Large scale DFT: why we need it and how to do it

David Bowler

Thomas Young Centre, UCL, Gower St, London, UK
London Centre for Nanotechnology, 17-19 Gordon St, London, UK
Department of Physics & Astronomy, UCL, Gower St, London, UK
WPI MANA, NIMS, Tsukuba, Ibaraki, Japan

david.bowler@ucl.ac.uk
http://www.atomisticsimulations.org/
http://www.order-n.org/
@MillionAtomMan
What is large scale?

• Large scale: 500-10,000 atoms \( (N_{\text{atoms}}/N_{\text{proc}} \approx 5-50) \)
• 100-1,000+ processes
• Linear scaling: 1,000-1,000,000 atoms

• Many atoms: length scales
• Many processes: time scales

• Almost all DFT calculations are under 500 atoms
  • Cubic scaling
  • Habit?
Why large-scale DFT?

• We have become very good at calculations using 100-1,000 atoms.
• But there are important larger-scale problems:
  • Large defects (dislocations, surfaces)
  • Amorphous & heterogeneous materials
  • Realistic concentrations (doping and solutions)
  • Biomolecules
  • Nanostructures
  • Testing and validation of NN potentials

Why large-scale DFT?

We now turn to the key electronic structure: DOS and the band gap, on which the slab thickness has a profound effect. In order to gain further insights into the influence of the slab thickness on three key properties: the bond length and buckling angle; and the maximum (VBM). As will become clear below, the VBM of thin slabs are dominated by the DB states, while those of thicker slabs are dominated by the VB states at different points in the Brillouin zone (BZ) to assign the edge, the VBM character of a clean Si(1 0 0) surface has not been well studied yet, and we will discuss this based on the calculated electronic structures of a set of 6, 14, 22, 38, 62, and 78 L thick slabs. All energies are shifted so that the bottom of the potential of Si is a reduction of the surface band gap. Figure 1(a) shows partial charge densities for each eigenvalue layer of each slab has the same value. (b) Surface (circle) and bulk (square) band gaps as a function of the slab thickness (table 3). The common features among various slab thicknesses. Figure 2(b) shows the DOS projected onto a Si dimer for slab 114 meV L thick slab 114 meV. We recall the discussion of the relative stability of the Si dimer bond length, buckling angle both converge at 14 d value. We chose carefully.

○ Figure 1.

○ Figure 2.

○ Figure 3.
Why not large-scale DFT?

• What is stopping us?
  • Lack of a “killer application”
  • Received wisdom/habit
  • Perception that big is hard
  • Familiarity with standard codes
  • Underestimation of need
Competing requirements

- Accuracy
- Functional
- System size
- Basis size
- MD time
- Solver
- Cost
How many atoms?

Efficiency
- Few
- Representation

Alloys (concentration)
Doping

Accuracy
- Many
- Realistic

Perfect crystals

Defects, dislocations etc
Solving for the ground state

- $N^3$
  - Direct diagonalisation
  - Variational solvers (orthogonalisation)
- $N^2$
  - PEXSI
- $N$: localisation
  - Iterative
  - Variational
  - Patching/subsystem
Why large-scale DFT?

- HPC centres give many 100,000s cores (exascale)
- Effective parallel scaling is key
  - Reduce time to ground state
  - Increase system size
  - Communication (CPU-CPU, CPU-GPU)
- Use of GPUs
- Multi-threading vs message passing
Large scale calculations

- CONQUEST (1,000,000)
- BigDFT (10,000)
- CP2K (1,000,000)
- OpenMX (10,000)
- ONETEP (21,000)
- MGmol (1,000,000)
- LDC-DFT (6,000,000)
- DFT-FE (10,000)
- PARSEC (26,000)
- DGDFT (2,500,000)
- RSDFT (107,000)
- Google TPU (31,000)
Large scale calculations

• CONQUEST (768,000)  
• CP2K (47,000)  
• ONETEP (100,000)  
• DGDFT (35,900,000)  
• RSDFT (664,000)

• MGmol (1,600,000)  
• LDC-DFT (786,000)  
• DFT-FE (156,000 + 22,800)  
• PARSEC (16,000)
Conquest
Extending DFT size

• Real-space locality
• Key to parallel efficiency \textit{and} scaling
• Standard implementations are \textit{non-local}: wavefunctions span whole system

• But electronic structure is local:
  • Charge density
  • Wannier functions
  • Density matrix decay
Key concept: locality

- Local basis functions are key
- Sparse matrices; reduced scaling
- Operations local to process
- Fits with nearsightedness
- Makes efficient parallelisation easier

\[ \psi_{nk}(\mathbf{r}) = \sum_{i\alpha} c_{i\alpha,nk} \phi_{i\alpha}(\mathbf{r}) \]

- Pulay forces
- Orthogonalisation?
- Systematic convergence?
Support Functions

- Support functions are represented in terms of a basis:
  \[ \phi_{i\alpha}(\mathbf{r}) = \sum_s c_{i\alpha s} \chi_s(\mathbf{r}) \]

- CONQUEST can use two different basis sets:
  - Pseudo-atomic orbitals (cf OpenMX, SIESTA)
  - B-splines or blips (cf wavelets, ONETEP psincs)

- PAOs: analytic operations, small basis, intuitive
- Blips: systematic convergence (to plane-wave accuracy)
Solving for DM

• Support functions give $H$ and $S$ (linear scaling)

• How do we find the density matrix?

  • Exact (SCALAPACK): 1-10,000 atoms
    • PAOs or multi-site support functions (MSSF)
  
  • $O(N)$ (range): 1,000-1,000,000+ atoms

• Complementary: choose appropriate method

• Aim: efficient, accurate solution & analysis
Diagonalisation: DM

• We build $\mathbf{K}$ from wavefunction coefficients

• Efficient memory use

$$\psi_{n_k}(\mathbf{r}) = \sum_{i_\alpha} c_{i_\alpha,n_k} \phi_{i_\alpha}(\mathbf{r})$$

$$K_{i_\alpha,j_\beta} = \sum_{n_k} w_{n_k} f_{n_k} c^*_{i_\alpha,n_k} c_{j_\beta,n_k}$$
**CONQUEST: Basic capabilities**

- Efficient parallelisation (from 1 to 200,000+ cores)
- Flexible (from 1 to 2,000,000+ atoms)
- Exact, analytic forces; MD on 32,000+ atoms
- ONCVPSP pseudopotentials (PseudoDojo)
- LDA, GGA, DFT-D2/3/TS, vdW-DF, EXX (partial)

Open source: [https://github.com/OrderN/CONQUEST-release/](https://github.com/OrderN/CONQUEST-release/)
Parallelisation

- Assign areas of space to MPI processes
- Care with communication, all-atom operations
- Increase processes with atom number
Pseudo-atomic orbitals
CONQUEST: PAOs

• Use PseudoDojo pseudopotentials
• Pseudo-atomic orbital basis sets
• Can we define reasonable default bases?
• How do we choose radii?
  • Siesta energy shifts
  • Adapt: large and small (2eV and 20meV)
  • Either share energies or average radii
  • Allows SZ up to TZTP automatic bases
Hamann NC pseudopotentials

PseudoDojo set (0.7 meV/atom)

As accurate as PAWs

High quality pseudopotentials
PAO generation

• Stand-alone CONQUEST code
• Reads output from Hamann code (small patch)
• Generates pseudo-atomic orbitals and ion files
• Automated basis sets possible
• Full user control available
CONQUEST basis set tests

Bulk Ge

<table>
<thead>
<tr>
<th></th>
<th>PW</th>
<th>CQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_0/V$</td>
<td>$B$</td>
</tr>
<tr>
<td>C</td>
<td>3.558</td>
<td>449</td>
</tr>
<tr>
<td>Si</td>
<td>5.431</td>
<td>93</td>
</tr>
<tr>
<td>Ge</td>
<td>5.676</td>
<td>67</td>
</tr>
<tr>
<td>STO</td>
<td>58.8</td>
<td>186</td>
</tr>
<tr>
<td>PTO</td>
<td>60.1</td>
<td>191</td>
</tr>
<tr>
<td>MgSiO$_3$</td>
<td>167.4</td>
<td>235.7</td>
</tr>
</tbody>
</table>

Basis set tests: SiO$_2$

SiO$_2$ (stishovite)  

Z. Raza (NIMS)

<table>
<thead>
<tr>
<th></th>
<th>a (Å)</th>
<th>c/a</th>
<th>% error in a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>4.177</td>
<td>0.638</td>
<td></td>
</tr>
<tr>
<td>VASP/PBE$^a$</td>
<td>4.225</td>
<td>0.637</td>
<td>+1.1%</td>
</tr>
<tr>
<td>SZ</td>
<td>4.476</td>
<td>0.638</td>
<td>+7.1%</td>
</tr>
<tr>
<td>SZP</td>
<td>4.278</td>
<td>0.638</td>
<td>+2.4%</td>
</tr>
<tr>
<td>DZP</td>
<td>4.270</td>
<td>0.638</td>
<td>+2.2%</td>
</tr>
<tr>
<td>TZTP</td>
<td>4.233</td>
<td>0.638</td>
<td>+1.3%</td>
</tr>
</tbody>
</table>

Basis set tests: SrTiO$_3$
How good is the density?

- Key quantity:
  - Energy
  - Polarisation, …
- Compare to PW
- 40Ha cutoff

$\Delta n(r)$ [electrons/Å$^{-3}$]

Electron. Struct. 2, 025002 (2020)
Figure 5. The convergence properties of plane-wave calculations where PAO calculations featuring the same error have been overlaid for comparison. Calculations were performed on the Pm\bar{3}m PTO structure. (a) Convergence with respect to the total electronic error integral of equation (1). (b) Convergence with respect to $\Delta E$, the energy difference between a given calculation and the energy obtained from the 40 Ha plane-wave cutoff. These are in good agreement with the values obtained from plane-waves. This once again emphasizes the electronic accuracy achievable with the default PAOs.

We have also examined the possibility that the errors in $q_B$ could be an artefact of pressure using the particular case of Pm\bar{3}m PZO. We can see from table 6 that the optimised lattice constants for the PAO bases overestimate the plane-wave result by 0.99\%, 0.46\% and 0.34\% for the SZP, DZDP and TZTP basis sets respectively. If we then perform simulations at the plane-wave lattice constant (as was done for the results in table 3) this imposes a isotropic pressure of $-5.28$ GPa (SZP), $-2.63$ GPa (DZDP) and $-1.46$ GPa (TZTP). Because of this fact, we calculated the Bader quantities once more using the zero pressure lattice constants. Remarkably, $q_B$ changes only marginally (no more than $\pm 0.01\text{e}$). Since we are working at a larger volume, the $V_B$ must of course increase, but, the ratios of the cation to anionic volumes remain constant. Since roughly the same amount of charge is now enclosed within a larger $V_B$, we naturally see an increase $n_B$ for all sites.

We see that the ratio of the cation to anion volumes is a decreasing function of basis set completeness, decreasing by $\approx 0.1$ for $V_{\text{B}}/V_O$ and $\approx 0.2$ for $V_{\text{Pb}}/V_O$ from SZP to TZTP. This implies that in the smaller basis set, O occupies a smaller ionic volume in comparison to the Pb and B-sites. This could result in small differences in lattice dynamics between the basis sets and could effect the Goldschmidt tolerance factor, depending explicitly on ionic radii [65, 66].

3.2. Soft-mode distortions

In this section we consider the soft-modes known to drive the phase transitions in PTO and PZO. We consider the amplitude of each identifiable irrep in the relaxed structures for each basis set. We also consider the degree of energy lowering associated with each of these irreps and define phase transition energies. We display the displacive modes in tables 4 and 5. Strain-mode influences the phase transition in PZO only by a small amount so we include only discussion of strain modes in PTO, coupling strongly to the displacive $\Gamma_{-4}$ mode.

The phase transition energies are quoted in table 7 and the linear evolution of mode energetics are shown in figure 6.

Before discussing mode amplitudes, we must first carefully define them. We do so following the format of the ISODISTORT amplitude normalised to the primitive cell [60]. Once an atomic displacement has been identified as belonging to a particular irrep, the displacement is calculated in fractional coordinates relative to the parent structure. This defines the amplitude of a specific displacement in the irrep. To calculate $A_p$ we now normalise the amplitude by a factor of $\sqrt{V_p/V_s}$ for supercell/primitive cell volumes $V_p/V_s$. Now to calculate the amplitude of the irrep as a whole, we take the square root of the sum of the squares for each irrep amplitude. Tables 4 and 5 are then tabulations of $A_p$.
Multi-site SFs
Multi-site SFs

• We want an accurate basis
• We want efficient performance
• Make a small set of SFs for each atom
• Build them from PAOs from several atoms
• Local filter diagonalisation

JCTC 10, 4813; PCCP 17, 31427
Multi-site functions

- Use PAOs from neighbouring atoms
- Effectively localised molecular orbitals
- Two ranges: MS and LFD
- This is accurate and efficient: small $H$ and $S$
- Follows method of Rayson & Briddon*
- Improvements: JCTC 10, 4813; PCCP 17, 31427

*M. J. Rayson and P. R. Briddon, Phys. Rev. B, 2009, 80, 205104
MSSF: bulk Si

- Lattice constant and bulk modulus good
- Band structure excellent
Optimising MSSF

• Original method performs filter diagonalisation
  • Large, primitive basis; small support function basis
• We diagonalise $\mathbf{H}$ for a small cluster
• Filter eigenstates onto MSSF using Fermi function
• Can we improve by *optimising* energy?
• Analytic energy gradients available
Optimising MSSF: bulk Si

• Little effect for large radii
• Significant improvement for small radii
• Promising approach
MSSF timing & scaling

<table>
<thead>
<tr>
<th></th>
<th>PAO</th>
<th>MSSF</th>
<th>PAO</th>
<th>MSSF</th>
<th>MSSF</th>
</tr>
</thead>
<tbody>
<tr>
<td># SF</td>
<td>27,192</td>
<td>8122</td>
<td>54,384</td>
<td>16,244</td>
<td>16,244</td>
</tr>
<tr>
<td># MPI</td>
<td>432</td>
<td>432</td>
<td>108</td>
<td>108</td>
<td>864</td>
</tr>
<tr>
<td>Matrix</td>
<td>64</td>
<td>400</td>
<td>156</td>
<td>1,455</td>
<td>406</td>
</tr>
<tr>
<td>Diag</td>
<td>1193</td>
<td>39</td>
<td>37,648</td>
<td>701</td>
<td>166</td>
</tr>
<tr>
<td>Total</td>
<td>1257</td>
<td>439</td>
<td>37,804</td>
<td>2156</td>
<td>572</td>
</tr>
</tbody>
</table>

• Graphene on Rh(1 1 1); 1,544 or 3,088 atoms
MSSF: Applications
Domains in PTO/STO

- PTO on STO
- Thin film (0-9 f.u.)
- Up to ~2,000 atom cells
- 3 cell film shown
  (a) Paraelectric
  (b) Monodomain FE (in plane)
  (c) Polydomain FE (out of plane)

Domains in PTO/STO

- Compare FE modes to AFD
- Mono vs polydomain
- Effect of thickness
- Monodomain[110] stable (3 fu)
- Polydomain stable (>3 fu)
- Competition between:
  - Ferroelectric modes
  - Anti-ferrodistortive modes

Domains in PTO/STO

- Polarization fields
  a) Thickness 9: flux closure
  b) Thickness 3: polar wave
- Competition between:
  - Ferroelectric modes
  - Anti-ferrodistortive modes
- Periodicity is $p(2 \times \Lambda)$

Doped Silicon Nanowires

- Pure Si NWs: size (radius, repeat length), doping
- \(~600-2,500\) atoms (MSSF)
- Electronic structure
• MSSF reproduces full lattice constant
• Gap reduces with increasing radius
• Converged by ~9nm diameter
• PDOS depends on proximity to surface
• First 2-3 layers mainly
Doping: As

- Largest NW (7.3nm)
- Eight layers along axis
- As at edge (1) & centre (2)
- Flatter band at centre
Core-shell NWs: VBM/CBM

Si-Ge circular NW (large)
Other applications

- Graphene on Rh(111)
- Topological defects in YGaO₃
- Au nanoparticles
Linear scaling
How to make it local?

Density matrix

\[ \rho(r, r') = \sum_n f_n \psi_n(r) \psi_n^*(r') \]

\[ \rho(r, r') \to 0, |r - r'| \to \infty \]

\[ \rho(r, r') = 0, |r - r'| > R_c \]
How do we represent it?

Support functions: local in space
Ranged matrix, $K$

$$
\rho(r, r') = \sum_{i\alpha, j\beta} \phi_{i\alpha}(r) K_{i\alpha, j\beta} \phi_{j\beta}(r')
$$
Linear scaling DFT

Minimise total energy with respect to $\phi_{i\alpha}(r)$ and $K_{i\alpha,j\beta}$ subject to:

1. Correct electron number
2. Self consistency (potential, charge density)
3. Idempotency of density matrix $\rho^2 = \rho$
Truncating density matrix: $O(N)$

- The support functions are confined within a sphere, radius $R_{\text{reg}}$
- The $K$ matrix is truncated
- All matrices are sparse, $\rho$ is local
- By increasing radii, approach exact result

\[ K_{i\alpha j\beta} = 0, \quad |\mathbf{R}_i - \mathbf{R}_j| > R_c \]
Linear scaling DFT

• How do we find $K$?
  • Direct methods
    • LNV
    • OMM
  • Iterative methods
    • McWeeny
    • TC etc
  • Divide-and-conquer
  • Recursion (Lanczos etc)
  • Penalty functionals
• Alternative: OFDFT
Truncation

• Spatial truncation
  • Approximate DM sparsity/range (imposed)
  • Variational
  • Consistent sparsity patterns (parallelisation)
  • LNV, OMM

• Numerical truncation
  • Approximate DM decay (impose tolerance)
  • Non-variational (?)
  • Varying sparsity patterns
  • McWeeny, TC2 etc

Idempotency

Hard to impose $K^2 = K$
Use approximation (McWeeny)
Minimising $K$: LNV

• Imposing idempotency is difficult
• We use the McWeeny transform:
  \[ \rho = 3\sigma^2 - 2\sigma^3 \]
• Here $\sigma$ is an auxiliary density matrix
• If $\lambda_\sigma$ lie in $[-0.5, 1.5]$ then $\lambda_\rho$ will lie in $[0,1]$
• Vary energy with respect to elements of $\sigma$
• During minimisation, $\rho$ tends towards idempotency
Issues to consider

• Prefactors
• Accuracy (truncation vs BZ sampling)
• Information required (eigenstates)
• Parallelisation
• Timescales (linear scaling MD and weak scaling)
Linear scaling: applications
Trenches in PTO

- Domain walls align parallel to surface trenches
- PTO film strained to STO
- $O(N)$ calculations
- Up to 5,136 atoms
- Most stable parallel to DC

PRL 127, 247601 (2021)
• Local polarisation fields for the three orientations
• Key: how polar texture adapts to trench
• PRL 127, 247601 (2021)
Trenches in PTO

- Vertical surface strain
- DC (top) and DW (bottom)
  - Pristine (solid line)
  - Trench (dashed line)
- Cooperative strain (DC)
- More disorder (DW)

PRL 127, 247601 (2021)
Heteroepitaxy: Ge/Si(001)

- Nanostructures & self-assembly fascinating
- Ge on Si(001) has 4.2% lattice mismatch
- After ~3ML, 3D hut clusters appear
- Is it driven by kinetics or energetics?
- Sergiu Arapan (NIMS)

![Imagery of 2xN, MxN, and Huts patterns]
Heteroepitaxy: Ge/Si(001)

- Hut clusters are complex, 3D objects
- Experimentally facets grow top-down
- What can we learn with O(N) DFT?

We performed structural relaxation of 200,000-atom systems with $O(N)$ DFT.
• How do huts grow?
• Experiment shows complete new facets
• We study dimer stability: where on the face?
• Elongated hut (blue: 8x13; yellow: 8x14)
• 6149 Ge atoms, 13824 Si atoms
Hut Cluster Growth

- Energy map for ad-dimers
- Red is stable, blue is less stable
- Short side only
- Two types of dimer: A and B
- A (generally) more stable
- Edges, base important
Electronic structure

- States within 0.01 eV of Fermi level
- Confined to hut
- 43 of 213,633 states

Linear scaling MD

- MD with $O(N)$ solver
- Core-shell NWs (~5,000 atoms) at 3000K
- Ge melts first
- NPT also possible
Conclusions

- Large-scale DFT with CONQUEST
  - Up to 1,000 atoms with no restriction
  - Up to 10,000 atoms with MSSF
  - Up to 1,000,000 atoms with $O(N)$ (so far)
- Application of DFT to PTO films
  - Very accurate basis sets
  - Polar domains in thin films
  - Interaction of domains with trenches
Acknowledgements

• Mike Gillan (CONQUEST originator)
• Tsuyoshi Miyazaki (CONQUEST co-leader)
• Jack Poulton, Shereif Mujahed, Zamaan Raza (PAOs)
• Ayako Nakata (MSSFs)
• Chathurangi Kumarasinghe (Si NWs)
• Jack Baker (PAOs, PTO/STO)
References

• Linear scaling:

• CONQUEST
  • J. Chem. Phys. 152, 164112 (2020)

• Ferroelectrics
  • Electron. Struct. 2, 025002 (2020)
  • Adv. Theory Simul. 3, 2000154 (2020)
  • PRL 127, 247601 (2021)

http://www.order-n.org/
http://www.atomisticsimulations.org/
https://github.com/OrderN/CONQUEST-release/