

Systematic prediction of crystal structures: An application to sp^3 -hybridized carbon polymorphs

Rachel T. Strong* and Chris J. Pickard
Cavendish Laboratory, Cambridge CB3 0HE, United Kingdom

Victor Milman
Accelrys Incorporated, 334 Science Park, Cambridge CB4 0WN, United Kingdom

Georg Thimm
School of MPE, Nanyang Technology University, Singapore 639798

Björn Winkler
Institut für Mineralogie, Johann Wolfgang Goethe Universität, D-60054 Frankfurt, Germany
 (Received 15 December 2003; published 2 July 2004)

A general systematic method of predicting hypothetical crystal structures could enable important advances in many areas of science. We describe a recently developed approach based on graph theory and density functional theory and apply it to enumerate systematically a number of sp^3 -hybridized carbon polymorphs with four atoms per unit cell. The calculations predict three unknown structures that are potentially metastable under appropriate pressure and temperature conditions. The theoretical properties of these hypothetical polymorphs and their relative stability with respect to diamond are discussed.

DOI: 10.1103/PhysRevB.70.045101

PACS number(s): 61.50.Ah

I. INTRODUCTION

The theoretical enumeration of crystal structures has been an area of considerable scientific interest for many years. The systematic prediction of molecules such as the fullerenes has been possible for some time.¹ However, methods used for the prediction of bulk structures are often based on isomorphic substitution of known structures or repetitions of known molecules.^{2–4} In other words they are not free of intuition and hence bias and do not necessarily lead to the complete enumeration of all crystal structures under a given set of criteria.

A potential approach to the systematic prediction of crystal structures, based on the tiling of minimal surfaces, is described by Friedrichs *et al.*⁵ However, their theory is considerably more complex than the approach presented below, is not generally applicable, and is inefficient for small structures. A sphere packing method developed by Kock and Fischer⁶ is based on the orbit of a point under all operations of a three-dimensional space group. All pairs of points of the orbit which are less than a specified distance apart are then connected by a line. Unfortunately, this method does not allow the straightforward generation of all structures with a specific number of atoms and bonds per unit cell.

Blatov⁷ independently made use of a reduced representation of nets, similar to the one described in this paper. His graph-theoretical method has been applied to the analysis of topological similarities of existing crystal structures⁸ and not to the prediction of new crystal structures. Curtarolo *et al.*⁹ recently proposed mining existing data libraries to investigate new crystal structures and hence incorporate the results of large numbers of quantum mechanical calculations into the prediction of structures.

The approach described in this paper is a combination of graph theory and quantum mechanics which enables the sys-

tematic prediction of crystal structures under a given set of constraints: for example, the number of atoms per unit cell and coordination of the atoms. Results for sp^3 -hybridized or fourfold-coordinated carbon polymorphs with four atoms per unit cell are presented and discussed in Sec. IV. Previous applications of this method to sp^2 -hybridized crystal structures with four and six atoms per unit cell are described by Winkler *et al.*^{10,11}

II. SYSTEMATIC PREDICTION OF BULK STRUCTURES

In 1977, Wells¹² showed that a system of interatomic bonds in a crystal structure could be completely represented by a three-dimensional infinite graph known as a *net*.

Each atom is represented by a vertex of the net, and each edge of the net represents a bond in the crystal structure. In the cases we study, the graphs describe bonding—i.e., we impose the sp^3 hybridization—and bonding is considered to be equivalent to connectivity (throughout this paper the terms are used as synonyms). For clarity, we would like to emphasize that *connectivity* is a geometrical feature, while *bonding* is a crystal chemical concept, and these need not be the same.

A net retains all the information on the connectivity of atoms within the crystal. In turn, a crystal may simply be thought of as an *embedded* net, in the sense that precise locations in Euclidean space are assigned to each vertex. A two-dimensional example of a crystal layer and the corresponding net are shown in Fig. 1; however, the theory may be applied equally well to three-dimensional crystals.

To enable the systematic generation of nets the method makes use of a *finite* or *reduced* representation of nets in the form of *quotient graphs*, described by Chung *et al.*¹³ A two-dimensional example of a net and the corresponding quotient

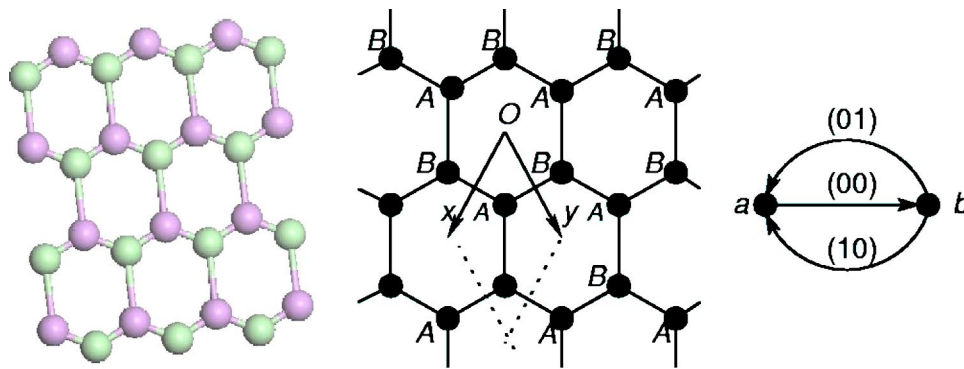


FIG. 1. (Color online) An illustration of a two-dimensional crystal layer, its corresponding net, and quotient graph.

graph are shown in Fig. 1. Each unit cell of the net, and hence the crystal it corresponds to, is made up of a pair of vertices labeled *A* and *B*. The basis vectors for the coordinate system, *x* and *y*, may be used to move from unit cell to unit cell, completely mapping out the net.

Each vertex of the quotient graph corresponds to a class of translationally equivalent vertices of the net. In other words, it is the image of a class of translationally equivalent atoms in the crystal. In the illustrated case, the vertices of the net labeled *A* are mapped onto vertex *a* of the quotient graph and those labeled *B* are mapped onto vertex *b*.

Similarly, each labeled edge of the quotient graph is the image of a class of translationally equivalent edges of the embedded net (or bonds within the crystal). The indices assigned to each edge of the quotient graph are the difference between the coordinates of the cell in which an arbitrary bond ends and the cell in which it starts, expressed in terms of the chosen coordinate system. The actual values of the indices vary depending on the structure and choice of basis vectors for the coordinate system. However, for real crystal structures, a coordinate system can be chosen such that the indices are from the domain $\{-1, 0, 1\}$. Parallels may be drawn between the relationship of a quotient graph and net and the relationship of a primitive unit cell and the corresponding crystal. More elaborate arrangements are also possible; for example, groups of atoms may be mapped onto a single vertex or an oxygen bridge could be mapped onto a single edge. However, such complex mapping is outside the scope of the structures studied in this paper.

The direct mapping between a crystal and its corresponding quotient graph is illustrated in Fig. 2. A reverse mapping from quotient graph to crystal is also feasible. However, the reverse mapping is not straightforward as the basis vectors for the coordinate system needed to map between the quotient graph and the crystal, and the positions of nodes within the cell are unknown and require optimization. Since all crystal structures can be mapped onto a quotient graph, in principle it is possible to reverse the procedure and generate possible crystal structures through the enumeration of quotient graphs.

The approach has the advantage that for a given connectivity and number of atoms per unit cell all finite labeled quotient graphs can be determined using combinatorial techniques, leading to the complete enumeration of all possible nets. The method used is based on the vector method described by Chung *et al.*¹³ The practical implementation and application of the technique to threefold-coordinated nets with four and six vertices per unit cell is discussed in Bader *et al.*¹⁴

The first stage of the approach consists of generating all of the finite graphs that will become quotient graphs, under the constraints of the crystal structures to be investigated. In other words, the number of vertices in the graph must be equivalent to the predicted number of atoms per unit cell, and the number and arrangement of edges must match the connectivity of the structures to be studied. The algorithm used to enumerate the graphs is quite complex, as additional steps and constraints are included to reduce the number of

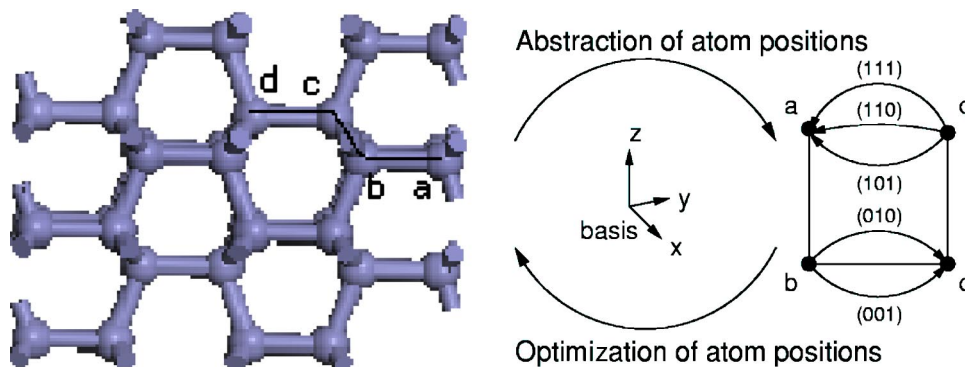


FIG. 2. (Color online) A schematic diagram of the relationship between the crystal, lonsdaleite, and its quotient graph.

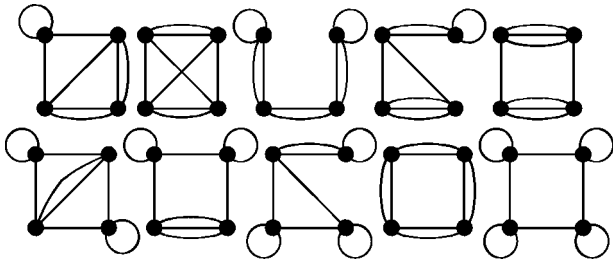


FIG. 3. Set of ten unlabeled quotient graphs with four fourfold-coordinated vertices.

redundant graphs. However, a basic summary of the algorithm follows.

(i) Enumerate all nonisomorphic graphs with the required number of vertices and further match the imposed constraints for connectivity.

(ii) Arbitrarily orientate all edges in the graph.

(iii) Complete the labeling of the graphs.

Edges of the finite graphs are labeled with all possible combinations of index triples (r, s, t) where $r, s, t \in \{-1, 0, 1\}$. In addition, loops may not have the label $(0, 0, 0)$ and no pair of parallel edges may have the same label or a label that results from multiplying all indices of the other label by -1 . Each finite graph may generate more than one labeled quotient graph leading to different nets.

(iv) Eliminate quotient graphs describing isomorphic nets. The labeled quotient graphs are checked for isomorphism and one representative is selected from each isomorphic group and taken forward to the next, embedding, stage.

There are also a few cases where isomorphic nets belong to nonisomorphic quotient graphs, but these may be recognized by studying invariant properties of the quotient graphs, such as cycle sequences.^{1,15}

A very high number of superfluous graphs are produced by this approach, hence the importance of producing a programmable algorithm to implement computationally the theory of Chung *et al.*¹³

The next stage of the method is the tentative embedding of each net obtained from the quotient graphs into Euclidean space. The idea of embedding is to approximately define the basis vectors for the coordinate system needed to map from the quotient graph to the crystal, producing a crystallographically viable net.

A net is said to have yielded an *acceptable* embedding if it meets the following requirements.¹⁴

(i) The distance between adjacent vertices of the net is 1.

(ii) The distance between nonadjacent vertices of the net is greater than 1.

The second requirement is due to the choice of investigating carbon polymorphs, where the connectivity described by the graphs actually corresponds to interatomic bonding. In a more general case the presence of complex coordination polyhedra including atoms in a second coordination shell could be required for a successful embedding. Those nets that meet the above criteria are potentially of crystallographic interest and are retained. An atom may then be added to the net at each vertex and the length of the edges of the net scaled so that the distance between atoms is average for the examined type of structure.

TABLE I. Structures, space groups, lattice energies relative to diamond (ΔE), and densities of the hypothetical carbon polymorphs.

Structure	Space group	Lattice parameters Cell angles (Å, deg)			Atom positions $x/a, y/b, z/c$			ΔE (eV/atom)	Density (g/cm ³)
A	$Fd\bar{3}m$	3.5658	3.5658	3.5658	0.0000	0.0000	0.0000	0.000	3.519
Diamond									
B	$Fd\bar{3}m$	3.5658	3.5658	3.5658	0.0000	0.0000	0.0000	0.000	3.519
Diamond									
C	$P(6_3/m)mc$	2.5074	2.5074	4.1735	0.6667	0.3333	0.4373	0.025	3.511
Lonsdaleite									
D	$I(4/m)mm$	4.3717	4.3717	2.5061	0.3197	0.3197	0.0000	0.198	3.331
E	$Imma$	4.9327	2.5434	4.1958	0.1629	0.2500	0.6042	0.364	3.031
F	$C2$	3.7694	6.2448	2.4757	0.0000	0.1116	0.5000	0.803	2.950
		90.00	68.17	90.00	0.3761	0.2442	0.3042		
					0.5000	0.4689	0.0000		
G	$P4_122$	3.1017	3.1017	2.4487	0.5000	0.1458	0.0000	1.101	3.387
H	$Fddd$	2.4304	4.9396	7.5260	0.0000	0.0000	0.1880	1.138	3.532
I	$C2$	3.9981	4.3078	3.8750	0.1674	0.4445	0.2274	1.328	2.865
		90.00	123.94	90.00	0.3893	0.2419	0.5945		
J	$I2_12_12_1$	6.5973	2.5921	2.7514	0.2500	0.3216	0.0000	1.579	3.391
					0.0766	0.0000	0.2500		
K	$I2_12_12_1$	5.9468	2.8971	2.9161	0.0000	0.2500	0.8719	1.677	3.176
					0.2500	0.3620	0.0000		

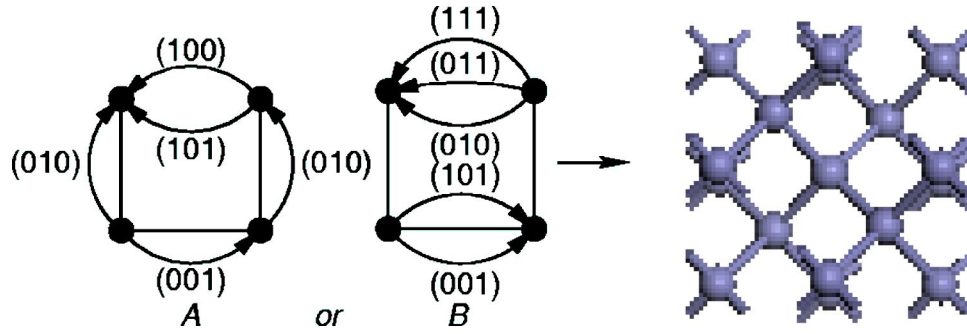


FIG. 4. (Color online) Diamond, produced by quotient graphs A and B.

The final stage of the method makes use of density functional theory (DFT) to optimize the geometry of acceptably embedded quotient graphs. DFT is an efficient way to use quantum mechanics to calculate the ground-state electron density and total energy of a system of interacting electrons for a given configuration of ions. No assumption is made about the nature of the system to be studied; instead, the approximations are based on general principles of physics. DFT is an accurate tool for the calculation of interatomic distances, bond angles, crystal density, and unit cell volume and properties such as the relative stability of structures, vibrational spectra, and elastic properties. More detailed reviews of the total energy pseudopotential method may be

found in Payne *et al.*¹⁶ and Segall *et al.*¹⁷ Any candidate structure that, after a full geometry optimization, is stable with respect to small strains is said to be a potential polymorph. Some nets produced by embedding quotient graphs in Euclidean space do not survive this final stage with their connectivity intact and are discarded.

III. COMPUTATIONAL DETAILS

In order to demonstrate the above method we investigated sp^3 -hybridized fourfold-coordinated framework carbon polymorphs with four atoms per unit cell. Any member of this class of structures can be completely mapped onto one of the

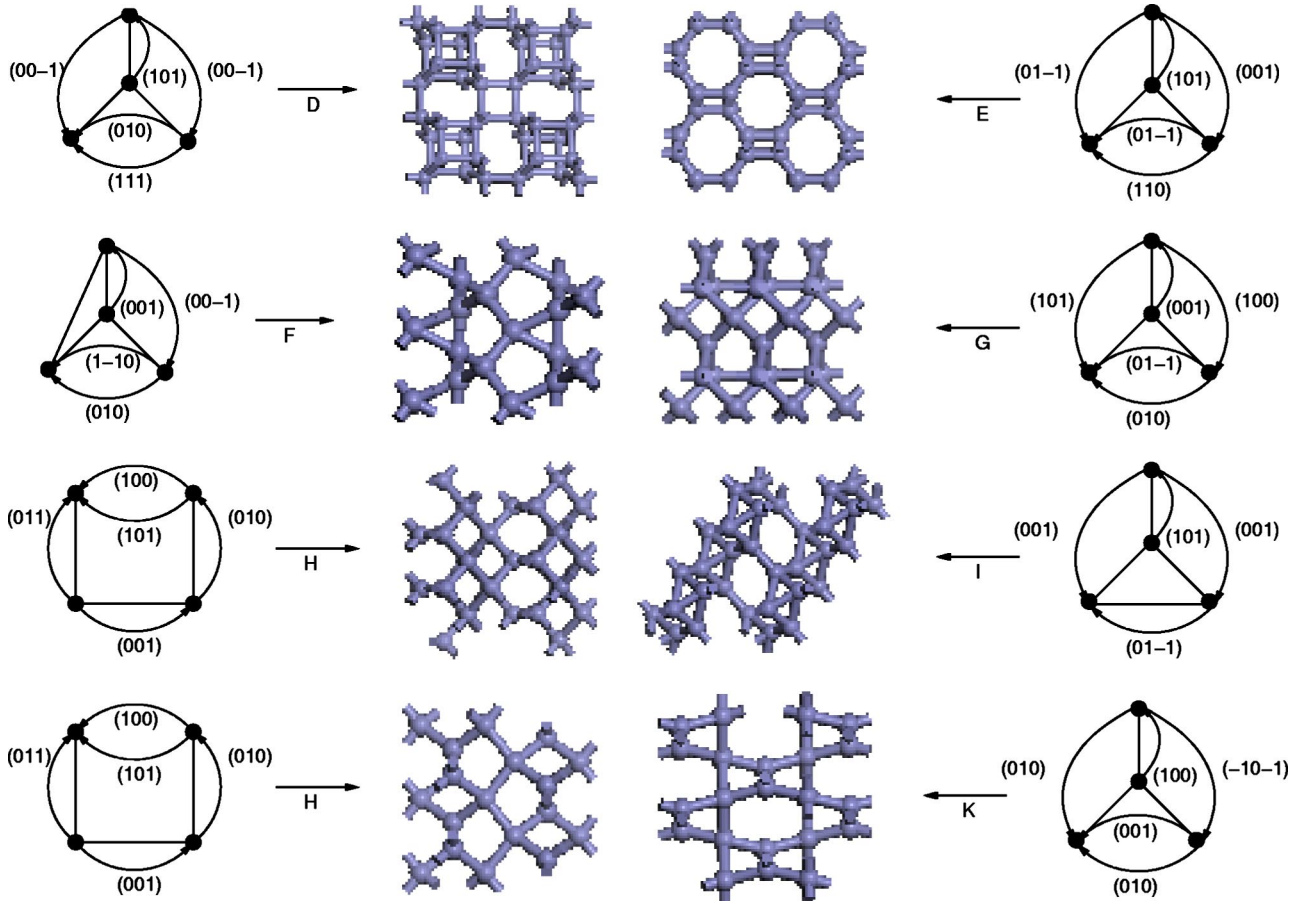


FIG. 5. (Color online) A selection of geometry-optimized structures of hypothetical carbon polymorphs.

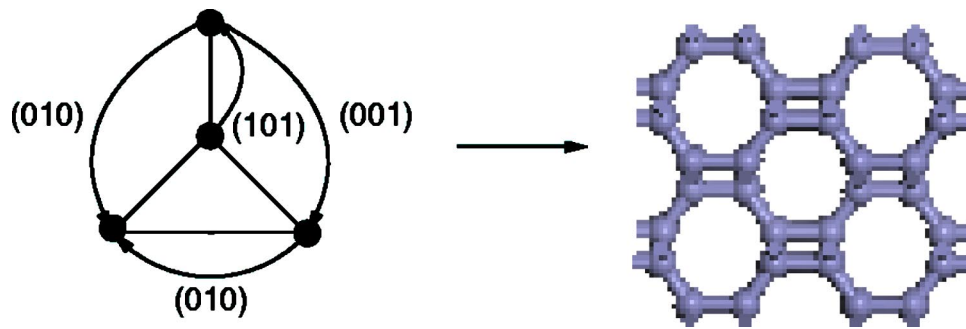


FIG. 6. (Color online) The quotient graph L , which did not retain the original connectivity.

ten as yet unlabeled quotient graphs shown in Fig. 3. Based on these graphs, the enumeration algorithm produced roughly 40×10^6 labeled quotient graphs, of which 67 507 describe distinct nets. Embedding the nets into Euclidean space produced 27 candidate structures suitable for geometry optimization.

The DFT calculations presented below were carried out using a DFT code CASTEP.¹⁷ CASTEP determines the electronic ground state of periodic systems by using a plane-wave basis set to represent the wave functions. Nonlocal ultrasoft pseudopotentials¹⁸ were used to describe the electron-ion interaction. Exchange-correlation effects in the electron gas were described in the generalized gradient approximation (GGA) due to Perdew *et al.*¹⁹ Reciprocal space was sampled using the Monkhorst-Pack scheme²⁰ with a distance between grid points of 0.05 \AA^{-1} . The plane-wave basis set cutoff energy was 550 eV.

A BFGS algorithm²¹ was used for geometry optimization. A structure was said to be successfully optimized when the maximum ionic force on all atoms was less than 0.05 eV \AA^{-1} and the maximum component of the stress tensor was less than 0.08 GPa.

IV. RESULTS FOR sp^3 -HYBRIDIZED CARBON POLYMORPHS

Eleven of the 27 candidate structures maintained their original connectivity as they were embedded into Euclidean space and optimized. The geometry, space group, density, and lattice energy for these structures are described in Table I. Structures were optimized in two different ways, either with the symmetry of the original structure constrained or with only the translational symmetry maintained. Each structure led to the same optimized result in both cases, suggesting that the final predicted polymorphs are all stable with respect to small distortions of the lattice.

Diamond and lonsdaelite were produced by the method. The crystal structures and original quotient graphs are illustrated in Figs. 4 and 2, respectively. Diamond appears twice as graph theory produced two nonisomorphic quotient graphs that both correspond to a double primitive cell of diamond.

Three of the unknown structures have a lattice energy within 0.803 eV/atom of diamond and may be metastable under certain temperatures and pressures. The five remaining structures are unlikely to be found in nature as their calcu-

lated lattice energies are more than 1.1 eV/atom greater than diamond. Figure 5 illustrates the unknown structures and the corresponding quotient graphs. All of the structures apart from structure H are less dense than diamond. Following geometry optimization, the carbon-carbon bonds range from 1.430 to 1.736 \AA , which are consistent with experimentally observed distances between sp^3 -hybridized carbon atoms. The most stable structures have bond angles between 90° and 120° .

Sixteen of the structures did not survive the final geometry optimization with their connectivity intact. Two of these structures did not relax to within the required stress constraint, suggesting that they would be unstable. The remaining structures contained some sp^2 -hybridized bonds or consisted purely of sp^3 -hybridized bonds with a different connectivity to the original quotient graph. Those structures that were purely sp^3 -hybridized were the same as some of the final optimized structures that had maintained connectivity throughout the optimization process.

Some of the quotient graphs studied were clearly likely to be very unstable from the outset as they contained three-membered rings. An example of a quotient graph that contained a triangular structure and the optimized structure that did not maintain connectivity are shown in Fig. 6. The quotient graph implies that the net contains neighboring three and four-membered rings; however, the optimized structure

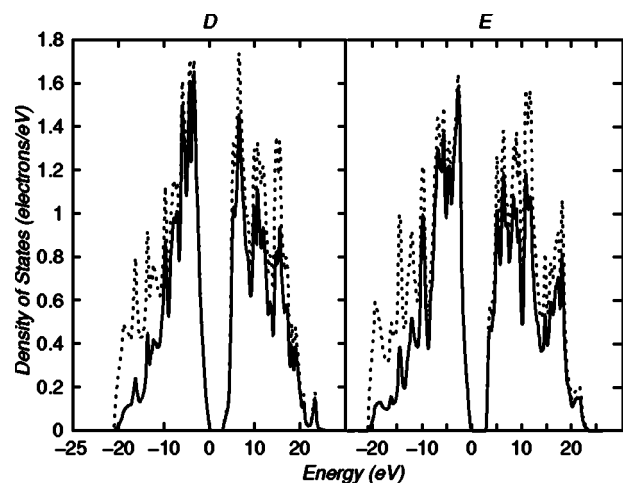


FIG. 7. Total density of states (thin line) and p -projected partial density of states (thick line) for two of the more stable carbon polymorphs.

TABLE II. Density functional theory band gaps and elastic constants for the hypothetical carbon polymorphs.

Structure	Indirect GGA band gap (eV)	Bulk modulus (GPa)	Compressibility (1/GPa)
A	4.5	402.7±1.0	0.00248
B	4.5	401.2±0.9	0.00249
C	3.7	405.3±1.1	0.00247
D	2.7	373.2±0.6	0.00268
E	2.8	324.7±0.5	0.00308
F	0.6	297.4±0.7	0.00336
G	1.1	336.2±0.5	0.00297
H	2.1	343.5±1.2	0.00291
I	0.9	154.0±0.6	0.00649
J	1.5	330.6±1.0	0.00302
K	None	333.2±5.2	0.00300

consists of four- and six-membered rings. Two such quotient graphs maintained their connectivity to produce structures *F* and *I*. However, extra care was needed during optimization. Quotient graphs containing four-membered rings produce stable structures in both the sp^3 and previously studied sp^2 -hybridized cases.¹¹ Four-membered carbon rings are found in cubane and related compounds,^{22,23} but have not as yet been found experimentally in crystals.

Figure 7 illustrates the total density of states (DOS) of two of the more stable polymorphs *E* and *D*. The p -projected DOS may be correlated to an experimentally measured electron energy loss spectrum (EELS), allowing the identification of local phases of polymorph within a sample. A range of DFT calculated elastic constants is shown in Table II. The error shown on the bulk modulus, which expresses the response of the material to an isotropic compression, is due to the fitting scheme used as opposed to an absolute error. It may be seen from the compressibility of the structures that none of them are expected to be stiffer than diamond. The computed indirect GGA band gaps of the set of polymorphs that are insulators range from a diamondlike 4.5 eV to a

siliconlike 1 eV, suggesting the possibility of band-gap engineering using the polymorphs. DFT has a tendency to calculate a theoretical band gap that is smaller than that found experimentally in certain systems;²⁴ hence, the band gaps have been compared to DFT-calculated values throughout.

Fyta *et al.*²⁵ investigated the energetics and thermodynamic stability of nanostructured amorphous carbon including an embedded sample of the sp^2 -hybridized structure polymorph previously predicted by Winkler *et al.*¹⁰ using the graph-theoretical method. The sp^2 -hybridized structure was considered as a possible candidate structure for the fragmentation and distortion of embedded nanotubes in an amorphous matrix. They report that it is possible that the sp^2 -hybridized structure could transform to metastable diamond under appropriate pressure conditions. This recent experimental work²⁵ shows that polymorphs predicted using graph-theoretical methods are of great potential interest in crystallography, chemistry, physics, and materials science.

V. CONCLUSIONS

The combination of graph theory and quantum mechanical calculations described in this paper allows the enumeration of crystal structures for a specified number of atoms per unit cell and coordination of atoms. This investigation has predicted three unknown potentially metastable sp^3 -hybridized carbon polymorphs with four atoms per unit cell; however, this approach could be applied equally well to other potential crystal structures.

While the current study cannot point towards the synthesis routes for such novel polymorphs, the structural and spectroscopic data provided here may aid in their identification. The current results can also be used to study hypothetical polymorphs in structurally related compounds such as silicon carbide and boron nitride.

ACKNOWLEDGMENTS

The work of R.T.S. and C.J.P. was supported by the EPSRC. R.T.S. acknowledges financial support from the Isaac Newton Trust, Cambridge. We thank the German Science Foundation for support from Grant No. Wi1232.

*Electronic address: rts21@cam.ac.uk

¹A. Balaban, X. Liu, D. Klein, D. Babic, T. Schmalz, W. Seitz, and M. Randic, *J. Chem. Inf. Comput. Sci.* **35**, 396 (1995).

²H. R. Karfunkel and T. Dressler, *J. Am. Chem. Soc.* **114**, 2285 (1992).

³R. Hoffman, R. Hughbanks, M. Kertesz, and P. H. Bird, *J. Am. Chem. Soc.* **105**, 4831 (1983).

⁴M. J. Bucknum and R. A. Hoffman, *J. Am. Chem. Soc.* **116**, 11 456 (1994).

⁵O. Friedrichs, A. Dress, D. Huson, J. Klinowski, and A. Mackay, *Nature (London)* **400**, 644 (1999).

⁶E. Kock and W. Fischer, *Z. Kristallogr.* **201**, 407 (1995).

⁷V. Blatov, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **56**, 178 (2000).

⁸E. Peresypkina and V. Blatov, *Acta Crystallogr., Sect. B: Struct. Sci.* **56**, 501 (2000).

⁹S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, *Phys. Rev. Lett.* **91**, 135503 (2003).

¹⁰B. Winkler, C. Pickard, V. Milman, W. Klee, and G. Thimm, *Chem. Phys. Lett.* **312**, 536 (1999).

¹¹B. Winkler, C. Pickard, V. Milman, and G. Thimm, *Chem. Phys. Lett.* **337**, 36 (2001).

¹²A. Wells, *Three-Dimensional Nets and Polyhedra* (Academic Press, New York, 1977).

¹³S. Chung, T. Hahn, and W. Klee, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **40**, 42 (1984).

¹⁴M. Bader, W. Klee, and G. Thimm, *Z. Kristallogr.* **212**, 553 (1997).

- ¹⁵G. Thimm and W. Klee, *Zeolites* **16**, 5 (1997).
- ¹⁶M. Payne, M. Teter, D. Allan, A. Arias, and J. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- ¹⁷M. Segall, P. Lindan, M. Probert, C. Pickard, P. Hasnip, S. Clark, and M. Payne, *J. Phys.: Condens. Matter* **14**, 2717 (2002).
- ¹⁸D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).
- ¹⁹J. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ²⁰H. Monkhorst and J. Pack, *Phys. Rev. B* **13**, 5188 (1976).
- ²¹W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge, England, 1992).
- ²²A. Bashir-Hashemi, *J. Am. Chem. Soc.* **110**, 7234 (1988).
- ²³R. Gilardi, M. Maggini, and P. E. Eaton, *J. Am. Chem. Soc.* **110**, 7232 (1988).
- ²⁴R. W. Godby, *Top. Appl. Phys.* **69**, 51 (1992).
- ²⁵M. G. Fyta, I. N. Remediakis, and P. C. Kelires, *Phys. Rev. B* **67**, 035423 (2003).