

## 7 SCIENTIFIC HIGHLIGHT OF THE MONTH: "Ab initio Random Structure Searching"

---

### *Ab initio* Random Structure Searching

Chris J Pickard<sup>1</sup> and R J Needs<sup>2</sup>

<sup>1</sup> Department of Physics and Astronomy, University College London,  
Gower St, London WC1E 6BT, United Kingdom

<sup>2</sup> Theory of Condensed Matter Group, Cavendish Laboratory, J J Thomson Avenue,  
Cambridge CB3 0HE, United Kingdom

#### Abstract

It is essential to know the arrangement of the atoms in a material in order to compute and understand its properties. The search for stable structures of materials using first-principles electronic structure methods, such as density functional theory (DFT), has grown rapidly in recent years. Here we describe our simple approach to searching for structures with DFT which we call *ab initio* random structure searching (AIRSS). Applications to discovering structures of solids, point defects, surfaces, and clusters are described.

#### 7.1 Introduction

Finding the most stable (lowest in energy or free energy) structure of a large assembly of atoms is a very difficult problem. The number of minima in the potential energy surface (PES) of a large system increases exponentially with the number of atoms. Finding the global minimum energy structure with certainty presumably involves visiting every local minimum and consequently the computational cost also increases exponentially with the number of atoms. This effectively prohibits an exact solution for large systems. Although the problem of structure prediction remains very difficult, steady progress has been made over the years. Advances in computing power, methods for calculating accurate energies of assemblies of atoms, and progress in searching methodologies has led to numerous successful predictions.

Predicting structure is important for a number of reasons. Structure prediction is relevant to all areas of science in which one would like to know the relative positions of atoms. Computational searching can be much easier and cheaper than experiments since a range of systems can quickly be searched, often obtaining interesting results and sometimes discovering promising new materials. The low-energy metastable minima are also interesting as they can be accessed at finite temperatures, or under pressure. Structures may also be trapped in metastable minima during growth or processing. Computational searches can augment experimental studies when the data is of poor quality or incomplete. For example, powder diffraction data may be insufficient for a

complete structural determination but may suffice to yield information such as the dimensions of the unit cell and an indication of its likely space group. The experimental data can then be used as constraints in a structural search. The positions of hydrogen atoms within a crystal cannot easily be determined from x-ray diffraction data, and here one can use the positions of the heavier atoms and the dimensions of the unit cell as constraints. Computational searches can also be used to investigate materials under conditions which cannot currently be accessed experimentally, for example, the pressures within the deep interiors of massive planets. Perhaps the most exciting possibility is the discovery of new materials in the computer which can be synthesised and have useful applications.

We have used our searching strategy, AIRSS, to predict stable and metastable structures of crystals and clusters and the atomic positions at point defects in solids, and we are beginning applications to surfaces and interfaces. Only fully quantum mechanical calculations suffice to deliver the required level of accuracy because of the wide range of inter-atomic bonding that may be encountered throughout the searches. We calculate the energetics using first-principles density-functional-theory (DFT) methods [1–3] which offer a high-level description of the electronic structure at a cost which is affordable for the many thousands of structures which must be considered in the course of a reliable search.

There is a rich literature on computational searching for structures. It is not our purpose here to review the entire field, although in Appendix A we briefly summarise other approaches to structure searching and give references to the literature. In this article we describe our preferred approach in detail, illustrating the discussion with a variety of examples.

## 7.2 Potential energy surfaces and the global searching problem

The exponential increase of the number of local minima with system size was derived and discussed by Stillinger [4]. The basic idea can be gleaned from the following simple argument. Suppose that a large system of  $N$  atoms can be divided into  $M$  equivalent subsystems, each of  $N/M$  atoms. If the subsystems are large enough they will have independent stable configurations. The total number of locally stable configurations of the system  $n_s$  therefore satisfies

$$n_s(N) = n_s^M(N/M) . \quad (1)$$

The solution to equation (1) is

$$n_s(N) = e^{\alpha N} , \quad (2)$$

where  $\alpha$  is a constant. Computational studies of Lennard-Jones clusters support the exponential dependence [5, 6].

The exponential increase in the number of local minima suggests that it will be very difficult to devise a reliable approach for finding the global minimum energy state of a large system. Perhaps clever methods can be found for eliminating the exponential scaling? Although it is not currently possible to give a definitive answer to this question, the prospects appear bleak. Determining the global minimum of a PES is classed as an NP-hard (non-deterministic polynomial-time hard) problem. These are problems for which it is widely suspected (but not proven) that it

is impossible to find an algorithm which works without fail in polynomial time. Reducing the strength of the exponential scaling (i.e., reducing the value of  $\alpha$  in equation (2)) is a more realistic goal, but theory also provides us with a warning about this. Wolpert and Macready have proved a “no free lunch theorem” for searching and optimisation which shows that all algorithms that search for the global minimum of an energy function perform exactly the same when averaged over all possible energy functions [7]. The implication is that it may be extremely difficult or even impossible to find a smart algorithm which works well in all circumstances.

We are interested in the energy functions which represent the PES of assemblies of atoms, and these form only a very small subset of all possible energy functions. Much of the PES of a reasonably large assembly of atoms corresponds to very high energy structures in which some atoms are much closer than an equilibrium bond length. This can readily be verified by calculating the energies of an ensemble of “random” structures, each formed by placing atoms at random positions within a box whose size gives a physically reasonable density. The average energy will be far higher than even the highest energy local minimum because of the strong short-range atomic repulsion. Other parts of the PES will correspond to fragmented structures. These may contain interesting energy minima, but if we are only interested in fully connected structures we can disregard them.

A basin of attraction of a PES is defined as the set of points for which downhill relaxation leads to the same energy minimum. A PES can therefore be divided into basins of attraction. Some rather general features of the PES of an assembly of atoms and its basins of attraction are known:

- (i) The substantial fraction of the PES in which some atoms are very close together contains almost no minima.
- (ii) The basins are normally arranged such that if one moves from a basin to a neighbour it is more likely that the neighbour will have a lower energy minimum if the barrier between the basins is small. This is a consequence of the relative smoothness of the PES at low energies and is related to the Bell-Evans-Polanyi principle which states that highly exothermic chemical reactions have low activation energies [8].
- (iii) Another implication of the Bell-Evans-Polanyi principle is that low energy basins are expected to occur near other low energy basins. Of course low energy basins can occur in widely separated “clumps”, which are normally referred to as “funnels”.
- (iv) Studies of Lennard-Jones clusters show that basins with lower energy minima tend to have larger hyper-areas than higher energy minima [9].
- (v) The probability distribution of the energies of the local minima of a PES is close to Gaussian for large systems, as would apply for the model which leads to equation (2).
- (vi) The hyper-areas of the basins of attraction appear to follow a power law distribution [10]. It seems that the power law behaviour must derive from some type of order in the arrangement of basins of different sizes, with smaller basins filling the gaps between larger ones [11]. The power law distribution does not occur in a simple model PES formed by arranging Gaussians of random width [11].
- (vii) Both very-low (and very-high) energy minima tend to correspond to symmetrical structures. This principle has been announced in many forms over the years and it is also supported by calculations [12, 13].

**(viii)** It has been observed that some space group symmetries are much more common than others in crystals formed from small organic molecules [14–16]. Inorganic systems show different space group frequencies [17, 18].

**(ix)** As well as general features of the PES of assemblies of atoms, there are particular features which arise from chemical considerations. In fact we normally know a great deal about the chemistry of the systems we study. We often know which atomic types prefer to bond to one another and the approximate lengths of the bonds, and the likely coordination numbers of the atoms.

### 7.3 Random Structure Searching

If nothing is known about the likely low-energy structures it is reasonable to start searching by relaxing random structures, which gives the widest coverage of the PES and an unbiased sampling. The notion of “random structures” is explored in Subsection 7.3.1, and it will turn out that we must impose limits on the initial structures for reasons of efficiency, so that our “random structures” might better be described as “random sensible structures”. Using random sensible structures is a useful approach which we have used successfully in several of the applications described in Section 7.6. The rather surprising degree of success of this approach derives from features **(i)**, **(iv)** and **(vi)** described in Section 7.2. These features imply that even random sampling has a good chance of finding low energy basins and that the wide coverage of the PES gives a chance of sampling the different “funnels” mentioned in **(iii)**. We exploit features **(vii)** and **(viii)** by imposing symmetry constraints as explained in Section 7.4. We make use of the proximity of low-energy basins of **(iii)** by “shaking” structures so that they fall into nearby minima, see Section 7.4. Following **(ix)**, we also make extensive use of chemical understanding of the system, as described in Section 7.4.

Our approach is very simple as it requires very few parameters and is very easy to implement. The biases are largely controllable, understandable, and based on sound principles. The searches run very efficiently on modern parallel computers. Our experience with the primitive method has been that we can perform highly reliable searches for the global minimum with up to at least 12 atoms (of one or two species) and sometimes more. When imposing constraints we can search successfully on much larger systems. Information from experiments, and chemical and structural information for the system in question or similar systems, and information generated by previous searches are combined to help design searches. The most successful approaches to searching are those which make the best use of the available information to bias the search towards finding the desired structures.

Our searches find many local minima, particularly if constraints are not imposed. As mentioned in Section 7.1, it is not only the ground state structure which is of interest, higher energy structures can also be important. For example, technologies such as molecular beam epitaxy (MBE) and Metal-Organic Chemical Vapour Deposition (MOCVD) allow controlled epitaxial growth of materials, which can result in structures far from equilibrium. Structure searching allows the discovery of many possible stable and metastable materials, which can then be ranked according to any property of interest such as the band gap or bulk modulus.

Random structure searching also teaches us chemistry! For example, we threw hydrogen atoms

(H) and oxygen atoms (O) in the ratio 2:1 into a box and relaxed, finding the most stable structures to consist of H<sub>2</sub>O molecules. Of course we expected this but, studying the higher-energy structures, we found other low-energy small molecules composed of H and O atoms [19].

### 7.3.1 Generating random structures

What do we mean by the term “random structure”? The arrangements of atoms in real materials are not at all random because the diameters of atoms and the bond lengths between them lie within a rather small range of roughly 0.75 to 3 Å. An assembly of atoms therefore has a “natural volume” which is proportional to the number of atoms present but only rather weakly dependent on the identities of the atoms and the external conditions. We start searches from fully-connected structures because separate fragments do not “see” each other and are unlikely to join up during relaxation. We adopt different procedures for generating initial structures for bulk solids, clusters and point defects in solids. Procedures can easily be devised for other purposes such as finding surface or interface structures, see Fig. 1.

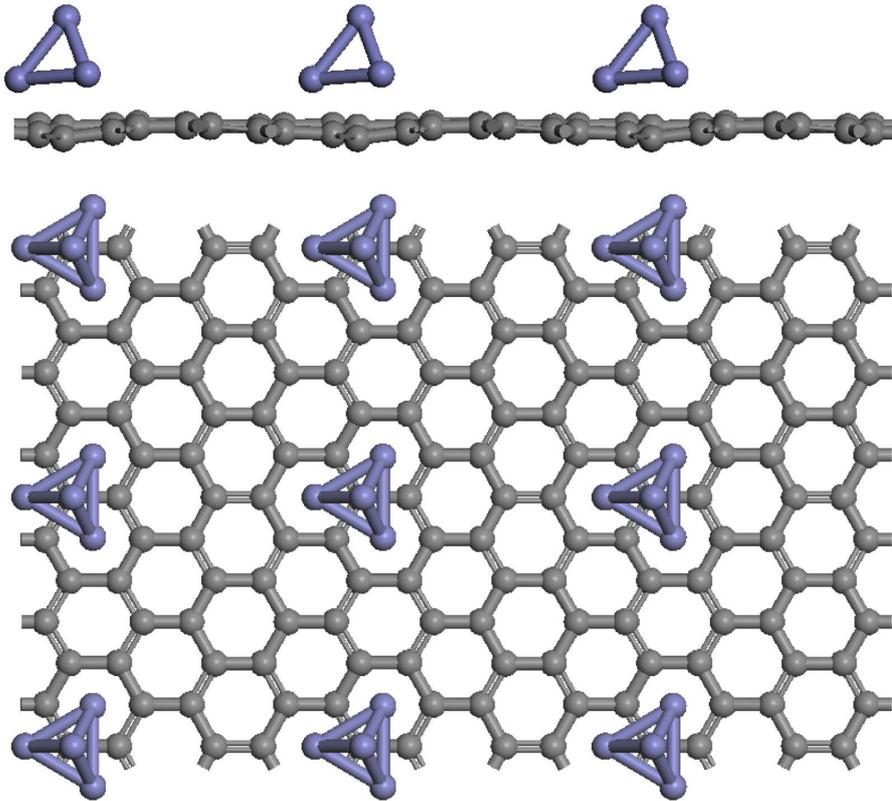


Figure 1: Iron clusters on graphene. A 24-atom supercell of graphene was set up and a (non-magnetic) four-atom iron cluster of random shape was placed at a random position on top of the graphene sheet, and all the atomic positions were relaxed. The lowest energy structure obtained after relaxing 69 structures is shown. Note the distorted tetrahedron of the iron cluster and how well the iron cluster matches the graphene lattice

**Periodic solids:** A random set of unit cell lengths ( $a, b, c$ ) and angles ( $\alpha, \beta, \gamma$ ) is chosen and the cell volume is renormalised to a random value within  $\pm 50\%$  (or thereabouts) of a chosen mean

volume. An appropriate mean volume can be determined from known structures composed of the same atoms, by adding up atomic volumes, or by relaxing a few “handmade” structures. The results are not very sensitive to the mean volume and range chosen. It turns out that a unit cell with very large or small angles can be transformed into an entirely equivalent unit cell with angles in the range  $60^\circ$ – $120^\circ$ . The more compact transformed cells are helpful for choosing efficient grids for Brillouin zone integrations and in visualising structures. We transform to more compact cells whenever possible.

**Clusters:** To generate initial structures for clusters we choose a box of a reasonable size to enclose the cluster and insert the atoms at random, as in a calculation for a periodic solid. We then place the box inside a considerably larger unit cell and impose periodic boundary conditions. An example of searching in clusters is described in Fig. 2.

**Point defects:** We start from a supercell of the perfect host crystal. In our work on defects in diamond-structure semiconductors (see Section 7.6) we have mostly used 32-atom supercells, although some defects may require larger cells. We remove a few neighbouring atoms from the crystal to make a “hole”, into which we place at random the desired host and impurity atoms.

**Keeping atoms/molecules apart:** Random structures may contain atoms which are very close together. Such occurrences are often harmless as the forces on the atoms are very large and they quickly move apart under relaxation. We have, however, sometimes encountered problems when transition metal atoms are nearly on top of one another which can make it very difficult to achieve self-consistency so that accurate forces cannot be obtained. A related problem occurs in searching for the structures of molecular crystals, where starting from randomly placed molecules can lead to unwanted chemical reactions. These difficulties can be avoided by rejecting starting structures in which atoms or molecules are too close. For very large systems the fraction of structures rejected will approach unity and a more efficient procedure should be used in which atoms or molecules are “nudged” apart.

## 7.4 Biasing the searches

**Choosing stoichiometries:** Does element A react with element B to form the compound AB, or perhaps  $A_2B$ , or  $A_2B_3$  etc., or is the compound  $A_2B_3$  unstable to the formation of  $A_2B + 2B$ , or  $2AB + B$ ? These questions can be answered by determining the energies of the most stable structures of each compound, which allows the thermodynamically most stable state of a mixture of A and B to be determined. This problem involves searching a larger space than is required for determining the most stable structure of a particular stoichiometry, but it can be tackled by carrying out structural searches for a range of stoichiometries. Searching with a particular stoichiometry may give hints about more stable stoichiometries as phase separation can occur within the unit cell. We have often noticed such behaviour although the limited size of the cells means that calculations with other stoichiometries and cell sizes may be necessary to unambiguously identify phase separation. An example of searches over different stoichiometries is described in Fig. 14. The first source of bias in studying a system is therefore the choice of stoichiometries.

**Choosing the number of units:** When searching for crystalline phases of a given stoichiom-

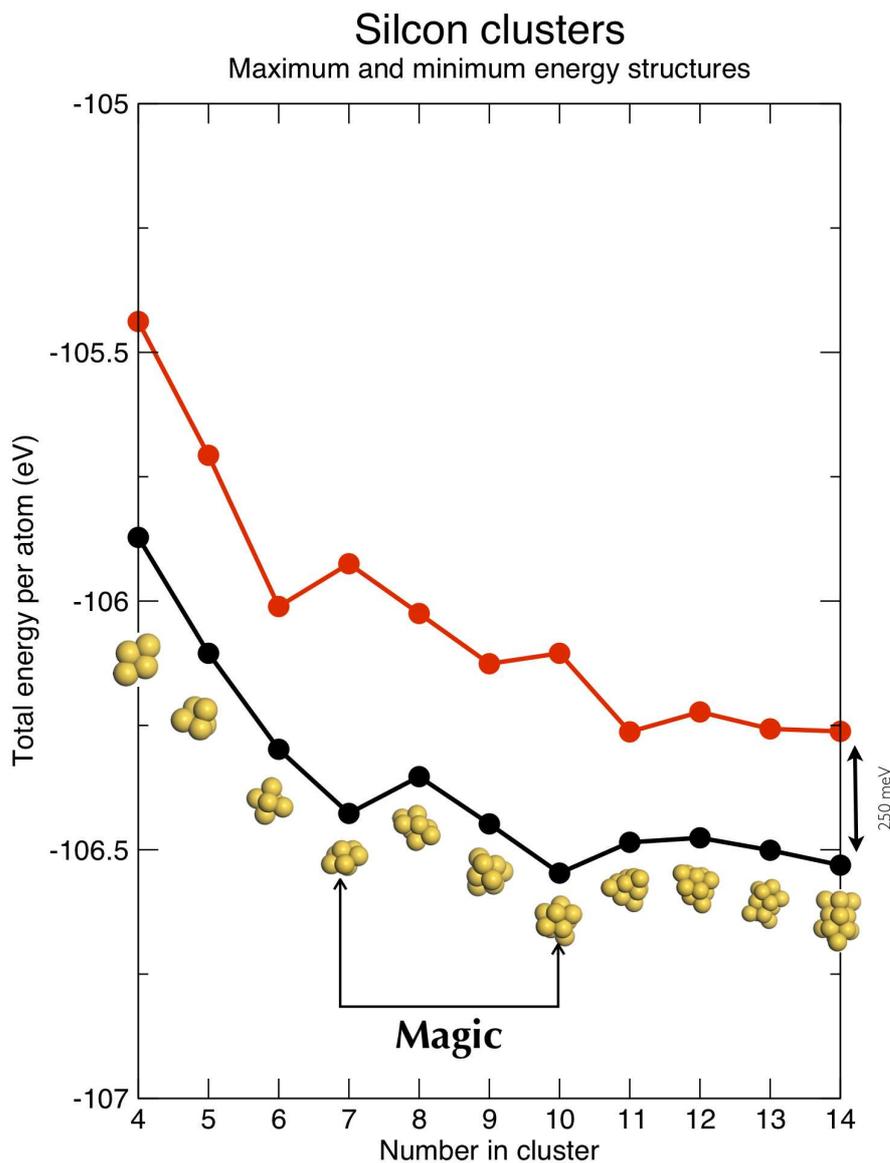


Figure 2: Silicon clusters were generated by placing atoms randomly within a small box inside a large unit cell and relaxing within DFT. The algorithm generated the same lowest-energy structures obtained in previous DFT studies [20], including the two “magic” number clusters with seven and ten atoms. We also found many local minima. The highest-energy minimum for each cluster size is only about 0.25 eV per atom higher in energy than the minimum energy structure

etry one does not *a priori* know how many formula units the primitive unit cell contains, and one should perform searches with different numbers of units. Searching using “usual” numbers of formula units, such as 2, 4, 6, and 8, will normally be an effective way to bias the search. However, it will preclude unexpected results, for example the 11 and 21 atom host-guest phases of aluminium (discussed in Section 7.6). We are fighting a computational cost which grows rapidly with system size and performing nearly exhaustive searches with more formula units rapidly becomes impracticable. Random structures are a perfectly reasonable starting point if one has no knowledge of the likely structures, but with a little thought one can often greatly

improve the efficiency of the search by biasing it towards finding low energy structures. This makes it possible to perform more comprehensive searches with larger numbers of atoms.

**Imposing chemical ideas:** Extensive knowledge of the chemistry of a system is often available, even if we know little about the actual structures which are favoured. Under these circumstances one can use chemical ideas to bias the searching. We already mentioned the idea of choosing initial structures composed of molecular units, and other examples of imposing chemical ideas are discussed in Section 7.3. Even if the system is non-molecular it is often possible to use chemical units to increase the efficiency of the search. For example, if one is interested in structures of gallium arsenide one can make initial structures from Ga–As units. This has the effect of making the densities of the Ga and As atoms much more uniform than a random structure, which becomes increasingly important for larger system sizes. Another important chemical idea is that of coordination number. For example, we can generate initial structures of carbon with  $sp^2$  bonding by creating random structures and rejecting all those which are not 3-fold coordinated, as illustrated in Fig. 3.

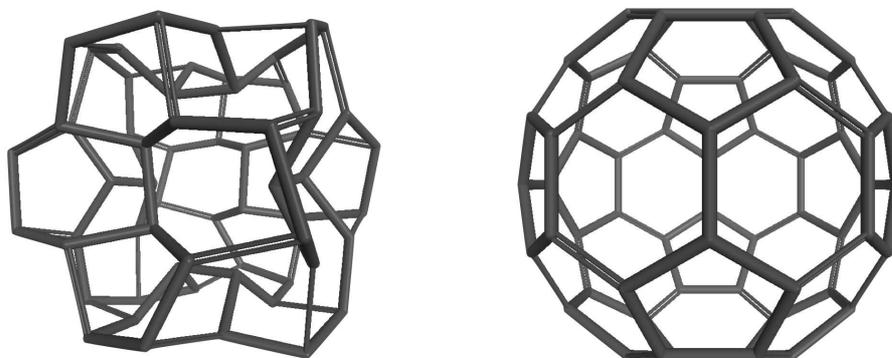


Figure 3: Left: A structure built by placing carbon atoms randomly within a small sub-box, subject to symmetry constraints. Random structures were generated and then screened to determine whether the atoms were three-fold coordinated. If not, the structure was rejected and another one was generated. Right: relaxation of this structure within DFT gave the well-known  $C_{60}$  "buckyball"

**Imposing symmetry:** As noted in Section 7.2, minima with very low or very high energies tend to correspond to symmetrical structures. Imposing a degree of symmetry on the initial structures and maintaining it during relaxation therefore eliminates a large amount of the PES while (hopefully) still allowing the global minimum energy structure to be found. We implement this strategy by searching randomly over all space groups with  $N_s$  symmetry operations. Such a search also allows structures to relax into space groups which are super-groups of those with  $N_s$  symmetry operations. Symmetry constraints have often been used in searching for crystalline polymorphs composed of small molecules such as the drug molecules developed within the pharmaceutical industry [21].

**Using experimental data:** We already mentioned the possibility of using experimental data to bias the searching in Sections 7.1 and 7.2. It may turn out that a powder diffraction spectrum is obtained with quite a few well defined peaks which, however, are insufficient for a full structural determination. In such cases it is often possible to determine the dimensions of the unit cell

and perhaps an indication of the most likely space groups from the data. Such information is extremely useful when performing a structural search, and an example of this type of constrained search is described in Section 7.6 for a high-pressure phase of ammonia monohydrate, and a test calculation for a dipeptide is illustrated in Fig. 4. Knowledge of the different space group frequencies, which we mentioned in Section 7.2, could also be used to bias searches.

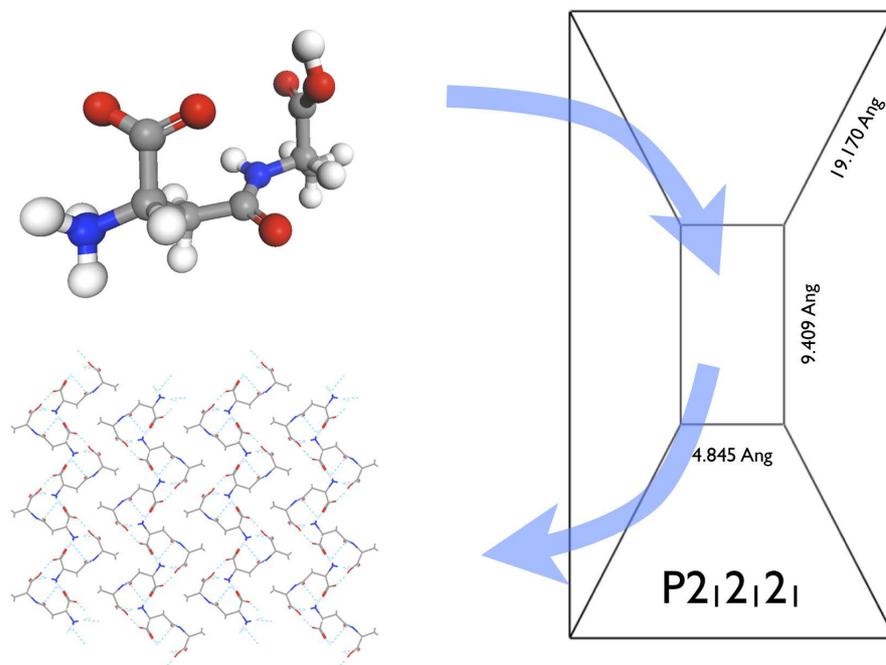


Figure 4: The crystal structure of the beta-L-aspartyl-L-alanine dipeptide is known experimentally. In this test we made structures from the experimental unit cell shown on the right and the  $P2_12_12_1$  space group of the crystal and the structure of the beta-L-aspartyl-L-alanine molecule (top left). Carbon atoms are shown in grey, oxygen in red, nitrogen in blue and hydrogen in white. A single molecule was placed randomly in the unit cell and the positions of the other three molecules were determined by the space group symmetry. The structure was rejected if two molecules overlapped and a new one was generated. Each non-overlapping structure was relaxed within DFT while maintaining the size and shape of the unit cell and the  $P2_12_12_1$  symmetry. The correct molecular packing (bottom left) was found after relaxing 18 structures

**Shaking:** In Section 7.2 we encountered the idea that low energy basins may be clustered together. This motivates the “shake”, a random displacement of the atoms and, if appropriate, a random adjustment of the unit cell. Atomic displacements of a large fraction of a bond length have a reasonable chance of pushing the system into a nearby basin of attraction. We have also used shaking to look for distortions of structures into doubled (or larger) unit cells. The shake is the same as a step in the basin hopping algorithm [22–24] (see also, Appendix A), although we have used it only with zero temperature and after considerable searching has already located low-energy structures.

A related idea is to calculate the harmonic phonon modes of a structure. The phonon modes at zero wave vector of a fully relaxed structure found from random searching must be stable, and the structure must also be stable against elastic distortions. The phonon modes at non-zero

wave vectors may, however, be unstable, so that the energy can be reduced by a distortion in a larger unit cell. Calculating the second derivatives of the energy to obtain the phonon frequencies and displacement patterns is expensive and we only perform such calculations on a few structures of interest after extensive searching. If unstable phonon modes are found then the energy-reducing distortions of the corresponding phonon eigenvectors can be followed to find more stable structures.

#### 7.4.1 Have we found the global minimum?

The searching is not exhaustive and therefore we cannot be sure that we have found the global minimum. One way to gauge the quality of a search is to look for known “marker” structures (if available). We happily terminate searches when the same lowest-energy structure has been found several times. This criterion is reasonable because we relax a very wide range of initial structures. When we apply constraints to the initial structures and maintain them during the relaxation we obviously cannot obtain structures which violate the constraints. When we apply constraints to the initial structures but allow free relaxation we are biasing the search, presumably towards structures which obey the initial constraints, but also perhaps in ways which we cannot predict. When we bias a search it is important to understand as well as possible which parts of the PES are being excluded or de-emphasised. This allows the user to assess the strengths and weaknesses of a search and, if required, to design further searches. It is therefore important that the effects of the “knobs” of the search (the parameters which can be varied) are as transparent as possible. We believe that the simplicity of our searching procedures results in a relatively small number of understandable and useful knobs.

#### 7.5 Some technical aspects of the calculations

**First-principles DFT calculations:** DFT calculations are much more expensive than empirical potential ones and the number of structures whose energies may be evaluated is therefore greatly fewer. Many first-principles DFT codes are available, and we use the CASTEP package [25] which uses a plane wave basis set, periodic boundary conditions, and pseudopotentials. The code returns the total energy of a structure and the forces on the atoms and stresses on the unit cell. We use the forces and stresses to relax structures to the nearest local minimum in the PES. The second derivatives of the energy may also be calculated, but this is much more expensive and although very useful in checking for unstable phonons/elastic distortions and in calculating thermal effects in stable structures, it is far too expensive to be used routinely as part of the search strategy.

**Pseudopotentials:** Accurate results at very high pressures can be obtained using pseudopotentials, but they must be constructed with sufficiently small core radii and with the appropriate electrons treated explicitly. The pseudopotentials provided with standard codes may be inadequate at the high pressures we often work at. Lithium is an unusually difficult case. It is standard to treat all three electrons of lithium explicitly, but the pseudopotential core radii must still be small [26] in high-pressure studies. We use ultrasoft pseudopotentials [27] and find them to be accurate when the distance between neighbouring atoms is about equal to or

greater than the sum of the core radii of the atoms. We recommend that pseudopotentials be tested for each application by generating them with different core radii and checking that energy differences are accurate for the shortest inter-atomic distances that will be encountered. For some of the applications described in Section 7.6 we have treated some core and semi-core states explicitly. For example, we used pseudopotentials with 11 electrons treated explicitly for our work on aluminium [28] and 16 electrons for iron [29].

**k-point sampling:** We use quite good Brillouin sampling and basis sets when searching because we find that poor quality calculations can lead to strong biases. We have come across modulated phases when searching in metals which went away when we relaxed them further with denser k-point sampling. We use Monkhorst-Pack (MP) meshes of k-points which are defined by choosing the smallest MP mesh where the smallest separation between k-points is less than some distance  $\Delta k$ . We often use  $\Delta k = 2\pi \times 0.07 \text{ \AA}^{-1}$  when searching and then perhaps  $\Delta k = 2\pi \times 0.03 \text{ \AA}^{-1}$  when refining the structures and their energetics. We deform the k-point mesh with the changes in the cell shape and occasionally recalculate the integer parameters of the MP mesh.

**Predicting stability over a range of pressures:** In our high-pressure studies we search at constant pressure, although one can just as easily search at constant volume. A search at pressure  $p_s$  may give many different structures. The structure with the lowest enthalpy  $H(p_s)$  is the most stable at  $p_s$ , but different structures may be more stable at another pressure  $p$ . To investigate this we can use the thermodynamic relation

$$H(p) \simeq H(p_s) + (p - p_s) \left. \frac{dH}{dp} \right|_{p_s} + \frac{1}{2} (p - p_s)^2 \left. \frac{d^2H}{dp^2} \right|_{p_s} \quad (3)$$

$$= H(p_s) + (p - p_s)V_s - \frac{1}{2} (p - p_s)^2 \frac{V_s}{B_s}, \quad (4)$$

where  $V_s$  and  $B_s$  are the volume and bulk modulus at  $p_s$ . Equation (4) can be used to approximate the enthalpy of each structure over a wide range of pressures. The simple linear approximation

$$H(p) \simeq H(p_s) + (p - p_s)V_s \quad (5)$$

is particularly convenient because the quantities required ( $H(p_s)$ ,  $p_s$ ,  $V_s$ ) are obtained directly from the search calculations. Suppose we apply equation (5) to two structures, A and B, found at  $p_s$ . Equation (5) tells us that if  $V_s^A < V_s^B$  then structure A will become more favourable with respect to structure B at  $p > p_s$  and less favourable for  $p < p_s$ . If B is more stable than A at  $p_s$  a phase transition from B to A could occur at some  $p > p_s$ . It may be possible to use the quadratic form of equation (4) with an empirical relationship between the bulk modulus and volume, but we have not explored this further. We find, however, that the linear approximation is very useful for estimating the stability regions of different phases over a wide range of pressures, which gives the approach a ‘‘far sightedness’’. Equations (4) and (5) can be used in both constant volume and constant pressure calculations, but we normally use a simple scatter-diagram representation in our constant pressure calculations, as explained in Fig. 5.

## 7.6 Brief survey of AIRSS calculations to date

**Silane:** In our first AIRSS paper we studied high pressure phases of silane ( $\text{SiH}_4$ ) [31]. This group IVB hydride is a metastable compound under ambient conditions, but above about 50 GPa

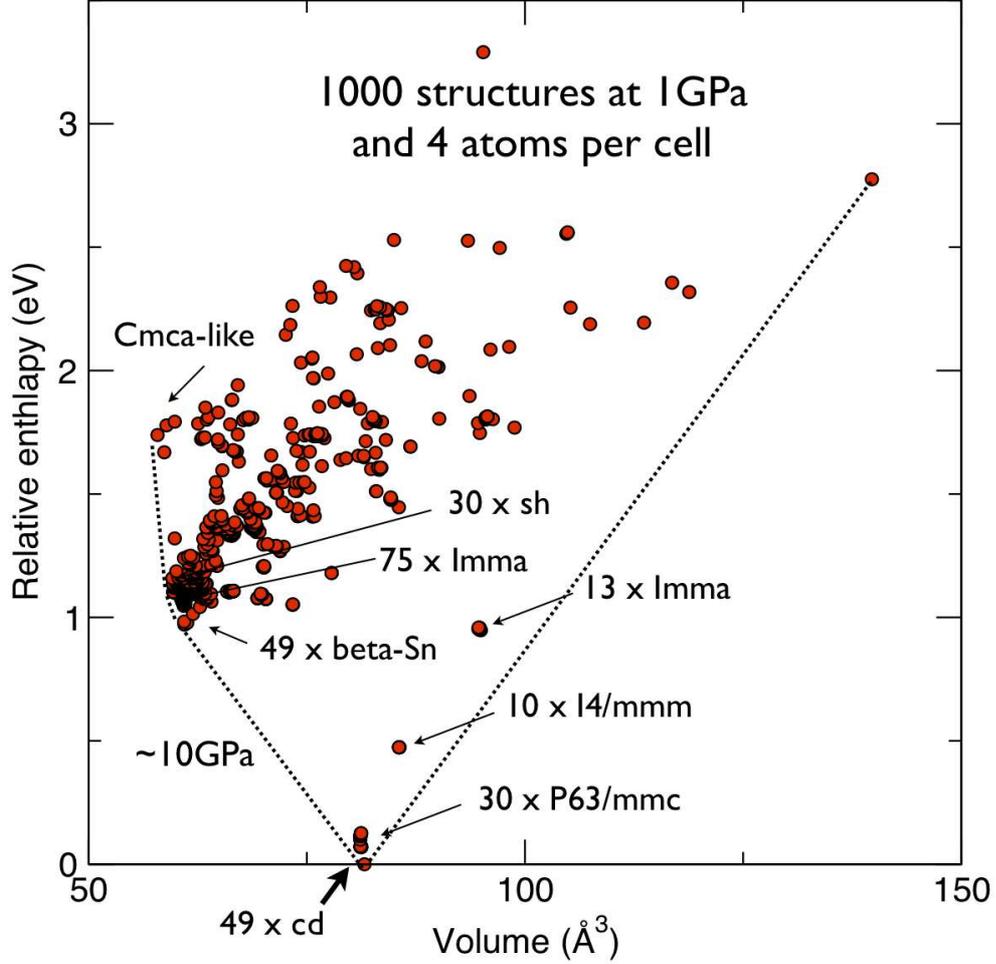


Figure 5: Scatter plot of the relative enthalpies against volume for a search at  $p_s = 1$  GPa with 4 silicon atoms per cell. The diamond structure (cd) is the most stable at this pressure and was found 49 times from a total of 1000 relaxed structures. The positions of the observed high-pressure phases [30], beta-Sn, *Imma*, sh (simple hexagonal) and *Cmca*-like, are also indicated on the figure. The hexagonal-close-packed (hcp) and face-centred-cubic (fcc) phases which are observed in experiments at pressures beyond the *Cmca* phase [30] were not found in the searches and we presume they are mechanically unstable at 1 GPa. Equation (5) shows that the stable phases can be found by drawing lines underneath the data points as shown in the figure. The stable phases at pressures greater than 1 GPa can then be read off the figure as those through which the dotted lines pass, and it can be seen that these are the experimentally observed ones at positive pressures where the dotted line has a negative slope. The slope of the line joining the cd and beta-Sn phases corresponds to a pressure of about 10 GPa, which is similar to the coexistence pressure [30]. The phases above the dotted lines are not the most stable at any pressure. The *P63/mmc* phase differs from cd only in the stacking of layers

it becomes stable against decomposition into its elements. Our work was motivated by a theoretical study [32] which used chemical intuition to predict interesting high pressure non-molecular phases of silane. We found more-stable phases, most notably an insulating phase of  $I4_1/a$  symmetry, shown in Fig. 6, which was the most-stable structure from about 50 GPa to over 200 GPa.

Each Si atom is bonded to eight H atoms which form bridges between neighbouring Si atoms. Each of the Si and H sites are equivalent in this high-symmetry structure. All of the bonds are electron-deficient three-centre-two-electron “banana” bonds, similar to those linking the boron atoms in diborane ( $B_2H_6$ ). Interestingly, Feng *et al* [32] predicted structures with some Si-H-Si banana bonds, and their chemical intuition was essentially correct, but our structure is totally bananas. The  $I4_1/a$  phase has subsequently been observed in x-ray diffraction studies [33] and its insulating behaviour was verified. We also found a slightly-less-stable phase of  $I\bar{4}2d$  symmetry only 0.1 eV per  $SiH_4$  unit above  $I4_1/a$  at 100 GPa. The  $I\bar{4}2d$  phase of silane has also been identified in experiments by Degtyareva *et al* [34]. An impressive debut for AIRSS!

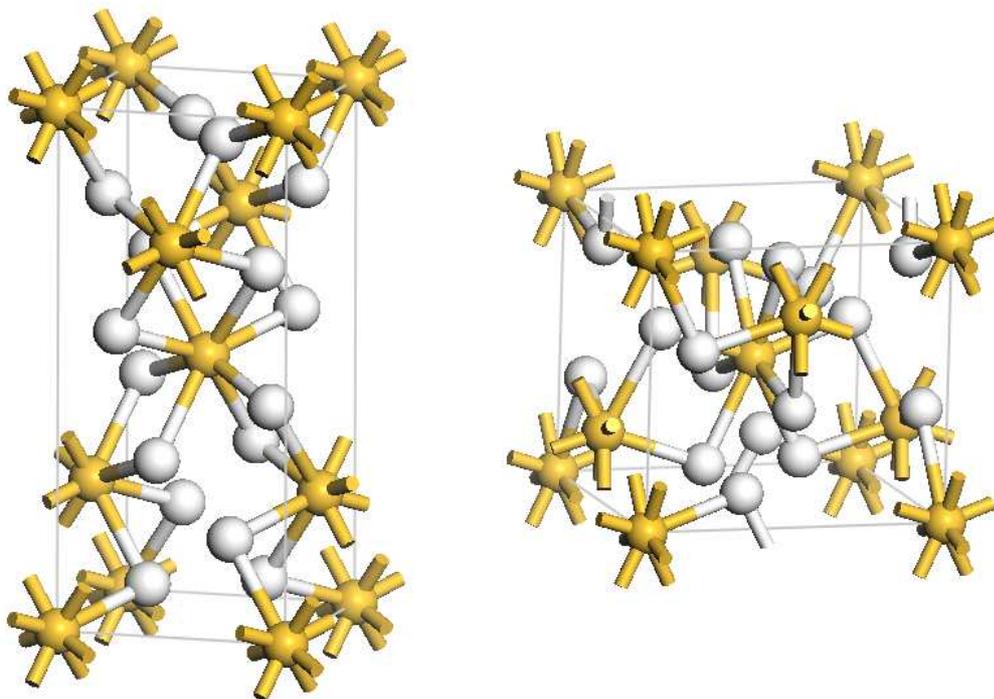


Figure 6: The  $I4_1/a$  structure of silane (left) and the slightly less stable  $I\bar{4}2d$  structure (right). Silicon atoms are shown in gold and hydrogen atoms are in white. All of the bonds in  $I4_1/a$  and  $I\bar{4}2d$  are of the Si-H-Si type. Both phases were subsequently found experimentally

**Aluminium hydride:** The silane studies were motivated by the quest for metallic hydrogen. Although metallic hydrogen has been formed fleetingly in shock wave experiments and must exist within planets such as Jupiter, it has not been produced in static compression experiments, where it could be studied in detail. Hydrides have been thought of as containing “chemically pre-compressed” hydrogen which might become metallic at pressures achievable in diamond anvil cells and might exhibit phonon-mediated high-temperature superconductivity [35]. The group IVB hydrides contain 80% hydrogen atoms, but the group IIIB hydrides contain nearly as much (75%). We studied aluminium hydride ( $AlH_3$ ) and predicted the stability of a metallic  $Pm\bar{3}n$  phase at pressures readily achievable in diamond anvil cells [36]. The structure of the  $Pm\bar{3}n$  phase is illustrated in Fig. 7. Hydrogen atoms are considerably more electronegative than aluminium ones, so the electron density on the hydrogen atoms is large, which suggests that the high-frequency hydrogen-derived phonon modes could provide substantial electron-phonon

coupling and promote superconductivity. However, the  $Pm\bar{3}n$  phase of  $\text{AlH}_3$  is a semimetal at the transition pressure with a relatively small electronic density of states at the Fermi energy, which strongly militates against superconductivity.  $Pm\bar{3}n$  develops a band gap on further compression but, on the other hand, reducing the pressure increases the density of states at the Fermi energy which would promote superconductivity. The semi-metallic  $Pm\bar{3}n$  phase was subsequently observed in high-pressure x-ray diffraction experiments [37], but it was not found to be a superconductor.

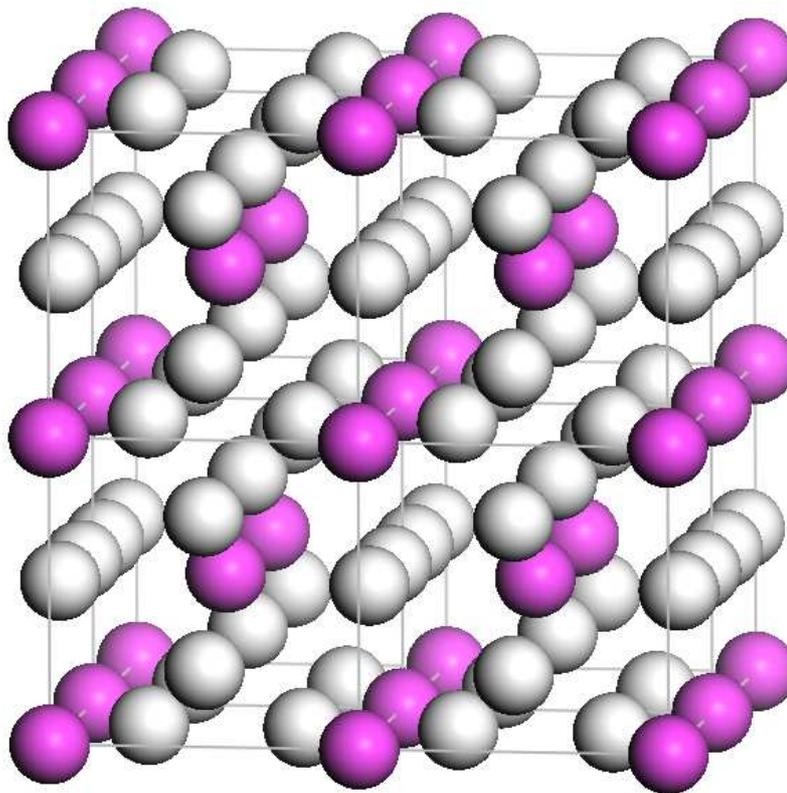


Figure 7: The  $Pm\bar{3}n$  phase of aluminium hydride. The Al cations are shown in purple and the H anions are in white. The linear chains of H atoms can clearly be seen. This structure is also adopted by niobium stannide ( $\text{Nb}_3\text{Sn}$ ) which is a superconductor used in high magnetic field applications

**Hydrogen:** Pure hydrogen has been compressed to over 300 GPa in a diamond anvil cell [38], but it stubbornly remains insulating. It is expected that a non-molecular and presumably metallic phase will become stable somewhere in the range 400-500 GPa [39], and such pressures will probably be achieved in static experiments in the near future. The metallic phase is expected to be a high-temperature superconductor, perhaps even a room-temperature superconductor. The structure of the low-pressure phase I of solid molecular hydrogen is well established [40]. Phase II is stable above 110 GPa, and probably consists of molecules arranged on a distorted close-packed lattice, and a molecular phase III of unknown structure appears above 150 GPa. Our AIRSS studies [41, 42] have shown there to be several candidate structures for phase II consisting of packings of molecules on distorted hexagonal-close-packed lattices. These structures are almost degenerate in enthalpy and quantum motion of the protons could mean that several significantly different local molecular configurations contribute to the overall structure of phase

II. Prior to our work, the DFT phase diagram showed a transition to a metallic phase below 200 GPa, in strong disagreement with experiment. We predicted new insulating molecular phases which are stable up to pressures well above 300 GPa. In particular, the predicted vibrational properties of our  $C2/c$  molecular phase (which has 24 atoms in the primitive unit cell and is shown in Fig. 8) agree with the available experimental data for phase III [41].

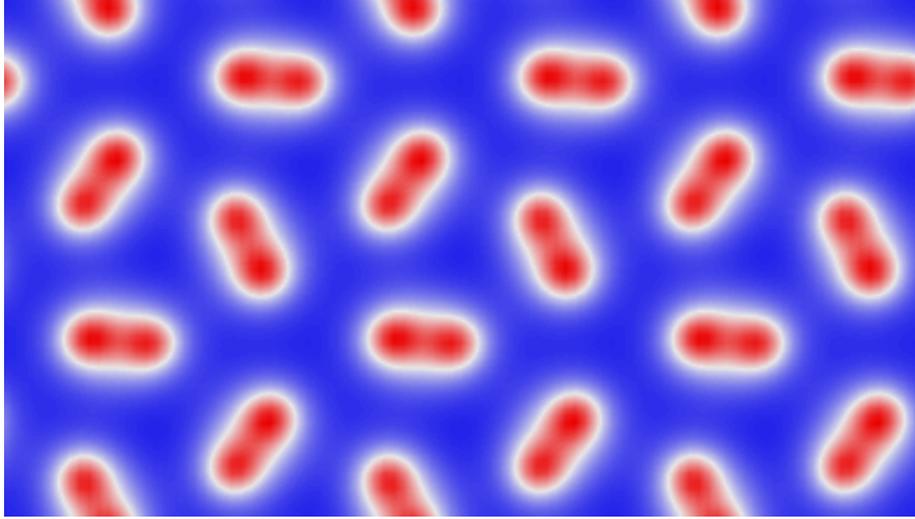


Figure 8: A slice through the charge density of a layer of the  $C2/c$  molecular hydrogen phase which we predicted to be the most stable in the pressure range 105–270 GPa [41]. Note that the two ends of the molecules are inequivalent so they have dipole moments and the crystal has infra-red (IR) active vibron modes. The calculations show intense IR vibron activity with strong absorption peaks which are close in frequency and would appear as a single peak in experiments [41]. The IR activity of the strong IR active vibrons in  $C2/c$  increases with pressure, as is observed in phase III [43]. The variation with pressure of the strong IR peak and the Raman active vibron frequency of  $C2/c$  are in good agreement with experiment [44]

**Nitrogen:** The phase diagram of nitrogen has been much studied, with a number of apparently stable and metastable molecular phases having been reported [45–47], although their structures are mostly unknown. We found a new class of molecular structures which we predicted to be more stable than previously suggested ones over a wide range of pressures [48]. The dissociation energy of a nitrogen molecule is more than twice that of a hydrogen molecule, and yet nitrogen molecules dissociate at far lower pressures [49]. The reason for this is simply that nitrogen atoms can form up to three covalent bonds so that molecular, polymeric and dense framework structures are possible, whereas a hydrogen atom can form only one covalent bond. “Polymeric” nitrogen can in fact be recovered to ambient conditions as a metastable high-energy-density material [50]. The structure of the high-pressure singly-bonded “cubic gauche” phase formed on molecular dissociation was in fact predicted using DFT calculations [51] over a decade before it was observed experimentally [50], a triumph for chemical intuition. Computational searches for the phases beyond cubic gauche have also been performed [48, 52]. Ma *et al* [52] used DFT and a genetic algorithm to predict the phase beyond cubic gauche to be a singly-bonded layered structure of  $Pba2$  symmetry with 16 atoms in the primitive unit cell. This structure is slightly more favourable than the very similar  $P\bar{4}2_1m$  structure we found with 8 atoms. Unfortunately

we did not perform searches with more than 12 atoms, so we could not have found the  $Pba2$  phase. This serves as a warning to all searchers - there could always be a better structure in a larger unit cell.

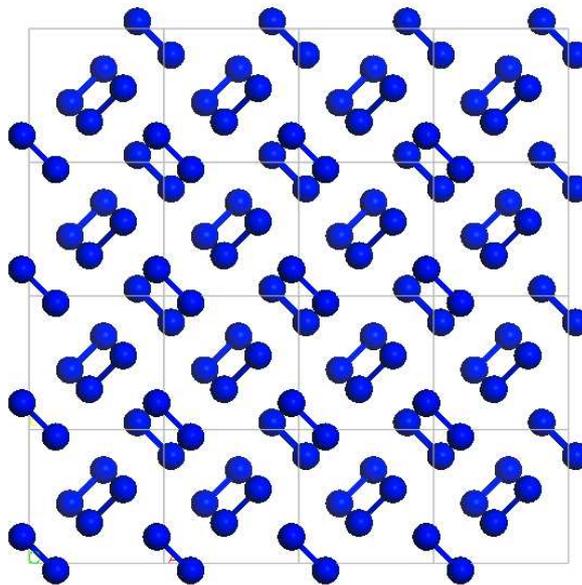


Figure 9: The  $P4_12_12$  molecular phase of nitrogen which we predict to be the most stable from  $\sim 9.5$  GPa up to molecular dissociation at about 56 GPa [48]

**Water:** Our work on structures of  $H_2O$  [19] was motivated by an experimental study [67] in which a new metastable form of  $H_2O$  was synthesised. Mao *et al* subjected water to an applied pressure of about 20 GPa and 10 keV x-ray radiation for many hours within a diamond anvil cell, producing a crystalline phase which does not consist of water molecules. Mao *et al* [67] concluded that they had synthesised an alloy of  $O_2$  and  $H_2$  molecules. We performed a AIRSS study at 20 GPa, finding that the structures obtained consisted almost entirely of weakly bonded  $H_2O$ ,  $H_3O$ ,  $H_2O_2$ ,  $H_2OH \cdots OH$ ,  $H_2$ , and  $O_2$  species. O–H bonds are the most energetically favourable at 20 GPa, so that the most stable phases consist of  $H_2O$  molecules and the highest enthalpy metastable phases consist of an “alloy” of  $H_2$  and  $O_2$  molecules (rocket fuel!). We argued [19] that the experimental x-ray diffraction, energy loss, Raman spectroscopy and other data were best rationalised not by an  $H_2/O_2$  alloy but by a much more stable mixture of  $H_3O$ ,  $O_2$  and  $H_2$  species, no doubt containing amounts of the other low-enthalpy species.

**Ammonia:** Compressed ammonia ( $NH_3$ ) plays a significant role in planetary science. Ammonia forms hydrogen-bonded solids at low pressures, but we predict that at high pressures it will form ammonium amide ionic solids [53]. These structures, consisting of alternate layers of ammonium cations ( $NH_4^+$ ) and amide ( $NH_2^-$ ) anions are expected to be stable over a wide range of pressures readily obtainable in diamond anvil cells, although experimental verification of our prediction is still lacking. The ionic  $Pma2$  phase, which is illustrated in Fig. 10, is predicted to be stable above 90 GPa. The driving force for the proton transfer reaction is that the ionic solid is substantially denser than the molecular one. The proton transfer costs energy under ambient conditions, but at high pressures the cost is overcome by the lower value of the  $pV$  term in the enthalpy. A proton transfer between water molecules, forming  $OH^-$  and  $H_3O^+$  ions, costs

more energy than in ammonia and water molecules pack better than ammonia molecules, so that proton transfer is not predicted to occur in compressed water. Proton transfer is even more favourable in water/ammonia mixtures which are expected to form  $\text{OH}^-$  and  $\text{NH}_4^+$  ions at moderate pressures [54].

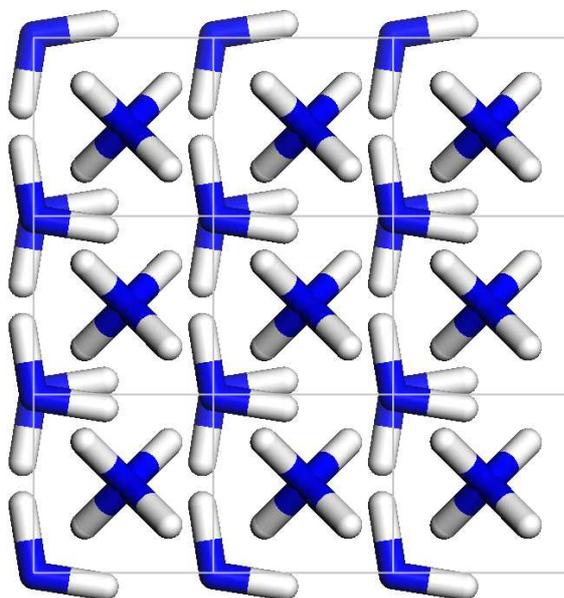


Figure 10: The ionic  $Pma2$  phase of ammonia is predicted to be stable above 90 GPa and consists of alternate layers of  $\text{NH}_4^+$  and  $\text{NH}_2^-$  ions. This view shows the three layers of the crystal structure. The top layer consists of  $\text{NH}_2^-$  ions with orientation  $\perp$ , the second layer consists of tetrahedrally bonded  $\text{NH}_4^+$  ions and bottom layer consists of  $\text{NH}_2^-$  ions with orientation  $\Gamma$

**Ammonia monohydrate:** The properties of compressed ammonia monohydrate ( $\text{NH}_3 \cdot \text{H}_2\text{O}$ ) are of direct relevance to models of the formation of Titan, Saturn’s largest moon. Fortes and coworkers performed neutron diffraction experiments under pressure which yielded the unit-cell parameters and the candidate space groups ( $Pcca$ ,  $Pnca$  and  $Pbca$ ) of phase II of ammonia monohydrate, which is formed at pressures of a few tenths of a GPa [55,56]. The cell parameters indicated that the unit cell contains 16  $\text{NH}_3 \cdot \text{H}_2\text{O}$  formula units, giving a total of 112 atoms. We performed AIRSS calculations using the experimental unit cell with the further assumption that the crystal consisted of weakly hydrogen-bonded  $\text{NH}_3$  and  $\text{H}_2\text{O}$  molecules. Each of the candidate space-groups contains eight symmetry operations, so the asymmetric unit contains two formula units. The initial structures were generated by inserting two  $\text{H}_2\text{NH} \cdots \text{OH}_2$  units at random, generating the rest of the structure using the symmetry operations and rejecting initial configurations in which the molecules overlapped strongly. Searches were performed using each of the three candidate space groups, and the lowest enthalpy structure was obtained with space group  $Pbca$ , which allowed a refinement based on the original data to be performed. These results motivated new experiments which yielded diffraction data which, with additional insights from our predicted structure, were of sufficient quality to allow a full structural determination. A structure of space group  $Pbca$  was determined whose hydrogen bonding network is almost identical to that of the computationally-derived structure [55,56]. Subsequent DFT calculations have shown that the experimentally determined structure is about 0.01 eV per seven-atom

formula unit lower in enthalpy than the theoretically predicted one [57]. This project shows the power of constrained searches. The size of the parameter space was enormously reduced by using the cell parameters and candidate space groups from experiment and the  $\text{H}_2\text{NH}\cdots\text{OH}_2$  unit assumed on chemical grounds. One can never be sure when it is safe to stop searching, and in this case the search was terminated before the correct structure was found. It would certainly have been possible to carry out many more searches in which the correct structure might well have been found, but the experimental determination made this redundant. This project is a nice example of synergy between experimental and computational structure determination.

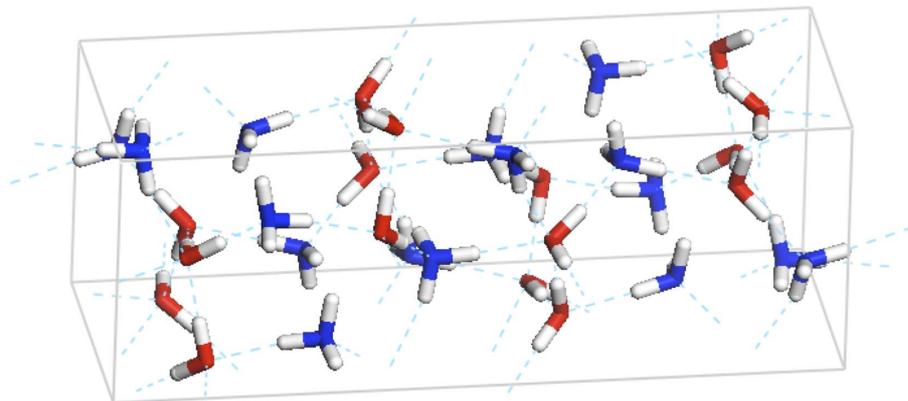


Figure 11: The structure of phase II of ammonia monohydrate predicted using AIRSS. Oxygen atoms are shown in red, nitrogen atoms in blue, and hydrogen in white. The dashed lines indicate close contacts between the molecules. The structure illustrated above and the structure obtained from the neutron diffraction data are very similar and both have  $Pbca$  symmetry, but they have slightly different proton orderings

**Graphite intercalation compounds:** Superconductivity was observed in some graphite intercalation compounds (GICs) in the 1960s. Interest in GICs was rekindled by the discovery of substantial superconducting transition temperatures in  $\text{C}_6\text{Ca}$  and  $\text{C}_6\text{Yb}$  which increase with pressure [58, 59]. The occupation of an inter-layer state is correlated with the occurrence of superconductivity [60]. Csányi *et al* [61] searched for low-enthalpy structures of  $\text{C}_6\text{Ca}$  under pressure. Energetically competitive structures were found at low pressures in which the six-membered rings of the graphene sheets buckle to accommodate Ca atoms within the troughs. Stone-Wales bond rotations [62] within the graphene sheets become favourable at higher pressures, leading to structures with five-, six-, seven- and eight-membered rings, with the Ca atoms sitting within the larger-diameter rings, see Fig. 12. The occurrence of large rings accommodating the intercalate atoms might be a general feature of highly-compressed GICs, and suggests a route to synthesising novel layered carbon structures.

**Hypothetical group IVB clathrate:** AIRSS produces many structures and the metastable ones are often interesting. Looking at the results of a search on carbon we noticed a low-density high-symmetry  $sp^3$ -bonded structure which was unfamiliar to us [63]. This structure (Fig. 13) has a six-atom primitive unit cell with all atoms equivalent, and it is chiral, so that it cannot be superimposed on its mirror image. We have named this the “chiral framework structure” (CFS). It is only 112 meV per atom higher in energy than carbon diamond, while in silicon it is 53 meV per atom higher in energy than the diamond structure [63]. Further investigation revealed

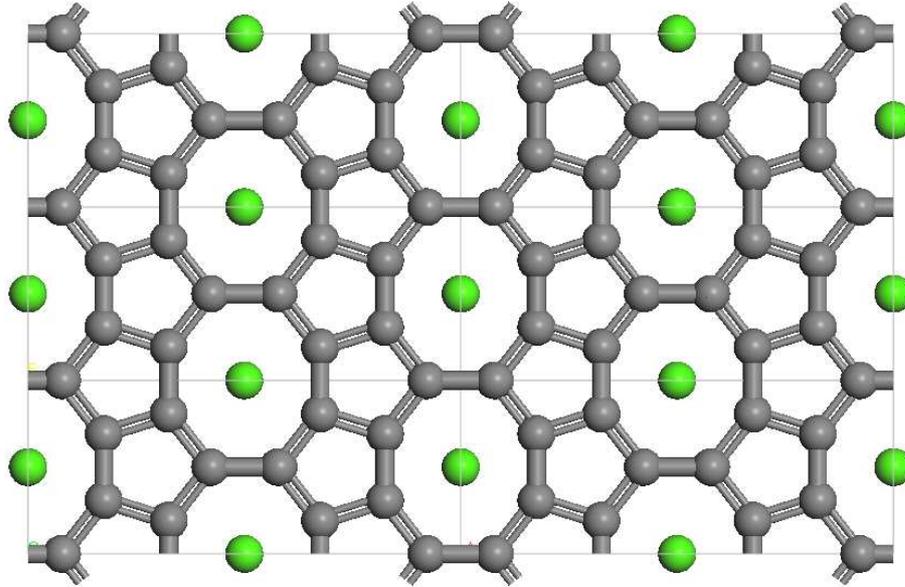


Figure 12: A  $\text{CaC}_6$  graphite intercalation compound of  $Cmcm$  symmetry. The carbon atoms are shown in grey and the calcium atoms in green. In the  $Cmcm$  structure the hexagonal rings of the graphene sheets are replaced by five- and eight-membered carbon rings. This phase is very favourable at high pressures because the cost of the Stone-Wales bond rotations is offset by a large volume reduction as the metal ions are accommodated within the larger rings [61]

it to be the elemental analogue of a zeolite-type structure and it is also related to clathrate structures. Clathrate structures of several different types have been synthesised consisting of silicon, germanium and tin (but not carbon) [64, 65]. The synthesis can only be performed by including “guest” atoms such as Na, K, Rb, Cs or Ba, which act as templates for the self-assembly of the nano cages forming the structures, although in some cases the guest atoms can largely be removed. The clathrate II structures of silicon and carbon are calculated to be about 52 meV per silicon and 72 meV per carbon atom higher in energy than the corresponding diamond structures. Considering that the silicon clathrate II structure has been synthesised [66], might it be possible to synthesise the silicon CFS? A suitable template would have to be found, but it is an intriguing possibility.

**Tellurium dioxide:** Metal dioxides with large cation radii often form cotunnite phases under high pressures, and presumably these transform to post-cotunnite structures at higher pressures. Tellurium dioxide ( $\text{TeO}_2$ ) is apparently the only dioxide in which a post-cotunnite phase has been observed [68], and it is therefore a candidate for the post-cotunnite structure of other metal dioxides. Unfortunately the quality of the x-ray diffraction data obtained by Sato *et al* for post-cotunnite  $\text{TeO}_2$  was insufficient to allow a structural determination, although it was possible to eliminate the known post-cotunnite structures of dihalides [68]. Our AIRSS study [69] found a transition to a post-cotunnite phase of  $\text{TeO}_2$  at 130 GPa, which is a little higher than the experimental transition pressure of 80-100 GPa. The calculated x-ray diffraction data for the predicted phase of  $P2_1/m$  symmetry is in reasonable agreement with experiment. Interestingly we found that the cotunnite phase shows re-entrant behaviour, becoming more stable than

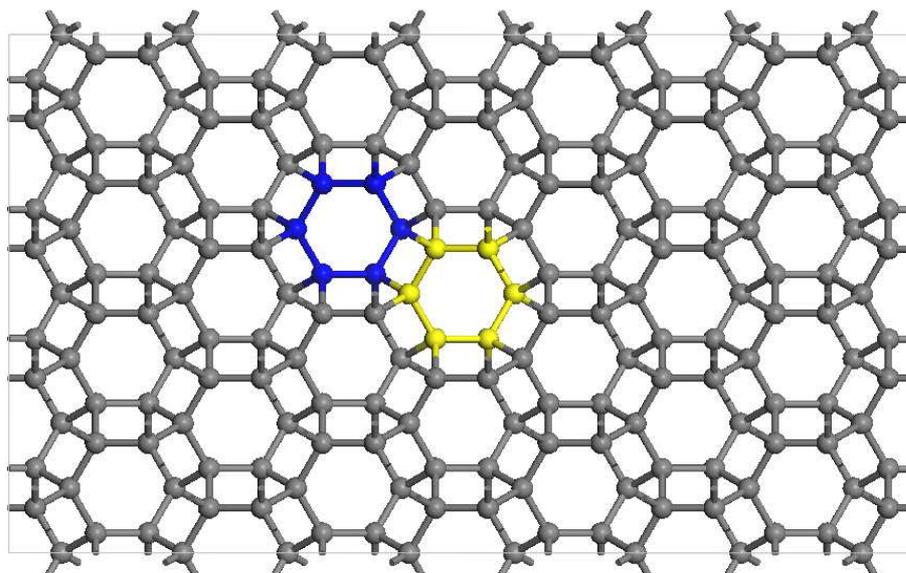


Figure 13: View of the “chiral framework structure” (CFS) along the axis of the helices [63]. The CFS is a low-energy hypothetical structure of group IVB elements which is only a little higher in energy than the diamond structure. The CFS has six atoms per primitive unit cell which are all equivalent by symmetry. The atoms are arranged in five-membered rings and are four-fold coordinated. The CFS has three bond angles slightly smaller than the perfect tetrahedral angle of  $109.5^\circ$  and one bond angle of about  $125^\circ$ . The structure consists of a hexagonal packing of helices which are crosslinked to satisfy four-fold coordination. The helices all twist either to the left or right, so that the crystal is chiral and cannot be superimposed on its mirror image

$P2_1/m$  again above 260 GPa. We tried our  $P2_1/m$  structure in other metal dioxides but it was never the most stable phase [69]. Higher quality x-ray diffraction data are required to test our identification of the  $P2_1/m$  structure as post-cotunnite  $\text{TeO}_2$ .

**Lithium-beryllium alloys:** Feng *et al* [70] used random structure searching to explore lithium-beryllium (Li-Be) alloys under pressure. These elements are immiscible under ambient conditions, but the calculations show they can react under pressure, with  $\text{LiBe}_2$  becoming more stable than the separated elements above about 15 GPa, and  $\text{Li}_3\text{Be}$ ,  $\text{LiBe}$  and  $\text{LiBe}_4$  having regions of stability at higher pressures. The electronic structure of the most stable  $\text{LiBe}$  compound shows two-dimensional character, with a characteristic step-like feature at the bottom of the valence band. The changes in the electronic structure which allow the formation of Li-Be alloys under compression arise from overlap of the Li  $1s$  core electrons which leads to charge transfer towards the Be atoms. In this work [70] the relative stabilities of the different stoichiometries was displayed using a “convex hull” diagram. An example of a convex hull diagram constructed using data obtained from our random searches for the Li-H system is shown in Fig. 14.

**Lithium:** One of the surprises in high pressure physics in recent years has been the discovery that *sp*-bonded elements often adopt complex non-close-packed structures under sufficient compression. The ionic cores take up a larger fraction of the total volume under pressure and some of the valence charge is pushed away from the atoms and into interstitial regions forming

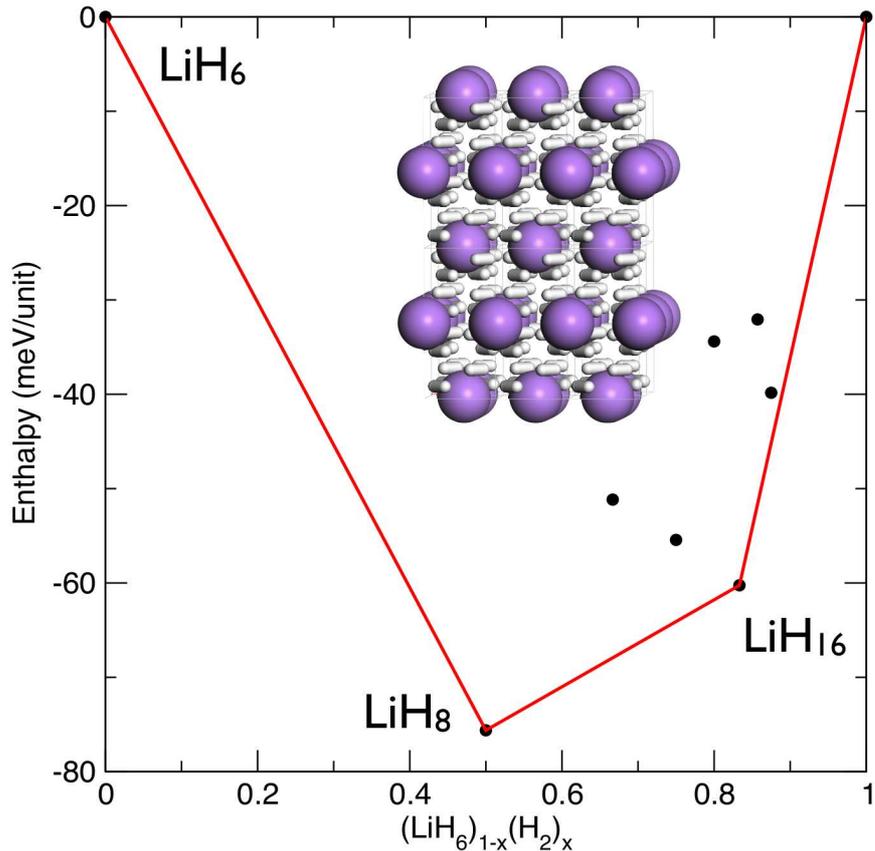


Figure 14: Zurek *et al* have found that “a little bit of lithium does a lot for hydrogen” [71]. In more extensive variable stoichiometry searches we find that even less lithium will do the trick of “metallising hydrogen”. We determined the low enthalpy structures of  $\text{LiH}_{2n}$  for  $n = 3 - 10$  at 100 GPa using AIRSS, displaying the results on a convex hull.  $\text{LiH}_{16}$  (shown) is stable against decomposition into  $\text{LiH}_8$  and  $\text{H}_2$ . It is metallic and is based on a body-centred-tetragonal (bct) packing of lithium atoms “coated” in  $\text{H}_2$  molecules

“blobs” which are rather isolated from one another. The resulting structure can be thought of as an “electride” in which the interstitial electrons are the anions. The valence electronic energy bands consequently become narrower than the free-electron bands [72, 73]. Lithium (Li) adopts the fcc structure under ambient conditions, but it transforms to a three-fold coordinated structure at about 40 GPa [74]. We searched for structures of Li at high pressures, finding two new candidate phases of *Pbca* and *Aba2* symmetry which are predicted to have small regions of stability around 100 GPa [26] and are distortions of the *Cmca-24* structure found in a previous theoretical study [75]. All of these structures have substantial dips in their electronic densities of states (e-DOS) around the Fermi level. This is consistent with, but does not fully explain, the significant increase in electrical resistivity and change in its temperature dependence near 80 GPa observed by Matsuoka and Shimizu [76]. The occupied valence bandwidths of the *Pbca*, *Aba2* and *Cmca-24* phases are substantially narrower than the corresponding free-electron values, demonstrating their electride nature. The low (three-fold) coordination number of these structures arises from Jahn-Teller-like distortions which lower the e-DOS around the Fermi level, and we predicted that the coordination will increase to four-fold above about 450 GPa [26], with

the diamond structure, see Fig. 15, becoming stable above  $\sim 500$  GPa. A first-principles study was also performed by Yao *et al* [77], who found similar results using random structure searching and an evolutionary algorithm. Overall we are, however, left with the impression that there are many nearly-degenerate structures around 100 GPa, and more twists in the story of compressed Li are likely.

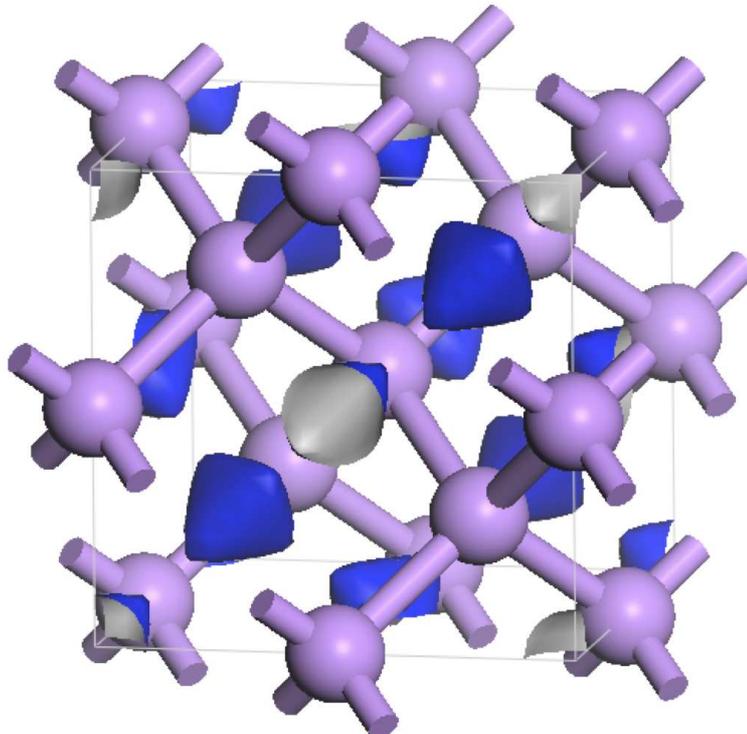


Figure 15: The diamond-structure electrider phase of Li, which is predicted to be stable above 483 GPa. The Li atoms are shown as purple balls and nearest neighbour contacts are shown as sticks. The charge isosurface in blue shows electrons also located on a diamond lattice, in the voids between the lithium ions

**Aluminium:** Aluminium is used as a standard material in shock wave experiments, for which purpose an accurate equation of state must be available. Aluminium adopts the fcc structure under ambient conditions and transforms to hcp at 0.217 terapascals (TPa) [78], and a further transition to a body-centred-cubic (bcc) structure has been predicted at 0.38 TPa using DFT methods [79]. Our searches have identified a transformation from bcc to the Ba-IV non-close-packed incommensurate host-guest structure at 3.2 TPa and a further transition to a simple hexagonal structure at 8.8 TPa [28]. The non-close-packed structures have smaller volumes than bcc and their occurrence significantly alters the high-pressure equation of state. An important feature of our searches was that we studied cells containing 2, 4 and 8–21 atoms. Such a systematic search can yield interesting results and we found commensurate analogues of the host-guest structures in cells of 11, 16, and 21 atoms. The physics behind the occurrence of non-close-packed structures in highly compressed aluminium is similar to that described above for lithium at much lower pressures. The simple hexagonal structure consists of alternate layers of aluminium ions and electrons. There are two “blobs” of electronic charge for every ion and,

considering the aluminium ions as the cations and the electron blobs as the anions, the structure is that of magnesium diboride ( $\text{MgB}_2$ ), which is well known in ionic compounds of  $\text{AB}_2$  stoichiometry. We described the stability of the different structures under pressure using empirical inter-atomic potentials to describe the aluminium ions and electron blobs. The potential parameters were tuned to stabilise the host-guest structure, and it then gave the bcc structure at lower pressures and the simple hexagonal structure at higher pressures. We also found a duality between the Ba-IV structure and the other incommensurate host-guest structure found in the elements, the Rb-IV structure, as explained in Fig. 16.

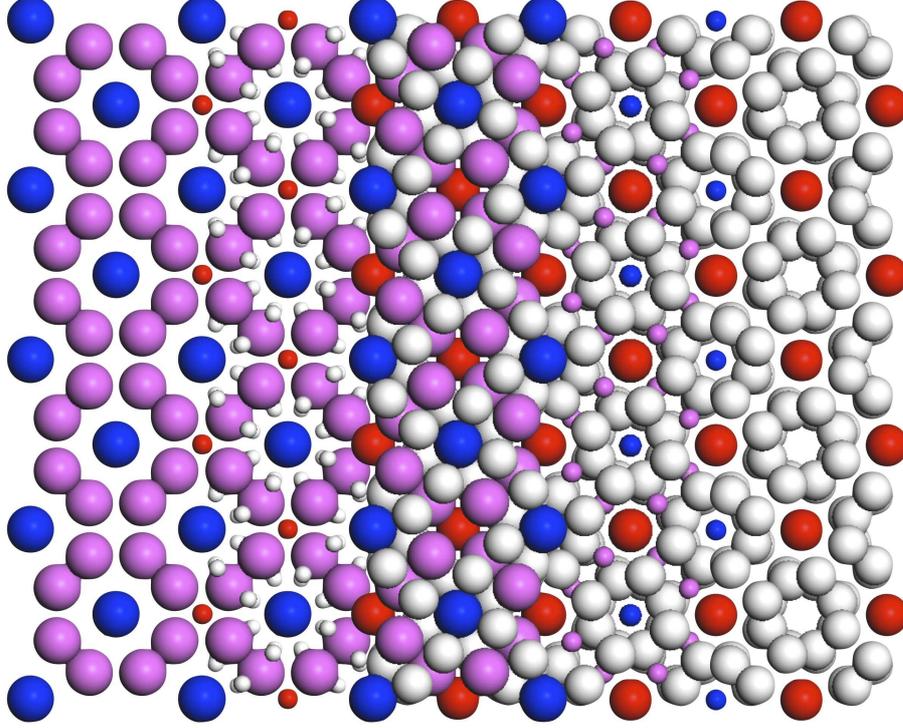


Figure 16: A representation of the Ba-IV incommensurate host-guest structure is shown on the left, with the host atoms in purple and the guest atoms in blue. The Ba-IV structure is also found in compressed Sr, Sc, As, Sb and Bi, and we predict it to be stable in aluminium in the range 3.2–8.8 TPa. A representation of the Rb-IV incommensurate host-guest structure is shown on the right with the guest atoms in red and the host atoms in white. The Rb-IV structure is found in Rb, K and Na at high pressures. Both structures consist of positively charged ions and negatively charged electron blobs located within interstitial regions. The Ba-IV and Rb-IV structures show a remarkable duality. The electron blobs in the Ba-IV structure occupy the atomic positions of the Rb-IV structure, while in the Rb-IV structure the electron blobs occupy the atomic positions of the Ba-IV structure [28]. The figure shows a view along the axis of the guest chains. As we scan the picture from left to right the structure changes from Ba-IV to Rb-IV

**Iron:** The Earth’s core is largely composed of iron. Other planets, including many of the recently-discovered extrasolar planets (or exoplanets), are expected to possess iron-rich cores. Pressures similar to those at the centre of the Earth have been achieved in static diamond anvil cell experiments, but the multi-terapascal (TPa) pressures expected at the centres of more massive planets can currently be achieved only in shock-wave experiments, which give

very limited structural information. Indeed, there are no materials whose structures have been determined experimentally at pressures of 1 TPa or more. At low pressures the electronic configuration of the iron atoms can be described as  $3d^64s^2$ , but the more extended  $4s$  orbitals are pushed up in energy with respect to the  $3d$  orbitals under compression and the  $4s$  charge slowly drains into the  $3d$  orbitals, leading to a  $3d^84s^0$  configuration at multi-TPa pressures. AIRSS showed that only the standard close-packed phases are energetically competitive at multi-TPa pressures [29], see Fig. 17. The bcc structure is stabilised at low pressures by its ferromagnetic spin ordering, but it transforms to a hcp structure at pressures well below 100 GPa. We found a transition from hcp to fcc and back to hcp at TPa pressures (see Fig. 17), although these structures have similar enthalpies in the range 5–30 TPa. The most outstanding result was our prediction that the bcc phase, and a small bct distortion of it, become much more stable than hcp and fcc at extremely high pressures [29]. The reason for this is that the density of bct/bcc is about 0.6 % higher than hcp at the phase transition, which amounts to a very large enthalpy gain at pressures of around 30 TPa. We also studied harmonic phonon modes and the effects of electronic excitations at finite temperatures, but the overall effect on the relative stabilities of the phases is not large [29].

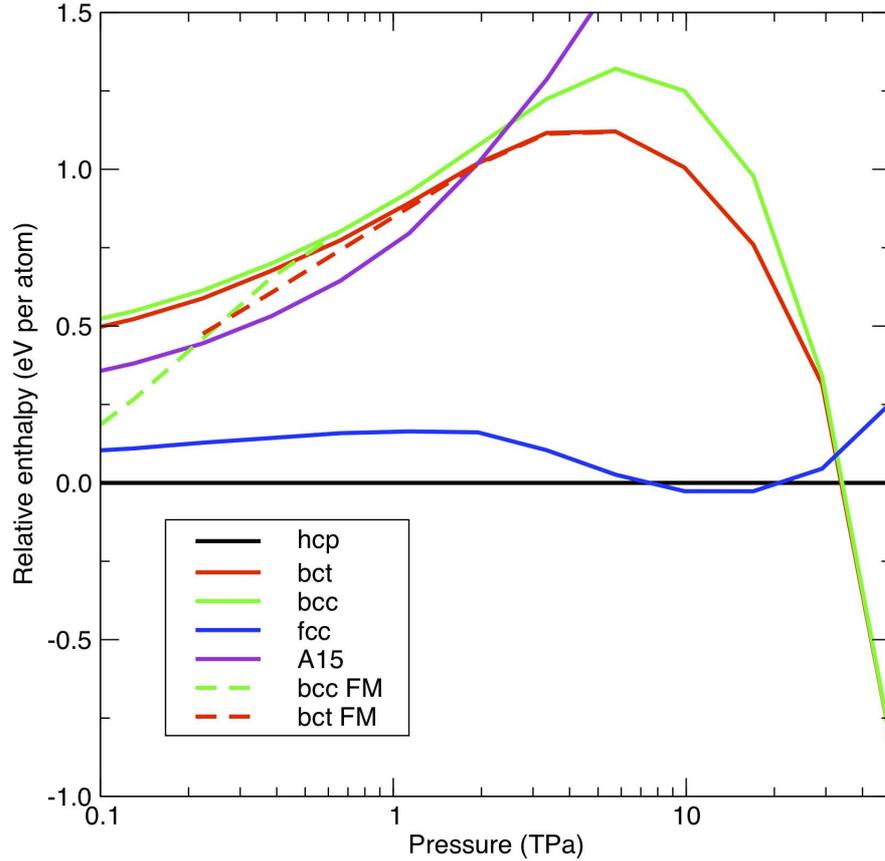


Figure 17: Variation with pressure of the enthalpies of various phases of iron with respect to the hcp phase. The dashed lines indicate ferromagnetic (FM) phases and the solid lines indicate non-magnetic phases

**Defects in silicon:** We have used AIRSS to study defect complexes in Si consisting of combinations of H, N, and O impurity atoms and Si self-interstitials and vacancies [80, 81]. Most

of the searches were performed with 32-atom supercells, although we used larger cells for a few searches. We embedded the most interesting defects in larger cells and relaxed them with a higher energy cutoff and better k-point sampling. We found almost all of the previously-known point defects containing these impurity atoms, and we also found a number of new lowest-energy defects for some stoichiometries, such as  $\{I, H\}$  (an interstitial Si atom and a H atom) [80], and  $\{3O\}$  (three interstitial O atoms) [81]. It is possible to automate the search procedure so that one needs specify only the host crystal, the impurity atoms to be included and the size and location of the “hole” in the host compound into which the impurity atoms are placed. The number of different combinations of impurity atoms need not be excessive. For example, using three different types of impurity atom and a total number of impurity atoms of  $\leq 4$  requires searching over only 34 possible cell contents, and using five different types of impurity atom and a total number of impurity atoms of  $\leq 4$  requires searching over only 125. We estimate that if we were presented with the crystalline structure of a new material containing up to, say, three atomic species and we took into account three possible impurity species (H, N, and O, for example), we could determine the important point defects and their physical and electronic structures within a few weeks. Of course we could also have predicted the structure of the host material.

## 7.7 Conclusions

The different searching methods which have been used in conjunction with DFT methods should be judged by the results obtained. We believe that the AIRSS results presented here are impressive and that they make a strong case for the method. Our approach is pragmatic, we start from the most random method for generating structures that we can think of and introduce biases based on chemical, experimental and/or symmetry grounds. The starting structures are then relaxed while preserving the experimental and symmetry constraints. Sometimes we perform shaking and/or phonon calculations on the relaxed structures to look for energy lowering distortions.

We like the simplicity of this approach, it is easy to understand what the various “knobs” of the method do, which makes it easier to decide which knobs to turn and how far to turn them. We concentrate our computational efforts on relaxing a very wide variety of initial structures, which means that our stopping criterion of obtaining the same lowest-energy structure several times gives a good chance of finding the global minimum of the PES.

Our searching strategy will work very well on the petascale computers which are becoming available now and the exascale computers which will be available in a few years time. Such computing resources will be able to generate enormous databases of structures which will be useful for many purposes, such as fitting and testing empirical force fields, determining structures from diffraction data and determining structures using data mining [82]. The efficient handling and analysis of the huge amounts of data produced by structure searches will pose challenges for the electronic structure community.

Searching for structures with first-principles electronic structure methods has already made an impact in various branches of science and we imagine that it will become an integral part of materials design and discovery. Indeed it is reasonable to suppose that it will become important in all fields in which it is relevant to know the relative positions of atoms.

## 7.8 Acknowledgements

This work has been supported by the Engineering and Physical Sciences Research Council (EPSRC) of the UK.

## A Summary of other computational searching methods

Although this article only deals with the AIRSS approach in detail, it is appropriate to mention other techniques which have been used to predict structures described by empirical or first-principles inter-atomic forces. There are many excellent reviews which describe structure prediction methods for clusters and solids [24, 83–87].

Simulated Annealing (SA) is a Monte Carlo technique devised by Kirkpatrick *et al* [88]. The name derives from an analogy with annealing in metallurgy, in which heating and cooling is used to remove defects from a metal. In this method the current approximate solution or state is replaced by a randomly chosen nearby state. The probability of accepting the new state is 1 if it is lower in energy than the initial state, and  $e^{-\Delta E/T}$  if it is higher, where  $\Delta E$  is the energy of the final state minus the initial state. If the temperature  $T$  is chosen to be zero then only states of lower energy than the current state are accessible and the algorithm normally becomes trapped in a local minimum. To avoid this, the temperature  $T$  is gradually reduced during the simulation and, if the cooling is slow enough, the system will eventually find the lowest energy state. SA with  $T > 0$  allows the system to jump out of local minima. However, the basic algorithm is normally inefficient as it often gets stuck in local minima and many variants of it have been devised and tested in the quest for higher efficiency. There is considerable freedom to alter the proposed moves and the form of the acceptance probability, and to use more complicated “annealing schedules” in which the temperature is sometimes raised during the run.

SA requires only the energies of different configurations of the system, energy derivatives (forces and stresses) are not required. It is, however, straightforward to calculate energy derivatives using empirical potentials and, with a little more effort, within first-principles methods. Energy derivatives can be used to replace the Monte Carlo algorithm by classical molecular dynamics (MD). The most widespread use of energy derivatives in structure searching is to relax a structure to the minimum of its basin of attraction.

Methods have also been devised which evolve ensembles of structures rather than evolving a single structure. The simplest such idea is to run entirely separate searches with different starting points. Ensemble SA methods have been developed in which an adaptive annealing schedule is controlled by ensemble averages of thermodynamic information [89]. Another idea is to use parallel runs at different temperatures, such as in the parallel tempering algorithm which derives from the work of Swendsen and Wang [90,91]. The particle swarm method was inspired by the collective behaviour of a flock of birds [92]. In this MD-based method each member of the ensemble or swarm is accelerated towards its own previous “best solution” and towards the swarm’s previous “best solution”.

Locating the global minimum is difficult because the energy surface contains many basins which may be separated by high barriers. One approach is to transform the energy surface to one which

is easier to search. Perhaps the simplest such idea is to increase the range of the inter-atomic potential [93] which has the effect of removing many local minima [94]. Such an unphysical potential may of course have a significantly different global minimum. Consider instead a transformed energy surface obtained by setting the energy throughout each basin of attraction to the minimum energy of the basin. Obviously this transformation does not affect the relative energies of the minima. We now have to search the transformed energy surface. A simple Monte Carlo procedure known as “basin hopping” [22–24] is to start at a random position, relax to the basin minimum, propose a random move and relax to the new minimum. The move is accepted if the energy is lowered and accepted with probability  $e^{-\Delta E/T}$  if the energy is raised. The simplest version of the algorithm has two parameters, the length of the move which may be adjusted to give a reasonable acceptance ratio, and the temperature  $T$ . The minima hopping method [95] is related to basin hopping.

Evolutionary algorithms (EA) are optimisation techniques inspired by biological evolution, involving concepts such as reproduction, mutation and recombination, fitness and selection [96]. Genetic algorithms (GAs) are a subset of EA in which a genetic representation of approximate solutions (structures) is used, normally a bit array [97]. An ensemble or population of structures is generated and each member is assigned a “fitness” which, for our purposes, is its energy or enthalpy. A fraction of the population is selected for reproduction, with a bias towards the fittest, and they are paired up for “recombination”, the splicing together of the parental genes. A “mutation” step may also be performed. The new population is then subjected to selection and the whole process is repeated. When using GAs for structure searching it is standard to relax structures to the minimum of their basin of attraction before reproduction, so that the inheritance might be described as Lamarckian rather than Darwinian. GAs have been applied to many optimisation problems, and a review of the design and use of GAs for determining the structures of atomic clusters described by empirical potentials is given by Johnston [83].

The set of algorithms for predicting structures described above is of course far from complete and interesting alternatives have been pursued. For example, crystalline network structures, such as zeolites and carbon polymorphs, have been enumerated systematically using graph theory [98–100]. Faken *et al* [101] have sought high-dimensional barrier-less pathways between local minima in the physical three-dimensional space, and methods using quantum delocalisation have also been investigated [102–104]. Metadynamics is a powerful sampling technique for reconstructing the free-energy surface as a function of a set of collective variables, and this method can be used to study phase transitions at finite temperatures [105, 106].

Some of the strategies described above can clearly be combined, and many additional refinements have been suggested. There is often a substantial overlap between the various different methods, and it can be difficult to determine where one method ends and the next begins. On reading the description of our AIRSS approach in Section 7.3, the reader will recognise elements from the searching methods described in this appendix.

Almost all of the methods described above were first used in searching for structures with empirical potentials, although they have since been used with first-principles methods. Jones and coworkers used molecular dynamics simulated annealing with first-principles DFT to study the structures of numerous clusters from the late 1980s [107, 108]. Deaven and Ho searched for

cluster geometries using a GA and a tight-binding model [109], and this work was important in bringing the possibilities of such methods to the attention of the “first-principles” community. Predicting crystal structures with first-principles methods is growing in popularity. Schön, Jansen and coworkers have used Hartree-Fock theory and DFT to search for stable structures and study the PES of various crystals [87, 110]. Zunger and coworkers [111, 112] and Oganov and coworkers [113, 114] have used GAs to search for crystal structures with DFT methods.

## References

- [1] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, 864 (1964).
- [2] W. Kohn and L. J. Sham, *Phys. Rev.* **145**, 561 (1965).
- [3] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- [4] F. H. Stillinger, *Phys. Rev. E* **59**, 48 (1999).
- [5] M. R. Hoare and J. McInnes, *Faraday Discuss. Chem. Soc.* **61**, 12 (1976).
- [6] C. J. Tsai and K. D. Jordan, *J. Phys. Chem.* **97**, 11227 (1993).
- [7] D. H. Wolpert and W. G. Macready, *IEEE Transactions on Evolutionary Computation* **1**, 67 (1997).
- [8] F. Jensen, *Computational Chemistry* (New York, Wiley, 1999).
- [9] J. P. K. Doye, D. J. Wales, and M. A. Miller, *J. Chem. Phys.* **109**, 8143 (1998).
- [10] C. P. Massen and J. P. K. Doye, *Phys. Rev. E* **75**, 037101 (2007).
- [11] C. P. Massen, J. P. K. Doye, and R. W. Nash, *Physica A* **382**, 683 (2007).
- [12] D. J. Wales, *Chem. Phys. Lett.* **285**, 330 (1998).
- [13] D. J. Wales, *Chem. Phys. Lett.* **294**, 262 (1998).
- [14] M. Nowacki, *Helv. Chim. Acta* **26**, 459 (1943).
- [15] A. D. Mighell, V. L. Himes, and J. R. Rodgers, *Acta Cryst. A* **39**, 737 (1983).
- [16] J. Donohue, *Acta Cryst. A* **41**, 203 (1985).
- [17] A. Mackay, *Acta Crystallogr.* **22**, 29 (1967).
- [18] V. S. Urusov and T. N. Nadezhina, *J. Struct. Chem.* **50**, S22 (2009).
- [19] C. J. Pickard and R. J. Needs, *J. Chem. Phys.* **127**, 244503 (2007).
- [20] <http://www-wales.ch.cam.ac.uk/~wales/CCD/Si.html>
- [21] S. L. Price, *Acc. Chem. Res.* **42**, 117 (2009).

- [22] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).
- [23] D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).
- [24] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
- [25] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. **220**, 567 (2005).
- [26] C. J. Pickard and R. J. Needs, Phys. Rev. Lett. **102**, 146401 (2009).
- [27] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
- [28] C. J. Pickard and R. J. Needs, Nature Materials **9**, 624 (2010).
- [29] C. J. Pickard and R. J. Needs, J. Phys.: Condensed Matter **21**, 452205 (2009).
- [30] A. Mujica, A. Rubio, A. Muñoz, and R. J. Needs, Rev. Mod. Phys. **75**, 863 (2003).
- [31] C. J. Pickard and R. J. Needs, Phys. Rev. Lett. **97**, 045504 (2006).
- [32] J. Feng, W. Grochala, T. Jaroń, R. Hoffmann, A. Bergara, and N. W. Ashcroft, Phys. Rev. Lett. **96**, 017006 (2006).
- [33] M. I. Eremets, I. A. Trojan, S. A. Medvedev, J. S. Tse, and Y. Yao, Science **319**, 1506 (2008).
- [34] O. Degtyareva, J. E. Proctor, C. L. Guillaume, E. Gregoryanz, and M. Hanfland, Sol. State Commun. **149**, 1583 (2009).
- [35] N. W. Ashcroft, Phys. Rev. Lett. **92**, 187002 (2004).
- [36] C. J. Pickard and R. J. Needs, Phys. Rev. B **76**, 144114 (2007).
- [37] I. Goncharenko, M. I. Eremets, M. Hanfland, J. S. Tse, M. Amboage, Y. Yao, and I. A. Trojan, Phys. Rev. Lett. **100**, 045504 (2008).
- [38] P. Loubeyre, F. Occelli, and R. LeToullec, Nature **416**, 613 (2002).
- [39] H.-K. Mao and R. J. Hemley, Rev. Mod. Phys. **66**, 671 (1994).
- [40] P. Loubeyre, R. LeToullec, D. Hausermann, M. Hanfland, R. J. Hemley, H.-K. Mao, and L. W. Finger, Nature **383**, 702 (1996).
- [41] C. J. Pickard and R. J. Needs, Nature Physics **3**, 473 (2007).
- [42] C. J. Pickard and R. J. Needs, Physica Status Solidi (b) **246**, 536 (2009).
- [43] R. J. Hemley, I. I. Mazin, A. F. Goncharov, and H.-K. Mao, Europhys. Lett. **37**, 403 (1997).
- [44] A. F. Goncharov, E. Gregoryanz, R. J. Hemley, and H.-K. Mao, Proc. Natl. Acad. Sci. USA **98**, 14234 (2001).

- [45] R. Bini, L. Ulivi, J. Kreutz, and H. J. Jodl, *J. Chem. Phys.* **112**, 8522 (2000).
- [46] E. Gregoryanz, A. F. Goncharov, R. J. Hemley, H.-K. Mao, M. Somayazulu, and G. Shen, *Phys. Rev. B* **66**, 224108 (2002).
- [47] E. Gregoryanz, A. F. Goncharov, C. Sanloup, M. Somayazulu, H.-K. Mao, and R. J. Hemley, *J. Chem. Phys.* **126**, 184505 (2007).
- [48] C. J. Pickard and R. J. Needs, *Phys. Rev. Lett.* **102**, 125702 (2009).
- [49] R. M. Martin and R. J. Needs, *Phys. Rev. B* **34**, 5082 (1986).
- [50] M. I. Eremets, A. G. Gavriliuk, I. A. Trojan, D. A. Dzivenko, and R. Boehler, *Nature Materials* **3**, 558 (2004).
- [51] C. Mailhot, L. H. Yang, and A. K. McMahan, *Phys. Rev. B* **46**, 14419 (1992).
- [52] Y. M. Ma, A. R. Oganov, Z. W. Li, Y. Xie, and J. Kotakoski, *Phys. Rev. Lett.* **102**, 065501 (2009).
- [53] C. J. Pickard and R. J. Needs, *Nature Materials* **10**, 757 (2008).
- [54] A. D. Fortes, J. P. Brodholt, I. G. Wood, L. Vocadlo, and H. D. B. Jenkins, *J. Chem. Phys.* **115**, 7006 (2001).
- [55] A. D. Fortes, E. Suard, M.-H. Lemé-Cailleau, C. J. Pickard, and R. J. Needs, *J. Am. Chem. Soc.* **131**, 13508 (2009).
- [56] A. D. Fortes, E. Suard, M.-H. Lemé-Cailleau, C. J. Pickard, and R. J. Needs, *J. Chem. Phys.* **131**, 154503 (2009).
- [57] G. I. G. Griffiths, R. J. Needs, and C. J. Pickard, unpublished.
- [58] T. Weller, M. Ellerby, S. S. Saxena, R. Smith, and N. Skipper, *Nature Physics* **1**, 39 (2005).
- [59] N. Emery, C. Hérold, M. d’Astuto, V. Garcia, Ch. Bellin, J. F. Marêché, P. Lagrange, and G. Loupiau, *Phys. Rev. Lett.* **95**, 087003 (2005).
- [60] G. Csányi, P. B. Littlewood, A. H. Nevidomskyy, C. J. Pickard, and B. D. Simons, *Nature Physics* **1**, 42 (2005).
- [61] G. Csányi, C. J. Pickard, B. D. Simons, and R. J. Needs, *Phys. Rev. B* **75**, 085432 (2007).
- [62] A. J. Stone and D. J. Wales, *Chem. Phys. Lett.* **128**, 501 (1986).
- [63] C. J. Pickard and R. J. Needs, *Phys. Rev. B* **81**, 154503 (2010).
- [64] J. S. Kasper, P. Hagenmuller, M. Pouchard, and C. Cros, *Science* **150**, 1713 (1965).
- [65] A. San-Miguel and P. Toulemonde, *High Pressure Research* **25**, 159 (2005).
- [66] M. Beekman and G. S. Nolas, *J. Mater. Chem.* **18**, 842 (2008).

- [67] W. L. Mao, H.-K. Mao, Y. Meng, P. J. Eng, M. Y. Hu, P. Chow, Y. Q. Cai, J. Shu, and R. J. Hemley, *Science* **314**, 636 (2006).
- [68] T. Sato, N. Funamori, T. Yagi, and N. Miyajima, *Phys. Rev. B* **72**, 092101 (2005).
- [69] G. I. G. Griffiths, R. J. Needs, and C. J. Pickard, *Phys. Rev. B* **80**, 184115 (2009).
- [70] J. Feng, R. G. Hennig, N. W. Ashcroft, and R. Hoffmann, *Nature* **451**, 445 (2008).
- [71] E. Zurek, R. Hoffmann, N. W. Ashcroft, A. R. Oganov, and A. O. Lyakhovc, *Proc. Natl. Acad. Sci. USA* **106**, 17640 (2009).
- [72] J. B. Neaton and N. W. Ashcroft, *Nature* **400**, 141 (1999).
- [73] B. Rousseau and N. W. Ashcroft, *Phys. Rev. Lett.* **101**, 046407 (2008).
- [74] M. Hanfland, K. Syassen, N. E. Christensen, and D. L. Novikov, *Nature* **408**, 174 (2000).
- [75] R. Rousseau, K. Uehara, D. D. Klug, and J. S. Tse, *Chem. Phys. Chem.* **6**, 1703 (2005).
- [76] T. Matsuoka and K. Shimizu, *Nature* **458**, 186 (2009).
- [77] Y. Yao, J. S. Tse, and D. D. Klug, *Phys. Rev. Lett.* **102**, 115503 (2009).
- [78] Y. Akahama, M. Nishimura, K. Kinoshita, and H. Kawamura, *Phys. Rev. Lett.* **96**, 045505 (2006).
- [79] M. J. Tambe, N. Bonini, and N. Marzari, *Phys. Rev. B* **77**, 172102 (2008).
- [80] A. J. Morris, C. J. Pickard, and R. J. Needs, *Phys. Rev. B* **78**, 184102 (2008).
- [81] A. J. Morris, C. J. Pickard, and R. J. Needs, *Phys. Rev. B* **80**, 144112 (2009).
- [82] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, *Nature Materials* **5**, 641 (2006).
- [83] R. L. Johnston, *Dalton Trans.* **22**, 4193 (2003).
- [84] F. Baletto and R. Ferrando, *Rev. Mod. Phys.* **77**, 371 (2005).
- [85] S. M. Woodley and R. Catlow, *Nature Materials* **7**, 937 (2008).
- [86] G. Rossi and R. Ferrando, *J. Phys.: Condensed Matter* **21**, 084208 (2009).
- [87] J. C. Schön, K. Doll, and M. Jansen, *Phys. Stat. Solidi B* **247**, 23 (2010).
- [88] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
- [89] G. Ruppener, J. M. Pedersen, and P. Salamon, *J. de Physique I* **1**, 455 (1991).
- [90] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [91] D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).
- [92] Y. H. Shi and R. Eberhart, *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 1998)*, Piscataway, NJ., 69 (1969).

- [93] F. H. Stillinger and D. K. Stillinger, *J. Chem. Phys.* **93**, 6106 (1990).
- [94] J. P. K. Doye and D. J. Wales, *J. Chem. Soc. Faraday Trans.* **93**, 4233 (1997).
- [95] S. Goedecker, *J. Chem. Phys.* **120**, 9912 (2004).
- [96] T. Bäck, *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*, (Oxford University Press, Oxford, 1990)
- [97] J. Holland, *Adaptation in Natural and Artificial Systems*, (University of Michigan Press, Ann Arbor, MI, 1975).
- [98] A. L. Mackay and H. Terrones, *Nature* **352**, 762 (1991).
- [99] O. D. Friedrichs, A. W. M. Dress, D. H. Huson, J. Klinowski, and A. L. Mackay, *Nature* **400**, 644 (1999).
- [100] B. Winkler, C. J. Pickard, V. Milman, and G. Thimm, *Chem. Phys. Lett.* **337**, 36 (2001).
- [101] D. B. Faken, A. F. Voter, D. L. Freeman, and J. D. Doll, *J. Phys. Chem. A* **103**, 9521 (1999).
- [102] P. Amara, D. Hsu, and J. E. Straub, *J. Phys. Chem.* **97**, 6715 (1993).
- [103] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, *Chem. Phys. Lett.* **219**, 343 (1994).
- [104] Y. H. Lee and B. J. Berne, *J. Phys. Chem. A* **104**, 86 (2000).
- [105] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002).
- [106] A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
- [107] D. Hohl, R. O. Jones, R. Car, and M. Parrinello, *Chem. Phys. Lett.* **139**, 540 (1987).
- [108] D. Hohl, R. O. Jones, R. Car, and M. Parrinello, *J. Chem. Phys.* **89**, 6823 (1988).
- [109] D. M. Deaven and K.-M. Ho, *Phys. Rev. Lett.* **75**, 288 (1995).
- [110] J. C. Schön and M. Jansen, *Z. Kristallogr.* **216**, 307 (2001).
- [111] G. L. W. Hart, V. Blum, M. J. Walorski, and A. Zunger, *Nature Materials* **4**, 391 (2005).
- [112] X. W. Zhang and A. Zunger, *Phys. Rev. Lett.* **104**, 245501 (2010).
- [113] A. R. Oganov and C. W. Glass, *J. Chem. Phys.* **124**, 244704 (2006).
- [114] A. R. Oganov, J. H. Chen, C. Gatti, Y. Z. Ma, Y. M. Ma, C. W. Glass, Z. X. Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko, *Nature* **457**, 863 (2009).