the Fermi energy are completely occupied and contribute fully to the density; bands above are empty and do not contribute at all.

k-points and metals



Sloshing

- Suppose band A is below the Fermi surface and B above. It is possible that by occupying band A we change the density such that bands A and B swap in energy.
- We now occupy band B instead and empty band A, whereupon the density changes such that the original ordering is restored.
- We will never converge to the groundstate! This is known as a sloshing instability.

The Mermin functional

- Sloshing arises from the discontinuous occupation of the eigenstates across the Fermi surface at zero temperature.
- Mermin smeared the Fermi surface, exactly like heating the system up.
- Eigenstates near the Fermi surface are now neither full nor empty, but partially filled.
- There is an entropy contribution to the energy from the partial occupancies.

Occupancies

- We assign the occupancies based on the eigenvalues and the Fermi energy.
- The density depends on the occupancies as well as the wavefunctions, so changing the occupancies changes the density, and hence the Hamiltonian.
- Changing the Hamiltonian changes the eigenvalues of the states, so yet again we need iterate to self-consistency.

Ensemble DFT (EDFT)

- In EDFT the density is reconstructed whenever the wavefunctions or occupancies change, and both the wavefunctions and occupancies are only changed when it would lower the total energy.
- Each SCF cycle consists of a number of wavefunction updates (usually five), and after each wavefunction update the occupancies are also updated.

Ensemble DFT (EDFT)

- Constructing the density requires a Fourier transform of the wavefunctions into real-space, and this can be slow (especially in parallel).
- EDFT is:
  - Variational (energy decreases at each step)
  - Stable
  - Slow

EDFT parameters

- ❒ EDFT is controlled by the .param file
  - elec_method: EDFT
- ❒ Several occupancy cycles are performed every step to find the optimum partial occupancies.
  - num_occ_cycles: 6

  for some systems you can reduce the number of occupancy cycles to speed up the calculation. 3 or 4 is often fine, even 2 is OK for simple systems.

Density mixing (DM)

- In DM the density is only reconstructed once per SCF cycle. To prevent sloshing instabilities this new density is mixed with a fraction of the previous densities.
- A typical DM SCF cycle will update each trial eigenstate two or three times, reassign occupancies, and only then construct the density.

Density mixing (DM)

- DM is:
  - Non-variational
  - Potentially unstable
  - Quick
- The key to the stability of DM is the manner of the mixing. The best schemes in Castep are the Pulay and Broyden schemes.

DM parameters

- ❐ DM is controlled by the .param file
  - elec_method: DM
- ❐ The number of past densities to store
  - mix_history_length: 7
- ❐ The DM scheme
  - mixing_scheme: Pulay or Broyden
- ❐ The initial mixing amplitudes
  - mix_charge_amp: 0.7 (0.5 in Materials Studio)
  - mix_spin_amp: 2.0

  decrease for a slight increase in stability.

DM parameters

❒ The screening wavevectors—the mixing schemes work by building up an approximation to the system's dielectric. The initial approximation is of the form

$$K(q) = \frac{q^2}{q^2 + q_c^2}$$

- mix_charge_gmax: 1.5 1/ang
- mix_spin_gmax: 1.5 1/ang

Large frequencies

- Recall that the sloshing instabilities were due to the low frequency (small q) components. What about the high-frequency components?
- Usually we accept the high frequency part of the new density without mixing it at all
- At what point is a frequency “high enough” to not mix? It’s controlled by the parameter
  - mix_cut_off_energy
- This defaults to the wavefunction cut-off (cut_off_energy).

General parameters

- ❒ The number of wavefunction updates per SCF
  - max_sd_steps: 1
  - max_cg_steps: 4

  DM may not perform all of these steps.
- ❒ The energy to smear the Fermi level by
  - smearing_width: 0.2 eV
- ❒ The number of extra bands to include
  - nextra_bands

  Including extra bands slows the SCF cycles, but may improve convergence (i.e. take fewer SCFs).

EDFT output (iprint: 1)

```
------------------------------------------------------------------------ <-- SCF
SCF loop      Energy           Fermi           Energy gain        Timer  <-- SCF
                               energy          per atom           (sec)  <-- SCF
------------------------------------------------------------------------ <-- SCF
Initial  -5.60447982E+003  6.04960700E+000                          5.57 <-- SCF
      1  -5.65487212E+003  8.54480709E+000   6.29903671E+000       34.10 <-- SCF
      2  -5.65488381E+003  8.55197524E+000   1.46193897E-003       62.10 <-- SCF
      3  -5.65488514E+003  8.54531708E+000   1.65970615E-004       90.12 <-- SCF
      4  -5.65488559E+003  8.54559484E+000   5.64376170E-005      120.43 <-- SCF
      5  -5.65488567E+003  8.54576989E+000   9.21133649E-006      149.52 <-- SCF
      6  -5.65488568E+003  8.54568880E+000   1.14975580E-006      179.38 <-- SCF
      7  -5.65488568E+003  8.54574147E+000   1.12351628E-007      210.01 <-- SCF
      8  -5.65488568E+003  8.54573805E+000   3.75314590E-008      240.19 <-- SCF
      9  -5.65488568E+003  8.54573963E+000   1.34522225E-008      270.22 <-- SCF
     10  -5.65488568E+003  8.54573949E+000   4.91008537E-009      300.98 <-- SCF
     11  -5.65488568E+003  8.54573975E+000   1.66028364E-009      330.80 <-- SCF
     12  -5.65488568E+003  8.54573998E+000   7.05528707E-010      358.88 <-- SCF
     13  -5.65488568E+003  8.54573998E+000   1.69276619E-010      387.17 <-- SCF
     14  -5.65488568E+003  8.54574005E+000   5.12373547E-011      416.85 <-- SCF
------------------------------------------------------------------------ <-- SCF
```

EDFT output (iprint: 2)

```
+------------------- WAVEFUNCTION LINE MINIMISATION --------------------+<- line
| Initial energy = -5.604480E+003 eV; initial dE/dstep = -3.982E+001 eV |<- line
|                                                                       |<- line
|            1st step    2nd step    3rd step    4th step    5th step   |<- line
+--------+--------------------------------------------------------------+<- line
| step   |  1.500E+00   2.352E+00* unnecessary unnecessary unnecessary  |<- line
| gain   |  4.069E+01   4.658E+01* unnecessary unnecessary unnecessary  |<- line
+--------+--------------------------------------------------------------+<- line
 * indicates the final, accepted state (should have the lowest energy)
```

```
+--------------------- OCCUPANCY LINE MINIMISATION ---------------------+<- EDFT
| Initial energy for occupancy cycle 1 = -5.651E+003 eV                 |<- EDFT
|                                                                       |<- EDFT
|         1st step     2nd step     3rd step     4th step    5th step   |<- EDFT
+------+----------------------------------------------------------------+<- EDFT
| step |  1.000E+00    6.200E-01    8.783E-01    8.656E-01* unnecessary |<- EDFT
| gain |  1.940E+00    1.807E+00    2.007E+00    2.007E+00* unnecessary |<- EDFT
+------+----------------------------------------------------------------+<- EDFT
 * indicates the final, accepted state (should have the lowest energy)
Final energy = -5.653E+003eV (change of  2.007E+000eV) for occ. cycle  1 <- EDFT
```

DM output (iprint: 1)

```
------------------------------------------------------------------------ <-- SCF
SCF loop      Energy           Fermi           Energy gain       Timer   <-- SCF
                               energy          per atom          (sec)   <-- SCF
------------------------------------------------------------------------ <-- SCF
Initial  -5.72401759E+003  7.33403352E+000                         4.93  <-- SCF
      1  -5.66310821E+003  7.15206775E+000  -7.61367268E+000       5.44  <-- SCF
      2  -5.66315105E+003  7.15066794E+000   5.35423262E-003       6.16  <-- SCF
      3  -5.65904032E+003  9.04183988E+000  -5.13841407E-001       7.43  <-- SCF
      4  -5.65484948E+003  8.35292985E+000  -5.23854073E-001       8.71  <-- SCF
      5  -5.65487927E+003  8.47113111E+000   3.72385144E-003      10.21  <-- SCF
      6  -5.65488530E+003  8.53910153E+000   7.53186999E-004      11.66  <-- SCF
      7  -5.65488567E+003  8.54432111E+000   4.59464638E-005      13.30  <-- SCF
      8  -5.65488568E+003  8.54547827E+000   1.23017408E-006      14.61  <-- SCF
      9  -5.65488568E+003  8.54569517E+000   4.93767620E-008      16.05  <-- SCF
     10  -5.65488568E+003  8.54573154E+000  -1.00753908E-009      17.34  <-- SCF
     11  -5.65488568E+003  8.54574500E+000  -6.34666478E-010      18.53  <-- SCF
     12  -5.65488568E+003  8.54574128E+000   2.38785408E-011      19.61  <-- SCF
     13  -5.65488568E+003  8.54574299E+000   8.01429567E-011      20.70  <-- SCF
------------------------------------------------------------------------ <-- SCF
```

DM output (iprint: 2)

```
+--------+-------+------------------------------+-----------------------+
|  Band  | Steps |          Eigenvalue          |     Change      stop  |<- line
|        |       |    Initial         Final     | (energy drop)   cond  |<- line
+--------+-------+------------------------------+-----------------------+<- line
|     1  |    2  | -3.309434E+01  -3.322212E+01 |   1.2778E-01     --X  |<- line
|     2  |    2  | -3.303377E+01  -3.322107E+01 |   1.8730E-01     --X  |<- line
|     3  |    2  | -3.299567E+01  -3.321013E+01 |   2.1446E-01     --X  |<- line
|     4  |    2  | -3.292534E+01  -3.314784E+01 |   2.2249E-01     --X  |<- line
|     5  |    2  | -3.281925E+01  -3.314753E+01 |   3.2828E-01     --X  |<- line
|     6  |    2  | -3.273043E+01  -3.312386E+01 |   3.9343E-01     --X  |<- line
|     7  |    2  | -3.267706E+01  -3.308311E+01 |   4.0606E-01     --X  |<- line
|     8  |    2  | -3.259183E+01  -3.308239E+01 |   4.9056E-01     --X  |<- line
|     9  |    2  | -3.256198E+01  -3.306716E+01 |   5.0518E-01     --X  |<- line
|    10  |    2  | -3.253910E+01  -3.306556E+01 |   5.2646E-01     --X  |<- line
|    11  |    2  | -3.252031E+01  -3.306529E+01 |   5.4498E-01     --X  |<- line
+--------+-------+------------------------------+------------------+++--+<- line
                                                                   |||   <- line
                      max no. steps performed <--------------------+||   <- line
                      elec_eigenvalue_tol reached <-----------------+|   <- line
                      gradient flattened out <-----------------------+   <- line
```

EDFT vs DM (Part I)

- Test system: 8-atom MgO (340eV)
- EDFT converges in 14 SCF cycles (416.9s)
  - 80 wavefunction updates
  - 328 density calculations
- DM converges in 13 SCF cycles (20.7s)
  - 50 wavefunction updates
  - 13 density calculations
- DM is the clear winner.

EDFT vs DM (Part I)

- Look at the initial energies; the Fermi energy in DM is much better than the EDFT one. Why?
- In DM the density is not obtained directly from the wavefunctions, so we can choose our initial density.
- We choose to start with the sum of the atomic densities. This makes the initial Hamiltonian pretty good compared to the random start EDFT has.

EDFT output (iprint: 1)

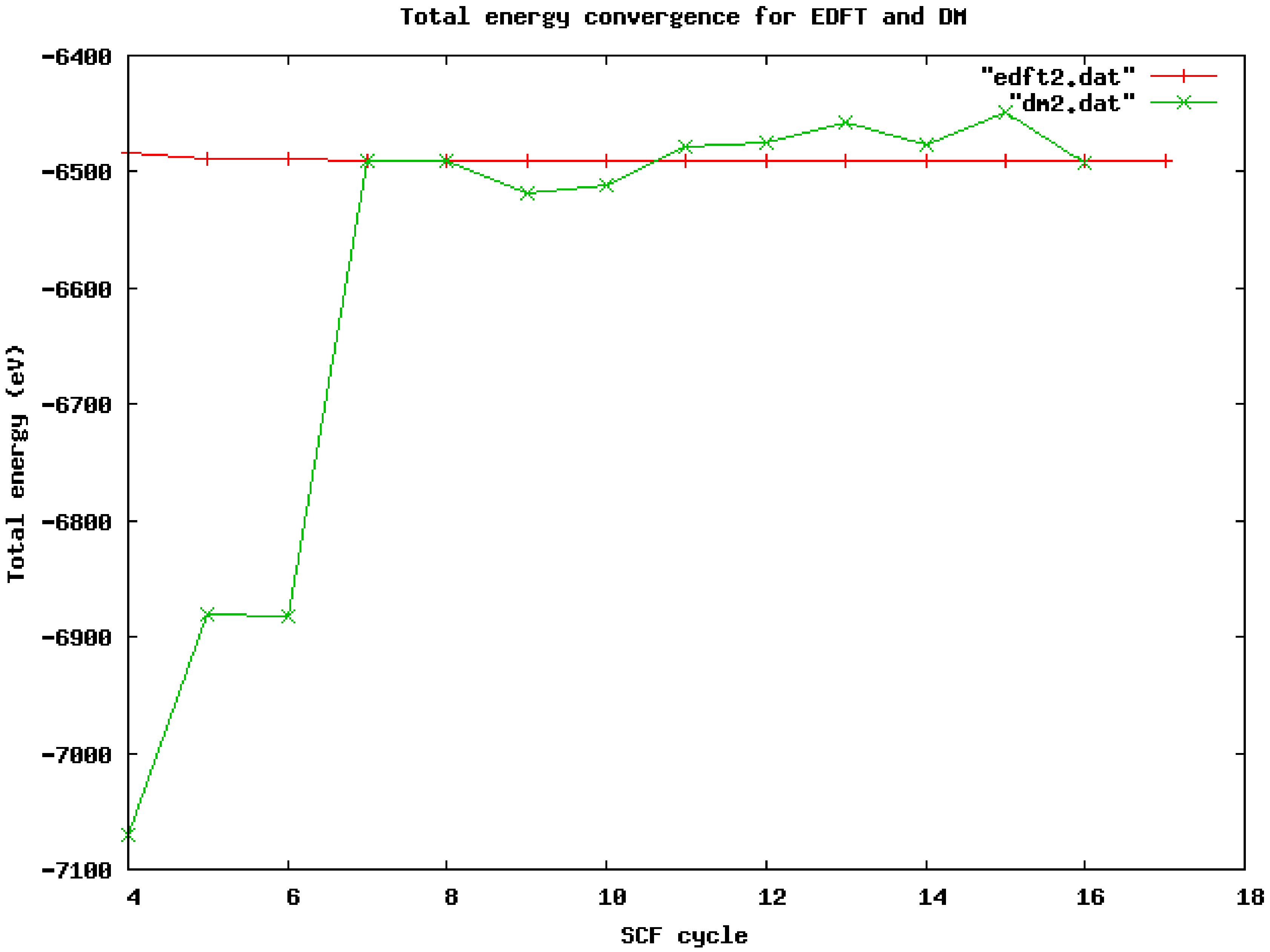
```
------------------------------------------------------------------------ <-- SCF
SCF loop      Energy           Fermi           Energy gain       Timer   <-- SCF
                               energy          per atom           (sec)  <-- SCF
------------------------------------------------------------------------ <-- SCF
Initial   5.22894717E+003  1.23054649E+002                        23.25  <-- SCF
      1  -1.34061398E+003  1.54028093E+002   4.69254368E+002     406.21  <-- SCF
      2  -6.24970545E+003  3.25481388E+001   3.50649391E+002     773.44  <-- SCF
      3  -6.44753749E+003  1.58875589E+001   1.41308601E+001    1122.97  <-- SCF
      4  -6.48366756E+003  8.21631220E+000   2.58071892E+000    1478.40  <-- SCF
      5  -6.48822994E+003  6.28831601E+000   3.25884507E-001    1829.99  <-- SCF
      6  -6.48935120E+003  5.38297421E+000   8.00902664E-002    2175.23  <-- SCF
      7  -6.48970423E+003  5.09069781E+000   2.52159741E-002    2529.80  <-- SCF
      8  -6.49027213E+003  4.60235674E+000   4.05645475E-002    2892.63  <-- SCF
      9  -6.49049787E+003  4.46602537E+000   1.61242962E-002    3252.49  <-- SCF
     10  -6.49069177E+003  4.32576189E+000   1.38501648E-002    3609.24  <-- SCF
     11  -6.49082746E+003  4.10031919E+000   9.69195382E-003    3981.20  <-- SCF
```

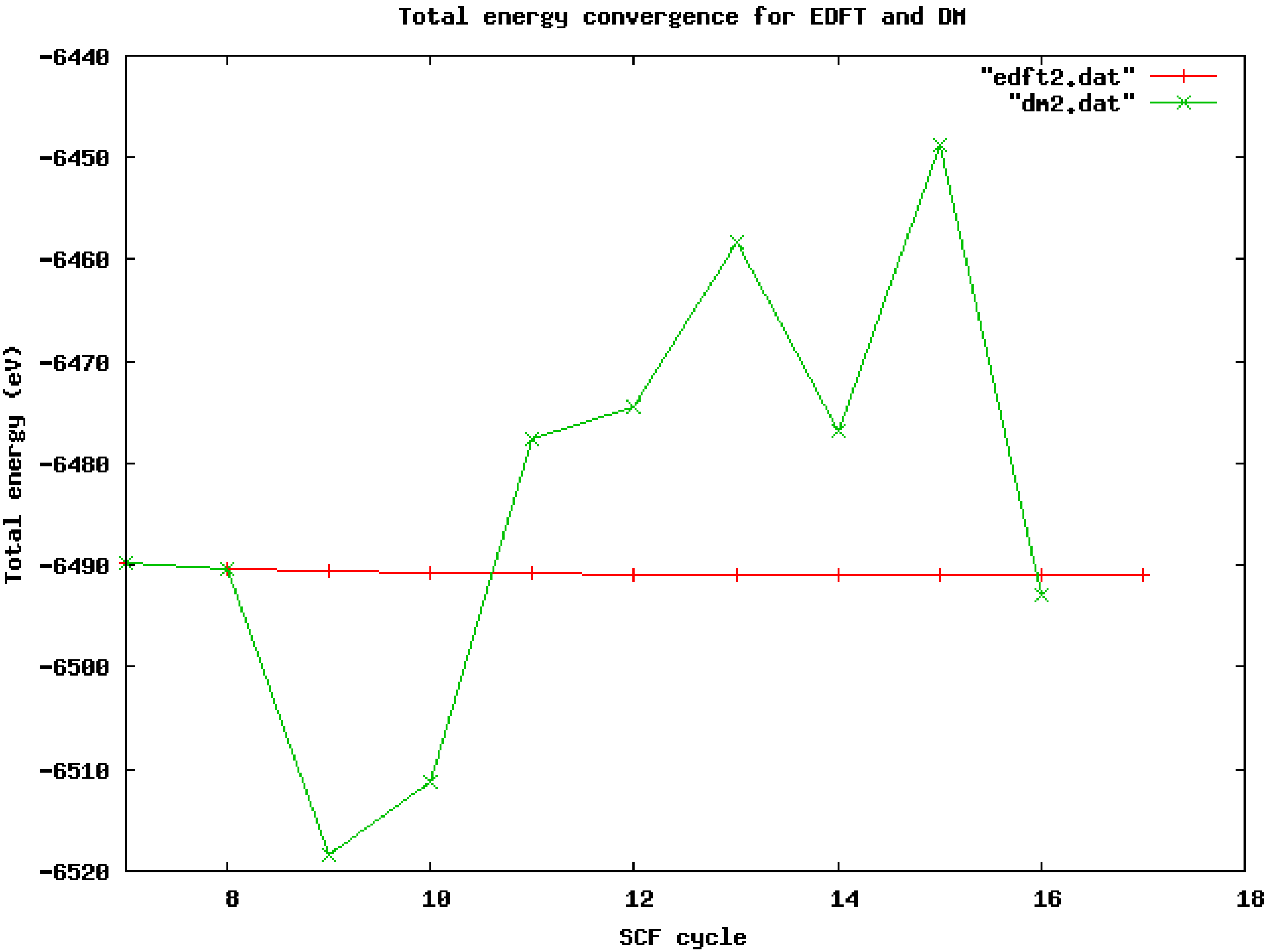
DM output (iprint: 1)

```
------------------------------------------------------------------------ <-- SCF
SCF loop      Energy           Fermi           Energy gain       Timer   <-- SCF
                               energy          per atom          (sec)   <-- SCF
------------------------------------------------------------------------ <-- SCF
Initial  -5.14317516E+003  5.12276157E+001                        12.03  <-- SCF
      1  -6.31741967E+003  6.63244417E+000   8.38746075E+001      20.60  <-- SCF
      2  -6.49931311E+003  3.75941681E+000   1.29923889E+001      29.84  <-- SCF
      3  -6.51038834E+003  3.54231581E+000   7.91087819E-001      39.34  <-- SCF
      4  -7.06998837E+003  2.22343892E+000   3.99714308E+001      55.71  <-- SCF
      5  -6.88101641E+003  3.72113246E+000  -1.34979974E+001      64.26  <-- SCF
      6  -6.88234462E+003  3.67678710E+000   9.48725740E-002      73.81  <-- SCF
      7  -6.48977868E+003  3.73752136E+000  -2.80404247E+001      84.70  <-- SCF
      8  -6.49044610E+003  3.74216565E+000   4.76729934E-002      93.59  <-- SCF
      9  -6.51835347E+003  4.26880418E+000   1.99338358E+000     107.69  <-- SCF
     10  -6.51118474E+003  4.28640804E+000  -5.12052231E-001     120.24  <-- SCF
     11  -6.47752722E+003  3.96827211E+000  -2.40410830E+000     133.02  <-- SCF
     12  -6.47436227E+003  4.17156369E+000  -2.26067919E-001     144.39  <-- SCF
     13  -6.45824918E+003  4.14265903E+000  -1.15093520E+000     155.99  <-- SCF
     14  -6.47675286E+003  4.55171279E+000   1.32169139E+000     167.02  <-- SCF
     15  -6.44884485E+003  4.41836929E+000  -1.99342920E+000     178.41  <-- SCF
     16  -6.49295003E+003  4.31979139E+000   3.15036998E+000     191.27  <-- SCF
     17  -6.49378104E+003  4.31843062E+000   5.93581671E-002     201.07  <-- SCF
     18  -6.44552085E+003  4.71784691E+000  -3.44715632E+000     213.32  <-- SCF
     19  -6.44589421E+003  4.73581949E+000   2.66682897E-002     222.73  <-- SCF
     20  -6.41032124E+003  5.23541858E+000  -2.54092638E+000     236.73  <-- SCF
```

EDFT vs DM (Part II)

- Test system: erbium and hydrogen in silicon (340eV)
- EDFT converges in 14 SCF cycles (4000s)
- DM doesn’t converge!
- EDFT is the clear winner.


Total energy convergence for EDFT and DM
"edft2.dat"
"dm2.dat"
Total energy (eV)
-6400
-6500
-6600
-6700
-6800
-6900
-7000
-7100
SCF cycle
4
6
8
10
12
14
16
18


Total energy convergence for EDFT and DM
"edft2.dat"
"dm2.dat"
Total energy (eV)
-6440
-6450
-6460
-6470
-6480
-6490
-6500
-6510
-6520
8
10
12
14
16
18
SCF cycle

DM's Problems

- What’s going wrong with DM?
- The larger the system, the smaller the smallest nonzero wavevector is, so sloshing is worse.
- Sloshing is made worse if the material has a large dielectric susceptibility, so metals can be especially problematic.
- Spin can also make matters worse, because there may be many metastable spin states, and the system can now spin slosh as well as charge slosh.

Castep scaling

- ❒ Constructing the density and applying the local potential require FFTs of each band from reciprocal to real-space. FFTs scale as NlogN.
  - In **G-vector parallel** calculations the prefactor increases with the number of processors.
- ❒ Orthonormalisation scales as N^3.
  - In parallel calculations the prefactor decreases with the number of processors.
- ❒ Parallel calculations scale almost perfectly when pure **k-point parallelisation** is used.

Summary

- EDFT is stable and fully variational, giving excellent forces and stresses; but it is slow.
- DM is fast, but non-variational. Since the density and wavefunctions are not quite consistent, better convergence tolerances are needed to get accurate forces and stresses.
- DM is usually the fastest, but can be unstable when there are large systems, band crossings, multiple spin states, and/or the dielectric susceptibility is large.