Practical calculations using first-principles QM
Convergence, convergence, convergence

Keith Refson
STFC Rutherford Appleton Laboratory

September 20, 2012
First-principles methods may be used for subtle, elegant and accurate computer experiments and insight into the structure and behaviour of matter.
Synopsis

This aims of this lecture are

1. To use the examples to demonstrate how to obtain converged results, i.e. correct predictions from the theory.
2. How to avoid some of the common pitfalls and to avoid computing nonsense.

Convergence

Results of First-Principles Simulations
Synopsis

Convergence
Approximations and Convergence
Convergence with basis set
Error Cancellation
Plane-wave cutoff
convergence/testing
Pseudopotentials and cutoff behaviour
FFT Grid parameters
Force and Stress
Iterative Tolerances
K-point convergence
Strategies for convergence testing
Golden rules of convergence testing

Structural Calculations
Summary
“Every *ab-initio* calculation is an approximate one”.

**Distinguish physical approximations**

- Born-Oppenheimer
- Level of Theory and approximate XC functional

**and convergable, numerical approximations**

- basis-set size.
- Integral evaluation cutoffs
- numerical approximations - FFT grid
- Iterative schemes: number of iterations and exit criteria (tolerances)
- system size

**Scientific integrity and reproducibility:** All methods should agree answer to (for example) “What is lattice constant of silicon in LDA approximation” if sufficiently well-converged.

**No *ab-initio* calculation is ever fully-converged!**
Basis set is fundamental approximation to shape of orbitals. How accurate need it be?

- The variational principle states that $E_0 \leq \frac{<\psi|H|\psi>}{<\psi|\psi>}$ so a more accurate representation of the orbitals $\psi$ arising from a more complete basis set will always lower the computed ground-state energy.

- corollary: Error in energy $\delta E_0 \propto \delta |\psi|^2$.

- Increase size and accuracy of plane-wave basis set by adding more plane waves with higher $G$ (increase $E_{\text{cut}}$).

- Increase size and accuracy of Gaussian basis set by adding additional orbitals, sometimes called “diffuse” or “polarization” functions. Ad-hoc because no single parameter to tune.

- Whatever the basis set it is prohibitively expensive to approach full convergence.

- Fortunately well converged properties may frequently be computed using an incomplete basis.

- Size of planewave basis set governed by single parameter cut_off_energy (in energy units) or basis_precision=COARSE/MEDIUM/FINE.

- Though $E$ is monotonic in $E_c$ it is not necessarily regular.
Consider energetics of simple chemical reaction
\[ \text{MgO}_\text{(s)} + \text{H}_2\text{O}_\text{(g)} \rightarrow \text{Mg(OH)}_2\text{(s)} \]

Reaction energy computed as
\[ \Delta E = E_{\text{product}} - \sum E_{\text{reactants}} = E_{\text{Mg(OH)}_2} - (E_{\text{MgO}} + E_{\text{H}_2\text{O}}) \]

Energy change on increasing \( E_{\text{cut}} \) from 500 → 4000eV
- \( \text{MgO} \) : -0.021eV
- \( \text{H}_2\text{O} \) : -0.566eV
- \( \text{Mg(OH)}_2 \) : -0.562eV

Convergence error in \( \Delta E \) : -0.030eV

Errors associated with H atom convergence are similar on LHS and RHS and cancel in final result.

Energy *differences* converge much faster than ground-state energy.

Always use same cutoff for all reactants and products.
Properties of a plane-wave basis.

- Cutoff determines highest representable spatial Fourier component of density.
- Energy differences converge faster than total energies. Electron density varies most rapidly near nuclei, where it influences bonding very weakly.

\[ E_{\text{cut}} \] (and \( G_{\text{max}} \)) depend only on types of atoms, not numbers.
- Simulation cutoff is maximum of individual elements/pseudopotentials.
- Required cutoff is \textit{system-independent} but not \textit{property-independent}.
With efficient pseudopotentials rate of convergence with cutoff is not always smooth.

- Can get plateaus or other non-asymptotic behaviour
- Sometimes cause is over-optimization of the pseudopotential too low a desired cutoff.
- In cases with this behaviour can choose criterion for convergence energy to be plateau values.
- Absolute energy convergence is rarely desirable. *Force and stress convergence is much more useful criterion.*

(Example is *not* a pseudopotential from the CASTEP database)
Some optimizations and tweaks of FFT grid dimensions ...

- FFT grid should be large enough to accommodate all G-vectors of density, $n(r)$, within cutoff: $G \leq 2G_{\text{MAX}}$.
- Guaranteed to avoid "aliasing" errors in case of LDA and pseudopotentials without additional core-charge density.
- In practice can often get away with $1.5G_{\text{MAX}}$ or $1.75G_{\text{MAX}}$ with little error penalty for LDA without core or augmentation charge.
- GGA XC functionals give density with higher Fourier components, and need $1.75G_{\text{MAX}} - 2G_{\text{MAX}}$.
- Finer grid may be needed to represent USP augmentation charges or core-charge densities.
- CASTEP incorporates a second, finer grid for charge density to accommodate core/augmentation charges while using $G_{\text{MAX}}$ for orbitals.
- Set by either parameter `fine_grid_scale` (multiplier of $G_{\text{MAX}}$) or `fine_gmax` (inverse length $G_{\text{FINE}}$).
- `fine_gmax` is property of pseudopotential and transferable to other cells but `fine_grid_scale` is not.
- (Rarely) may need to increase `fine_grid_scale` for DFPT phonon calculations using GGA functionals for acoustic sum rule to be accurately satisfied.
Frequently need to find crystal structure at *mechanical equilibrium*.

- Given guessed or exptl. initial structure, seek local minimum of Born-Oppenheimer energy surface generated by K-S functional.
- Energy minimum implies forces are zero but *not* vice versa.

- CASTEP provides for “geometry optimization” using *quasi-newton* methods.
- Need to worry about *accuracy* of forces.
- Forces usually converge at *lower cut-off* than total energy because density in region of nucleus unimportant.

- different physical quantities converge at different rates.
- Always test convergence specifically of *quantities of importance to your planned calculation*
How to control the *iterative solvers*?
- Parameter `elec_energy_tol` specifies when SCF energy converged.
- Optimizer also exits if max cycles reached – *always check that it really did converge*.
- How accurate does SCF convergence need to be?

**Energetics**: same accuracy of result.

**Geometry/MD**: much smaller *energy* tolerance needed to converge forces.
- Cost of higher tolerance is only a few additional SCF iterations.
- Or use `elec_force_tol` to exit SCF only when force convergence criterion satisfied.

- Inaccurate forces are common cause of geometry optimization failure.
Brillouin Zone sampling is another important convergence issue. Usually set in `<seed>.cell` with keyword `kpoint_mp_grid p q r` with optional offset `kpoint_mp_offset`.

- Convergence is **not** variational and frequently oscillates.
- Even simple metals like Al need dense meshes for primitive cell.
- Finite-temperature *smearing* can accelerate convergence, but must extrapolate the result back to 0K.
- In case of insulators some k-point error cancellation occurs but only between *identical* simulation cells.

Consequently comparative phase stability energetics and surface energetics frequently demand high degree of k-point convergence.

![Graph](image-url)
Convergence tests should be designed to be as quick and simple as possible.
- Make use of transferrability of cutoff and fine grid scale, and length scaling for k-point convergence.
- Test cutoff convergence on *small* model systems with same elements. If target system is zeolite, test plane wave convergence on related oxides.
- When testing plane-wave cutoff use a coarse k-point mesh.
- When testing k-point convergence use a low cutoff.
- Geometry optimizations unduly complicate convergence testing. Better to use (cheaper) single-point calculations and plot forces and stress.
- Accuracy of forces also useful as proxy measure of convergence in phonon calculations - but not completely reliable. Again consider testing a related but simpler system.
- Finish with small number of tests on full system and final property of interest.
Golden rules of convergence testing

1. Wherever possible use single-point energy calculations.
2. Test PW convergence on *small* model system(s) with same elements.
3. Only change *one parameter at a time*.
4. Use knowledge of transferability of PW cutoff and fine_gmax.
5. Don't be fooled into choosing too high a plane-wave cutoff when you are really just converging fine FFT grid. Check fine_gmax for USPs.
6. Use physical $\Delta k$ spacing to scale k-point sampling from small system to large.
7. Use *forces*, *stresses* and other cheap properties as measure of convergence.
8. Write your convergence criteria in the paper, **not** just “250eV cutoff and 10 k-points in IBZ”.
9. Convergence is achieved when value stops changing with parameter under test, **NOT** when the calculation exceeds your computer resources and **NOT** when it agrees with experiment.
Sometimes we need to compute a non-periodic system with a PBC code.

- Surround molecule by vacuum space to model using periodic code.
- Similar trick used to construct slab for surfaces.
- Must include enough vacuum space that periodic images do not interact.
- To model surface, slab should be thick enough that centre is "bulk-like".
- Beware of dipolar surfaces. Surface energy does not converge with slab thickness.
- When calculating surface energy, try to use same cell for bulk and slab to maximise error cancellation of k-point set.
- Sometimes need to compare dissimilar cells – must use absolutely converged k-point set as no error cancellation.
Two ways to evaluate equilibrium lattice parameter - minimum of energy or zero of stress.

Stress converges less well than energy minimum with cutoff.

Incomplete basis error in stress approximated by Pulay stress correction.

“Jagged” E vs V curve due to discreteness of $N_{PW}$. Can be corrected using Francis-Payne method (J. Phys. Conden. Matt 2, 4395 (1990)).

Finite-size basis corrections for energy and stress automatically computed by CASTEP.
Two possibilities for variable-cell MD or geometry optimization when using plane-wave basis set.

- In fixed basis size calculation, plane-wave basis $N_{PW}$ is constant as cell changes.
- Cell expansion lowers $G_{max}$ and K.E. of each plane wave, and therefore lowers effective $E_c$.
- Easier to implement but easy to get erroneous results.
- Need very well-converged cutoff for success.
- fixed cutoff calculations reset basis for each volume, changing $N_{PW}$ but keeping $G_{max}$ and $E_c$ fixed.
- This is almost always the correct method to use.
Phase Equilibrium
Consider simple structure such as rocksalt (B1)

one of simplest calculations is to compute E as function of V

B1 rocksalt every atom is on a crystallographic high-symmetry site. No geometry optimisation needed!

**Equation of state** (Really PV curve, which is commonly measured experimental quantity.)
Real EOS are PVT relations, but this lecture restricted to easy case, T=0.

- Model EOS for fluids, eg Van der Waals, Redlich-Kwong
- There are several common models of EOS for solids based on elasticity theory
- Birch Murnaghan (Phys. Rev 71 809 (1947)) is most commonly used EOS

\[
E(V) = E_0 + \frac{9V_0B_0}{16} \left\{ \left[ \left( \frac{V_0}{V} \right)^{\frac{2}{3}} - 1 \right]^3 B_0' + \left[ \left( \frac{V_0}{V} \right)^{\frac{2}{3}} - 1 \right]^2 \left[ 6 - 4 \left( \frac{V_0}{V} \right)^{\frac{2}{3}} \right] \right\}
\]

- Two definitions of bulk modulus \(B_0\); \(B_0 = -V_0 \frac{d^2P}{dV^2}\) and as a fit to B-M or other EOS.
- These are not equivalent.
- Compression/EOS experiments use a B-M fit whereas ultrasound methods measure 2nd derivative
Birch-Murnaghan Equation of State

- **range** of data changes fit result.
- Only $V \leq V_0$ accessible to experiment.
- $V > V_0$ also accessible to calculation.
- $B_0$ should be compared using fit over same or similar range to expt.
- Need to understand how $B_0$ was measured to make valid comparison.
- $B_0$ from ultrasonic expt. should be compared to $-V_0 \frac{d^2P}{dV^2}$

**Figure:** Birch-Murnaghan fits to CaO EOS

- **Castep calculation** $V_0 = 110.2$, $B_0 = 28.4$ GPa, $B'_0 = 4.16$
- $V_0 = 110.3$, $B_0 = 27.9$ GPa, $B'_0 = 4.23$
Phase equilibrium occurs when 2 criteria met simultaneously

- Pressures of 2 phases are equal, ie \( P = -\frac{dE_1}{dV} = -\frac{dE_2}{dv} \).
- Enthalpies are equal \( E_1 + PV_1 = E_2 + PV_2 \Rightarrow E_2 - E_1 = P (V_2 - V_1) \).

Then \( P_{eq} \) is gradient of common tangent

Alternatively use enthalpy-pressure plot, where most stable phase is that of lowest enthalpy. Need accurate calculation of pressures for this approach to work, so higher degree of PW convergence needed.
Zero-point energy

- Zero-point energy contribution \( \sum_i \frac{\beta \hbar \omega_i}{2} \) can be significant and is in general different for different phases.
- Can calculate if vibrational phonon spectrum or DOS known.
- Use lattice dynamics calculation or (less satisfactory) experimental data.
- Harmonic Free energy is
  \[
  F = E + \frac{1}{\beta} \int g(\omega) \log \left(2 \sinh \left(\frac{\beta \hbar \omega}{2}\right)\right) \, d\omega
  \]
  where \( g(\omega) \) is the phonon density of states (DOS).
- To calculate DOS need phonon frequencies at all \( q \) in BZ.

**MgH2 phonon DOS**
Quasi-harmonic approximation assumes that phonon frequencies depend only on the cell parameters.

- Ignores intrinsic anharmonic thermal effects on DOS.
- Works for relatively harmonic systems
- Valid for $T \lesssim 0.5 T_m$.
- QHA is cheapest way of extending ab-initio to $TT > 0$

In Zero Static Internal Stress Approximation geometry optimise co-ordinates at each volume (or strain)

- Quasi-Harmonic Free energy geometry optimization possible with empirical force fields but not yet ab-initio.
Elastic strain theory gives

\[ E(\epsilon) = E_0 + V/2 \sum_{ij}^6 C_{ij} \epsilon_i \epsilon_j + O(3) \]

“Programmed” strains \( \epsilon \) may be used to extract individual elastic constants, e.g. if

\[ \epsilon = \frac{1}{2} \begin{bmatrix} 0 & \delta & \delta \\ \delta & 0 & \delta \\ \delta & \delta & 0 \end{bmatrix} \]

then

\[ E = E_0 + \frac{3}{2} VC_{44} \delta^2 + O(3) \]

for a cubic crystal.


For low symmetry crystals more efficient to compute stress and use

\[ \sigma_{\alpha \beta} = C_{\alpha \beta \gamma \delta} \epsilon_\gamma \epsilon_\delta. \]

But need very well converged stress.

No automated calculation built into CASTEP; available in Materials Studio.
Elastic properties of CaO

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Theory</th>
<th>Expt (1)</th>
<th>Expt (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>4.784 ± 0.001</td>
<td>4.8105</td>
<td>4.8105</td>
</tr>
<tr>
<td>$K_0$</td>
<td>111.9 ± 1.5</td>
<td>117</td>
<td>112.5 ± 0.6</td>
</tr>
<tr>
<td>$K'_0$</td>
<td>4.3 ± 0.3</td>
<td>5.9 ± 2</td>
<td>4.8 ± 0.1</td>
</tr>
<tr>
<td>$C_s$</td>
<td>83.1 ± 0.9</td>
<td>81.9 ± 3</td>
<td>82.0 ± 0.3</td>
</tr>
<tr>
<td>$C_{11}$</td>
<td>223.1 ± 1.9</td>
<td>226.2 ± 0.9</td>
<td>221.9 ± 0.6</td>
</tr>
<tr>
<td>$C_{12}$</td>
<td>56.9 ± 1.6</td>
<td>62.4 ± 0.9</td>
<td>57.8 ± 0.7</td>
</tr>
<tr>
<td>$C_{44}$</td>
<td>83.9 ± 0.4</td>
<td>80.6 ± 0.3</td>
<td>80.3 ± 0.1</td>
</tr>
</tbody>
</table>

Surfaces
Surfaces can be modelled as a *slab* cleaved from bulk crystal. Can calculate

- Surface *energy* or *free energy*
- Energies of steps
- Adsorption energies and structures of adsorbates
- Surface chemical reaction energies.
- Choice of slab: usually need to make 2 surfaces identical.
- Surfaces related by symmetry operation are more easily geometry optimised.
- Simulation cell should not be optimised. (use \textit{fix\_all\_cell} = \textit{T}).
- In-plane cell parameters usually fixed at values from relaxed bulk crystal.
Surface free energy defined as

\[ E_{\text{surf}} = \frac{(E_{\text{slab}} - E_{\text{bulk}})}{A} \]

A is total area of both surfaces.

- \( E_{\text{surf}} \) is sensitive quantity requiring well-converged total energies.
- Can sometimes gain some \( k \)-point error cancellation between slab and bulk calcs by using non-primitive bulk cell with same in-plane vectors as slab. (not \( \text{CaCO}_3 \) 10\( \bar{1} \)4 shown).
- Sometimes need to compare dissimilar cells – must use absolutely converged \( k \)-point set as no error cancellation.
- Need only 1 \( k \)-point in direction perpendicular to slab. Any dispersion in bands is error due to insufficient vacuum gap - no point in calculating accurately!
- Need to test convergence with both slab thickness and vacuum gap.
Electric dipoles perpendicular to surface raise theoretical difficulties.

Energy *does not converge* with slab thickness.


In classical charge model, Type III unstable and must always reconstruct.

In *ab initio* calculation, surfaces can instead become metallic.

Polar molecules on surface also raise dipole-dipole problems.

Can sometimes use double surface with inversion symmetry.
Defect calculations

- Only need to converge energy to a few meV, but still need accuracy in forces to correctly describe strain relaxation.
- Strain relaxation. Local strain around defect decreases as 1/R. Can model long-range strain relaxation using classical models if suitable potential exists.
- Charged defects can be modelled using periodic interaction correction terms (M. Leslie and M. Gillan, *J. Phys. Cond. Mat.* 18, 973 (1985))
Summary
- Used with care, first principles simulations can give highly accurate predictions of materials properties.
- Full plane-wave basis convergence is rarely if ever needed. Error cancellation ensure that energy differences, forces and stress converge at lower cutoff.
- Convergence as a function of adjustable parameters must be understood and monitored \textit{for the property of interest} to calculate accurate results.
- Don’t forget to converge structural/finite size/statistical mechanics as well as the electronic structure if the calculation demands it.
- A poorly converged calculation is of little scientific value if the convergence error is unknown.
- An over-converged calculation could be costly computer resources. You should match degree of convergence to scientific question you seek to answer.