Addition patterns, codes and contact graphs for fullerene derivatives

P. W. Fowler,* B. de La Vaissière,* and M. Deza†

*School of Chemistry, University of Exeter, Stocker Road, Exeter, UK EX4 4QD
†CNRS and DMI, Ecole Normale Supérieure, 45 rue d'Ulm, 75230 Paris, France

The mathematical concept of the d-code and its associated contact graph give a model for sterically constrained addition patterns in fullerene derivatives $C_{60}X_m$ and $C_{70}X_m$. In combination with simple electronic arguments, the stoichiometries, symmetries, and location of addends can be predicted, yielding a small number of candidates for further study. For example, sterically and optimal solutions $C_{60}X_m$ with pairwise separation of d bonds between addends are found at m(d) = 24(2), 12(3), 6(4,5), 2(6 to 9). The solution for $C_{60}X_{24}$ is unique, and the model selects 12 candidates for $C_{60}X_{12}$ from a starting set of 11661527060 possibilities. © 2001 by Elsevier Science Inc.

Keywords: fullerenes, derivative, addition patterns, polyhedra, code

INTRODUCTION

Known as a mathematical object since ancient times, the truncated icosahedron is a late-comer to chemistry, with C_{60} appearing first as a theoretical speculation,[1,2] then as the Nobel-Prize-winning explanation of a mass spectrum,[3] and finally as a material in macroscopic quantity.[4] Geometry, symmetry, and graph theory have all played their parts in the construction of systematic theories for the fullerene class of molecules.[5] Now, chemical functionalization of the fullerenes[6,7] brings new challenges and opportunities for making connections between mathematics and chemistry. The sheer number of hypothetically possible addition patterns for C_nX_m ($\sim n! /[(n-m)!m! |G|]$, where $|G|$ is the order of the point group of the bare fullerene) outpaces exhaustive calculation on the fastest computer and makes it urgent to identify broad structural principles to reduce the numbers of structures under consideration to manageable proportions. A combination of steric and electronic arguments has been shown to rationalize the stoichiometry and structure of $C_{60}Br_{24}$, picking out the experimental

isomer from $\sim 3 \times 10^{14}$ possibilities by graph theory alone.[8] A similar model [9] predicts a stoichiometry $C_{70}Br_{26}$ for the end-point of bromination of C_{70}. The present investigation uses some ideas from coding and graph theory to classify the hierarchy of structures for $C_{60}X_m$, $C_{70}X_m$ that is predicted when the dominant factor is the increasing steric bulk of the addend.

BACKGROUND

Fullerenes are pseudospherical polyhedral cages in which each of the *n* constituent carbon atoms has three directly bonded neighbors, and every face is either a pentagon (1 of 12) or a hexagon (1 of *n*/2–10). Thus, the associated chemical graphs, in which vertices stand for atoms and edges for bonds, are *cubic* and *polyhedral*, with face cycles of sizes 5 and 6. *Distance* between any two vertices of a graph is defined as the number of edges in the shortest path between them, and the *diameter*, D, is the largest pairwise distance in the graph (see, e.g., Matoušek and Nešetril[10] for standard graph theoretical terminology).

Though formally distinct, graph theoretical distance and ordinary geometrical separation can be expected to correlate in more-or-less spherical structures, and this suggests the use of the former to deal with steric effects in fullerenes. In earlier studies[8], the idea of independence number was used in this connection. The *independence number* of a graph G is the maximum size of a subset H of vertices of G, such that no two vertices of H are neighbors in G. Clearly, this matches the chemical situation where the bulk of a given addend prevents addition to an immediately neighboring site. The notion of independence number is generalized by the *d-code*.

A *d*-code (distance code) in G is a subset of vertices such that all pairwise distances within the set are $\geq d$. This terminology comes from the theory of error-detecting and error-correcting codes,[11] and usually applies to subsets of vertices of the *n*-dimensional cube, but the idea can be used for any graph. In these terms, an independent set is a 2-code, and the independence number is the cardinality (size) of the largest 2-code. The general *d*-code represents the situation where a ligand casts a steric shadow that precludes addition to all vertices at distances $\leq d - 1$. In the remainder of the paper, the term *d*-code will be taken to mean a largest *d*-code unless otherwise specified.

Corresponding author: P.W. Fowler, School of Chemistry, University of Exeter, Stocker Road, Exeter, UK EX4 4QD. Tel.: 44-1392-263466; fax: 44-1392-263434.

E-mail address: p.w.fowler@exeter.ac.uk (P.W. Fowler)

A given graph G has a series of invariants $|C_d|$, the sizes of its d-codes from $d = 1$ to $d = D$. Trivially, $|C_1|$ for an n-vertex graph and $|C_D| = 2$ for an antipodal graph. Intermediate values of d for a fullerene graph may admit multiple distinct solutions, each of which is the basis for a plausible chemical structure, when steric bulk is the only consideration.

A useful way to visualize these distinct d-codes is by their *contact graphs*. The contact graph for a d-code can be defined as follows: it has the same vertices as the code, and an edge for every pair of vertices at distance d. When taken with a metric realization of a fullerene graph, the contact graph gives a 3D visualization of the packing of the bulky ligands around the central carbon cage. In some case, the contact graph will be polyhedral, but in many others it will have divalent, monovalent, or even isolated vertices. It is easy to see that there will always be at least one *connected* (in the graph theoretical sense) d-code for a given d; graphs with *disconnected* (isolated) vertices correspond to distinct chemical isomers with some atoms force to 'rattle' inside the 'box' imposed by local steric constraints.

Calculation of the d-codes for $d > 2$ reduces to the problem of finding the independence numbers of series of graphs. The original graph G is augmented with extra edges for all pairwise distances $<d$, and the 2-code of the new graph is the d-code of G. Calculation of the independence number and enumeration of the distinct independent sets is an NP-complete problem[12], but the fullerenes of interest, C_{60} and C_{70}, are small enough. The remainder of the present paper deals with the d-codes, their corresponding addition patterns and contact graphs, for these two archetypal fullerenes.

RESULTS

Straightforward calculation gives the diameters of C_{60} and C_{70} graphs as 9 and 10, respectively. Construction of the adjacency lists and application of Hansen and Mladenovic's program for stable sets[12] gives the sizes of the various d-codes in Table 1.

For each value of d, d-codes were generated and sorted to give complete lists of symmetry-distinct isomers, $C_{60}X_{|C_d|}$ and $C_{70}X_{|C_d|}$ with the appropriate minimum separation of addends. The numbers and symmetries of these isomeric patterns are listed in Table 2 and Table 3. Perhaps most striking is the way in which imposition of the steric criterion leads to a reduction of many orders of magnitude in the isomer count. In the case of the $C_{60}X_{24}$, the $\sim 3 \times 10^{14}$ ways of attaching 24 atoms to the cage are reduced by the steric 2-code criterion to 1085, and for $d > 2$ the number of codes never exceeds 12. Similar reductions are seen for $C_{70}X_{|C_d|}$.

From a purely steric point of view, the listed d-codes solve completely the problem of maximum functionalization of C_{60} and C_{70}. However, for practical applications, it is necessary to take account of another important consideration: the

π-electronic configuration of a candidate molecule. The π electrons of the fullerene C_n occupy $n/2$ molecular orbitals distributed over the n vertices and $3n/2$ edges of the molecular graph. Each addition removes a π electron from the count and cuts out one vertex and three edges from the graph. After successive additions, the molecule has several separate π systems, corresponding to disjoint subgraphs of unoccupied carbon vertices. If the final product is not to be a radical, the addition pattern should have an overall closed-shell π structure, with all electrons of each subgraph paired up in orbitals corresponding to positive eigenvalues of the local adjacency matrix. Barring energetically unlikely zwitterionic charge transfer between subsystems, the overall pattern should obey the rule that each separate π system in the addition pattern should have a closed π shell, i.e., an adjacency spectrum for each subgraph split into equal halves across a nonzero gap. A corollary is that each subgraph should contain an even number of 'empty' vertices. Therefore, d-codes of odd size are automatically excluded.

Tables 2 and 3 give the numbers of survivors on application of the even-subgraph rule. This rule eliminates 1084 of the 1085 2-codes $C_{60}X_{24}$, both of the 4-codes $C_{60}X_7$ and all the 6-codes $C_{60}X_3$. For C_{70}, the 2-codes, 3-codes, and 7-codes are all discarded. In the case of the C_{70} 2-code, it is necessary to drop back to 26 addends to obtain an independent set that has a closed electronic shell: 26 is said to be the *closed-shell independence number* for the C_{70} fullerene cage.[9] In the other cases where $|C_d|$ is odd, it is easy to find nonradical molecules with $|C_d| - 1$ addends and the same minimal pairwise separation. Tables 2 and 3 list the counts for these isomers, which represent the best compromise between steric and electronic effects.

In some cases, codes for which all subgraphs are of even size are still unfavorable on π-electronic grounds, for example when they have a small or vanishing gap in Hückel theory, but as this is a model-dependent circumstance, it may be risky to exclude such isomers. The differences $N_e - N_c$ in Tables 2 and 3 give the numbers of isomers affected by this choice.

(i) C_{60} codes: the unique closed-shell solution to the 2-code problem for C_{60} is the T_h-symmetric $C_{60}X_{24}$, which corresponds exactly to the experimental structure of $C_{60}Br_{24}$.[8] The 3-codes for C_{60} have not been considered before. Figure 1 shows the full set of twelve 3-codes $C_{60}X_{12}$ selected from $\sim 10^{10}$ by this simple mathematical criterion. Isomers 2, 5, and 12 each have a doubly degenerate nonbonding level in the 48-atom π system produced by the addition pattern and are thus 'accidental' open-shell molecules. By $d = 4$, addends have become so large that no ring of C_{60} (hexagonal or pentagonal) may include more than one functionalized position. All 4-codes of C_{60} are intrinsically radical $C_{60}X_7$ molecules, but many

Table 1. Fullerene d-codes

d	2	3	4	5	6	7	8	9	10
$|C_d|$:C_{60}	24	12	7	6	3	2	2	2	
$|C_d|$:C_{70}	29	15	8	6	4	3	2	2	2

At each distance d, $|C_d|$ gives the size of the code, i.e. the maximum number of sterically independent addends, for C_{60} and C_{70} cages.

Table 2. Sterically favoured isomers of $C_{60}X_m$

d	2	3		4		4		5	6	6	7	8	9
$\|C_d\|$	24	12		7		6		6	3	2	2	2	2
N_{all}*	*~10^{14}	*~10^{10}		3220218		418470		418470	303	23	23	23	23
T_h	1	1	0	0	0	0	0	0	0	0	0	0	0
D_{3d}	0	0	0	0	0	1	1	1	0	0	0	0	0
T	1	0	0	0	0	0	0	0	0	0	0	0	0
D_{2h}	0	0	0	0	0	0	0	0	0	0	0	0	0
S_6	3	0	1	0	0	2	1	0	0	0	0	0	0
C_{3v}	0	0	0	0	0	1	1	0	1	0	0	0	0
D_3	1	0	0	0	0	1	1	0	0	0	0	0	0
C_{2h}	0	0	0	0	0	1	1	0	0	1	1	1	1
C_{2v}	0	0	0	0	0	0	0	0	0	1	1	1	0
D_2	0	0	0	0	0	0	0	0	0	0	0	0	0
C_3	17	0	0	0	0	3	3	0	1	0	0	0	0
C_i	5	0	1	1	0	2	2	0	0	0	0	0	0
C_s	0	0	0	0	0	4	4	0	1	1	0	0	0
C_2	20	0	3	3	0	17	17	0	0	3	3	1	0
C_1	1037	0	7	5	2	151	151	0	3	3	1	0	0
N_d	1085		12		2	183		1	6	9	6	3	1
N_e	1		12		0	183		1	0	9	6	3	1
N_c		1		9			182	1	0	9	6	3	1

For each distance d and number of addends $\|C_d\|$, there are N_{all} symmetry distinct isomers $C_{60}X_m$ in the absence of further restriction. Of these, N_d obey the steric condition (i.e. are d-codes), N_e are d-codes with no odd sub-graphs, and N_c are d-codes with formal closed shells in Hückel theory. A breakdown by symmetry group is given for N_d and the other sets where necessary. When $\|C_d\|$ is odd, the table also shows a breakdown of closed-shell isomers $C_{60}X_m$ at the highest value of m compatible with both steric and electronic conditions, where the definitions of N_d, N_e and N_c are now extended to these non-sterically optimal sets.
* Exact values: 24: 300436595453640 and 12: 11661527060.

solutions $C_{60}X_6$ are available (Table 2). Codes with $d > 4$ are perhaps of more limited chemical interest and are not discussed in detail here.

(ii) C_{70} codes: C_{70} has odd values $\|C_2\|$ and $\|C_3\|$. The case of the 2-code has been discussed in detail elsewhere[9]: a closed shell with all distances > 1 demands that

Table 3. Sterically favoured isomers of $C_{70}X_m$

d	2	3	3			4	5	6	7	7	8	9	10
$\|C_d\|$	29	15	14			8	6	4	3	2	2	2	2
N_{all}*	~10^{18}	~10^{13}	~10^{12}			~10^{8}	6561107	46275	2792	143	143	143	143
D_{5h}	0	0	0	0	0	0	0	0	0	0	0	0	0
D_5	0	0	0	0	0	0	0	0	0	0	0	0	0
C_{5v}	0	0	0	0	0	0	0	0	0	0	0	0	0
C_{5h}	0	2	0	0	0	0	0	0	0	0	0	0	0
C_5	0	1	0	0	0	0	0	0	0	0	0	0	0
C_{2v}	0	0	7	0	0	2	0	2	0	3	2	1	0
C_2	0	0	28	21	20	8	5	2	0	9	5	5	1
C_s	118	0	11	0	0	2	1	1	0	2	1	0	0
C_s'	0	16	58	32	24	6	0	0	0	1	1	0	0
C_1	2854	36	1769	685	645	226	6	12	2	28	14	4	0
N_d	2972	55	1873			244	12	17	2	43	23	10	1
N_e	0	0		738		244	12	17	0	43	23	10	1
N_c	0	0			689	244	12	17	0	43	23	10	1

Notation and conventions are as listed in the caption of Table 2. The subgroup C_s retains one of the five equivalent 'vertical' mirror planes of D_{5h} and C_s' is based on the unique 'horizontal' mirror plane, For $d = 2$ there are only 10 closed-shell isomers $C_{70}X_{26}$[9]: 2 of C_2 symmetry, 1 of C_s symmetry, with C_1 symmetry.
* Exact values: 29: 2024917322816219232, 15: 36074046545320, 14: 9662695546292, 8: 472064572.
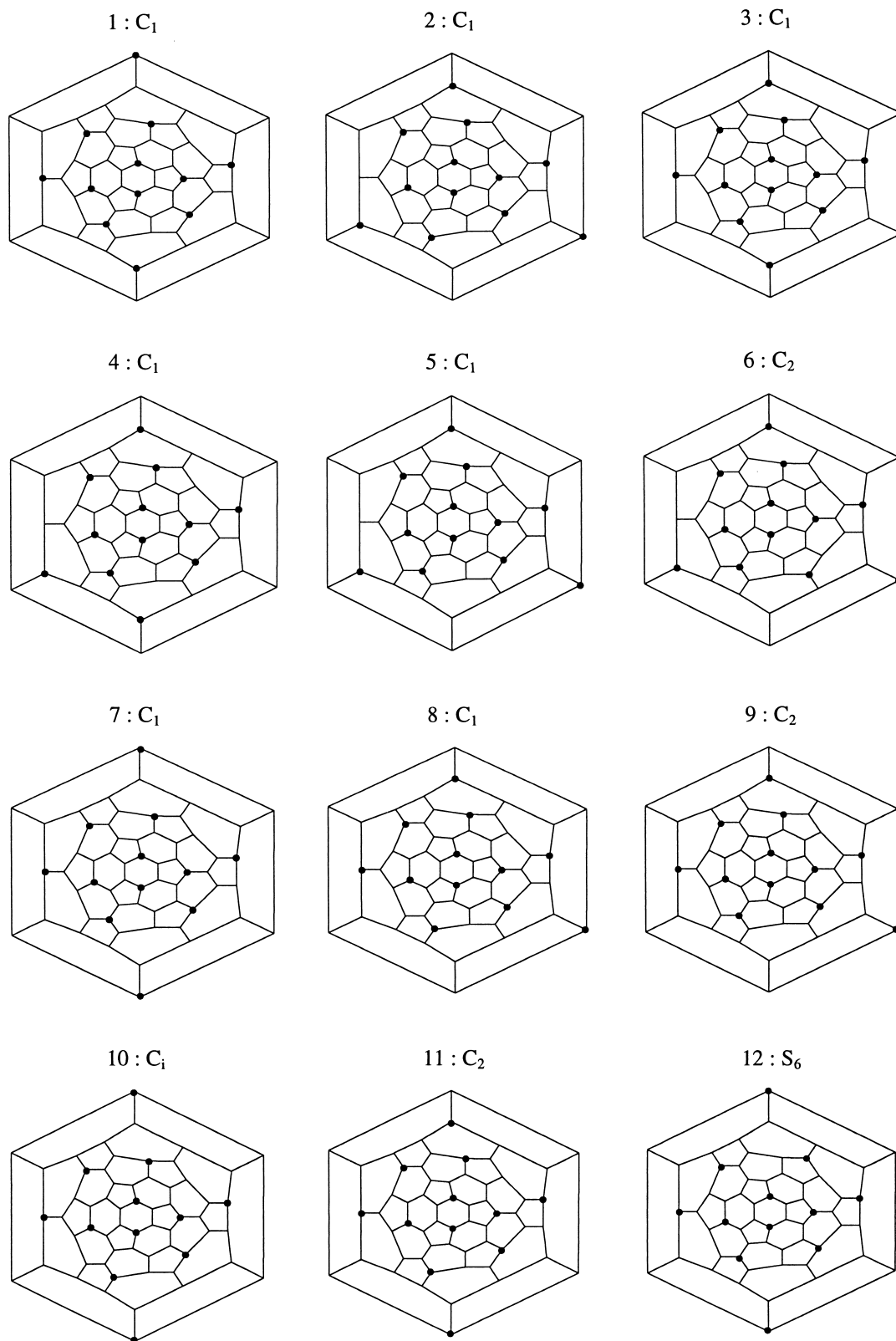
Figure 1. Sterically optimal isomers $C_{60}X_{12}$. Each isomer is shown as a Schlegel diagram, with filled circles representing addition sites. All X pairs are separated by at least three bonds, i.e., these are 3-codes of the truncated icosahedron.

addition stops at $C_{70}X_{26}$, three steps below $|C_2| = 29$. For the 3-codes, the electronic factor again induces a shortfall, this time of only 1 addend. Multiple solutions with all distances >2 for $C_{70}X_{14}$ are available (Table 2). For $d = 4$ all 244 codes $C_{70}X_8$ are closed-shell.

Some further insight into the steric implications of d-codes
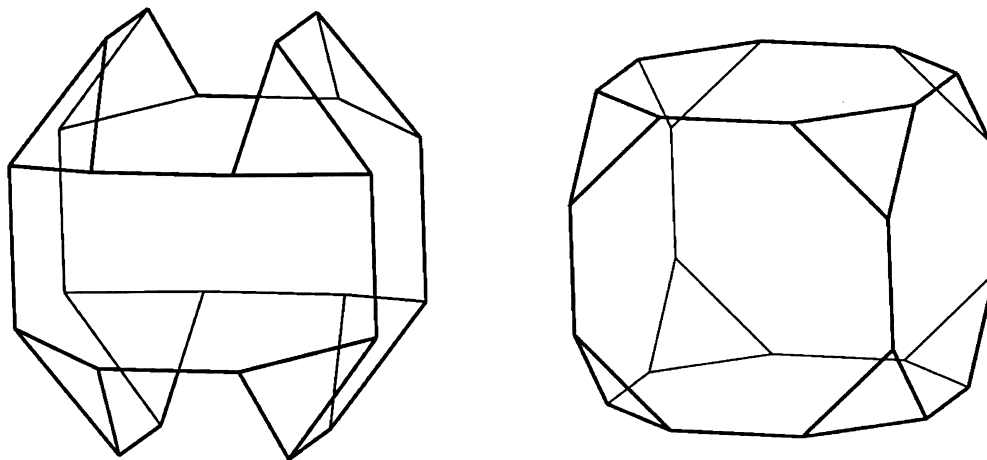
Figure 2. Contact graph of the 2-code $C_{60}X_{24}$ (T_h isomer). These graphs have 24 vertices of degree 3 and 36 edges. When drawn with the physical coordinates of the sp^3 sites (left), the contact graph has T_h symmetry; when realized with planar equilateral equiangular faces (right), it is revealed as the truncated cube with O_h symmetry.

can be gained from their contact graphs. As defined earlier, two vertices of a d-code are connected in the contact graph when their distance is exactly d, notionally indicating that the corresponding addends are in van der Waals contact. Some codes are more efficient than others in achieving this packing of the addends. For example, of the 1085 2-codes on C_{60}, the one with the minimal number of edges in its contact graph is the unique T_h isomer. This graph has 24 vertices of degree 3 and hence 36 edges; a realization as a truncated cube is shown in Figure 2. Addition of 12 edges linking vertices at distance 3 would give a rhombicuboctahedron. Both polyhedra have realizations with O_h symmetry, even though O_h is not a subgroup of the I_h group of the underlying C_{60} cage. In general, contact graphs may have higher symmetry than the molecule they represent.

Only one other 2-code on C_{60} has a regular contact graph; this is the T isomer $C_{60}X_{24}$ in which all vertices of the contact graph are of degree 4. A realization of this graph as a polyhedron (of maximal symmetry T_d) is also shown in Figure 3. Whereas in the truncated cube each triangular face arises from 1, 3, 5 addition to a hexagon of C_{60}, in this T_d polyhedron, 8 of the triangles arise from such additions, but 12 arise from isolation of a fullerene vertex by 3 addends at mutual distance 2. If a contact graph is assigned a vertex-degree signature $\{V_1, V_2, V_3, V_4, V_5\}$, 155 different signatures are found among the 1085 C_{60} 2-codes. The maximum number of edges in these graphs is 48 and all have $V_1 = 0$, $V_2 \leq 6$, $V_3 \leq 24$, $V_4 \leq 24$, and $V_5 \leq 10$.

Other interesting contact graphs include the unique 5-code, which is a D_{3d}/O_h octahedron, and the 7-vertex graphs for the two 4-codes, which are isomorphic edge-bridged tetrahedra (Figure 4). In the IUPAC numbering sheme.[13] the addend positions for the two 4-codes are 1, 8, 20, 28, 33, 45, and n, where n is either 51 or 55, and for the 5-code the addends occupy sites 1, 16, 29, 38, 43, and 60.
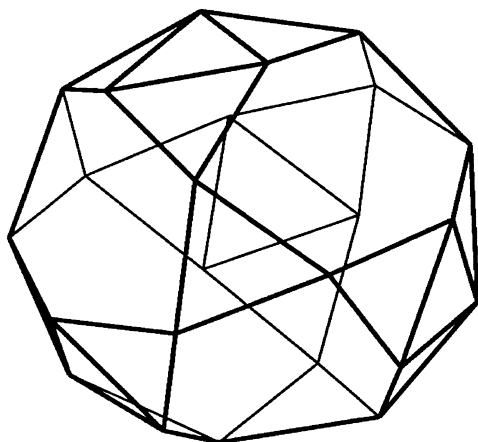

Figure 3. Contact graph of the 2-code $C_{60}X_{24}$ (T isomer). This graph has 24 vertices of degree 4 and 48 edges. If physical coordinates of addend atoms are used, it has only T symmetry, but the equilateral polyhedron has maximal T_d symmetry. The polyhedron is formally derived by truncating the cube on a tetrahedral set of vertices and then constructing the medial.
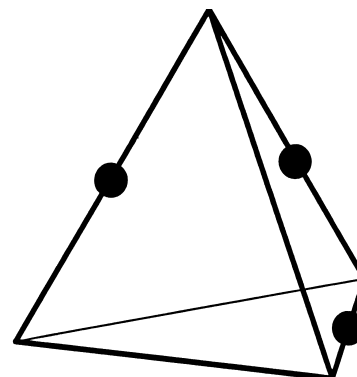

Figure 4. Contact graph of both 4-codes $C_{60}X_7$. This graph has 7 vertices and 9 edges. In its most symmetrical form it is a tetrahedron with 3 edges bridged by divalent vertices.

CONCLUSION

The *d*-code gives a framework for treating steric effects in fullerene derivatives. When combined with simple π-electronic theory it results in a purely graph theoretical model for selecting chemically plausible candidates from this otherwise intractable plethora of possibilities.

2-codes have already proved valuable for the case of $C_{60}Br_{24}$, where the model gives a unique prediction of the experimental isomer. It will be interesting to see if the structures predicted at the 3-code level and beyond find a place in fullerene chemistry in the future.

ACKNOWLEDGEMENT

REFERENCES

1 Osawa, E.**.** *Kagaku.* 1970, 25**,** 654 ; Yoshida, Z., and Osawa, E. In: *Aromaticity.* Eds Kagakudojin, Kyoto, 1971, 194

2 Davidson, R.A. Unified theory of graph spectral reduction networks. Spectral analysis of graphs by cyclic automorphism subgroups. *Theor. Chim. Acta.* 1981, **58**, 193–231

3 Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., and Smalley, R.E. C_{60}-Buckminsterfullerene. *Nature*, 1985, **318**, 162–163

4 Krätschmer, W., Lamb, L.D., Fostiropoulos, K., and Huffman, D.R. Solid C_{60} —A new form of carbon. *Nature*, 1990, **347**, 354–358

5 Fowler, P.W., and Manolopoulos, D.E. *An atlas of fullerenes.* Clarendon Press, Oxford, 1995

6 Taylor, R. *Lecture notes on fullerene chemistry.* Imperial College Press, London, 1999

7 Hirsch, A. *The chemistry of fullerenes.* Thieme Medical Publishers, New York, 1994

8 Fowler, P.W., Hansen, P., Rogers, K.M., and Fajtlowicz, S. $C_{60}Br_{24}$ as a chemical illustration of graph theoretical independence. *J. Chem. Soc., Perkin Trans. 2*, 1998, 1531–1533

9 Fowler, P.W., Rogers, K.M., Somers, K.R., and Troisi, A. Independent sets and the prediction of addition patterns for higher fullerenes. *J. Chem. Soc., Perkin Trans. 2*, 1999, 2023–2027

10 Matoušek, J., and Nešetril, J. *Invitation to discrete mathematics.* Oxford University Press, Oxford, 1998

11 MacWilliams, F.J., and Sloane, N.J.A. *Theory of error-correcting codes.* North Holland, Amsterdam, Second edition, 1978

12 Mladenovic, N., and Hansen, P. Variable neighborhood search. *Computers and operations research*, 1997, **24**, 1097–1100

13 Godly, E. W., and Taylor, R. Nomenclature and terminology of fullerenes: A preliminary survey. *Fuller. Sci. and Techn.* 1997, **5**, 1667–1708