A Real-Space Genetic Algorithm for Crystal Structure Determination

N.L. Abraham and M.I.J. Probert
Department of Physics, University of York, Heslington, York, YO10 5DD
http://www-users.york.ac.uk/~nla101/

Abstract

There has been much interest in using genetic algorithms for determining the ground-state structure of clusters [Deaven and Ho, 1995] and more recently silicon surfaces [Chuang et al., 2004]. We present a real-space encoded genetic algorithm which is suitable not only for surface structure calculations, but also for bulk crystal structure determination. This algorithm makes use of a novel crossover technique in the generation of offspring. The method is also suitable as a polymorph search, and is flexible enough that population members can have different supercells. We will present results from a variety of empirical and *ab initio* systems, where all calculations have been performed using the CASTEP [Segall et al., 2002] code.

1 Introduction

Genetic Algorithms (GAs) are becoming widely used to determine the minimum energy configurations of atomic clusters [Deaven and Ho, 1995, Johnston, 2003], nanowires [Wang et al., 2001], and surfaces [Chuang et al., 2004, 2005]. These previous studies made use of a planar cut in the crossover operation, slicing each parent in two, and then swapping halves to generate offspring. However, a planar cut does not take the system periodicity into account.

The prediction of crystal structures from first principles has long been recognised as one of the outstanding challenges in solid state physics [Maddox, 1988, VanDeWalle, 2005]. The most recent methods of cluster expansion [Blum et al., 2005, Hart et al., 2005] assume the lattice structure of the crystal. In this poster we demonstrate a new method for unbiased *ab initio* crystal structure determination using a novel Genetic Algorithm which makes no assumptions of atom number, unit cell or lattice structure.

2 Genetic Algorithms

- Genetic Algorithms are a stochastic minimisation technique inspired by Darwin's Theory of Evolution [Holland, 1992] (see figure 1).
- The "population" is made up of a number of members, each of which is a viable solution to the problem being studied.

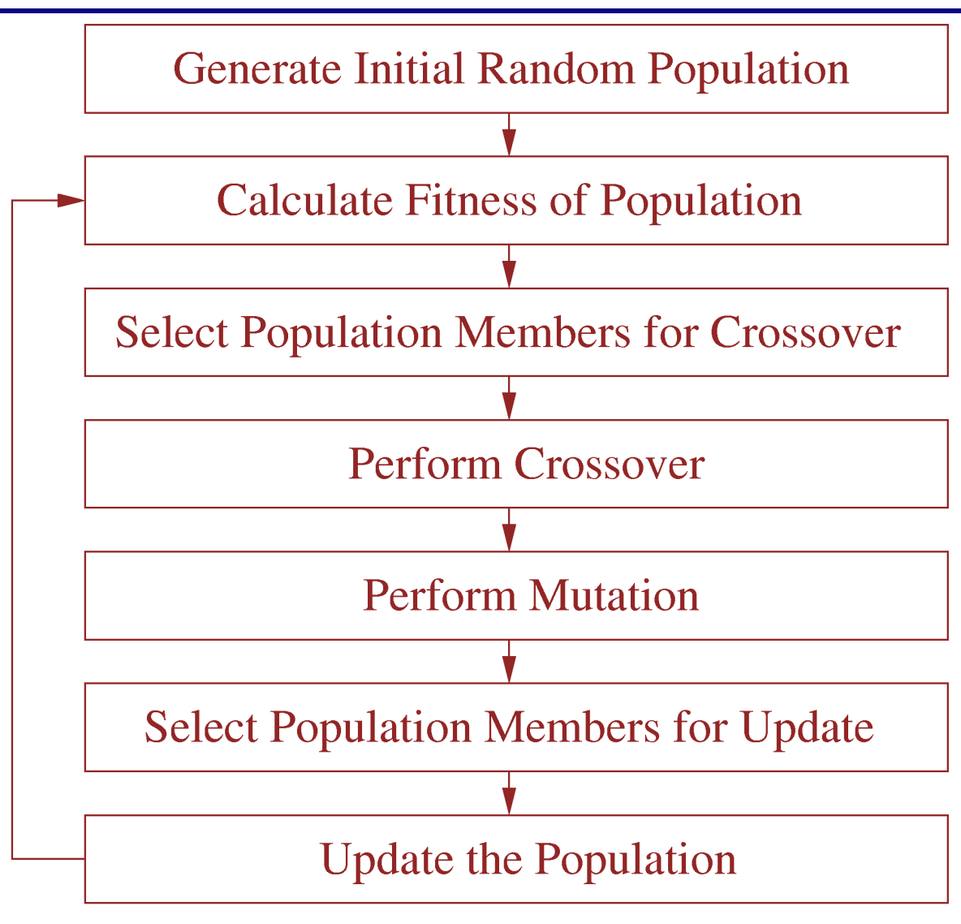

Figure 1: *A Genetic Algorithm*

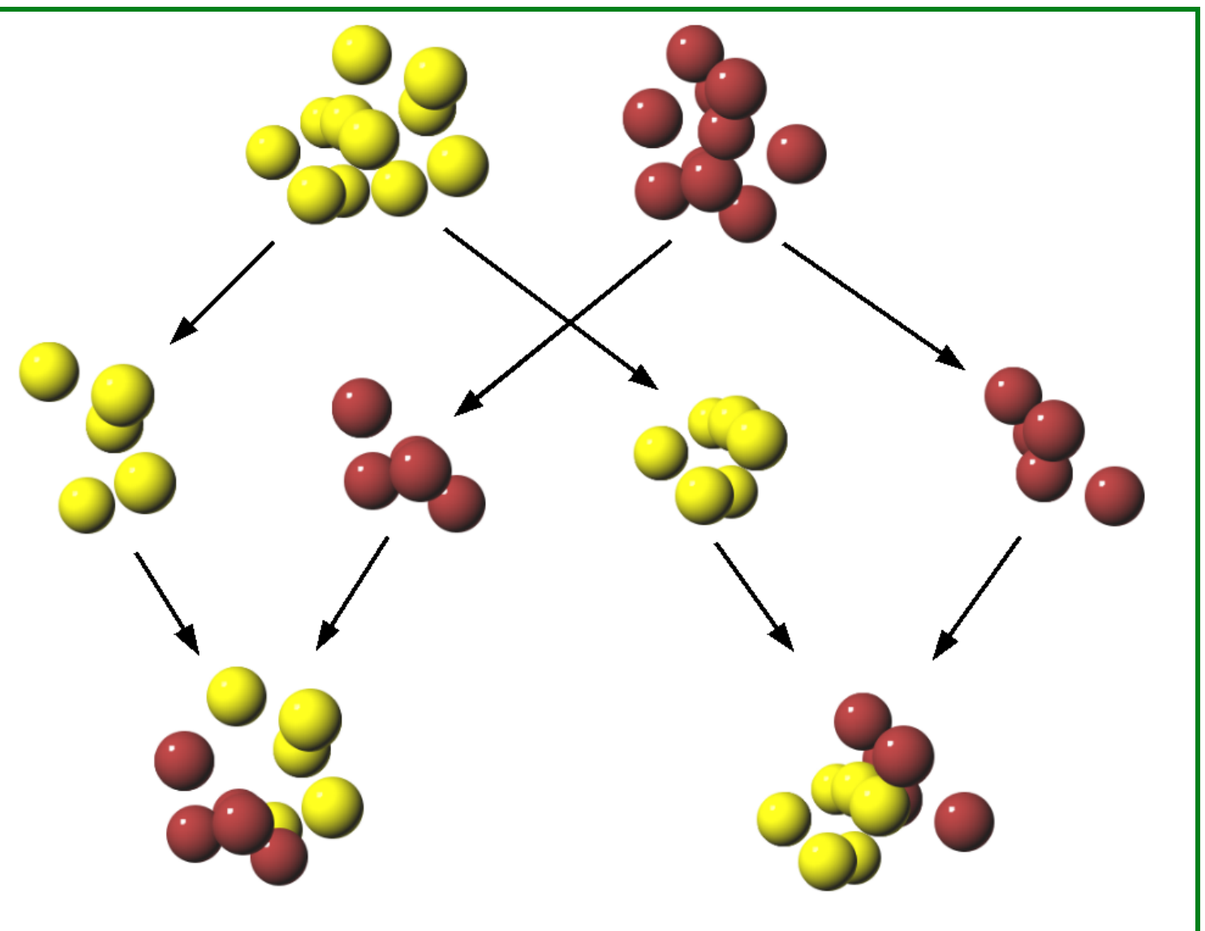

Figure 2: *Crossover*

- Typically, a GA requires a number of things: representation, calculation of fitness (dependent on the problem studied), selection for crossover (reproduction) and update, and how crossover and mutation are performed.
  - originally a binary string was used to represent a member of the population
  - crossover was performed by swapping sections of two (or more) parent strings to produce offspring
  - mutation was incorporated by flipping a percentage of bits in the string
- For physical systems, binary strings are not the most efficient representation [Deaven and Ho, 1995].
  - members are represented by an array of atomic position vectors
  - crossover is done by taking a plane through the centre of mass of the two parent structures and swapping the halves (see figure 2)
  - mutation and selection can be done in a number of methods
  - after crossover/mutation a direct minimiser is applied to relax the offspring
- Fitness is calculated from energy. "Fitter" members are more likely to be selected for crossover/update.

3 Crossover in Periodic Systems

- Most of the previous applications of GAs to atomic configurations have focused on non-periodic clusters [Deaven and Ho, 1995, Johnston, 2003].
- Only more recently have surfaces with periodic boundary conditions been studied using these methods [Chuang et al., 2004, 2005].

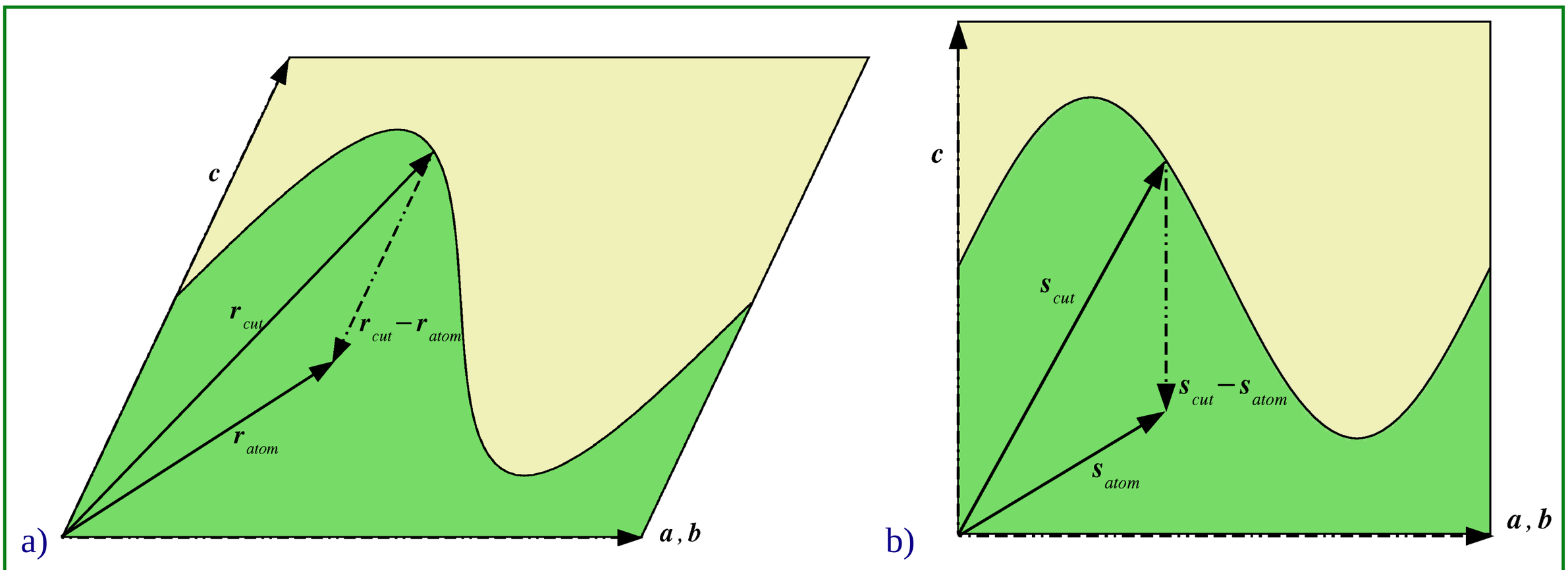

Figure 3: Diagram showing crossover in a) absolute representation and b) fractional representation

- We have developed a general method for crossover which can be used to periodic systems.
  - crossover is performed in fractional coordinates (see figure 3)
  - the cut is defined by any function with the same periodicity as the super-cell, $\mathbf{s}_{cut} = \mathbf{f}\left(\mathbf{s}_{atom}^{(\zeta,\eta)}\right)$, ($\mathbf{s}_{atom}^{(\zeta,\eta)}$ is the fractional position vector for each atom along the $(\zeta,\eta) = (\mathbf{a},\mathbf{b})$, $(\mathbf{b},\mathbf{c})$ or $(\mathbf{c},\mathbf{a})$ directions)
  - the metric tensor $\underline{\underline{g}} = \underline{\underline{h}}^{\mathrm{T}}\underline{\underline{h}}$ (where $\underline{\underline{h}} = [\mathbf{a},\mathbf{b},\mathbf{c}]$) is used to calculate the product α_{cut} (see equation 1) where $\mathbf{X} = \mathbf{c}$ when $(\zeta,\eta) = (\mathbf{a},\mathbf{b})$ etc.
  - α_{cut} is then used to generate halves to create offspring (see equation 2)

$$\alpha_{cut} = (\mathbf{s}_{cut} - \mathbf{s}_{atom})^{\mathrm{T}}\ \underline{\underline{g}}\,\mathbf{X} \tag{1}$$

$$\alpha_{cut} \begin{cases} > 0 & \text{the atom is } \textit{"above"} \text{ (outside) the cut} \\ \leq 0 & \text{the atom is } \textit{"below"} \text{ (inside) the cut} \end{cases} \tag{2}$$

3.1 Bulk Systems

- In this case two cuts are required to prevent a mismatch along the periodic boundary in the $(\zeta\,\eta)$-plane where $\left|\mathbf{s}^{\mathbf{X}}_{atom}\right| \equiv 0 \equiv 1$ (see figure 4).

- Each crossover operation has an equal probability of the cuts being made in reference to either $\mathbf{a}$, $\mathbf{b}$ or $\mathbf{c}$ directions.

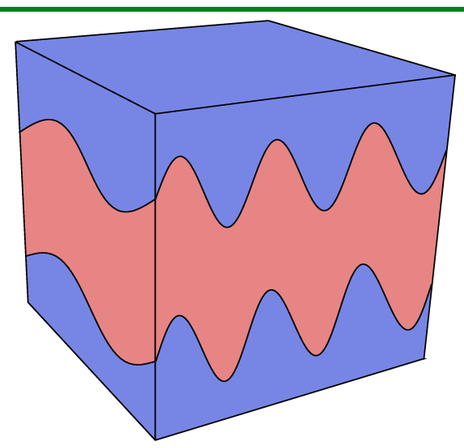

Figure 4: Crossover operation for bulk after crossover

4 Results

- All results presented were calculated using the *ab initio* planewave DFT code CASTEP [Segall et al., 2002] which had the Empirical Lennard-Jones potential added to allow the algorithm to be tested easily.
- Fitness is determined by the relative energies of the population members, and each population member was chosen for crossover based on its fitness using roulette wheel selection [Johnston, 2003].
- An atom is mutated with probability m_R by randomly moving its position inside a box of side $2m_A$ centred on the original atomic coordinate.
- In the update procedure, the fittest population member was allowed to proceed, with *Roulette Wheel* selection being used for the remaining members.
- Another method for updating is to only allow the M fittest population members to proceed to the next generation, from a super-population of $2M$ parents and offspring, called *Elitist* selection.
- The number of atoms could be kept fixed during the whole calculation, or could be allowed to vary after crossover.

4.1 Lennard-Jones Potential

This method has been applied to atoms acting under the empirical Lennard-Jones potential [Lennard-Jones and Ingham, 1925] in the shifted-force formulation [Stoddard and Ford, 1973]

- This has a HCP ground state structure [Pollack, 1964] which is almost degenerate with the FCC structure (energy difference from HCP is $< 0.1\%$ [Kane and Goeppert-Mayer, 1940]).
- Calculations were performed with 150 atom supercell and commensurate with both HCP and FCC
- Simulations started from a *random configuration of atomic positions* (see figure 5)
- We compared the efficiency the traditional planar cut with the periodic cuts proposed here.
- The mutation rate was kept fixed at 10% ($m_R = 0.10$), and the mutation amplitude at $m_A = 2.5$Å for all 120 simulations.

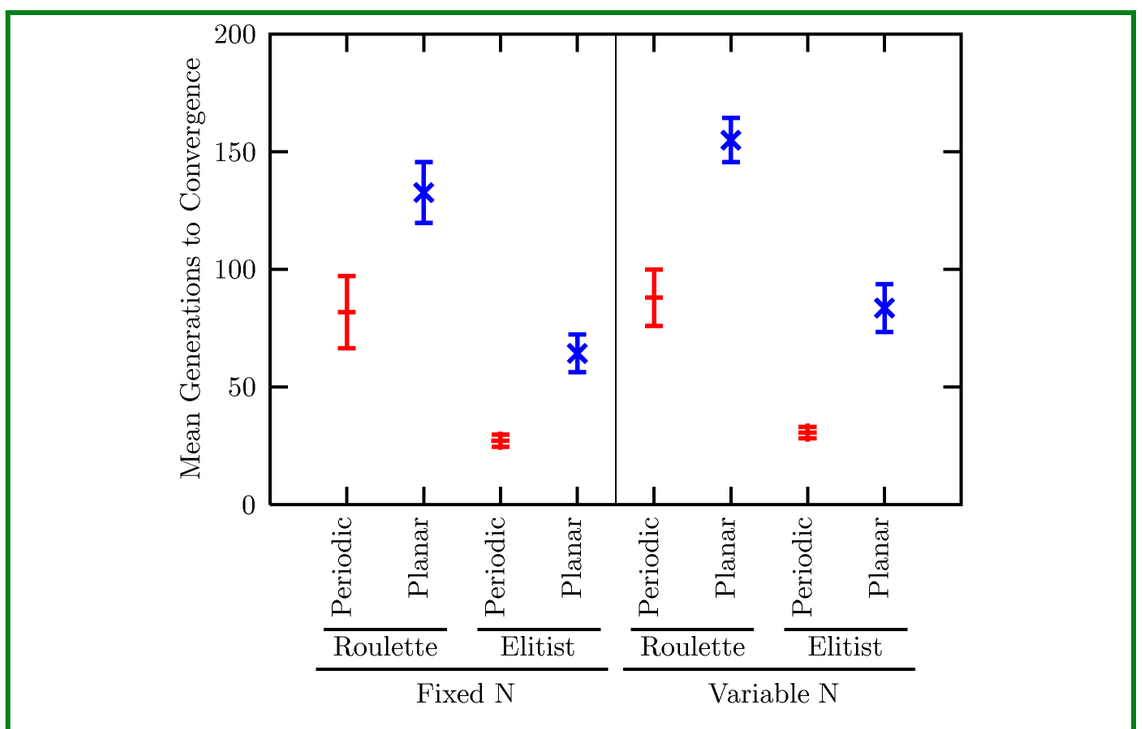

Figure 5: Summary of the 120 calculations performed under different conditions.

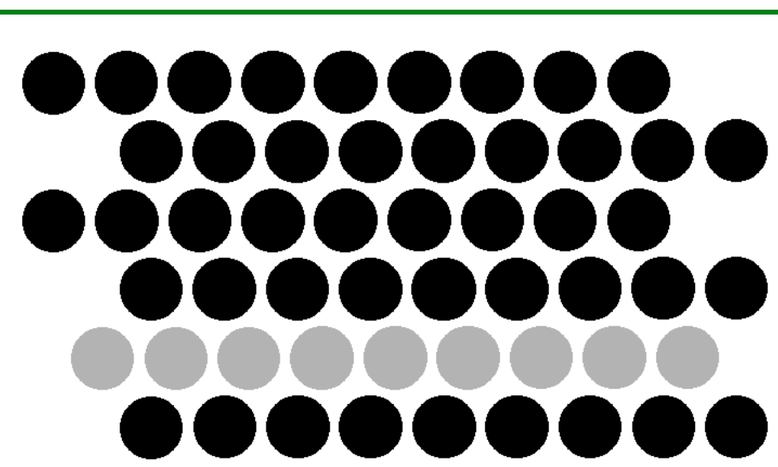

Figure 6: Minimised structure from one of the calculations shown summarised figure 5. This a HCP structure with one FCC stacking fault shown in grey. This was a variable atom number calculation made using periodic cuts in the crossover operation and using Roulette Wheel selection in the update procedure.

- None of the calculations performed in the fixed atom approach successfully minimised to an ideal crystal structure, instead the system froze into local minima.
- Allowing the atom number to vary improved convergence to ideal structures, although the simple update scheme with periodic cuts also froze structures into local minima.

4.2 *Ab Initio* Carbon Polymorph Search

- An alternate application of this GA as a tool for predicting the structure of different polymorphs.
- There has been recent interest in *ab initio* carbon polymorphs using systematic search techniques
  - sp^3-hybridised structures with four atoms per unit cell [Strong et al., 2004]
  - sp^2-hybridised structures with four or six atoms per unit cell [Winkler et al., 1999, 2001]
- Such calculations either require a graph-theoretical enumeration of possible configurations, resulting in an exponentially growing search space as the number of atoms per unit cell increases.
- A GA approach is not restricted to any particular form of bonding or chemical intuition, and is very efficient at searching high dimensional spaces.
- We chose to study the four atom per unit cell carbon polymorphs using density functional theory (DFT).
  - A planewave basis set was used
    - 400 eV cutoff
    - 0.05 Å^{-1} sampling of reciprocal space using the Monkhorst-Pack scheme [Monkhorst and Pack, 1976, Pack and Monkhorst, 1977]
  - Non-local ultrasoft psuedopotentials [Vanderbilt, 1990] were used to describe the electron-ion interaction
  - local density approximation (LDA) [Ceperley and Alder, 1980] was used to treat exchange correlation effects
  - The positions of the ions and the unit cell vectors were simultaneously optimised using the quasi-Newton method of Pfrommer et al. [1997]
    - reciprocal space sampling density and effective planewave cutoff energy maintained at all times
- After 10 generations there were two distinct types of structure
  - 5 were graphite-like (see figure 7)
  - 3 were Lonsdaleite-like (see figure 8)

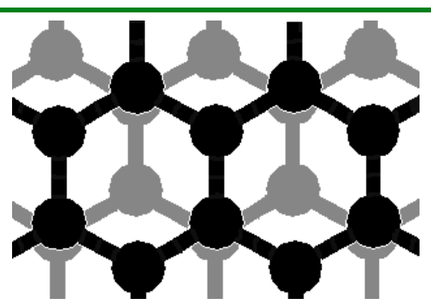

Figure 7: A graphite-like structure.

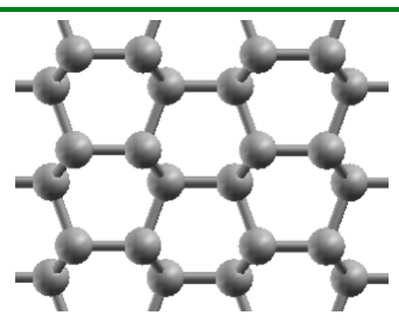

Figure 8: A Lonsdaleite-like structure.

5 Conclusions

By performing all crossovers in fractional coordinates each population member may be allowed to have a different unit cell, and if the local minimiser also optimises the cell vectors then the optimisation process will not be biased either by initial atom number or unit cell.

We have demonstrated a general method for optimising periodic systems using a genetic algorithm, which makes use of the inherent periodicity in the system in the calculation of crossover between parent members of the population. This method can also be used as a polymorph search technique. Future work will include surface calculations, multi-species calculations and phase transition searches.

For more information please see

A Periodic Genetic Algorithm with Real-Space Representation for Crystal Structure and Polymorph Prediction, N.L. Abraham and M.I.J. Probert, *Phys. Rev. B.* (2006) (submitted)

References

D. M. Deaven and K. M. Ho. Molecular-geometry optimization with a genetic algorithm. *Phys. Rev. Lett.*, 75: 288–291, 1995.

F. C. Chuang, C. V. Ciobanu, V. B. Shenoy, C. Z. Wang, and K. M. Ho. Finding the reconstructions of semiconductor surfaces via a genetic algorithm. *Surf. Sci.*, 573:L375–L381, 2004.

M. D. Segall, P. L. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne. First-principles simulation: Ideas, illustrations and the CASTEP code. *J. Phys.: Cond. Matt.*, 14(11):2717–2743, 2002.

R. L. Johnston. Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries. *Dalton Trans.*, pages 4193–4207, 2003.

B. L. Wang, S. Y. Yin, G. H. Wang, A. Buldum, and J. J. Zhao. Novel structures and properties of gold nanowires. *Phys. Rev. Lett.*, 86:2046, 2001.

F. C. Chuang, C. V. Ciobanu, C. Predescu, C. Z. Wang, and K. M. Ho. Structure of si(114) determined by global optimization methods. *Surf. Sci.*, 578:183–195, 2005.

J. Maddox. Crystals from 1st principles. *Nature*, 335:201, 1988.

A. VanDeWalle. Ab initio modelling - genesis of crystal structures. *Nature Materials*, 4:362, 2005.

V Blum, G. L. W. Hart, M. J. Walorski, and A. Zunger. Using genetic algorithms to map first-principles results to model Hamiltonians: Application to the generalized Ising model for alloys. *Phys. Rev. B*, 72:165113, 2005.

G. L. W. Hart, V. Blum, M. J. Walorski, and A. Zunger. Evolutionary approach for determining first-principles Hamiltonians. *Nature Materials*, 4:391, 2005.

John H. Holland. *Adaption in Natural and Artificial Systems*. MIT Press/Bradford Books Edition, 1992. ISBN 0-262-58111-6.

J. E. Lennard-Jones and A. E. Ingham. On the calculations of certain crystal potential constants, and on the cubic crystal of least potential energy. *Proc. Royal Soc.*, A107:636–653, 1925.

S. D. Stoddard and J. Ford. Numerical experiments on stochastic behavior of a Lennard-Jones gas system. *Phys. Rev. A*, 8:1504, 1973.

G. L. Pollack. The solid state of rare gases. *Rev. Mod. Phys.*, 36:748, 1964.

G. Kane and M. Goeppert-Mayer. Lattice summations for hexagonal close-packed crystals. *J. Chem. Phys*, 8: 642, 1940.

R. T. Strong, C. J. Pickard, V. Milman, G. Thimm, and B. Winkler. Systematic prediction of crystal structures: An application to sp(3)-hybridized carbon polymorphs. *Phys. Rev. B*, 70:045101, 2004.

B. Winkler, C. J. Pickard, V. Milman, W. E. Klee, and G. Thimm. Prediction of a nanoporous sp(2)-carbon framework structure by combining graph theory with quantum mechanics. *Chem. Phys. Lett.*, 312:536, 1999.

B. Winkler, C. J. Pickard, V. Milman, and G. Thimm. Systematic prediction of crystal structures. *Chem. Phys. Lett.*, 337:36, 2001.

H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:5188, 1976.

J. D. Pack and H. J. Monkhorst. "special points for Brillouin-zone integrations"-a reply. *Phys. Rev. B*, 16: 1748, 1977.

D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41: 7892, 1990.

D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45: 566, 1980.

B. G. Pfrommer, M. Cote, S. G. Louie, and M. L. Cohen. Relaxation of crystals with the quasi-newton method. *J. Comp. Phys.*, 131:233, 1997.