Graph Theoretic and Electronic Properties of Fullerenes & Biasing Molecular Modelling Simulations with Experimental Residual Dipolar Couplings

Lukas N Wirz

2015

A thesis presented in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
at
Massey University, Albany
New Zealand

Contents

Acknowledgements

There are several people I have to thank for making my last 3.5 years possible and enjoyable: Peter Schwerdtfeger, for giving me the opportunity to come to Massey University, and shielding all of us from the administrators. Thank you for support and a lot of freedom. · Jane Allison, for being extremely motivating, always taking time to answer and discuss my questions, and for the cake! · James Avery, for teaching me how to write less bad code, how to think and write as a mathematician (for chemists) and patience when either failed. Thank you for inviting me to stay in Copenhagen for two-and-a-half enjoyable and productive months. · Florian Senn and Lukáš Pašteka for useful discussions, advice, and encouragement when I was in need of it. · Andrew Punnett for taming the bits and electrons (as well as explaining how to) on smaller and larger computers, whenever and in whichever way they misbehaved. · Jayson Cosme, Lukáš Pašteka, and Lukas Trombach for the daily fix of ping pong (table tennis, to be accurate). · And everyone else in CTCP for the throughout friendly and encouraging atmosphere.

List of Publications

[1] Schwerdtfeger P, Wirz L, Avery J. Program fullerene: A software package for constructing and analyzing structures of regular fullerenes. *J Comput Chem* 2013, 34:1508–1526. doi:10.1002/jcc.23278.

[2] Wirz LN, Tonner R, Avery J, Schwerdtfeger P. Structure and properties of the nonface-spiral fullerenes T-C_{380}, D_3-C_{384}, D_3-C_{440}, and D_3-C_{672} and their halma and leapfrog transforms. *J Chem Inf Model* 2014, 54:121–130. doi:10.1021/ci4005578.

[3] Schwerdtfeger P, Wirz LN, Avery J. The topology of fullerenes. *WIREs Comp Mol Sci* 2015, 5:96–145. doi:10.1002/wcms.1207.

[4] Wirz LN, Allison JR. Fitting alignment tensor components to experimental RDCs, CSAs and RQCs. *J Biomol NMR* 2015, 62:25–29. doi:10.1007/s10858-015-9907-x.

[5] Wirz LN, Tonner R, Hermann A, Sure R, Schwerdtfeger P. From small fullerenes to the graphene limit: A harmonic force-field method for fullerenes and a comparison to density functional calculations for Goldberg–Coxeter fullerenes up to C_{980}. *J Comput Chem* 2016, 37:10–17. doi:10.1002/jcc.23894.

[6] Wirz LN, Allison JR. Comment on "a tensor-free method for the structural and dynamic refinement of proteins using residual dipolar couplings". *J Phys Chem B* 2015, 119:8223–8224. doi:10.1021/acs.jpcb.5b02801.

[7] Wirz LN, Babić D, Avery J, Schwerdtfeger P. Toward tight upper and lower bounds for Hamilton cycles in fullerene graphs 2015.

[8] Sundholm D, Wirz LN, Schwerdtfeger P. Novel hollow all-carbon structures. *Nanoscale* 2015, 38:15886–15894. doi:10.1039/C5NR04370K.

Abstract

In this thesis two different models, that is levels of abstraction, are used to explore specific classes of molecular structures and their properties.

In part I, fullerenes and other all-carbon cages are investigated using graphs as a representation of their molecular structure. By this means the large isomer space, simple molecular properties as well as pure graph theoretical aspects of the underlying graphs are explored. Although chemical graphs are used to represent other classes of molecules, cavernous carbon molecules are particularly well suited for this level of abstraction due to their large number of isomers with only one atom type and uniform hybridisation throughout the molecule.

In part II, a force field for molecular dynamics, that is the step wise propagation of a molecular structure in time using Newtonian mechanics, is complemented by an additional term that takes into account residual dipolar couplings that are experimentally measured in NMR experiments. Adding this force term leads to more accurate simulated dynamics which is especially important for proteins whose functionality in many cases crucially depends on their dynamics. Large biomolecules are an example of chemical systems that are too large for treatment with quantum chemical methods but at the same time have an electronic structure that is simple enough for accurate simulations with a forcefield.

Part I

Graph theoretical and electronic properties of fullerenes

The following six chapters investigate graph theoretical, topological, and electronic properties of fullerenes and related carbon cages. Fullerenes are molecular cages, that consist exclusively of trivalent carbon atoms; all their faces are either pentagons or hexagons. As the molecules increase in size, the number of their isomers grows rapidly, giving rise to a vast number of fullerenes with a wide range of properties, some of which only emerge in very large graphs for the first time.

All of the following chapters are based on previously published/prepared articles. The contribution of the authors is detailed on pages 299 ff.

Chapter 1 is an introduction to fullerenes including historical and experimental aspects but with a focus on their graph theoretical and topological properties.

In chapter 2 a program is introduced for creating, modifying and characterising fullerene structures and graphs. This allows for the efficient exploration of this large field of molecules and underlying graphs. The functionality and usage of the program are described as well as implementational details.

The remaining four chapters of this part are on specific topics that were explored making use of the program *Fullerene*. Results on these topics are presented and the respective algorithms employed are explained in more detail.

Chapter 3 explores the properties of fullerene graphs and the respective fullerenes that do not allow for any face spirals. A generalised face spiral algorithm that can be applied to any polyhedral graph and its implementation are explained.

Chapter 4 explains in detail the two force-field implementations in the *Fullerene* program. One of the implementations is parametrised against calculated vibrational frequencies. The performance of the resulting force field is assessed by comparison with quantum chemically optimised structures and extrapolation of properties to the graphene limit.

Chapter 5 explores the number of Hamilton cycles in fullerene graphs by counting their number for fullerenes up to C_{120} and establishing upper and lower bounds. The exact number of Hamilton cycles is derived for (5,0) and (6,0) nanotubes, thus proving the exponential growth.

In chapter 6 several novel classes of carbon cages are introduced that locally resemble graphynes and graphdiynes. A constructive algorithm is described for creating all members of these molecule classes. Example structures are optimised with a force field and at DFT level of theory. Subsequently electromagnetic properties are calculated.

1 Introduction[a]

Buckminsterfullerene C_{60}-I_h is a convex and "spherical" molecule (in the sense that the atoms lie on the surface of a sphere) with a highly symmetric icosahedral structure. It was originally conjectured independently by Ōsawa in 1970[2,3] and Stankevich, Bochvar and Gal'pern in 1973,[4,5] discovered by mass spectrometry in 1985 by Kroto et al. through laser evaporation of graphite,[6–9] and synthesized in larger amounts by Krätschmer et al. in 1990[10] (the first mass-spectroscopic studies on small carbon cluster were performed by Mattauch et al.[11] in 1943). Fullerenes are the polyhedral analogs to the two-dimensional graphene sheets.[b] They occur in nature and have been detected in interstellar space, albeit in minute amounts.[12–15] In general, (classical) fullerenes are cage-like, hollow molecules of pseudo-spherical symmetry consisting of pentagons and hexagons only, resulting in a trivalent (and in the most ideal case) convex polyhedron with exactly three edges (bonds) joining every vertex occupied by carbon, idealized as sp^2 hybridized atoms.[16] In graph theoretical terms, fullerenes belong to the class of cubic, planar, 3-connected and simple graphs, consisting of 12 pentagons only (the *12 Pentagon Theorem*) and F_6 hexagons. This gives the general formula of C_{20+2F_6} ($F_6 \geq 0$ and $F_6 \neq 1$) for carbon fullerenes, with the number of hexagonal faces $F_6 = 20$ for C_{60}. The smallest possible fullerene is C_{20} ($F_6 = 0$), a dodecahedron consisting of 12 connected pentagons,

[a] This chapter is composed of sections that are taken from the previously published article "The topology of fullerenes" by Schwerdtfeger et al.[1] and is reproduced with kind permission from the authors and John Wiley & Sons, Ltd. The candidate's contribution to the article is itemised on pages 299 ff. Of the article, three sections are omitted from this chapter because their topics are dealt with in more depth in chapters 3, 4, and 5 of this thesis. Three more sections of the article are omitted for brevity and because they do not support the original work presented in the following chapters. The introduction to graph theory (section 1.1) has been extended relative to the article.

[b] If we grow Goldberg-Coxeter transforms of C_{20} to infinity we obtain spherical graphene sheets connected by 12 pentagons. For definitions see below.

and the only Platonic solid in the family of fullerene polyhedra.[17] The truncated icosahedron C_{60} belongs to the class of Archimedean solids.[18] C_{60}-I_h is the minimal arrangement such that no two pentagons share an edge (isolated pentagons): The edges of each pentagon join only hexagons, and the edges of each hexagon alternately join pentagons and hexagons. Thus, it is the smallest member in the class of fullerenes fulfilling the so-called *isolated pentagon rule* (IPR), which brings thermodynamic stability to a fullerene cage.[19]

There are infinitely many fullerenes, the number of isomers growing as $\mathcal{O}\left(N^9\right)$ for N carbon atoms as shown in Figure 1.1.[20] The reason for the N^9 growth comes from algebraic geometry, and is discussed below. These isomers come in many different shapes depending on the distribution of the pentagons as shown in Figure 1.2.[c] As experimental techniques advance to synthesize new fullerene isomers[22] with many interesting chemical and physical applications,[23–26] see for example the recent bottom-up synthesis of C_{60} by Scott et al.[27] and Kabdulov et al.,[28] one naturally wants to know more about their thermodynamic stability and electronic properties.[29] As the isomer space rapidly grows into the millions and beyond, this becomes a daunting task for larger fullerenes. Even when the search is restricted to fullerenes that have isolated pentagons, known to yield the best stability, the number of candidates for most thermodynamically stable isomers is huge, many orders of magnitude too large for quantum chemical methods to be a viable means of analyzing stability. But, remarkably, the fullerene graphs contain all the information we need, and we are able to sort through the millions of isomers, finding a few candidates for the most stable, by way of simple, easily computed topological indicators,[30,31] such as the neighbor indices for pentagons or hexagons,[32] or the incremental assignment of heat of formations to certain face patches.[33,34]

There has been an intense activity in the field of topological and graph theoretical descriptions of molecules such as fullerenes[30,32,37–41] over the past 20 years, to the extent that it has become a major sub-discipline within mathematical chemistry. It comes perhaps as a surprise that graph theory applied to molecular structures has not yet made it into main-stream university teaching. Furthermore, since the publication of the standard reference

[c] An interesting side aspect is that fullerene-like shapes are realized also in viral shapes.[21]

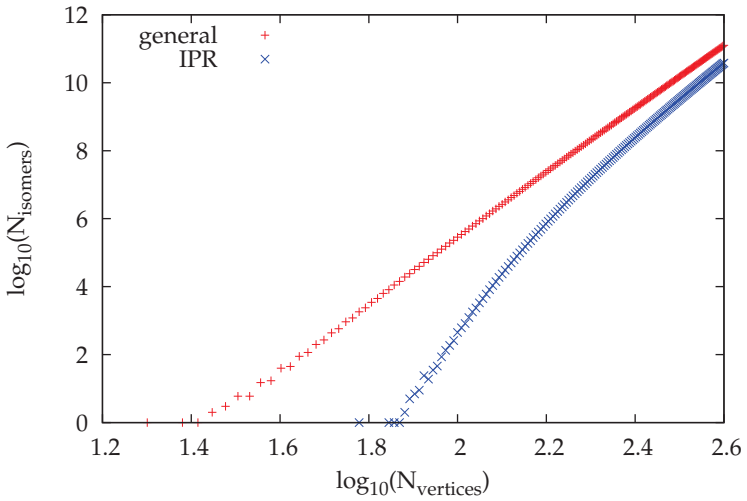


Figure 1.1 Number of distinct (non-isomorphic) fullerene isomers C_N (with and without fulfilling the IPR) with increasing number of carbon atoms N up to $N = 400$ (double logarithmic scale). Data taken from the House of Graphs (Ref.[35]).

work in this area, *An Atlas of Fullerenes* by Fowler and Manolopoulos,[32] there has not been a comprehensive review on this subject highlighting the many activities in recent times. In contrast, there are excellent reviews and books available on the chemistry and physics of fullerenes[23,24,42–48] or on their electronic structure calculations.[29] In this overview we aim to close this gap and report on the many new and exciting developments in the topological and graph theoretical treatment of fullerenes, which have taken place over the past decade or so.[d]

[d]Most of the useful graph theoretical and topological aspects discussed here are implemented in a Fortran/C++ program called *Fullerene*, an open-source code freely available at the Massey University website.[36]

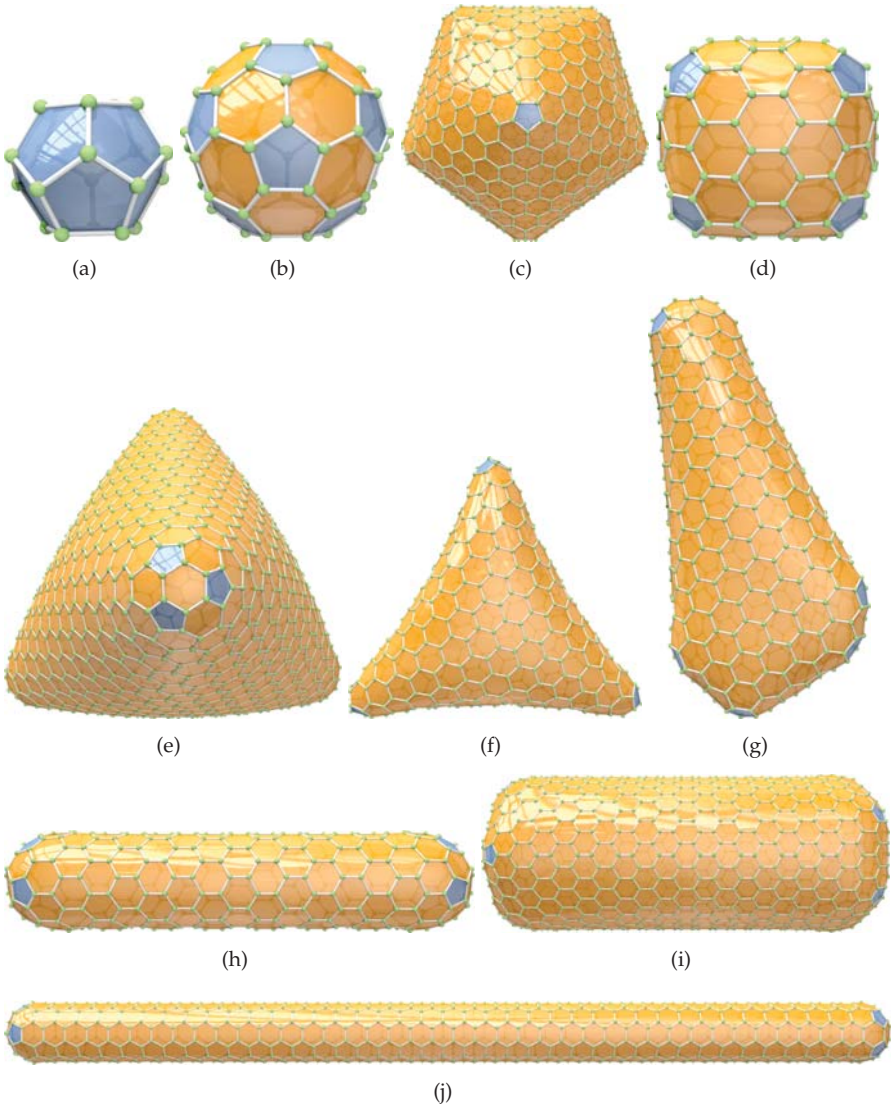


Figure 1.2 A selection of different 3D shapes for regular fullerenes (distribution of the pentagons D_P are set in parentheses). "Spherically" shaped (icosahedral), e.g., (a) C_{20}-I_h, (b) C_{60}-I_h and (c) C_{960}-I_h ($D_P = 12 \times 1$); barrel shaped, e.g., (d) C_{140}-D_{3h} ($D_P = 6 \times 2$); trigonal pyramidally shaped (tetrahedral structures), e.g., (e) C_{1140}-T_d ($D_P = 4 \times 3$); (f) trihedrally shaped C_{440}-D_3 ($D_P = 3 \times 4$); (g) nano-cone or menhir C_{524}-C_1 ($D_P = 5 + 7 \times 1$); cylindrically shaped (nanotubes), e.g., (h) C_{360}-D_{5h}, (i) C_{1152}-D_{6d}, (j) C_{840}-D_{5d} ($D_P = 2 \times 6$). The fullerenes shown in this figure and throughout the paper have been generated automatically using the *Fullerene* program.[36]

1.1 A Short Introduction to Graph Theory and Embeddings

A *graph* is a pair $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ defines a finite and non-empty set of vertices, and $\mathcal{E}$ is a (possibly empty) set of edges, each an unordered pair connecting two distinct vertices. I.e. one edge cannot connect one vertex with itself and a pair of vertices cannot be connected by more than one edge. Vertices and edges are commonly visualised by dots and lines connecting dots, respectively.

A graph in which edges are not required to be distinct, i.e., in which two vertices can be connected by more than one edge, is called a *multigraph*. If, furthermore, the vertices in an edge need not be distinct, i.e., edges may form loops, the graph is called a *general graph*. A *subgraph* SG of a graph G is a graph whose vertex set and edge set are subsets of the respective vertex set and edge set of G. A graph in which edges are ordered pairs of vertices is a *directed graph*, or *digraph*. Directed edges can be visualised as arrows.

A sequence of distinct vertices $v, \ldots, v_l, v_k, \ldots$ where any consecutive pair of vertices is joined by an edge is called a *path*. If a path consists of directed edges in their correct orientation it is called a *directed path*. If the first and last vertex of the path or directed path coincide it is called a *circuit* or *cycle*. If there exists a path between any two vertices in a graph, the graph is called *connected*, otherwise it is *disconnected*. A digraph is called *strongly connected* if there is a directed path between any two vertices. If every vertex in a connected graph has at least degree n the graph is called *n-connected*. This is equivalent to the statement that at least n edges need to be removed to disconnect an n-connected graph.

A path (circuit/cycle) that includes every vertex of a graph is called a *Hamiltonian path* (*Hamiltonian circuit/cycle*). A graph that features at least one Hamiltonian circuit, is called a *Hamiltonian graph*.

A *perfect matching* is a pairwise assignment of every vertex in a graph to another vertex with which it shares a bond such that no vertex is left.

If a graph can be drawn in the two-dimensional plane without edge crossings, then it is called *planar*, and in this case it can also be drawn using only straight lines. A *planar embedding* is a map $\mathcal{V} \to \mathbb{R}^2$ assigning a 2D-coordinate to each vertex such that straight-line drawings of the edges do

not cross. For planar graphs in general, planar embeddings are not unique. But whenever at least three vertices must be removed before the graph separates into two disconnected components, the graph is called *3-connected*, and there is essentially only one way[e] to embed it in the plane.[49] Because of this, a 3-connected planar graph has a well-defined set of *faces*, $\mathcal{F}$. We can therefore represent 3-connected planar graph as $G = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ without ambiguity, and will do so where convenient.

According to Kuratowski's theorem, 'a graph is planar if and only if it either has no subgraph homeomorphic to K_5 or $K_{3,3}$ or it has no subgraph that can be contracted to K_5 or $K_{3,3}$',[50] where K_5 is the complete graph of five vertices (i.e., in which every vertex is connected to every other vertex) and $K_{3,3}$ is the complete bipartite graph of six vertices which can be divided into two sets of three and where each vertex is connected to all vertices of the other set and none in its own set.

Whenever a graph is planar, it can also be embedded onto a surface of a sphere without edge crossings, and vice versa. Once again, this embedding is unique (in the same sense as for planar embeddings) if and only if the graph is 3-connected, and we obtain the same faces. In fact, the graphs describing 3-dimensional polyhedra are exactly the 3-connected planar graphs,[51] and are therefore called *polyhedral graphs*.

A *polyhedron* is a three dimensional object that is bounded by planar faces. In a polyhedron of genus 0 (which is isomorphic to sphere) the faces fulfil the following formula,

$$\sum_{i}^{\mathcal{F}} (6 - \text{size}(F_i)) = 12, \tag{1.1}$$

where $\text{size}(F_i)$ is the number of vertices of a given face.

The number of neighbors to a vertex v, i.e., the number of edges incident to v, is called its *degree*, abbreviated as $\deg(v)$. If every vertex of a graph G has the same degree k, then G is called *k-valent* or *k-regular*. The special

[e] The embedding is unique in the following sense: All the planar embeddings of a 3-connected graph are topologically equivalent, which means that they can be continuously deformed into each other without crossing any edges. Consequently, every planar embedding of such a graph defines the same set of faces. The converse it true as well: only 3-connected graphs have this property.

case of 3-valent graphs is called *cubic*, and has a host of attractive mathematical properties, not least of which is their deep connections to algebraic geometry.[20]

Planar connected graphs fulfil *Euler's polyhedron formula*,

$$N - E + F = 2 \tag{1.2}$$

with $N = |\mathcal{V}|$ being the number of vertices (called the *order* of the graph), $E = |\mathcal{E}|$ the number of edges, and $F = |\mathcal{F}|$ the number of faces (for fullerenes these are hexagons and pentagons). The hand-shaking lemma in graph theory tells us that

$$\sum_{i=1}^{N} \deg(v_i) = 2E. \tag{1.3}$$

Since for fullerenes $\deg(v_i) = 3$ for all vertices, we obtain[52]

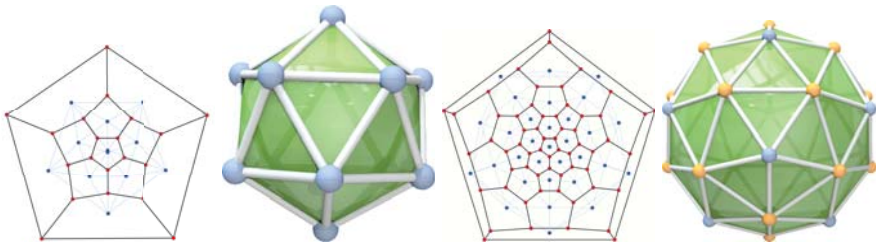
$$E = \frac{3}{2}N, \qquad F = \frac{1}{2}N + 2 = \frac{1}{3}E + 2, \quad \text{and} \quad E = 3F - 6. \tag{1.4}$$

Hence, for fullerenes and fulleroids[f] one has an even number of vertices. The total number of faces is $F = F_5 + F_6$, where F_5 and F_6 are the number of pentagons and hexagons respectively. From this we derive from eq. 1.4 $E = 3F_5 + 3F_6 - 6$ and $N = 2F_5 + 2F_6 - 4$. On the other hand, each pentagon (hexagon) has five (six) edges, which gives $E = \frac{5}{2}F_5 + 3F_6$. From these equations together with eq. 1.2 we obtain $F_5 = 12$ for regular fullerenes (the "12 Pentagon Theorem" for fullerenes), which gives the number of hexagons as $F_6 = (N - 20)/2$ with $N \geq 20$, and the general formula C_{20+2F_6} for fullerenes. Hence, from the fact that we only allow for cubic planar graphs consisting of pentagons and hexagons, we restrict the number of pentagons to exactly 12 (no more and no less!), and every addition of a hexagon adds two new vertices. Moreover, $N = 22$ with just 1 hexagon and 12 pentagons is not valid, but every number of hexagons two or greater yields at least one fullerene.[53] The most stable C_{22} molecule is, however, a cage-like fulleroid with a four-membered ring.[54]

In the present paper, we will often work with the *dual graph*. The dual graph G^* of a planar connected graph G has a vertex corresponding to each

[f] Fulleroids are cubic graph fullerene-like structures where we allow also for other polygons than pentagons or hexagons.

face of G, and an edge joining two neighboring faces in G. By this process, the old vertices become the new faces. The dual of a planar graph is itself a planar graph, and for fullerenes it is a triangulation with 12 vertices of degree 5 and the remaining of degree 6. For planar graphs in general, the dual is not unique, but as discussed above, it is unique for 3-connected graphs such as fullerenes, and hence the dual operation is well defined.[49] The dual operation is an involution, i.e., $(G^*)^* = G$, so we can think of the dual as just another representation of the same graph. Figure 1.3 shows the duals of two representative fullerenes with their corresponding 2D graph representation discussed in the next section.



(a) (b)

Figure 1.3 Planar embeddings of fullerene graph and dual (blue color and dotted lines for the dual representation), and 3D embeddings of the duals: (a) C_{20}-I_h, for which the dual is the icosahedron; (b) C_{60}-I_h, for which the dual is the pentakis-dodecahedron.

A useful representation of a graph is the *adjacency matrix* A_{ij} well known to chemists, as it is widely used in Hückel theory of conjugated organic molecules. The adjacency matrix is a symmetric matrix defined by $A_{ij} = 1$ if the graph G contains an edge v_i—v_j, and $A_{ij} = 0$ otherwise. It is equivalent to the edge-set representation, and is much easier to work with in many settings. As the number of edges for fullerenes is $E = 3N/2$, the adjacency matrix is very sparse. If two graphs G_1 and G_2 are isomorphic, the corresponding adjacency matrices A_1 and A_2 are related by vertex permutation, and thus share the same set of eigenvalues and eigenvectors. Matrices play a special role in chemical graph theory, and the reader is referred to Ref.[55] for more details.

1.2 Drawing fullerene graphs: Methods for planar embedding

Planar embeddings can be drawn on paper and give us a way to visualize the structures of planar graphs. However, not all choices of embeddings are equally informative. We want the drawings to be simple, not too cluttered, and to expose as much as we can of the graph's structure and symmetry. In general the full symmetry group can not be realized in two dimensions, but we can find embeddings that show as much of it as possible. Figure 1.4 shows planar drawings of three icosahedral fullerenes, C_{20}-I_h, C_{60}-I_h, and C_{540}-I_h, exposing 10-fold D_5-symmetry out of the full I_h group. In this section, we will discuss methods for constructing good planar embeddings.

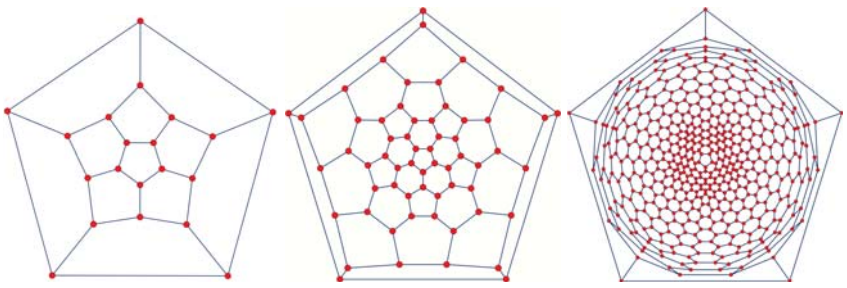
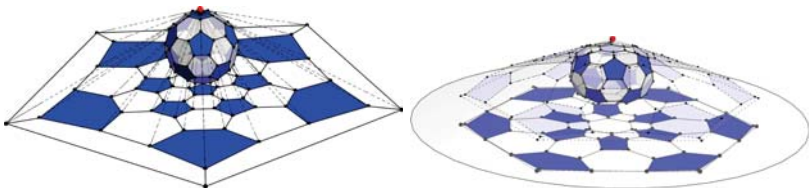


Figure 1.4 Planar embeddings (no crossing edges) for three different fullerenes obtained from a perspective projection. (a) icosahedron C_{20}-I_h; (b) truncated icosahedron C_{60}-I_h; (c) C_{540}-I_h.

While a small graph can easily be drawn by hand, larger graphs such as C_{540} shown in Figure 1.4(c) require a computer algorithm and a drawing program. The first method we will discuss is called the *Schlegel projection*, and requires first an embedding of the graph as a 3-dimensional polyhedron. In 1886, Victor Schlegel showed that a planar graph can be obtained from a 3-dimensional polyhedron by projection in the following way: A point $\boldsymbol{x}_0$ is placed slightly above one of the faces, and a plane P is chosen to lie below the polyhedron. Lines are then drawn from $\boldsymbol{x}_0$ through each of the polyhedron's vertices. The intersection of each line $\boldsymbol{x}_0$—$\boldsymbol{x}_v$ with P defines

the 2D coordinate of the vertex v. If $\boldsymbol{x}_0$ is chosen sufficiently close to the face, and if the points of the polyhedron lie on a sphere, the resulting 2D embedding will be planar. While Schlegel only proved that the projection produces a planar layout when the vertices lie on a sphere, it still works in many cases where this is not the case, and the method is widely used to the extent that planar graph embeddings sometimes (incorrectly) are called Schlegel diagrams. However, Schlegel diagrams are specifically the embeddings that result from the Schlegel projection. The Schlegel projection is illustrated in Figure 1.5 (a).

A problem with the Schlegel projection is that it often fails, and produces crossing edges, when used with many non-spherical polyhedra or when the parameters are not well chosen. The cone-projection method, shown in Figure 1.5 (b), which projects the vertices out on an enveloping cone of the fullerene polyhedron and then down onto a plane, seems to be more robust.[36]



(a) (b)

Figure 1.5 (b) Schlegel projection of C_{60}-I_h. (b) Cone projection of disk-shaped C_{72}-D_{6d}.

Both the cone-projection and the Schlegel projection require first a polyhedral embedding of the graph, and on top of that often necessitate a bit of hand-tuning to produce a planar embedding. The third method we mention is called the *Tutte embedding*,[56] and is guaranteed to always yield a planar embedding for 3-connected planar graphs. In the Tutte embedding, one fixes the coordinates of a single, outer face, and solves a sparse linear system

$$\sum_j T_{ij}\boldsymbol{x}_j = \boldsymbol{y}_i \tag{1.5}$$

such that in the result, every vertex is placed in the barycenter of its neigh-

bors. For fullerenes, $T_{ij} = \delta_{ij} - A_{ij}/3$ and $\boldsymbol{y}_i = 0$ for every i where v_i is not a fixed vertex, and for the vertices v_i on the outer face, $T_{ij} = \delta_{ij}$ and $\boldsymbol{y}_i$ is the fixed coordinate. While the Tutte embedding is guaranteed to be planar, the solution to the linear system results in exponential crowding of the vertices when embedding large graphs, making the result difficult to interpret. Hence, it is useful to follow the Tutte embedding by an optimization to make the face sizes more even.[36] We note that there are many different graph drawing algorithms available, each with advantages and disadvantages, but none that are guaranteed to always yield good results.[56–59]

1.3 The large isomer space

How does the number of C_N-isomers grow as we increase the number of vertices N? Naively one might assume a hypothetical sphere of "connected" hexagons (which is of course impossible because of Euler's theorem) and substitute 12 hexagons by pentagons to obtain a set containing all possible fullerene isomers fulfilling Euler's theorem (just like playing Lotto), we obtain an upper bound for the number of isomers of the order $\mathcal{O}\left(N^{12}\right)$. This is, however, three orders of magnitude too large, but the proof is nontrivial. In 1998, William Thurston showed that all the triangulations of the sphere with vertex degree at most 6 form a lattice of integer points in a complex hyperbolic space $C(1,9)$ through a parameterization as described in Ref.[20]. The triangulations with N vertices or less correspond to the lattice points inside the intersection between a certain cone and a ball of radius $\sqrt{N}$ in this space. Hence their number is roughly proportional to N^{10}, and the number of triangulations with exactly N vertices is roughly proportional to

$$\Delta(N^{10}) = N^{10} - (N-1)^{10} = 10N^9 - 45N^8 + 120N^7 + \cdots \qquad (1.6)$$

or, in short, $\mathcal{O}\left(N^9\right)$. Because they are the largest class of the finitely many triangulation classes with vertex degree six or less, the fullerenes have the same asymptotic count. Hence, the number of C_N isomers grows as $\mathcal{O}\left(N^9\right)$. We note in passing that the connection of fullerenes to combinatorial manifolds is a deep one and gives rise to many of the beautiful mathematical properties exhibited by fullerenes.

The high order polynomial growth of fullerene isomer counts makes it difficult to establish a complete database for fullerene graphs up to high vertex numbers.[35] Moreover, the proportion of IPR isomers to the total number of isomers grows with increasing N, gradually approaching 1 with increasing N, as is shown on Figure 1.1.

1.4 Generation of fullerene graphs

In order to explore chemical, physical or graph theoretical properties for a wide range of fullerenes, it is important to have access to a list of (stored) fullerene graphs. For this, one requires an exhaustive and efficient generator for all fullerene isomers of a given vertex number N. The general face spiral algorithm is well suited for compactly storing and recreating specific fullerene graphs. It also allows for sorting all isomers of C_N for a given vertex number N according to their FSPIs including information about jumps if required. It is, however, extremely inefficient to generate all isomers C_N directly through the a face spiral algorithm. In the non-general case one attempts to create graphs from $\binom{F}{12}$ spiral strings while the number of non-isomorphic graphs only grows $\sim N^9$. A large and quickly growing portion of all potential spiral strings are either non-canonical spirals, duplicates or are not successful for coding a fullerene. Including general spirals increases the overhead by two orders of magnitude, as jumps could be inserted at any step and with arbitrary lengths $l < N$.

A different approach to generating fullerene graphs is by adding faces to an existing graph, while considering different sites for addition at each step. Liu et al.[60] introduced a method which grows a graph starting with a single face and thus only reaches a fullerene graph in the last step. Brinkmann et al.[61] use patches, rather than single faces, as building blocks. These patches can be obtained by subdividing existing fullerene graphs.

Further development in efficient graph generation came from transforming an existing fullerene graph into a new, larger one by adding faces. Using C_{24} as a 'seed', Brinkmann et al.[62] were able to generate almost all fullerene graphs up to C_{200}. However, three graphs were not accessible by the applied set of transformations. To cure this shortcoming, Hasheminezhad et al.[63] defined a set of patch replacements (or growth operations) which, starting

from either C_{20} or C_{28}-T_d, yield every fullerene in a systematic way. The set consists of one single operation that elongates a nanotube with minimal caps, one class of linear generalized Stone-Wales transformations[64,65] that depend on one parameter, and one class of bent generalized Stone-Wales transformations that depend on two parameters. Brinkmann et al.[66] used this set of growth operations to define a fast and complete algorithm that recursively generates all fullerene isomers up to a given maximal N. The algorithm prunes the recursion tree in such a way that the only one representative of each isomorphism class is ever considered. Without such a scheme, exhaustively generating all fullerene graphs would succumb to combinatorial explosion.[67] The result is an incredibly efficient algorithm which has been used to generate an exhaustive database up to C_{400}, which is available at the House-of-Graph website.[35] Sometimes one is not interested in generating exhaustive isomer lists (which are huge for large N), but rather to generate in a directed fashion fullerenes that have specific properties. For this, one may start with one specific isomer C_N and derive others by certain transformations. This will be discussed in the following section.

1.5 Transformation of fullerene graphs

Transformations of fullerene graphs can be divided into local transformations that leave all but a certain region of the graph unchanged, and global transformations such as the Goldberg-Coxeter transformation.[68,69] A planar graph can be transformed locally by replacing one fullerene patch[65,70–75] by either a different patch or the same patch in a different orientation. A *patch* is a set of faces that is bounded by a simple cycle,[73,76] i.e., a cycle that traverses no vertex or edge twice. Some authors relax this condition and allow cycles that are non-selfintersecting, but are permitted to traverse any edge at most twice. This definition includes patches that have bridges, and is used in the complete characterization of fullerenes by Hasheminezhad et al.[63]

Two patches can be exchanged if they share the same *boundary code*. The boundary code of a patch is the sequence of free valencies of the vertices that lie on the boundary (in case of cubic graphs, a sequence of zeros and ones). A few patch replacements are shown in Figure 1.6.

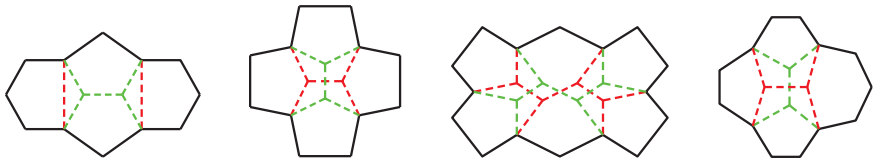


Figure 1.6 (a) Endo-Kroto (EK) 2-vertex insertion (b) Stone-Wales (SW) transformation (c) Example of an extended Stone-Wales transformation (d) A patch replacement which introduces one heptagon

Replacing patches of equal size can be understood as a formal isomerization, while replacement by larger or smaller patches, referred to as vertex insertion or deletion, or growth operations, formally derives a molecular graph of a different size. Two fullerene graph patches with the same boundary code contain the same number of pentagons, as can be seen from Euler's polyhedron formula. The most prominent example for formal isomerizations is the *Stone-Wales* (or pyracylene) *transformation* (SW),[77] which can be generalized[64,78] by allowing a variable distance between the two pentagons. Astakhova et al.[79] extended this rearrangement mechanism even further to higher numbers of pentagons. Brinkmann et al.[74] generated a catalogue of isomerization pairs with up to five pentagons.

Isomerization operations can be reformulated as sequences of vertex-pair switches, called Seidel switches or 2-switches.[80] According to Berge's switching theorem,[81] the whole isomer space of a given C_N is accessible by repeated application of 2-switches. However, intermediate structures of this procedure might have faces different from pentagons and hexagons, or result even in non-planar graphs.

Growth operations can be classified according to the number of pentagons and the number of vertices that are added. There are no growth operations for fullerene graphs that involve no or only one pentagon.[72] The smallest example is the *Endo-Kroto transformation* (EK),[82] which involves two pentagons and adds two vertices. Additional noteworthy examples are the addition of four and six vertices at a patch that contains three pentagons and has C_3 symmetry.[83] Brinkmann et al.[75] compiled an extensive list of growth pairs. The three classes of patch replacements defined by

Hasheminezhad et al.[63] form the basis of the currently fastest fullerene graph generator, as discussed in a previous section.[66]

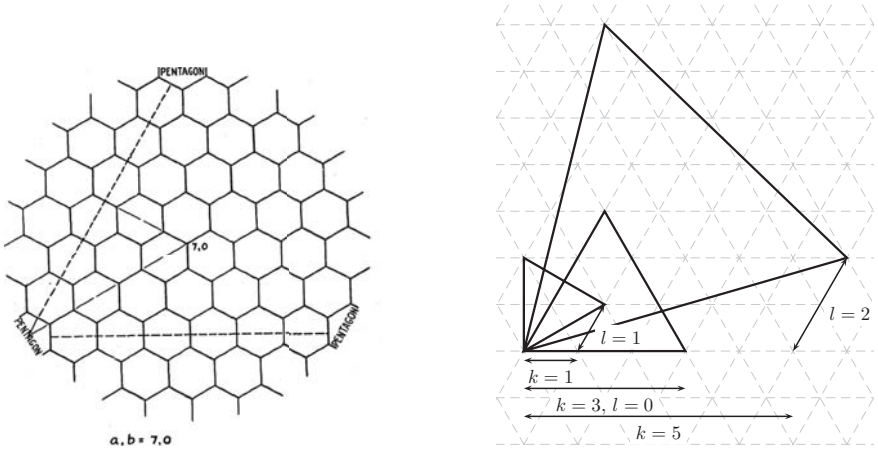
With respect to the polyhedral representation of a fullerene graph, growth operations can be understood as the capping of a domain by additional vertices. As a result the pentagons in that domain move towards each other. Patches in which all pentagons are fused cannot be capped. Conversely, the inverse of a growth operation corresponds to the truncation of a domain of high curvature: The distances between the affected pentagons increase.

While patch replacements are useful from a graph theoretical point of view to obtain new fullerene isomers, the EK C_2 insertion and the SW transformation have also been suggested to resemble viable reaction pathways.[77,82,84–86] As Stone and Wales pointed out in their initial article, a concerted mechanism for the SW transformation ([2+2]) is Woodward-Hoffmann forbidden in the electronic ground state.[77] Other mechanisms that involve the breaking of one bond[84,87,88] or the catalysis by additional carbon atoms[89] are therefore more likely. For a discussion of different mechanisms see Ref.[84]. For every suggested pathway, however, the activation barrier is so high that SW transformations are only feasible at very high temperatures.[84,85] The SW transformation is believed to be one of the main mechanisms by which fullerene cages equilibrate during formation to form the most stable isomer.[90,91] Starting at C_{60}-I_h, 1709 out of the 1811 other isomers are accessible by consecutive SW transformations.[92]

Endo and Kroto proposed a concerted mechanism for what has since been known as the EK C_2 addition.[82] However, more recent experimental and theoretical investigations show, that the concerted reaction leads to an unstable adduct and is followed by C_2 ejection.[93] For a summary of alternative mechanisms of the C_2 addition to fullerene cages see Ref.[86].

Given any fullerene, it is possible to construct an infinite series of larger ones with (essentially) the same three-dimensional shape. This is possible through the *Goldberg-Coxeter* transformation, named after an infinite series of icosahedral-symmetry polyhedra described by Goldberg in 1937.[68] By superimposing a hexagonal mesh on the surface of the dodecahedron, as shown in Figure 1.7 (a), a new polyhedron is obtained with the same number of pentagons and an increased number of hexagons. The number of vertices in the new polyhedron is $k^2 + kl + l^2$, where k and l are integers describing the scale and orientation of the mesh, as shown for the dual in Figure 1.7 (b).

The construction used by Goldberg was discovered independently, and applied to the shapes of vira, by Caspar and Klug in the 1960s, and later popularized by Coxeter[69].



(a) (b)

Figure 1.7 Illustration of GC-transform. (a) Original figure of a hexagonal sheet from Goldberg's paper[68]. (b) The GC-transform acting on a face in the fullerene dual for various values of k and l.

While originally only defined for C_{20}, yielding exactly all the fullerenes of icosahedral symmetry, Dutour and Deza have shown that it is well defined for all cubic planar graphs.[94] However, it is not trivial to determine how to automatically perform the general transform, the main difficulty lying in how to "glue" the graph back together when the new edges cross the transformed triangles in complex ways. The general construction by Dutour and Deza is quite algebraically heavy handed, and is not easily understood, nor lends itself easily to implementation. In the case of fullerenes, however, it is quite easy to give a procedural description of the transformation in terms of the fullerene dual: Because a fullerene dual is a triangulation of the sphere with no vertices of degree more than 6, it can be unfolded onto the plane of equilateral triangles, called the *Eisenstein plane*. The unfolded

surface forms a polygon in the Eisenstein plane with every degree 5 vertex on the polygon periphery, and with each edge on the periphery appearing twice, once in each direction.

The vertices in the Eisenstein triangulation is a sub-ring of the complex plane, and we can write

$$(a, b) = a + b\omega \quad \text{where} \quad \omega = e^{i\frac{2\pi}{6}} \tag{1.7}$$

Because it is a sub-ring, multiplication by Eisenstein integers (a, b) are exactly the operations that bring $(1, 0)$ into every other vertex point. Because $\omega^2 = \omega - 1$, we get the multiplication rule

$$(a, b) \cdot (c, d) = ac + (ad + bc)\omega + bd\omega^2 = (ac - bd, bc + (a + b)d) \tag{1.8}$$

and the magnitude of (a, b) is $(a + b\omega)(a + b\omega^{-1}) = a^2 + ab + b^2$. Hence, if T is an equilateral triangle with unit area, the area of $(k, l)T$ is $k^2 + kl + l^2$. This is no coincidence: multiplication in the Eisenstein ring exactly corresponds to the Goldberg-Coxeter transformation in the dual formulation. This yields a simple and practical formulation of the Goldberg-Coxeter transform of any triangulation of the sphere with $\deg(v) \leq 6$:

$$\texttt{GC}_{k,l}(G^*) = \texttt{fold}((k + l\omega) \cdot \texttt{unfold}(G^*)) \tag{1.9}$$

The procedure is illustrated in Figure 1.8. Details for how to efficiently perform the `fold` and `unfold` operations are given in Ref.[95]. Note that two successive GC-transformations on a graph G^* can be expressed as a single GC-transformation by multiplying the two Eisenstein numbers together. Similarly, a GC-transformation can be inverted by dividing instead of multiplying. For any particular fullerene, we can even use Euclid's algorithm to find out whether it is a GC-transform of a smaller fullerene, and to find the smallest such "parent" fullerene.[96]

Notice, that if one were to cut out the two diagrams in Figure 1.8, gluing together the edges so that the numbers on the vertices match, one obtains the three dimensional structure of the given C_{32} and C_{224} fullerenes. The reader is invited to do so.

There are two special cases of the Goldberg-Coxeter transformation that warrant closer study: the case $l = 0$, named the *Halma transform* (or chamfer

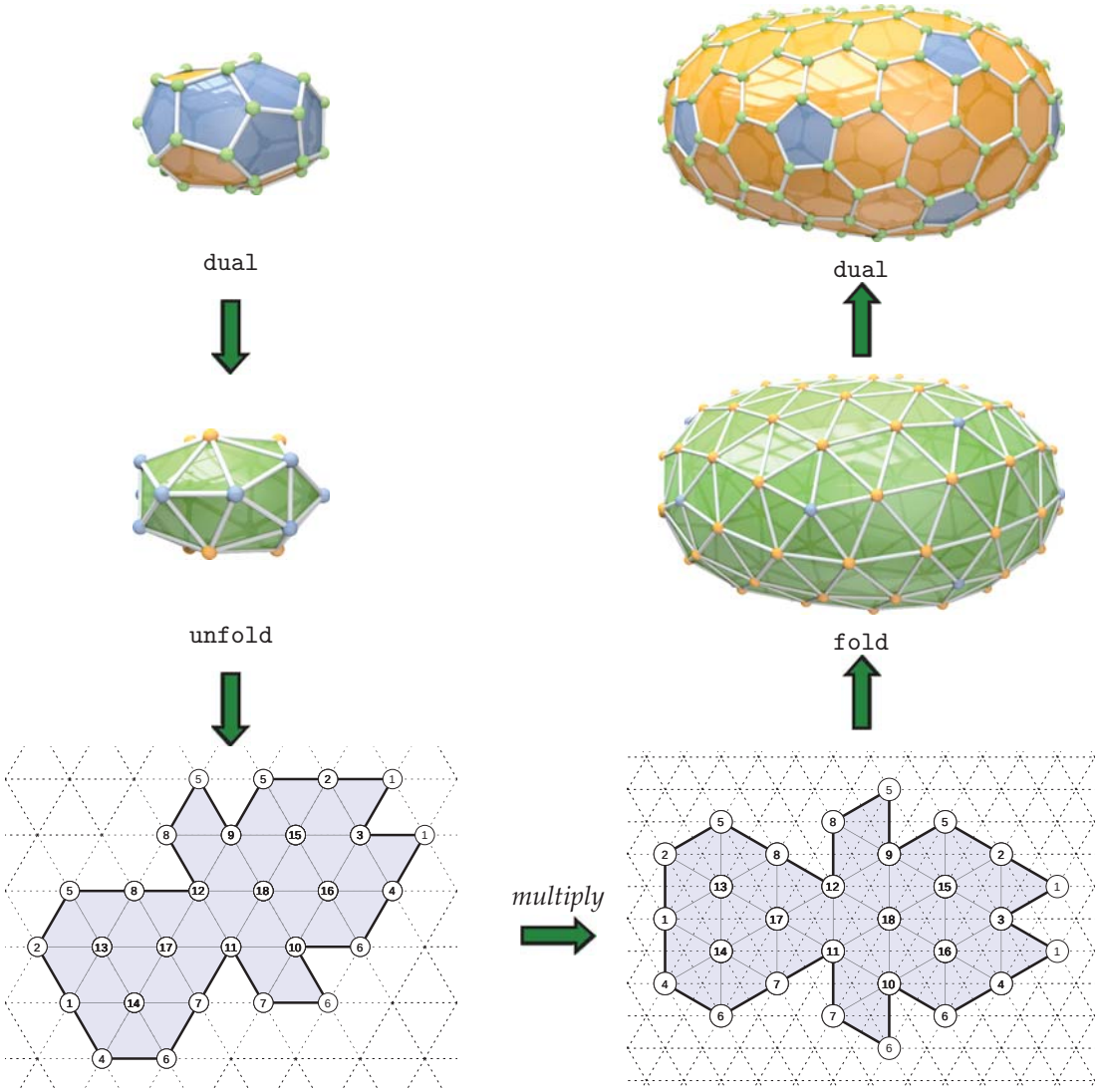


Figure 1.8 The steps in the $GC_{2,1}$ Goldberg-Coxeter transformation from C_{32}-$D_{3h}(5)$ to C_{224}-D_{3h}. The transform and the diagrams of the unfolded fullerenes were automatically generated using the program *Fullerene*.[36]

transform) for the construction's similarity to the game board of Halma, or "Chinese Checkers"; and the case $k = l$, called the *leapfrog transform*. Both of these types of Goldberg-Coxeter transformations are simple to understand: every triangle is subdivided and reconnected the same way, as illustrated on Figure 1.9. While the GC-transform in general can introduce, reverse, or remove chirality (but otherwise preserves symmetry), both the Halma and leapfrog transformation always preserve symmetry.

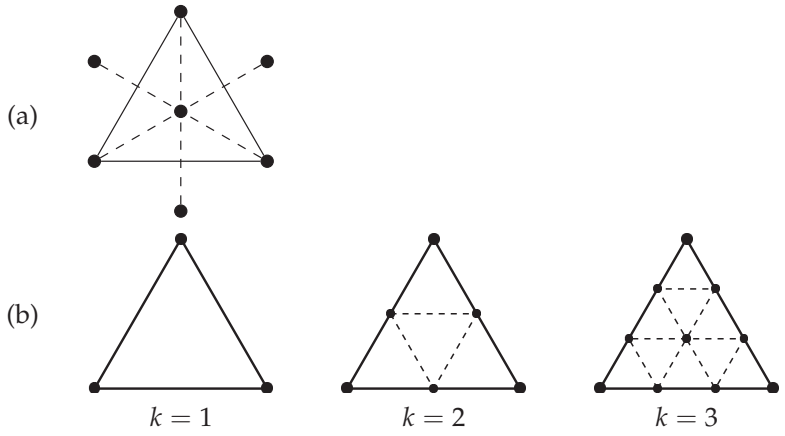


Figure 1.9 Subdivision of a triangle in the dual graph; (a) for the leapfrog transformation ($k = l = 1$), and (b) for the lowest few Halmas ($l = 0$).

1.6 Geometry of Fullerenes

Many of the beautiful properties of fullerenes derive from their relation to algebraic and differential geometry. This relation is mostly out of scope for this review, but in this section we will touch on the subject informally. The subject is treated in great depth by Thurston[20] and others.

An important quantity for understanding the geometry and shapes of fullerenes is the *Gaussian curvature*. The Gaussian curvature is the product of the two principal curvatures, which for each point on the surface are the maximal and minimal curvatures in any direction through that point. By

the Bertrand-Diquet-Puiseux theorem, the Gaussian curvature in a point is the same as the difference between 2π and the angle required to make a circle in a neighbourhood around this point. Figure 1.10 illustrates three categories of surfaces with zero, positive, and negative Gaussian curvature respectively. If the Gaussian curvature is zero in a point, the surface only bends in one direction around that point. A surface that has zero Gaussian curvature everywhere is characterized by being "flat" in the sense that it can be unwrapped onto a plane without tearing. Positive Gaussian curvature bends the same way in all directions. A positive curvature surface can be cut open and unwrapped onto a plane. Finally, negative curvature around a point constitutes a saddle point, yielding a wobbly pringle-like surface. Negative curvature surfaces cannot be unwrapped onto a plane, because there is no room: around a negative curvature point, the angle of a circle is greater than 2π. If one were to make a cut in a negative curvature surface, it could only be flattened out by allowing parts of it to overlap itself.

The Gaussian curvature only depends on the topology of the surface, and is independent of how it is isometrically embedded in space. In the same way, the surface metric (or Riemann metric), which determines the geodesics and distances between points along the surface, can be derived directly from the graph,[97] and is also independent of the spatial embedding. However, while we do not require a 3D embedding to analyze these surface properties, they themselves do determine the possible ways that the surface *can* embed isometrically into space, and what is its "natural" three-dimensional shape.

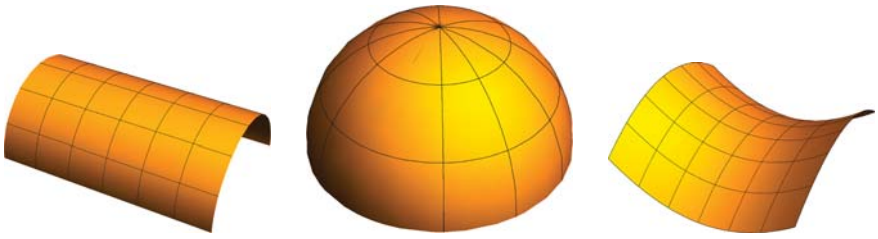


Figure 1.10 Examples of surfaces with zero (a), positive (b), and negative (c) Gaussian curvature around a point

Fullerenes are all closed surfaces with non-negative Gaussian Fullerenes are all closed surfaces with non-negative Gaussian curvature everywhere. Surprisingly, this is simply due to them having faces no larger than hexagons! In fact, the slightly larger class of 3-connected cubic planar graphs with largest face size 6, and their dual triangulations, are in a sense *all* the positive curvature genus 0 surfaces.[g]

It is most easy to understand the geometry of fullerenes when considering their duals. These are equilateral triangulations of a closed surface. The equilateral triangle plane, also called the Eisenstein plane, is the dual of a hexagonal mesh. Hexagon planes (like graphene) and their dual degree-6 triangulations are *flat* in the sense that they have Gaussian curvature 0 everywhere: any surface of equilateral triangles with vertex degree only six can be unwrapped into the plane without stretching or tearing. What happens now if we set the degree to 5 of a single vertex in the plane? To do this, we must cut out an infinite triangular wedge of angle $2\pi/6$. Gluing together the edges forms an infinite cone with Gaussian curvature zero everywhere except the degree-5 vertex at the cone's apex, which introduces the positive curvature $2\pi/6$. Every time we introduce a vertex of degree-5 (or a pentagon in the hexagonal plane), it introduces a surface cut that is glued together to locally introduce the positive Gaussian curvature $2\pi/6$ at the vertex. After 12 pentagons we reach 4π, the Gaussian curvature of the sphere, closing the surface.[h] Unfolding again along the 12 cuts would result in a polygon similar to the diagrams shown in Figure 1.8. Had we instead introduced e.g. degree-4 vertices, each would contribute with curvature $2\pi/3$, requiring only 6 to close the surface. The positions at which we placed the 12 pentagons determine where and how strongly the surface bends, and through that the natural shape of the fullerene, giving rise to the many interesting polyhedral shapes shown in Figure 1.2.

Because fullerenes always have exactly 12 curvature-$2\pi/6$ bends, with

[g] In the mathematical sense that they are dense in this set, i.e., any smooth non-negative curvature genus 0 surface is the limit of a sequence of genus 0 triangulations with vertex degree-6 and less.

[h] The interested reader is directed to the Gauss-Bonnet theorem, which generalizes Euler's theorem to arbitrary Riemann surfaces. In the discrete case they are the same, except that Gauss-Bonnet is stated in terms of the Gaussian curvature. From this identification, we can see that every n-gon, or vertex of degree-n in the dual, contributes the angle $\frac{2\pi(6-n)}{6}$ to the Gaussian curvature, which must sum to $2\pi(2-2g)$, where g is the surface's genus.

the remaining surface of Gaussian curvature zero, they adopt polyhedral shapes rather than being spherical in nature (the latter defined as atoms lying on a sphere), and the most symmetric shapes that they can achieve have icosahedral symmetry. The non-spherical nature of fullerenes was pointed out already by Bakowies and Thiel in 1995.[98,99]

The curvature of cubic graph surfaces also provides an intuitive understanding of why fullerenes, having only hexagon and pentagon faces, would be more prevalent in nature than other polyhedral carbon structures: smaller face sizes introduce much sharper bends, reducing stability. Heptagons by themselves introduce no sharper bends than pentagons, but each heptagon necessitates an extra pentagon to balance out the negative Gaussian curvature and reach a closed surface, increasing the total number of bends by two. However, with the right distribution of heptagons and pentagons, it is possible to construct quite smoothly curving near-spherical fulleroid surfaces that could potentially be stable in nature. We discuss fulleroids, a generalization of fullerenes that allow arbitrary face sizes, in the final section.

1.7 Generating accurate 3D geometries[i]

The Cartesian coordinates for the ideal molecular geometry of small highly symmetric fullerenes such as C_{20}-I_h or C_{60}-I_h can be calculated directly using basic geometry.[100] This is not the case for fullerenes in general: They come in many different shapes, most of which do not have a simple mathematical formula. However, we do know that the ideal geometry (in which the faces of the dual polyhedron are equilateral triangles) can be derived directly from the fullerene graph. This is because of their special properties discussed above. However, while we know that a fullerene graph determines the ideal 3D structure as a conformal, isometric embedding of the surface into space, it is not yet known how to compute it directly. However, we can use numerical optimization methods to obtain good 3D structures – both

[i] In the original article[1] this section encompasses algorithms for obtaining an initial 3D structure from the fullerene graph as well as the force field for the consecutive optimisation of the Cartesian coordinates. The force field is described in more detail in chapter 4 and is therefore omitted from this section.

efficiently and with results that are very close the the physical molecular geometry.

1.7.1 Generating initial structures

Before optimizing a fullerene structure by for example a force-field method, we need a reasonable initial structure. Specifically, we need an embedding of the graph as a polyhedron (i.e., no crossing edges) such that its faces are the same ones as in the final polyhedron, and such that the magnitude of the force experienced by the vertices does not vary too greatly, causing slow convergence or numerical instability. We here outline two methods for obtaining initial geometries that are suitable for input to the fullerene force-field optimization.

The *Fowler-Manolopoulos 3D embedding algorithm* (also called Adjacency Matrix Eigenvector or AME algorithm) diagonalizes the $N \times N$ adjacency matrix A_{ij} of the fullerene graph to obtain eigenvectors $\boldsymbol{x}_i$ ($i = 1, \ldots, N$), of which three "P-type eigenvectors" with one nodal plane are chosen to construct the Cartesian coordinates (termed *topological coordinates* by Manolopoulus and Fowler) for the initial structure.[32,101] This often yields suitable coordinates because fullerenes can be modelled as spherical aromatic systems fulfilling approximately the Laplace differential equation with real spherical harmonics $Y_{lm}(\theta, \varphi)$ as eigenfunctions, of which the three Y_{1m} ($l = 1$) are associated with the coordinates (x_i, y_i, z_i).[32,102] Cartesian coordinates constructed in this way already reflect the space symmetry of the fullerene. The P-type eigenvectors can easily be found for near-spherical fullerenes like C_{20} or C_{60}-I_h, because the sequence of eigenvalues mimics the spherical harmonics. However, for distorted fullerenes, or for very large fullerenes where the spectrum of the adjacency matrix becomes very dense with quasi-degeneracies, it becomes difficult to find the appropriate eigenvectors, and the assumption that the eigenvectors resemble spherical harmonics becomes decreasingly valid. For example, for large fullerene nano-tubes aligned around the z-axis, such as the one shown in Figure 1.2, the eigenvalues belonging to the P_x- and P_y-type eigenvectors do not lie close to the P_z eigenvalue. Hence, one must search for the appropriate set of eigenvectors, and identifying them is not always easy.[103–105] Second, the bond lengths tend to vary substantially in the constructed fullerene, leading

to unreasonable large bond distances once the minimum bond distance is set to $\approx 1.4\,\text{Å}$, and a scaling procedure for the eigenvectors may only solve part of this problem.[32] Finally, because it is often necessary to search for the right eigenvectors and hence calculate a large number of them, the algorithm scales up to $\mathcal{O}\left(N^3\right)$, which quickly becomes a heavy operation as N grows.

We have found it useful to sacrifice some accuracy in the initial geometry for a method that is more stable, and which scales approximately as $\mathcal{O}\left(N\right)$. The fullerene force-field optimization is insensitive enough to the starting geometry that the difference is not noticeable. The method starts with any planar embedding of the fullerene graph. We choose the *Tutte-embedding*, which is guaranteed to be planar, and can be computed in near-linear time by solving two sparse linear systems of equations.[56] The planar embedding is then mapped onto the surface of a sphere (or e.g., ellipsoid, if the fullerene is a priori known to be elongated) in the following way: For every vertex v, let the topological distance from the outer face (the "depth" of the vertex) be d_v, and its coordinate in the Tutte-embedding be $\boldsymbol{x}_v$. Let D be the maximum depth, and $\boldsymbol{x}_C$ be the barycenter in the embedding. We assign one angle $\phi(v) = \frac{(d_v+\frac{1}{2})\pi}{D+1}$, placing vertex layers equidistantly along the ϕ-angle according to their depth. The second angle $\theta(v)$ is the angle of $\boldsymbol{x}_v$ around $\boldsymbol{x}_c$. Then assigning the angles $v \mapsto (\theta(v), \phi(v))$ yields an embedding on the sphere with no crossings, since the Tutte-layout is guaranteed to be planar, but which avoids the exponential crowding of the planar Tutte-embedding, such that the vertices are distributed in a reasonable manner across the surface of the sphere.[j] It remains, however, a challenge to embed the fullerene graph on a more realistic surface.

1.8 Fullerene symmetry

There are two symmetry groups associated with a fullerene: the *ideal* or *topological symmetry group* of the fullerene graph, and the *real* or *physical symmetry group* of the molecule in 3D space. The physical symmetry group is a subgroup of the ideal one, due to (first or second order) Jahn-Teller

[j] When there are multiple exponentially crowded centers, the Tutte-embedding is still guaranteed planar, but the mapping to the sphere can have crossings. In practice, however, it turns out to not matter.

distortions, variation of bond-lengths, or other effects from external perturbations. But because the shape is often only mildly altered from the ideal shape, the ideal symmetry group carries useful information about the physical system.

A surprising theorem by Mani[106] shows that any 3-connected cubic planar graph can be embedded in space as a convex polyhedron, the point group of which realizes the full symmetry group of the graph. That is: every graph automorphism of a fullerene is also a rotation or reflection of its ideal polyhedral shape. In addition, it can be shown[107] that the point group for a fullerene must be one of the 28 point groups listed in Table 1.1. The largest group is the icosahedral group I_h of order 120, and each of the 27 remaining attainable point groups are subgroups of I_h.

Can we automatically determine the symmetry group for the fullerene directly from the graph, without referring to spatial coordinates? (The latter requires first finding a maximally symmetric 3D embedding) For general graphs, this is a difficult problem. However, in the case of cubic polyhedral graphs, for which general face spirals determine isomorphism, it is both simple and efficient to do so.

Order	Point groups	Order	Point groups	Order	Point groups
120	I_h	60	I	24	T_d, T_h, D_{6h}, D_{6d}
20	D_{5h}, D_{5d}	12	T, D_6, D_{3h}, D_{3d}	10	D_5
8	D_{2h}, D_{2d}	6	D_3, S_6, C_{3h}, C_{3v}	4	D_2, S_4, C_{2h}, C_{2v}
3	C_3	2	C_2, C_s, C_i	1	C_1

Table 1.1 The 28 possible point groups for fullerenes sorted according to their order $|\mathcal{G}|$ of the group $\mathcal{G}$.[32]

Before moving on, we take note of an ambiguity in terminology. The terms *symmetry group* and *automorphism group* often refer simply to the *group* formed by operations that leave the system invariant. In the group theoretical sense, a group is fully defined by its multiplication table. *Point groups*, however, are not just groups, but are *isometries of space* and carry additional information about how they transform space. For example, the point groups D_6, C_{6v}, D_{3d}, and D_{3h} are all the exact same group, but differ

as point groups. It is unfortunate for this reason that "point groups" are called "groups". To make the distinction explicit, we will use the term *abstract group* for the symmetry group in the group theoretical sense, and either *point group* or *symmetry group* for the isometries.

Assume that we are given the fullerene graph dual G^*, constructed from a generalized spiral $S = (d_1, d_2, \ldots, d_F)$, and wish to compute the automorphism group $\mathcal{G}$ of G^*. Since G^* is constructed from S, the entries in S are the degrees of vertex number $1, 2, \ldots, F$ in G^*. For every vertex v of degree d_1, we have $2d_1$ different spiral starts: d_1 for clockwise and d_1 for counter-clockwise traversal. If a spiral start (f_1, f_2, f_3) unwinds G^* to the input spiral S, there is an automorphism of G^* that maps $(1, 2, 3) \mapsto (f_1, f_2, f_3)$. These are *all* the automorphisms, and the number of starts that unwind to S is the order $|\mathcal{G}|$ of the automorphism group. This construction lets us easily obtain the actual group: The spiral start determines the entire spiral, and by simply writing down the vertex names in G^* while unwinding the spiral, we obtain a permutation representation of the group element

$$\pi_F(g) = \begin{pmatrix} 1 & 2 & 3 & \cdots & F \\ f_1 & f_2 & f_3 & \cdots & f_F \end{pmatrix} \tag{1.10}$$

This is a faithful (but not irreducible) representation of the group, and if we wish, we can easily build the multiplication table by composing all pairs of the permutations. Similarly, we can calculate characters, irreducible representations, and all other properties, and we can identify the group. There are a number of ways to make this efficient. In Ref.[108], we present a general, but efficient, $\mathcal{O}(N)$-algorithm that works for all cubic polyhedral graphs, including all fullerenes and fulleroids.

The permutation representation π_F allows us to not only find the abstract symmetry group, but contains enough information to identify the point group, i.e., the isometries of the 3-dimensional fullerene polyhedron. Fowler and Manolopoulos published a program for identifying the point group of any spiralable fullerene in Ref.[32], based on site symmetries and counting group orbits. The *symmetry points* of interest in a fullerene are the vertices, midpoints of edges, the barycenter of the polygons and the whole cage, the latter having the full symmetry of the point group. They have certain site symmetries according to the rotational axes or mirror planes going through these symmetry points,[32] which are collected in Table 1.2. The full isometry

group $\mathcal{G}$, which correspond to the rotations, reflections, roto-inversions, and inversions that leave the ideal polyhedron invariant, also act as permutations of the symmetry points,[101,109,110] the action always being a subgroup of $\mathcal{G}$. The permutations of the symmetry points in fact completely determine the point group symmetry, and we can find the full point group of the graph G (or, equivalently, its dual G^*) as follows: In step 1, one computes the face permutation representation π_F of the abstract group, as described above. From this one derives vertex and edge permutation representations π_V and π_E by acting with every group operation on the dual graph.

Symmetry Points	Site Symmetries (order)
vertices	$C_{3v}(6)$, $C_3(3)$, $C_s(2)$, $C_1(1)$
edge centers	$C_{2v}(4)$, $C_2(2)$, $C_s(2)$, $C_1(1)$
pentagon centers	$C_{5v}(10)$, $C_5(5)$, $C_s(2)$, $C_1(1)$
hexagon centers	$C_{6v}(12)$, $C_6(6)$, $C_{3v}(6)$, $C_3(3)$, $C_{2v}(4)$, $C_2(2)$, $C_s(2)$, $C_1(1)$
cage center	full point group

Table 1.2 The site symmetries of the local symmetry points in a fullerene.[32] The order of the site-symmetry group is given in parentheses (the maximum value of any site symmetry group order is 12).

In step 2 one computes the vertex, edge, and face *group orbits* by acting with the permutation representations on every vertex, edge, and face.[111] Each orbit belongs to a certain site-symmetry group as shown in Table 1.2. The site-symmetry groups are determined by the orbit sizes. For example, the site-symmetry group of the face f_i has order $|\mathcal{G}_{f_i}| = |\mathcal{G}|/|\mathcal{G}f_i|$. By counting the number of sites belonging to each site-symmetry group, we obtain a signature that uniquely identifies the point group:

$$m_F(k) = \left| \left\{ f \in \mathcal{F} \,\middle|\, k = \frac{|\mathcal{G}|}{|\mathcal{G}f|} \right\} \right|, \tag{1.11}$$

$$m_E(k) = \left| \left\{ e \in \mathcal{E} \,\middle|\, k = \frac{|\mathcal{G}|}{|\mathcal{G}e|} \right\} \right|, \quad \text{and} \tag{1.12}$$

$$m_V(k) = \left| \left\{ v \in \mathcal{V} \,\middle|\, k = \frac{|\mathcal{G}|}{|\mathcal{G}v|} \right\} \right| \tag{1.13}$$

The information can be condensed to a *site-symmetry count* for each site-group order:

$$m_S(k) = m_F(k) + m_E(k) + m_V(k) \tag{1.14}$$

In the final step, the point group is determined by the site-symmetry counts by the decision tree structure shown in Figure 1.11. The method can be extended to any fullerene and fulleroid by using general spirals. However, for every possible point group, the site-symmetry signature must be worked out and added to the decision tree.

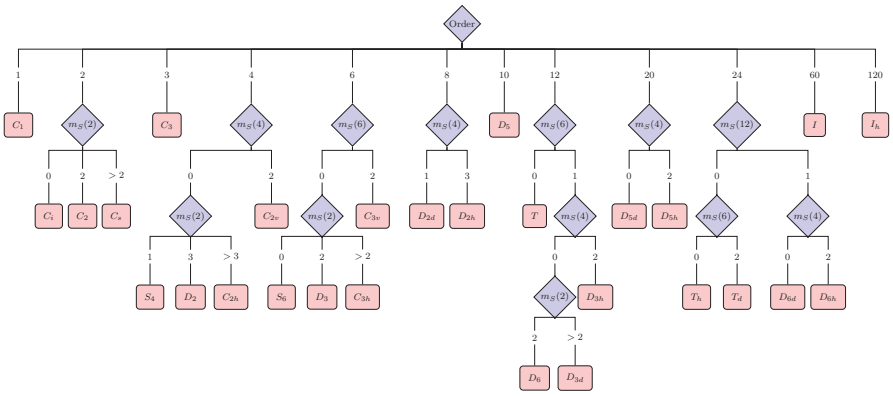


Figure 1.11 The decision tree for determining the symmetry point group for any fullerene from the group order and orbit counts.

Once the point group has been determined, the number of infra-red and Raman active lines, as well as the ^{13}C NMR pattern can be derived.[32,101] Moreover, point groups lacking an inversion center are further divided into polar and chiral point groups. A *chiral point group* is one without any roto-inversion symmetry elements, and a *polar point group* allows for the fullerene to have a dipole moment. A point group with an inversion center or a mirror plane perpendicular to the axis of rotation cannot be polar. The nine chiral point groups for fullerenes are $I, T, D_6, D_5, D_3, D_2, C_3, C_2$, and C_1.[32] The polar fullerenes belong to either of the point groups $C_{3v}, C_3, C_{2v}, C_2, C_s$ or C_1.[32] For larger fullerenes, the fraction of low-symmetry to high-symmetry isomers grows rapidly, and the C_1 point group increasingly dominates. This

can be intuitively understood combinatorially from distributing pentagons on a sphere. Already at C_{100}, more than 99% of the isomers are C_1.

The separation of spiral starts into equivalence classes by their spirals yields an interesting relation:

$$|\mathcal{G}| = \frac{6N}{N_S^s} \tag{1.15}$$

where N_S^s is the number of symmetry distinct (general) spirals. For I_h symmetry we have $|I_h| = 120$, which gives $N = 20N_S^s$. Hence, I_h-fullerenes can only occur when N is a multiple of 20. In a similar way, for I symmetry we have $N = 10N_S^s$, and for T_d, T_h, D_{6h}, or D_{6d} we obtain $N = 4N_S^s$. This explains why some of these high symmetry groups are not found in certain isomer lists.

In the case of icosahedral symmetry, we have a full characterization of when they occur: As proved by Dutour and Deza,[94] every fullerene of I_h or I symmetry is a Goldberg-Coxeter transform of C_{20}. This means that they occur exactly when $N = 20(k^2 + kl + l^2)$ for integers k and l. It is also possible to determine when these are of I_h and when they are of I symmetry. Halma and leapfrog transformations, which correspond to Goldberg-Coxeter transforms of $l = 0$ and $k = l$, respectively, both preserve the symmetry.[112] Hence, there is a fullerene with I_h symmetry at $N = 20j^2$ and $N = 20(3j^2)$ for every $j \in \mathbb{N}$, corresponding to a single Halma or leapfrog transformation on C_{20}. Consecutive application of the two shows that isomers with I_h symmetry are found for every $N = 20(3^i j^2)$ ($j \in \mathbb{N}$ and $i = 0, 1$). Moreover, general Goldberg-Coxeter (k, l) transforms with $k \neq l$ and $l \neq 0$ ($k > 0$) break horizontal mirror plane symmetry.

1.9 Shapes: Volume and surface area, sphericity and convexity

For ideal C_{60}-I_h, where all edges are of equal length, the geometric volume V and surface area A can be obtained through simple algebraic and geometric considerations. The volume and surface are of a regular 20-sided polyhedron (icosahedron, see Figure 1.3) made out of equilateral triangles with edge

length a is,

$$V_{\text{ico}} = \frac{5}{12}(3+\sqrt{5})a^3 \quad \text{and} \quad A_{\text{ico}} = 5\sqrt{3}a^2 \tag{1.16}$$

Cutting out 12 pentagonal pyramids at length $a/3$ gives the ideal truncated icosahedron C_{60}-I_h with

$$V_{C_{60}} = \frac{1}{4}(5^3+43\sqrt{5})r_e^3 \quad \text{and} \quad A_{C_{60}} = 3r_e^2\left(\sqrt{5^2+10\sqrt{5}}+10\sqrt{3}\right) \tag{1.17}$$

where $r_e = a/3$ is the length of the edge (C–C bond length). This gives the ratios $V_{ico}/V_{C_{60}} \cong 1.065$ and $A_{ico}/A_{C_{60}} \cong 1.073$ between the two volumes and surface areas. C_{60} has, however, two different bond lengths as the crystal structure shows,[113] $r_5 = 1.455$ Å (edge at pentagon-hexagon fusion) and $r_6 = 1.391$ Å (edge at hexagon-hexagon fusion). A smaller bond distance in the hexagon implies that more of the original icosahedron is cut off,

$$V_{C_{60}} = \frac{5(3+\sqrt{5})}{12}(2r_5+r_6)^3 - \frac{(5+\sqrt{5})}{2}r_5^3 \quad \text{and} \tag{1.18}$$

$$A_{C_{60}} = 3r_5^2\sqrt{5^2+10\sqrt{5}}+5\sqrt{3}\left(r_6^2+r_5^2+4r_5r_6\right) \tag{1.19}$$

For $r_5 = r_6$ we obtain eq. 1.16. If $r_6 = 0$ ($r_5 = a/2$) the hexagons become equilateral triangles and for $r_5 = 0$ we just get a hollow icosahedral C_{12} cluster. For C_{20} (equal edge lengths) the vertices lie in the center of each face of an icosahedron, and the volume and surface area are easily obtained,

$$V_{C_{20}} = \frac{(15+7\sqrt{5})}{4}r_e^3 \quad \text{and} \quad A_{C_{20}} = 3r_e^2\sqrt{5^2+10\sqrt{5}} \tag{1.20}$$

How can we get the volume V and surface area A for any fullerene isomer? As fullerene cages are not guaranteed to have planar faces, their volume or surface area are only approximately defined. There are, however, a number of definitions according to which we can express both quantities. We could triangulate all faces by adding a barycenter $\boldsymbol{b}_j$ to each face with lines to each vertex of that face. We call the polyhedron obtained in this way a *triangulated face polyhedron*, TFP. For a fullerene, the graph representing the TFP is identical to the dual structure of its leap-frog transform. The total surface area A is obtained by summing over all areas A_i of the triangles

obtained, which works even for faces where the vertices do not lie on a plane. Using Gauss' theorem (divergence theorem), the volume V is found by summing over the face normals,

$$V = \frac{1}{3}\sum_i A_i(\boldsymbol{p}_i \cdot \boldsymbol{n}_i) \tag{1.21}$$

where $\boldsymbol{p}_i$ is any point on face i, and A_i is the area of the face.

The different volumes and surfaces areas can be used for calculating some important measures for fullerenes, such as the *sphericity S* (how spherical a fullerene is) and *convexity C* (how convex a fullerene is). The surface of a triangulated fullerene may not be convex. We can measure the non-convexity by comparing to the *convex hull* of the fullerene cage, which is the smallest convex polyhedron that contains all the points. The convex hull (CH) is uniquely defined, and there are several algorithms available, such as the incremental 3D convex hull algorithm.[114] A measure for convexity C is obtained from the proportion of "empty space" between the non-convex polyhedron and the convex hull by comparing the volumes or surface areas, i.e., $C_A = A/A_{\mathrm{CH}}$ and $C_V = V/V_{\mathrm{CH}}$.

The simplest measure of sphericity is the *isoperimetric quotient* q_{IPQ},[115] defined for a polyhedron as

$$q_{\mathrm{IPQ}} = 36\pi\frac{V^2}{A^3} \quad \text{with} \quad q_{\mathrm{IPQ}} \in [0,1] \quad \text{and} \quad D_{\mathrm{IPQ}} = (1 - q_{\mathrm{IPQ}}) \tag{1.22}$$

where $q_{\mathrm{IPQ}} = 1$ for an ideal sphere and D_{IPQ} is a measure of the deviation from an ideal spherical shape. D_{IPQ} is shown for several fullerenes in Figure 1.12.

We usually regard C_{20}-I_h and C_{60}-I_h as almost spherical since all vertices lie on a (covering) sphere. We may therefore analyze whether or not the vertices of a given fullerene lie on a sphere. For this, we define the *minimum covering sphere* as a sphere of minimum radius that encloses all vertices in the polyhedral embedding (J. J. Sylvester, 1857).[116,117] Mathematically, this translates into finding the solution to the problem[118,119]

$$R_{MCS} = \min_{\boldsymbol{c}_{\mathrm{MCS}}} \max_i \|\boldsymbol{v}_i - \boldsymbol{c}_{MCS}\|\,, \tag{1.23}$$

where $\|\cdot\|$ denotes the Euclidian norm, $\boldsymbol{v}_i$ is the location of the vertex v_i in 3D space, and $\boldsymbol{c}_{\mathrm{MCS}}$ is the center of the MCS with radius R and is usually close

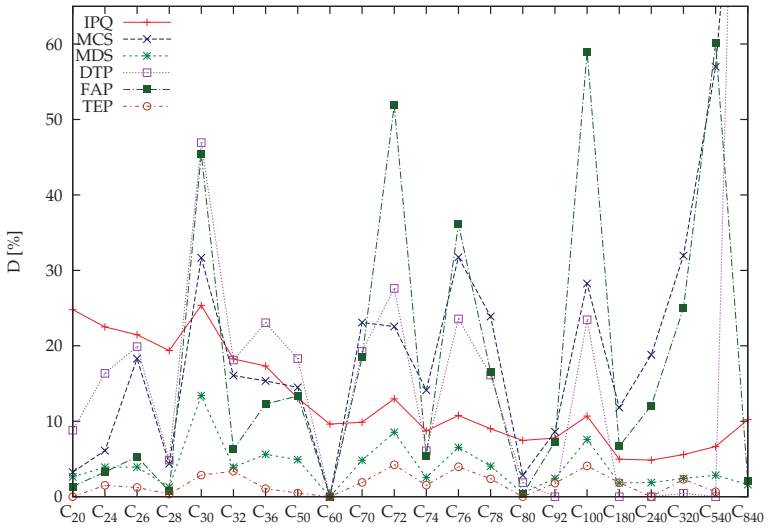


Figure 1.12 Various deformation parameters D (in percent) for a series of fullerenes selected according to stability. For larger fullerenes, Goldberg-Coxeter transformed structures of C_{20} were chosen. Geometries were obtained from DFT (up to C_{540}) or force field optimizations. IPQ: isoperimetric quotient; MCS: Minimum covering sphere; MDS: Minimum distance sphere; DTP: Diaz-Tendero parameter; FAP: Fowler asymmetry parameter; TEP: $(1-\rho)$ from the topological efficiency parameter ρ.

to the barycenter of the fullerene. The MCS is uniquely defined[118,119] and can be obtained using an efficient algorithm, such as the one proposed by Yıldırım.[120] The convex hull of a 3D embedding of a graph G is contained entirely withing the MCS. For fullerenes with an ellipsoidal shape like C_{72}-D_{6d}, or with a cylindrical shape such a nanotube[121] (see Figure 1.2), it is more appropriate to use a minimum covering ellipsoid or cylinder instead of a sphere. The ellipsoidal problem has been addressed recently, and is known as the *minimum volume axis-aligned ellipsoid problem* (MVAE).[122–125]

We can now define a number of useful measures for sphericity. The *MCS distortion* (normalized to the smallest C–C distance in the fullerene) is given by,

$$D_{\mathrm{MCS}} = \frac{1}{N r_{\min}} \sum_{i=1}^{N} \left(R_{\mathrm{MCS}} - \| \boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MCS}} \| \right) \tag{1.24}$$

The MCS definition for the distortion is biased to the case of few atoms sticking out on a sphere and another measure may therefore be more appropriate. We define the minimum distance sphere as

$$D_{\mathrm{MDS}} = \min_{c_{\mathrm{MDS}} \in \mathrm{CH}(S)} \frac{1}{N r_{\min}} \sum_i |R_{\mathrm{MDS}} - \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MDS}}\||| \tag{1.25}$$

with

$$R_{\mathrm{MDS}} = \frac{1}{N} \sum_i \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MDS}}\| \tag{1.26}$$

A similar definition that uses the mean deviation from the average distance taken from the barycentric point has been introduced by Nasu et al.[126]

Some other useful sphericity parameters are the Diaz-Tendero (DTP) parameter,[127,128]

$$q_{\mathrm{DTP}} = \sqrt{(a-b)^2 + (a-c)^2 + (b-c)^2}/a \tag{1.27}$$

where $a \geq b \geq c$ are the rotational constants, and the Fowler asymmetry parameter (FAP),

$$\lambda_{\mathrm{FAP}} = \sum_{i=1}^{n_v} \frac{(R_i - R_{\mathrm{av}})^2}{R_{\mathrm{av}}^2} \tag{1.28}$$

where R_i is the radial distance of atom i from the barycenter, and R_{av} is the average distance.[129] If all atoms lie on a sphere, we have $\lambda_{\mathrm{FAP}} = 0$.

Figure 1.12 shows a comparison of sphericity parameters for a number of stable fullerenes. As C_{20} is slightly deformed due to Jahn-Teller distortion, the vertices do not lie exactly on a sphere anymore, and $D_{\mathrm{MDS}} \neq 0$. In contrast, the purely topological efficiency parameter (TEP, $1 - \rho$) (see discussion in the topological indicator section) is exactly zero for C_{20} (and for C_{60}, C_{80} and C_{240}). For C_{60}, all vertices lie on a sphere and all deformation parameters are zero, except for the IPQ, which except for its simple definition is perhaps not the best measure for sphericity. All deformation parameters reveal the highly deformed fullerenes (spikes in Figure 1.12).

1.10 Topological and chemical indicators

Chemists can deduct many useful properties of a chemical system just by looking at its structure. For fullerenes, the distribution of the 12 pentagons

on a surface, e.g., a sphere, can tell us qualitatively how stable the fullerene is or how it would pack in the solid state. Moreover, the symmetry of the underlying structure determines many useful spectroscopic properties. It is perhaps a realistic goal to connect the graph theoretical properties of a fullerene directly with its physical properties by mapping the fullerene graph G into a number describing that property. In more general terms, we define a *topological indicator* as a map τ_I from the graph G into a finite series of numbers,

$$\tau_I : G \mapsto \{x_1, \ldots, x_n\} \tag{1.29}$$

where the numbers x_i are called *topological indices* and can be chosen to be integers, rationals, or reals. A topological index is called a *chemical index* if it is related to a chemical (or physical) property. Topological indicators may be placed in the same category as crude chemical bonding models, except that they can be strictly defined in graph theoretical terms and sometimes have interesting mathematical properties. There is no restriction that the mapping between fullerene graphs and any topological indicator be one-to-one, and indeed, most of the commonly used topological indicators are the same for many different isomers. Of course, any chemical or physical property is ruled by its underlying electronic structure governed by the Schrödinger equation. However, solving the electronic structure problem for any large fullerene is a daunting task and therefore, topological indices that are easily obtained can be very useful as we shall see.

To give some (trivial) examples, we can list all the possible point groups for fullerenes and define $x_i = 0$ for the ideal point group not matched by the fullerene graph, and $x_i = 1$ if the point group is. The order of the ideal point group $|\mathcal{G}|$ as shown in Table 1.1 is also a topological index. Another example for a topological index is the 12 face spiral pentagon indices as described above.

The first topological index of chemical relevance was introduced by Wiener in 1947.[130] He defined a so-called path number (*Wiener index*) as the sum of the entries in the *topological distance matrix* d_{ij}, consisting of the length of the shortest path between every pair of vertices in the chemical graph representing the non-hydrogen atoms in the molecule,

$$W(G) = \frac{1}{2} \sum_{i,j \in \mathcal{V}} d_{ij} \tag{1.30}$$

where $d_{ii} = 0$. The Wiener index provided a good measure of compactness for acyclic alkanes and gave a reasonable correlation to boiling points.[130] The first application of the Wiener index to fullerenes came from Ori and co-workers, who obtained $W = 8340$ for C_{60}-I_h.[131] For fullerenes, low Wiener indices provide a measure of high compactness of the cage. This can be seen from the scaling law, which has been found to behave like $O(N^{5/2})$ for "spherical" fullerenes[132] and $O(N^3)$ for fullerene nanotubes. For carbon nanotubes, analytical formulae in terms of polynomials in the vertex number are available, e.g., for the smallest D_{5h}/D_{5d} nanotubes we have $W(k) = \frac{1}{3}(100k^3 + 1175k - 2010)$ with $N = 10k$ and $k \geq 5$,[133] and for the smallest D_{6h}/D_{6d} nanotubes we have $W(k) = 12(4k^3 + 69k - 136)$ with $N = 12k$ and $k \geq 7$.[134] It is conjectured that the Wiener index grows like $O(N^{2+1/d})$, where d describes the dimensionality of the system (*the Wiener dimensionality*).[132] The almost cubic growth with increasing number of vertices can make the Wiener index unwieldy for larger fullerenes. Therefore, Ori and coworkers defined a *topological efficiency index* ρ, derived from the Wiener index[135,136]

$$\rho = 2\frac{W(G)}{NW_{\min}} \quad \text{with} \quad W_{\min} = \min_i \left\{ \sum_{j \in \mathcal{V}} d(i,j) \right\} \tag{1.31}$$

which has the advantage that, unlike the Wiener index, it does not grow to large numbers. Small values of $\rho \geq 1$ indicate topologically efficient structures, e.g., both C_{20}-I_h and C_{60}-I_h have $\rho = 1$.[134] It can be seen as a measure of sphericity that does not involve 3D Cartesian coordinates, as shown nicely in Figure 1.12. Vukičević et al. showed that among the 4478 isomers of C_{66}, the joint information of the Wiener index and the topological efficiency index correctly identify the C_{66}-$C_{2v}(11)$ as the most stable molecule[135] (as detected experimentally in endohedral $Sc_2@C_{66}$[137]).

Many different topological indices have been introduced and studied since, mainly for the structural and statistical analysis of molecules, polyhedra, and graphs in general, often yielding interesting mathematical properties (for a list of topological indices see for example Ref.[138]). The distance matrix d_{ij} gives rise to a number of very useful topological indices including the topological radius R and diameter D,[139] Hosoya polynomials[140] and related Wiener indices, the Szeged index S_z,[141] and the

Balaban index J,[30,142] to name but a few (for fullerenes one has $W \sim S_z$ and $W \sim -J$.).[30,143] There are also a number of topological indices connected to the adjacency matrix (or Hückel matrix), for example Hückel orbital energies and related properties (occupation numbers, band gap,[32] spectral moments[144] etc.),[32] and the Estrada index.[145,146] For a more detailed discussion of topological indices see Refs.[30, 80, 139, 147].

Some of these topological indices can directly be related to the stability of the fullerene cage. For example, the *Fowler-Manolopoulos pentagon indices* $\{p_n | n = 0, \ldots, 5\}$ define the number of pentagons attached to n other pentagons (n is called the neighbor index),[32] i.e., for IPR fullerenes we have $p_0 = 12$ and $p_n = 0$ for $n \geq 1$. For any fullerene, we of course must have $p_1 + \cdots + p_5 = 12$. *Fowler-Manolopoulos hexagon indices* are similarly defined and useful for IPR fullerenes, i.e., $\{h_i | i = 0, \ldots, 6\}$, where h_n defines the number of hexagons with neighbor index n. In an IPR fullerene every hexagon is adjacent to a minimum of three others and we can restrict the list of hexagon neighbor indices for example to (h_3, h_4, h_5, h_6).[32]

We may contract the neighbor indices to one useful topological index describing the stability of fullerenes.[32] The single *pentagon signature* P_1 is defined as

$$P_1 = \frac{1}{2} \sum_{k=1}^{5} k p_k \tag{1.32}$$

$\max(P_1) = 30$ (for C_{20} only), and for IPR fullerenes such as C_{60} we have $P_1 = 0$. For IPR fullerenes a more useful single topological index is defined through the hexagon signatures. The standard deviation σ_h of the hexagon neighbor index distribution is defined as

$$\sigma_h = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} \tag{1.33}$$

where

$$\langle k^n \rangle = \frac{H_n}{H_0} \quad \text{with} \quad H_n = \sum_{k=0}^{6} k^n h_k \in \mathbb{N} \quad \text{and} \quad H_0 = \sum_{k=0}^{6} h_k = \frac{N}{2} - 10 \tag{1.34}$$

We call the topological index H_n the *n-th moment hexagon signature*. For general fullerene isomers, low P_1 values and high H_i values correlate with high thermodynamic stability.[32] It turns out, however, that H_1 stays constant

for a given vertex count in IPR fullerenes, as this index is related to certain face patterns as we shall see. Fowler et al. advocated the use of the index σ_h instead, as low values of σ_h are better predictors for stability, and the strain is minimized when all hexagon neighbor indices are as similar as possible.[32] For example, for C_{60} we have $\{h_i\} = \{0, 0, 0, 20, 0, 0, 0\}$ giving H_1 = 60, H_2 = 180 and σ_h = 0. Fowler et al. devised the following scheme for estimating the internal order among fullerenes sorted by stability:[139] First order the isomers according to lowest P_1-value (each pentagon-pentagon fusion carries a penalty of about 20–25 kcal/mol). If multiple IPR fullerenes are present, i.e., $P_1 = 0$, we order these first by largest H_2-value, and finally by lowest Wiener index W (which indicates high compactness of the fullerene cage). The result is an efficient but effective method to find the most stable fullerene isomers.

Ju et al. presented a relationship between the Fowler-Manolopoulos hexagon indices and a particular hexagon structure count to provide a graphical interpretation.[149] The complete list of different pentagon and hexagon combinations up to three connected faces are shown in Figure 1.13. Note that for connecting two or three faces we have

$$n_{55} + n_{56} + n_{66} = E = \frac{3N}{2} \quad \text{and} \quad n_{555} + n_{556} + n_{566} + n_{666} = \frac{11}{2}N - 30 \tag{1.35}$$

where the notation is described in Figure 1.13. For example n_{556} is the sum of all occurring ring patterns containing the combination of two pentagons and one hexagon. Ju et al. showed that for the second moment hexagon signature one gets[149]

$$H_2^{\text{IPR}} = 2n_{66} + 6n_{c666} + 2(n_{l666} + n_{b666}) = 2(n_{l666} + n_{b666}) + 9N - 480 \tag{1.36}$$

and

$$H_1^{\text{IPR}} = 2n_{66} = 3(N - 40) \tag{1.37}$$

Hence for IPR fullerenes, σ_h and H_2 are related through the simple equation

$$\sigma_h^{\text{IPR}} = (N - 20)^{-1} \left(2(N - 20)H_2 - 36(N - 40)^2\right)^{1/2} \tag{1.38}$$

and one requires only H_2 to discuss the stability of IPR fullerenes. Stevanović

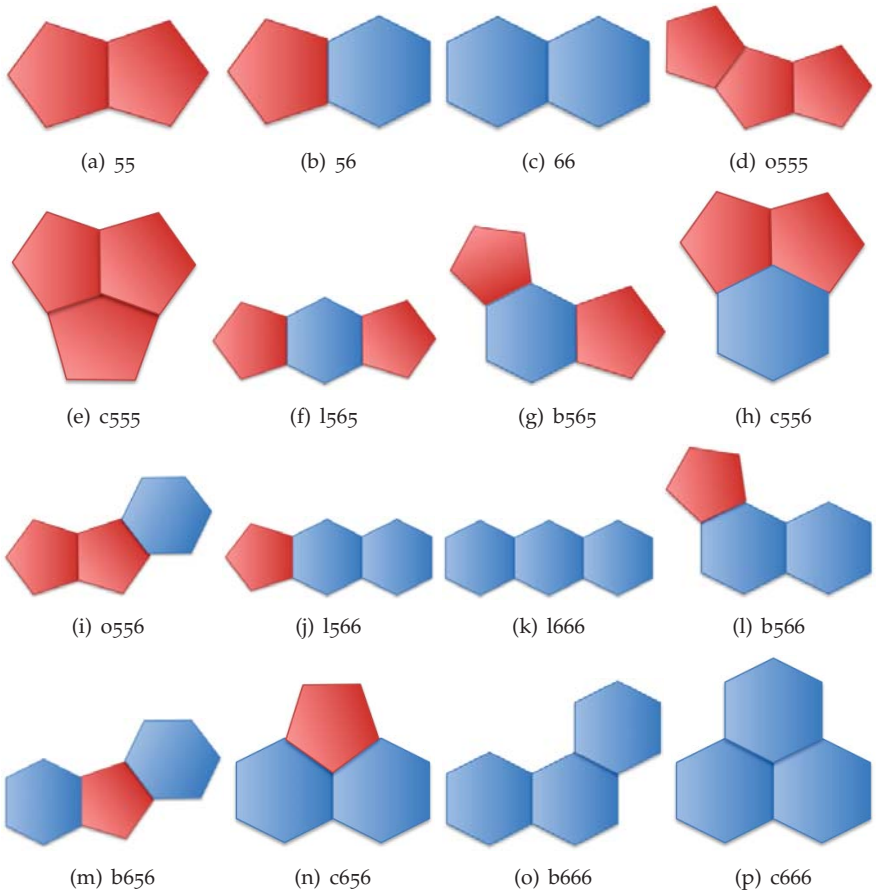


Figure 1.13 The 16 basic shapes for the two- and three-ring (face) adjacencies on the surface of a fullerene. (l) denotes linear, (b) bent, (o) open and (c) closed ring patterns, and the numbers are the face sizes in the patch (see Ref.[148]).

extended these relations to general fullerenes (IPR or not),[150]

$$H_1 = 3(N - 20) - n_{56} \tag{1.39}$$

and

$$H_2 = 18(N - 20) - (6n_{56} + 2n_{c666} + n_{l666} + n_{b666}) \tag{1.40}$$

This not only illustrates nicely the current development in the area of topological indices, it also suggests that the stability of fullerenes can be approximated by counting different face patterns.[148]

Indeed, Cioslowski et al.'s incremental scheme for the heat of formation ΔH_f^o for IPR fullerenes[34] uses 25 linearly independent face patterns (structural motifs) of up to 13 connected hexagons and pentagons to calculate the energy E_{FP}, including an additional curvature term E_C,

$$\Delta H_f = E_{FP} + E_C = \sum_{i=1}^{25} m_i \epsilon_i - 8050.751(N - 30.050)^{-1} \text{ [kcal/mol]} \tag{1.41}$$

where m_i stands for the total count of a specific face pattern (FP) that appears in the fullerene, ϵ_i is the FP energy contribution, and the second term in eq. 1.41 is the curvature term (C), which is repulsive for C_{20} and approaches zero for $N \to \infty$. For example, for C_{60}-I_h we just have one pattern in Cioslowski et al.'s scheme consisting of a hexagon surrounded by 3 pentagons and 3 hexagons. There are 20 of these in C_{60}, and with $\epsilon = 44.281$ kcal/mol, we obtain $\Delta H_f = 616.814$ kcal/mol for C_{60}.[34] This compares rather well with the estimated NIST listed heat of formation for C_{60} (612±25 kcal/mol), although this value should probably be corrected downwards to a value of 602.7 kcal/mol as recently pointed out by Karton et al. using high-level theoretical procedures.[151] The rms error for the standard enthalpies of formation for the 115 IPR isomers chosen was 4.0 kcal/mol compared to B3LYP/6-31G* calculations.

Alcami et al. devised a similar scheme for general fullerenes.[33] By analyzing the most stable structures from B3LYP-DFT calculations of C_{20} to C_{72}, and using only connections between four faces with varying number of pentagons with a total of 9 different motifs, an expression similar to Cioslowski et al. was used for E_{FP}, but without the curvature term E_C. This scheme gives $\Delta H_f = 654.0$ kcal/mol for C_{60}-I_h. For all fullerenes studied

the rms error is 15.4 kcal/mol for the enthalpy of formation compared to DFT calculations, or 0.31 kcal/mol per carbon atom.

The question is: How well are topological indices suited to describe fullerene stability? Table 1.3 shows a comparison of stabilities for all C_{34} isomers obtained from DFT geometry optimizations (using the PBE functional[152] and a def2-SVP basis set for carbon[153]) in comparison with several topological indicators.[36] According to the pentagon signature, isomer C_{34}-$C_2(5)$ has the lowest P_1 value, but with a very small HOMO–LUMO gap obtained from Hückel theory. This is indeed the case as the DFT results show. Alcami et al.'s stability indicator correctly predicts the most stable ($P_1 = 14$) and the least stable isomer ($P_1 = 17$). For the four isomers with pentagon signature $P_1 = 15$, the two results (ΔE^{DFT} and ΔE^{A}) do not agree in their sequence (note that Alcami et al.'s scheme was not designed to distinguish between different isomers of a fixed vertex count). The different fullerene isomers are all very similar in surface area and volume. Concerning the HOMO–LUMO gap, the most stable isomer does not correspond to the one with the largest HOMO–LUMO gap. Moreover, these values are very sensitive to the method applied, and a better measure for the gap given by the singlet-triplet separation. The fullerenes with the smallest HOMO–LUMO gaps (isomers 2 and 6) prefer in fact the triplet ground state. The Hückel HOMO–LUMO gap on the other hand is only good for a qualitative discussion. We note that the two different fullerene isomers 3 and 5 have identical Wiener indices, and the isomer 6 with the smallest Wiener index is not the most stable isomer.

The first halma transform of C_{20} is C_{80} and therefore of I_h symmetry. There are seven IPR isomers out of a total of 31,924 isomers as possible candidates for the energetically most stable one. Table 1.4 shows the properties of all IPR isomers of C_{80}. With larger fullerenes the band gap becomes very small (graphene is a semi-metal or zero-gap semiconductor[155]), and one has to check for states of higher spin multiplicity. For C_{80}, Hückel theory indicates an open shell case for the isomers 6 and 7. In fact, the results show that four of the isomers have a triplet ground state, with the singlet and triplet states for C_{80}-I_h being quasi-degenerate. The DFT calculations further show that the most stable fullerene is not the high symmetry C_{80}-I_h isomer, but the first one in the IPR isomer list, which is of D_{5d} symmetry and has the largest H_2 value (previously, isomer 2 has been predicted to be the most stable isomer[156]). The results show that the first three isomers are

L_I	PG	P_1	ΔE^{DFT}	ΔE^{A}	$\Delta\epsilon^{\mathrm{H}}_{\mathrm{HL}}$	$\Delta\epsilon^{\mathrm{DFT}}_{\mathrm{HL}}$	$\Delta E^{\mathrm{DFT}}_{\mathrm{ST}}$	A	V	W	N_{HC}	N_{PM}
1	C_2	17	74.93	13.5	0.069	0.424	0.0267	81.41	60.30	1978	28	212
2	C_s	15	28.10	4.2	0.120	0.150	-0.153	81.27	61.70	1975	58	219
2(T)	C_s	15	24.55	4.2	-	-	-	81.32	61.69	1975	58	219
3	C_s	15	31.48	4.0	0.368	0.406	0.0945	81.42	61.80	1973	52	196
4	C_2	15	15.19	5.5	0.343	0.619	0.2229	81.32	62.06	1976	33	229
5	C_2	14	0	0	0.006	0.480	0.1333	81.28	62.27	1973	42	204
6	C_{3v}	15	33.70	7.8	0	0.225	-0.094	81.42	61.79	1971	66	195
6(T)	C_{3v}	15	31.54	7.8	-	-	-	81.40	61.75	1971	66	195

Table 1.3 Stability and topological indicators for the 6 isomers C_{34}. L_I is the numbering scheme according to the lexicographically ordered face spiral pentagon indices, PG the ideal point group, P_1 the pentagon signature, energy differences (in kcal/mol) ΔE^{DFT} the DFT(PBE) energy difference to the most stable isomer (kcal/mol), ΔE^{A} the Alcami energy difference to the most stable isomer (kcal/mol), energy gaps (in eV) $\Delta\epsilon^{\mathrm{H}}_{\mathrm{HL}}$ the Hückel HOMO–LUMO gap, $\Delta\epsilon^{\mathrm{DFT}}_{\mathrm{HL}}$ the PBE HOMO–LUMO gap, $\Delta E^{\mathrm{DFT}}_{\mathrm{ST}}$ the singlet-triplet gap (in eV), A the surface area (Å^2), V the volume (Å^3), W the Wiener index (see Ref.[154]), N_{HC} the number of Hamiltonian cycles, and N_{PM} the number of perfect matchings. The symbol (T) for isomers 2 and 6 in the first column indicates that the triplet electronic state is taken instead of the singlet state.

very close in energy and it requires perhaps a more sophisticated electron correlation treatment to sort out the sequence in stability. Isomer 1 also has the highest count in perfect matchings amongst all IPR isomers (although the non-IPR C_{60}-$D_{3d}(4)$ nanotube has a much higher perfect matching count of $N_{\mathrm{PM}} = 524{,}250$). The volumes and surface areas are all very similar with the highest symmetry isomer C_{80}-I_h being largest.

The results shown here clearly demonstrate that the topological indicators help enormously to sort out the most stable isomers. The situation is often quite complicated as seen from the DFT calculations here, or for example from the work of other authors who compared stabilities within a list of isomers.[157–159] In the next section we introduce topological indicators that are connected to π-electron resonance structures (Kekulé structures) in fullerenes.

L_I	PG	H_2	ΔE^{DFT}	ΔE^{A}	ΔE^{C}	ΔE^{HRE}	$\Delta\epsilon^{\mathrm{H}}_{\mathrm{HL}}$	$\Delta\epsilon^{\mathrm{DFT}}_{\mathrm{HL}}$	$\Delta E^{\mathrm{DFT}}_{\mathrm{ST}}$	A	V	W	N_{HC}	N_{PM}
1	D_{5d}	500	0	0	4.34	2.12	0.220	0.355	0.168	203.70	255.16	17340	10450	270153
2	D_2	496	2.63	3.0	0.75	0.0	0.528	0.460	0.293	203.85	257.72	17352	10642	237585
3	C_{2v}	488	1.81	10.5	0.04	14.7	0.102	0.009	0.091	203.92	260.80	17412	9918	201623
4	D_3	492	7.03	7.5	7.17	6.70	0.408	0.098	-0.036	203.87	258.71	17368	9906	222588
4(T)	D_3	492	6.21	7.5	7.17	-	-	-	-	203.85	258.40	17368	9906	222588
5	C_{2v}	484	6.33	16.5	2.92	24.7	0.298	0.038	-0.090	203.99	262.48	17454	9004	182555
5(T)	C_{2v}	484	4.25	16.5	2.92	-	-	-	-	203.98	262.41	17454	9004	182555
6	D_{5h}	480	6.81	22.5	0	37.7	0	0.081	-0.002	204.01	263.50	17500	10970	169375
6(T)	D_{5h}	480	6.38	22.5	0	-	-	-	-	204.00	263.49	17500	10970	169375
7	I_h	480	17.49	30.0	14.2	63.4	0	0.115	-0.003	204.02	263.61	17600	10500	140625
7(T)	I_h	480	17.43	30.0	14.2	-	-	-	-	204.02	263.63	17600	10500	140625

Table 1.4 Stability and topological indicators for the 7 IPR isomers of C_{80}. L_I is the numbering scheme according to the lexicographically ordered IPR face spiral pentagon indices, PG the ideal point group, H_2 the second moment hexagon signature, energy differences (in kcal/mol) ΔE^{DFT} the DFT(PBE) energy difference to the most stable isomer, ΔE^{A} the Alcami energy difference to the most stable isomer, ΔE^{C} the Cioslowski energy difference to the most stable isomer, ΔE^{HRE} the Hückel resonance energy (taking the resonance value as β = -60 kcal/mol), energy gaps (in eV) $\Delta\epsilon^{\mathrm{H}}_{\mathrm{HL}}$ the Hückel HOMO–LUMO gap, $\Delta\epsilon^{\mathrm{DFT}}_{\mathrm{HL}}$ the PBE HOMO–LUMO gap, $\Delta E^{\mathrm{DFT}}_{\mathrm{ST}}$ the singlet-triplet gap (in eV), A the surface area (Å^2), V the volume (Å^3), W the Wiener index, N_{HC} the number of Hamiltonian cycles, and N_{PM} the number of perfect matchings. The symbol (T) indicates that the triplet electronic state is taken instead of the singlet state.

1.11 Kekulé structures and perfect matchings, Clar and Fries structures, and the Pauling bond order

Every chemist learns how to draw double bonds into an aromatic system called a Kekulé (or resonance) structure, sometimes also called a benzenoid graph.[160] For benzene there are only two possible Kekulé structures with three double bonds as we all know, but for C_{60} there are as many as 12,500 Kekulé structures with 30 double bonds (a fullerene has exactly $n_{\mathrm{DB}} = N/2$ double bonds), of which only 158 are symmetry distinct, i.e., non-isomorphic.[161–163] A *Kekulé structure* is the same as a *perfect matching* in the chemical graph $G = (\mathcal{V}, \mathcal{E})$, where the edges of the matching correspond to double bonds. We write $\mathcal{K}(G)$ for the set of all perfect matchings of the graph, and the number of different Kekulé structures $K = |\mathcal{K}(G)|$ is called the *Kekulé number*. In graph theory, a perfect matching is a selection of edges such that every vertex of the graph G is part of exactly one edge in the

matching. The edges of the matching correspond to the double bonds.

Chemists know from basic Hückel theory that for two different benzenoid graphs B_1 and B_2, $K(B_1) > K(B_2)$ implies that B_1 is more stable than B_2 because of resonance stabilization. Schmalz et al. calculated the Kekulé number for several small fullerenes up to C_{84}.[52] Their work showed that for C_{60}, the least stable isomer C_{60}-$D_{5h}(1)$ has K = 16,501, while for the most stable isomer C_{60}-$I_h(1812)$ we have K= 12,500.[52] In fact, Austin et al. showed that 20 isomers of C_{60} have a higher Kekulé numbers than C_{60}-I_h.[164] The Kekulé number is therefore not a good indicator for fullerene stability,[164] but useful for rationalizing the different bond lengths in fullerenes as we shall see.

All cubic graphs have exponentially many perfect matchings and hence exponentially many Kekulé structures.[165] While the theoretical lower bounds proved in Refs.[165] and [166] have small exponents, the actual exponential behavior kicks in rapidly, as seen in Figure 1.14: The mean number of perfect matchings for a C_N isomer is approximately $20 \cdot 2^{(n-20)/4}$. In general, calculating the number of perfect matchings for graphs is intractable,[k] but for planar graphs (such as fullerenes) the Kekulé number can be computed in $\mathcal{O}\left(N^3\right)$ time using the Pfaffian of the adjacency matrix (the Fisher-Kasteleyn-Temperley algorithm).[167]

Pauling, Brockway, and Beach introduced bond orders derived from Kekulé structures in order to rationalize the differences in the C–C bond lengths encountered for polycyclic aromatic compounds.[168] The *Pauling bond order* (PBO) P_{ij} of an edge v_i—v_j is defined as the sum over all appearances of double bonds in perfects matchings

$$P_{ij} = \frac{1}{K} \sum_{M \in \mathcal{K}(G)} \beta_{ij} \tag{1.42}$$

where $\beta_{ij} = 1$ if the edge v_i—v_j has a double bond in the perfect matching M, otherwise $\beta_{ij} = 0$. This gives $0 \leq P_{ij} \leq 1$, with 0 being a pure C–C single bond and 1 a pure C=C double bond. In C_{60}-I_h there are two different

[k] Counting perfect matchings for general graphs is #P-complete. #P is the complexity class of counting solutions for decision problems in NP. While only exponential algorithms exist both for NP-complete and #P-complete problems, in practice #P-complete problems are dramatically harder than NP-complete.

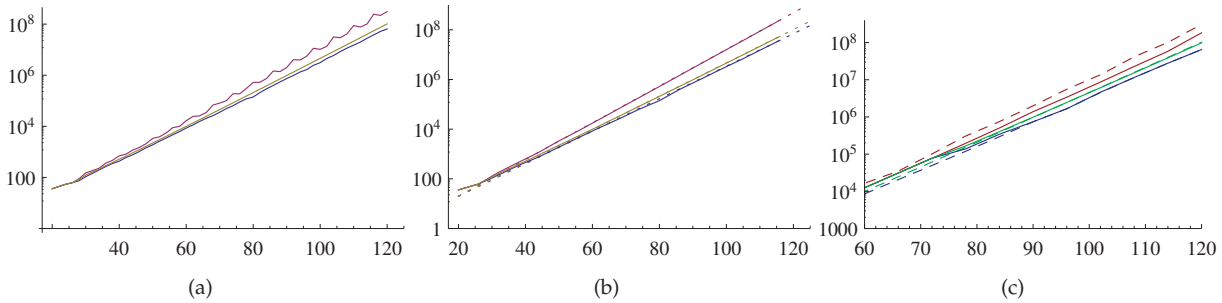


Figure 1.14 Minimum, median, and maximum perfect matching count for all isomers of C_N up to C_{120}. (a) When looking at the plot for all N, we notice three distinct series depending on the value of $N \bmod 6$. (b) For each of the three series, both the maximum and minimum counts follow a simple exponential function (shown as dotted lines). Here, the "peak series" at $N \bmod 6 = 2$ is shown. (c) The bounds for the IPR isomers (solid curves) are shown together with the bound for all isomers (dashed curves) for the series $N \bmod 6 = 0$. The other two series behave similarly.

type of bonds and therefore two different PBOs, depending if the edge shares two hexagons ($P_{hh} = 11/25 = 0.44$) or a hexagon and a pentagon ($P_{hp} = 7/25 = 0.28$).[169] Note that the *Hückel bond orders* derived from the eigenvectors of the adjacency matrix are generally larger with $P_{hh} = 0.6010$ and $P_{hp} = 0.4758$ for C_{60}-I_h.[170]

Narita et al. derived a linear relationship between the PBO and the bond distance in a fullerene obtained from X-ray diffraction,[169]

$$r_{ij}[\text{Å}] = 1.554 - 0.399 P_{ij} \tag{1.43}$$

Figure 1.15 shows an AFM picture of one of the symmetry equivalent hexagons in C_{60} clearly showing the two different type of bonds. The measured bond lengths are $r_{\text{hh}} = 1.38(2)$ Å and $r_{\text{hp}} = 1.454(12)$ Å,[171] in excellent agreement to the values obtained from eq. 1.43 ($r_{\text{hh}} = 1.378$ Å and $r_{\text{hp}} = 1.442$ Å).

Kekulé structures containing benzenoid moieties are of special interest. These are the Fries and Clar structures. *Fries structures* are Kekulé structures with maximum number of benzenoid hexagons. The *Fries number* $Fries(G)$ of a benzenoid graph is the maximum number of benzenoid hexagons over all of its Kekulé structures, and a Fries structure is a perfect matching

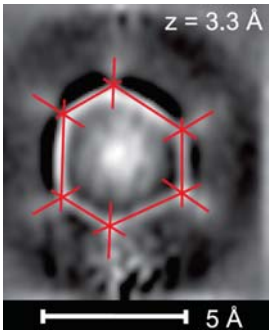


Figure 1.15 Atomic force microscopy (AFM) image for C_{60} by Gross and co-workers[171] at a tip height of $z = 3.3$ Å showing the different bond orders of individual carbon-carbon bonds in a hexagon. (Copyright © 2012, American Association for the Advancement of Science).

that realizes this maximum (there could be many). Similarly, the *Clar number* $Clar(G)$ is the largest number of independent sets of benzenoid faces (separated from each other) over all Kekulé structures, and a *Clar structure* is a perfect matching that realizes this (there could be many). It is clear that a Clar structure localizes benzenoid structures in fullerenes. For C_{60}-I_h we have $Fries(G) = 20$ and $Clar(G) = 8$, and an example for a Fries and Clar structure of C_{60}-I_h is shown in Figure 1.16. One might naively assume that a set of Clar structures form a subset of Fries structures, but this is generally not the case for fullerenes.[172]

We expect maximum stability for fullerenes with the highest Clar number. For example, C_{60}-I_h is unique amongst all other isomers in that it has a Fries structure where all hexagons contain three double bonds and all pentagons none. This is also seen as a reason for the unique stability of C_{60}. However, finding the Clar number is not a trivial problem as it is computationally NP-hard.[173] Ye and Zhang recently showed that for C_{60} there are exactly 18 fullerenes with maximal Clar number of 8.[174] Some of them contain as many as 6 fused pentagons. Hence the Clar number alone is not a good measure for stability. However, out of the 18 fullerenes with Clar number 8, C_{60} has the largest Kekulé count.[175] Fowler showed that leapfrog transforms of fullerenes not only are closed shell, but also have the maximum proportion of benzenoid hexagons.[176]

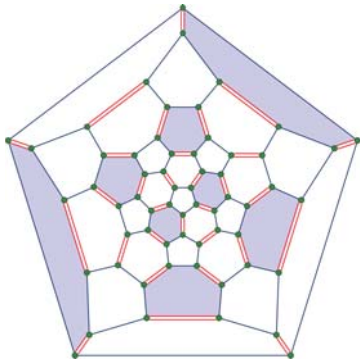


Figure 1.16 One of the Fries structures of C_{60}, and superimposed the structure with highest Clar number ($Clar(C_{60}) = 8$). Double bonds are shown in red and isolated aromatic hexagons are shaded in.

Figure 1.14 (c) shows that the expected number of perfect matchings is no different for IPR fullerenes than it is for all fullerenes, except for small fullerenes for which it is larger than average. However, the maximum number is always significantly lower for IPR fullerenes. One could assume that the larger the Kekulé number is the more Clar sextets can be found and the more stable the fullerene is. While this has not been explored in detail, Table 1.4 shows that for C_{80} the most stable D_{5h} isomer has the lowest perfect matching count. There are, however, some good lower and upper bounds of these topological indices known. Došlić showed that the number of Kekulé structures grows exponentially with a lower bound of $2^{N/8}$.[177] This is not a very tight bound, however, as for C_{60} it predicts at least 181 Kekulé structures, and there are as many as 12,500. The upper bound for Fries numbers is well known, $Fr(G) \leq N/3$, which is seen to be the graphenic limit. Fries structures with a maximum $N/3$ Fries number are called *complete Fries structures*[178] or *perfect Clar structures*. This happens exactly for Clar structures that use every vertex. In that case, the Clar and the Fries structure is the same, and corresponds to a "soccer ball coloring" of the faces. An interesting result comes from Pisanski who proved in a paper with Fowler that the fullerenes with maximum Fries number $N/3$ are exactly the class of leapfrog fullerenes.[179] Given a leapfrog, we can find a perfect Clar structure, and given a perfect Clar structure, we can derive an

inverse leapfrog transformation. This gives an easy way to test whether a particular isomer admits a perfect Clar structure: simply test whether the inverse leapfrog operation is successful. An upper bound for Clar numbers has been obtained by Zhang and Ye, $Clar(G) \leq (N/6 - 2)$.[180] Both IPR fullerenes, C_{60}-I_h and C_{70}-D_{5h} with $Clar(C_{70} = 9)$ achieve this upper bound. Such fullerenes are called *extremal fullerenes*.[180]

Finally we mention that the Kekulé structures can be used to determine the resonance stability of fullerenes. To start with a very simple model, Randić et al. took the Kekulé structures of fullerenes ranging from C_{20} to C_{72}, and calculated the *π-content* P_π of hexagons and pentagons, which is obtained by summing all the Pauling bond orders in a specific ring.[163] For C_{60}-I_h this gives $P_\pi^h = 3P_{hp} + 3P_{hh} = 54/25 = 2.16$ for a hexagon, and $P_\pi^p = 5P_{hp} = 35/25 = 1.40$. High hexagon π-count together with low pentagon π-count gives stability to a fullerene. High hexagon π-count increases the aromaticity in the system (for more details see Ref.[163]). Klein et al. applied Herndon's resonance theory[181,182] using counts of $2m$-resonance cycles to obtain the resonance energy.[161] They correctly pointed out that the reduced p_π-orbital overlap in strongly curved fullerenes will reduce the resonance energy and has to be taken into account.[161] Here we should mention again that obtaining all Kekulé structures for a fullerene is an exponential problem, and it is more advantageous to apply methods that scale polynomially in time, such as Hückel or more sophisticated (semi-empirical) theories.

1.12 Thermodynamic stability and the graphene limit

The spherical shape of C_{60}-I_h with no adjacent pentagons is seen as the main reason for its unusual stability,[183] which underlines the importance of Kroto's isolated pentagon rule (IPR).[19] For example, Nagase and co-workers investigated the stability of potential candidates for the lowest energy structure of C_{74}, with the sole IPR D_{3h} isomer being 16 kcal/mol lower in energy than the C_2 isomer containing two fused pentagons.[184] It is therefore convenient to compare the stability of a fullerene to C_{60}-I_h (relative

fullerene stability, RFS) by using the isodesmic reaction $C_{60}/60 \rightarrow C_N/N$,

$$\Delta E^{\mathrm{RFS}}(N) = \frac{1}{N}E[C_N] - \frac{1}{60}E[C_{60}{-}I_h] \tag{1.44}$$

where $E[C_N]$ is the total energy for a C_N fullerene. For $N \rightarrow \infty$ we obtain the graphene limit, which can be estimated to be approximately $\Delta E^{\mathrm{RFS}}(\infty) = -9$ kcal/mol.

As $E \sim N$, E/N approaches a constant value for infinite systems, e.g. the graphene limit if the pentagons are as separated as possible from each other, so that the polyhedral surface is mainly constituted of flat graphene sheets. Alternatively, for nanotube fullerenes with pentagon caps, the infinite particle limit is just the corresponding infinite nanotube. For the graphene limit, the next dominant term in a $1/N$ expansion will be the surface tension or curvature term, which is $\sim N^{-1}$ as for example used by Cioslowski et al., see eq. 1.41.[34] (although a $N^{-1/2}$ law has been also suggested before[96]). Indeed, as Figure 1.17 shows, the stability follows approximately a linear behavior in our $1/N$ fit (we set $\Delta E_{\mathrm{RFS}} = 0$ for $N = 60$),

$$\Delta E^{\mathrm{RFS}}(N) = 527.4N^{-1} - 8.79 \quad (\mathrm{kcal/mol}) \tag{1.45}$$

that is fullerenes become more energetically stable with increasing vertex number N. It was already shown experimentally that C_{70} is more electronically stable than C_{60}.[185] Alcami et al.'s scheme predicts the correct graphene limit ($N \rightarrow \infty$), and Cioslowski's scheme performs well for medium sized IPR fullerenes, but has an incorrect convergence behavior towards the graphene limit.[33,34] C_{20} is the smallest member of the fullerene family, and as seen from Figure 1.17 also the least electronically stable one. It has been detected in 2000 by Prinzbach[17] (see also Ref.[186]), and a current review on the state of affair concerning C_{20} is given by Fei et al.[187]

It becomes more and more difficult to extract the most stable isomers from the huge isomer space as the size of the fullerene increases. Besides using topological indicators, Hückel theory can be used to determine the resonance stability of a fullerene. It requires diagonalization of the adjacency matrix A_{ij} (the Hückel matrix is defined as $H_{ij} = \alpha\delta_{ij} + \beta A_{ij}$, where α is the Coulomb and β the resonance integral, and δ_{ij} is the unit matrix), and the occupied Hückel orbitals determine the *Hückel resonance stability* (HRS) of a

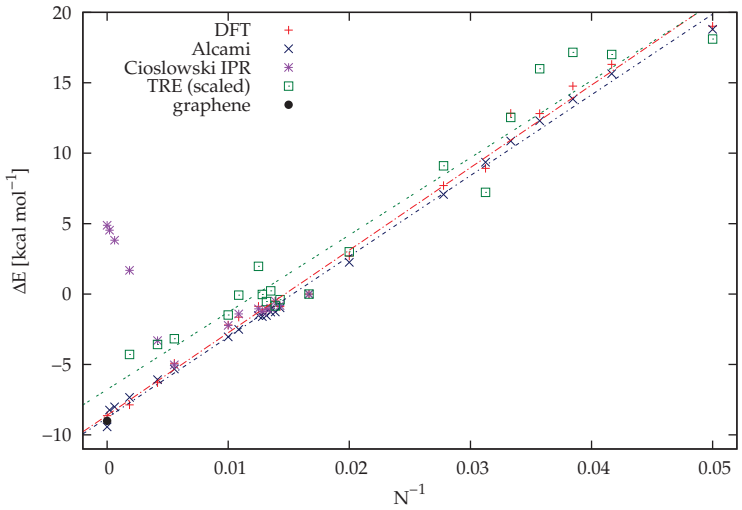


Figure 1.17 The stability of fullerenes C_N in comparison to C_{60} obtained from density functional calculations up to the graphene limit ($N \to \infty$). Topological stability indices from resonance energies[188] (TRE using $\beta = -216$ kcal/mol) or ring patterns by Alcami[33] and Cioslowski[34] are also shown. The graphene limit is estimated from the heat of formation of C_{60}.[151]

fullerene (in units of β),

$$\Delta E^{\mathrm{HRS}}(N) = E_\pi - N \quad \text{with} \quad E_\pi = \sum_{i=1}^{\mathrm{occ}} n_i \epsilon_i \quad (\epsilon_i \in [-3, +3]) \qquad (1.46)$$

where $n_i = 0, 1, 2$ are the occupation numbers, ϵ_i the orbital energy or Hückel orbital i, and the last term comes from subtracting the non-resonant ethene value (single double bond). This gives $E_{\mathrm{HRS}} = 33.161\beta$ for C_{60}. Per π-electron this gives a resonance energy of 0.5527β much higher compared to benzene with exactly $\beta/3$. This should indicate that C_{60} is more aromatic than benzene, which has been a matter of some debate in the past (see the review by Bühl and Hirsch[189]). The problem is that the π-overlap in non-planar systems is not ideal and there is substantial mixing (hybridization) with the p_σ orbitals of carbon, thus questioning the σ-π-separability in Hückel theory. Hence there should be a correction accounting for the curvature on each carbon atom. Such curvature corrections, for example

through using the π-orbital axis vector method (POAV), have been considered before with some success,[190,191] and the different approaches are discussed in detail by Bakowies and Thiel.[192] For example, Haddon et al. showed that C_{240} is significantly more stable than C_{60},[191] which is indeed the case as Figure 1.17 shows. Clearly, as the fullerene system becomes larger, the curvature term becomes smaller. Moreover, for similar shaped fullerenes the Hückel resonance energy might still be a good approximation. Indeed, as Table 1.4 for C_{80} shows, the two first isomers are the most stable ones predicted by the simple Hückel resonance energy term in agreement with DFT calculations, and isomer 7 is the least stable one. The Hückel method only requires a diagonalization of the adjacency matrix, and for obtaining bond orders and resonance energies it is computationally more efficient than sorting out Kekulé structures.

We mention that the Hückel spectrum $\{\epsilon_i\}$ for fullerenes shows some interesting mathematical properties.[193,194] The largest eigenvalue (which translates into the lowest occupied level as the resonance integral β is negative) in a fullerene is always $\epsilon_{\max} = +3$. The smallest eigenvalue (highest unoccupied level) is largest for C_{20} with $\epsilon_{\min} = -\sqrt{5}$, and for the IPR isomers C_{60} we have $\epsilon_{\min} = -(1+\sqrt{5})^2/4$. As a geometric consequence, the famous golden ratio $(1+\sqrt{5})/2$ appears here as it does in the volume and surface area calculations discussed above. We also mention that in general, different graphs can have the same Hückel spectrum,[195] and this is the case also for fullerenes: the eigenvalues do not uniquely determine the fullerene graph.[102]

It is well known that the Hückel scheme for canonical resonance structures is of rather limited use, and Aihara[196] and Gutman et al.[197] therefore introduced a different reference system by defining the topological resonance energy (TRE) for a graph G as

$$\Delta E^{\mathrm{TRE}}(G) = E^{\pi}(G) - E^{\mathrm{MP}}(G) \tag{1.47}$$

where the reference energy $E^{\mathrm{MP}}(G)$ is the sum of all roots given by the *matching polynomial*, which for a fullerene is

$$P^{\mathrm{MP}}(G) = \sum_{k=0}^{N/2} (-1)^k \, c(G,k) x^{N-2k} \quad \left(c^{\mathrm{MP}}(G,k) \in \mathbb{Z} \right) \tag{1.48}$$

Here the polynomial coefficients $c(G,k)$ define the number of ways of choosing k non-adjacent edges from the graph G. Balasubramanian gave polynomial coefficients $c(G,k)$ for a number of fullerenes up to C_{50},[198] Babić determined the coefficients for C_{70},[199] and Salvador for C_{60} and C_{70} to C_{100}.[200] The first few polynomial coefficients are known as they are independent of the isomers for a specific vertex count, i.e., $c(G,0) = 1$, $c(G,1) = -3N/2$, $c(G,2) = 3N(3N-10)/8$, $c(G,3) = -(9N^3 - 90N^2 + 232N)/16$.[198] The last coefficient $c(G,N/2)$ is just the number of perfect matchings in the graph. The computation of these coefficients soon becomes computationally intractable.[200] Babić et al. found, however, a good correlation between E^{MP} and E^{π} values for fullerenes, and the E^{π} value suffices to calculate the topological resonance energy approximately as (in units of β),[188]

$$\Delta E^{\mathrm{TRE}} = 1.024296E^{\pi} - 1.512148N \tag{1.49}$$

In order to give reasonable results for the isodesmic reaction energy (1.44), a rather low value of $\beta = -216$ kcal/mol has to be applied. This scaled TRE plot is shown in Figure 1.17, which is in reasonable agreement with the results obtained from DFT calculations.

The five or six Pauling bond orders in a pentagon or hexagon respectively can be added to give the Pauling ring bond order.[201] For C_{60}-I_h this gives 1.40 ($5 \times 11/25$) for a pentagon and 2.16 ($3 \times 11/25 + 3 \times 7/25$) for a hexagon compared to 3.00 for benzene. According to Randić this suggests only moderate aromaticity for C_{60}-I_h.[201] We are not going into the controversy of aromaticity of fullerenes,[202] but refer to a recent paper by Schleyer and co-workers,[203] who find that both C_{20}-I_h and C_{60}-I_h are not spherically π aromatic[204] but spherically π anti-aromatic, which for C_{60} explains the large heat of formation. In this respect it is interesting that Schleyer and co-workers conclude that fullerenes are not highly stable molecules,[203] which perhaps is in line with the data shown in Figure 1.17 showing no "magic" stability for C_{60} compared to the other fullerenes. We finally mention that fullerene cage abundance is not only guided by thermodynamics, but mostly by kinetic stability. For example, the relative isomer abundance of fullerenes and carbon nanotubes correlates well with kinetic stability.[205]

1.13 Electronic aspects to structure and stability

So far we have discussed topological aspects to determine the structure and stability of fullerenes and basic Hückel theory gives us a first insight into the electronic structure.[206] For example, Manolopoulos et al. have shown that leapfrog fullerenes adopt a closed-shell structure,[179,207,208] and have equal numbers of positive and negative eigenvalues.[209] For a more detailed discussion on this subject see Ref.[210]. However, as we have already seen, more detailed quantum chemical calculations are required to describe the bonding in fullerenes accurately (see also the work by Thiel and co-workers[211–213]). For example, the IPR isomer C_{80}-$C_{2v}(5)$ shown in Table 1.4 prefers a triplet ground state (at the DFT-PBE level of theory) contrary to Hückel theory, which predicts a singlet state. This is perhaps not surprising as already Hückel theory predicts a small HOMO/LUMO gap. As the band gap closes further with an increasing number of vertices (see for example the recent paper by Noël et al.[214]), we cannot expect anymore to predict the correct spin ground state from simple topological arguments. Moreover, the system becomes more multi-reference in nature, making even a quantum theoretical treatment difficult. As an example, C_{50}-$D_{5h}(271)$ was investigated by Lu et al. who showed that two singlet states with different HOMO symmetries are quasi-degenerate.[215]

In addition, some fullerenes may undergo first- or second-order Jahn-Teller (JT) distortions (which could, however, be very small and almost undetectable for larger fullerenes), i.e., they can distort to subgroups of the ideal point group symmetry given by the fullerene topology.[216,217] For example, C_{20} is well known to distort away from the ideal I_h symmetry, which has led to much discussion and debate over the correct electronic ground state.[218] The JT distortion in C_{20} is a consequence of two electrons being distributed over four energetically degenerate g_u-orbitals, resulting in a rather complicated topology of the JT energy hypersurface.[215,217,219–221] JT distortions for highly symmetric open- or closed-shell fullerenes may require a multi-reference treatment. It is therefore often difficult to predict the correct electronic ground state and corresponding physical point group symmetry of a fullerene.

1.14 The gas phase formation of fullerenes

Since the discovery of fullerenes almost 30 years ago, there has been considerable activity, both from the experimental and theoretical side, to gain a detailed understanding of fullerene formation in the gas phase. However, the formation mechanism and especially the high yield of C_{60}-I_h and C_{70}-D_{5h} remains elusive and somewhat controversial.

Fullerenes can be produced by (a) evaporating a carbon target (graphite, amorphous carbon, fullerenes), optionally with addition of metal oxides with a laser,[6] (b) an electric arc between carbon electrodes,[10] or (c) by partial combustion of carbon rich organic compounds.[222] Each of these methods can be adjusted by several experimental parameters (gas pressure, carrier gas, rod feeding rate, amount of oxygen, etc., and therefore indirectly the carbon vapor concentration, expansion rate, annealing time, etc.). Furthermore, fullerenes are found in space,[15] at meteor impact sites, after lightnings and bush-fires, and soot from household candles.[90] These conditions are similar in that carbon vapor is formed at very high temperatures with a deficit of possible reactants like oxygen or hydrogen. The distribution of yielded fullerene cage sizes depends on the production method and the above-mentioned experimental parameters; however, C_{60}-I_h and C_{70}-D_{5h} are always among the most abundant species.

A large number of formation mechanisms have been proposed.[90] First, inspired by their resemblance to graphene sheets or nanotubes, it was suggested that graphene curls or nanotubes break apart to form fullerene cages, followed by various bottom-up strategies that suggest the successive addition of faces (party line[223,224], pentagon road[224]) or polyyne rings (ring stacking[225,226]) before the cage is closed, the folding of chains to form a cage (ring fusion zipper[227]), or the growth of already existing cages (fullerene road[228], closed network growth[89]), and many others. The currently most widely accepted mechanism comes from Irle, Zheng, Wang and Morokuma, and is called the 'shrinking hot giant road'.[90] It is based on the concepts of self organizing structures under non-equilibrium conditions, backed by extensive QM/MD simulations.

This mechanism can be divided into five phases: First, linear polyyne chains and cycles form. In the second stage ('nucleation'), entangled carbon chains rehybridize and form faces. It is to be noted, that pentagons and

hexagons are close in energy at the given temperatures,[90] however, smaller and larger faces are formed as well. Third, more carbon dimers attach to the side chains of an existing nucleus ('growth'), allowing for the formation of additional faces. Fourth, after the formation of sufficiently many faces of sizes < 6, the cage may close spontaneously ('cage closure'), leaving a carbon cage with face sizes not restricted to 5 and 6 and with polyyne side chains attached to it. The final step is the ejection of carbon dimers off side chains and the cage in combination with rapid isomerization of the cage structure resulting in a fullerene without side chains and faces of sizes 5 and 6 only. Fullerenes can not only shrink but also grow in steps of C_2 [86,89,229], and the existence of C_2 and C_3 fragments is backed by spectroscopy[230]. The addition and ejection of small carbon fragments (especially C_2) to and from forming fullerene cages is a fast equilibrium. While the non-cage carbon concentration is high addition prevails; as the carbon vapor expands ejection begins to dominate.[86] The first four steps are exothermic, while the last step is endothermic but increases the overall entropy.

As shown in Figure 1.17, large fullerenes are more electronically stable than small ones, with graphene having a lower energy than any fullerene. In an equilibrium one would therefore anticipate the formation of graphene—the formation of strained cages and especially the high yield of C_{60}-I_h and C_{70}-D_{5h} require further explanation. The experimental conditions of the cooling and expanding carbon vapor are, however, far from an equilibrium, and the whole formation must be understood as a process of self-organization that is governed by kinetics moreso than thermodynamics. Once fullerenes have formed they are subject to restricted equilibration only. Curl et al. have shown numerically that exchange of carbon dimers between existing cages is sufficient to explain the high yield of C_{60}-I_h and C_{70}-D_{5h} as well as a broad distribution of larger cages.[231] They show, that the driving forces for the C_2 exchange are the energy difference between C_N and $C_{N\pm2}$ and the high stability of C_{60}-I_h and C_{70}-D_{5h} relative to their isomers rather than entropy; even at 4000 K entropic effects are not strong enough to explain the existence of small cages.[231]

The experimental observation of endohedral metallofullerenes supports a top-down formation mechanism in two ways: As any enclosed fragments need to enter the cage before it is closed, fullerenes containing fragments that use up most of the space available in the carbon cage must either be formed

top down or their existence implies the breaking and reformation of carbon bonds to open and close the cage after it was formed. Secondly, enclosed metal fragments may stabilize otherwise unstable fullerenes: Zhang et al.[232] report the finding of non-IPR $M_2C_2@C_{84}(51383)$, which they interpret to be a missing link in the top-down road, and which would have ejected further C_2 fragments without the stabilizing metal carbide.

2 Program Fullerene: A Software Package for Constructing and Analyzing Structures of Regular Fullerenes[1]

2.1 Introduction

Since the discovery of buckminsterfullerene C_{60} by Kroto, Smalley and Curl in 1985,[6,9] and its subsequent synthesis in macroscopic amounts by Krätschmer *et al.* in 1990,[10] the chemistry and physics of fullerenes have advanced remarkably in the past two decades.[233,234] There are, however, many open problems related to the structure and stability of fullerenes that remain to be solved. While there exists a number of heuristics that estimate the thermodynamic stability of a particular fullerene, for example by way of spectral analysis of its graph, it is not yet feasible to do so systematically for all isomers of large fullerenes. The number of regular fullerene isomers $N_{\mathrm{iso}}(C_N)$ for a specific number of carbon atoms N increases as $N_{\mathrm{iso}}(C_N) \sim \mathcal{O}\left(N^9\right)$. As an illustration of the growth, C_{60} has 1,812 isomers, and C_{180} has 79,538,751 isomers.[35] The order N^9-growth follows from

[1] Sections 2.1 to 2.13 of this chapter have previously been published by Schwerdtfeger et al.[36] and are reproduced with kind permission from the authors and John Wiley & Sons, Ltd.. The candidate's contribution to the article is quantified on pages 299 ff. The last section of the article (Examples) is omitted. Section 2.14 is added to describe in more detail the structure of the program, in particular functionality that has been added to the back end since this article was published. Furthermore it gives an outlook on possible future extensions of the program.

Thurston's parameterization of triangulations on geodesic domes.[20] Even though the number of isomers is reduced considerably when pentagons are attached to hexagons only—the so-called isolated pentagon rule (IPR), which increases the thermodynamic stability of fullerenes considerably[19,52]—the number of IPR isomers still seem to increase asymptotically as N^9, as shown on Figure 1.1. This is likely due to the proportion of IPR to non-IPR fullerenes growing with increasing N: Out of the 1,812 C_{60} isomers only one is IPR, but out of the isomers for C_{180}, 4,071,832—one in twenty—are IPR.[35] As a consequence, for large fullerenes it becomes difficult to find the thermodynamically most favored isomer. Moreover, graph theoretical indicators yield only rough estimates for stability. Once candidates for stable isomers have been narrowed down, the 3D fullerene structure must be generated before they can be subjected to accurate quantum theoretical treatment.

To obtain a reasonably accurate structure for larger fullerenes is nontrivial. For this, the ring-spiral algorithm, developed by Fowler and Manolopoulos, with a subsequent structure generation using the adjacency matrix eigenvector algorithm (AME) helped to generate and list approximate fullerene structures up to C_{50}, and for IPR fullerenes up to C_{100}, creating "An Fowler-atlas-2006 of Fullerenes" and the first useful database.[32] More recently, Brinkmann *et al.* constructed fullerene isomers up to C_{400}, using the *Buckygen* program.[35,66,235] They also list special fullerenes without ring spirals starting at a pentagon and fullerenes without any ring spiral at all up to C_{400}.[35,236]

The initial structure obtained through various embedding algorithms such as the AME[32,102] or the 3D-TE method (described later) is usually far from the physical 3D structure and requires further optimization, for example by carefully adjusted force fields.[237–239] Standard force fields usually distinguish between single and double bonds. However, the assignment of double bonds (Kekulé structures) to fullerenes is equivalent in graph theory to finding perfect matchings,[240] of which fullerene graphs have exponentially many.[166] While the exponential theoretical lower bound for the number of perfect matchings in fullerenes only kicks in at C_{380}, the rapid growth in their numbers starts much sooner: The IPR fullerene I_h-C_{60} (I_h-C_{70}) has as many as 12500 (52168) Kekulé structures, of which 158 (2780) are non-isomorphic.[163,241] As this number grows, searching through all Kekulé structures for the most stable one becomes impractical, and force

fields for fullerenes should therefore not be designed with double bonds in mind. While it is easy to find *one* perfect matching in $\mathcal{O}(n^2)$ time, if it exists, finding the optimal Clar structure with the maximum number of isolated sextets requires searching through the exponentially many perfect matchings.[242,243]

A similar situation is found if one searches for the lexicographically smallest of the longest carbon chains in a fullerene, required for example for the IUPAC nomenclature of organic compounds.[199] This requires searching through all Hamiltonian cycles[m] in the fullerene graph.[245] Since determining whether a Hamiltonian cycle in a planar cubic graph exists is an *NP*-complete problem, finding all Hamiltonian cycles is at least #*P*-hard and likely cannot be solved in polynomial time. It is not known whether all fullerenes have Hamiltonian cycles, although most 3-regular graphs are Hamiltonian.[246] Brinkmann, Goedgebeur, and McKay demonstrated that Hamiltonian cycles exist for all fullerenes up to C_{316} by generating all isomers and testing for Hamiltonicity.[66]

There are many more open problems concerning fullerene structures related mostly to topological and graph theoretical aspects.[41] We only mention here the problem to find all fullerenes that cannot be generated from ring spirals, the two smallest ones being C_{380} and C_{384},[236] and the close packing problem of fullerene polyhedra to estimate the solid-state 3D structure, which is related to one of Hilbert's famous problems in mathematics and to Kepler's conjecture.[247]

While properties of smaller fullerenes have been explored extensively in the past by both experimental and computational methods,[248] structural, topological and electronic properties of larger fullerenes remain relatively unexplored. Although there are several other computer programs already available to deal with fullerene structures and graphs like *CaGe*(*plantri*,[249] *fullgen*,[250] *Buckygen*[66,235]),[251] *Vega*,[252], *FuiGui*,[253] or the routines introduced in the book by Fowler and Manolopoulos,[32] a comprehensive computer program addressing most of the problems mentioned using effi-

[m] Or longest non-repeating paths if no Hamiltonian cycle exists: it is still an unsolved problem whether all fullerenes have Hamiltonian cycles or not. If Barnette's conjecture[244] is true, it may be possible to construct fast algorithms for counting Hamiltonian cycles in fullerenes. However, it is still an open problem, and consequently the only known algorithms take exponential time.

cient algorithms (of course within the limitations of #*P*-hard or *NP*-complete problems) is to our knowledge not available. We therefore felt that there is a need to develop a general purpose and open-source program whose feature set is a superset of what other programs can do. The program introduced in this paper aims to fill this role.

2.2 General structure and history of the program *Fullerene*

Fullerene is written in standard Fortran and C++ and can easily be installed on a Linux or UNIX environment using the GNU compiler collection (gfortran and g++). The first version of the program *Fullerene* was written in Fortran only for calculating the surface area and volume of a fullerene to determine the corresponding changes due to endohedral incorporation of rare gas atoms in a fullerene cage.[254] A much improved version 2 allowed for the creation of fullerene structures using the ring-spiral algorithm introduced in the book by Fowler and Manolopoulos,[32] and version 3 was already dealing with some topological and graph theoretical properties analyzing basic ring patterns. It soon became clear that many open problems remain in fullerene structure generation and corresponding topological properties, which needed to be addressed from a computational point of view. It was therefore decided to extend version 3 to a general purpose program package for fullerenes and to make it available to the scientific community.

In the current version 4.4, only regular fullerenes are considered (i.e., of genus 0 and consisting of pentagons and hexagons only) fulfilling Euler's polyhedral theorem,

$$n_v - n_e + n_f = 2 \tag{2.1}$$

where n_v is the number of vertices (carbon atoms) ($n_v \geq 20$ and $n_v \neq 22$),[53] $n_e = 3n_v/2$ the number of edges (C–C bonds), and $n_f = n_v/2 + 2$ the number of faces (pentagons or hexagons). From Euler's polyhedral theorem it follows that the number of pentagons in a regular fullerene is exactly 12, no more and no less, known as the 12 pentagon theorem.[255] The main task of the program is to determine the structure of a specific fullerene and to calculate its topological properties. The results can be

used for plotting planar drawings of fullerene graphs (also called Schlegel diagrams) and 3D structures, and serves as a good starting point for further quantum theoretical treatment. A number of approximate physical and chemical properties are calculated as well. A flow diagram for the main steps performed in the program is depicted in Figure 2.1.

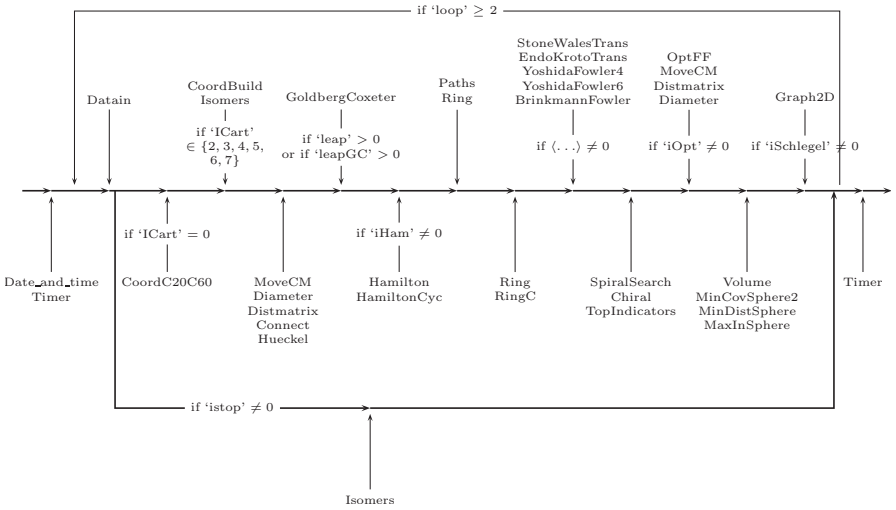


Figure 2.1 Flow diagram for the main tasks of the program *Fullerene*.

2.3 3D structure generation

The first step in the program is to obtain an approximate 3D fullerene structure with correctly assigned bonds between the carbon atoms, *i.e.*, to determine the fullerene graph $G = (V, E)$ with vertex set V and edge set E, and the molecular 3D geometry arising from it. A particular fullerene isomer is uniquely determined by its underlying *fullerene graph*: Any 3-connected planar cubic graph with only pentagons and hexagons as faces defines a unique fullerene, and for any fullerene, the graph defined by the bond structure in turn defines a 3-connected planar cubic graph with only

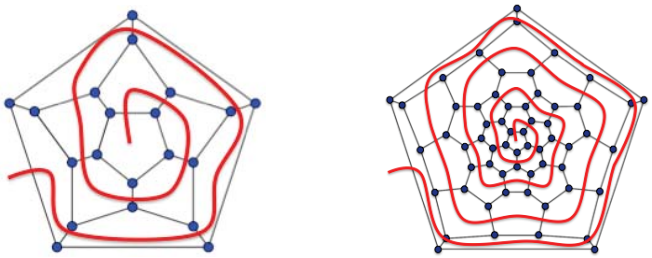
hexagon and pentagon faces. The properties of a fullerene isomer can all be derived starting from its fullerene graph.

Program *Fullerene* allows for a number of different ways to define a fullerene. The user can define the cubic graph $G = (V, E)$ by its adjacency matrix in sparse row format, i.e. for each vertex listing the indices of its three neighbors. Alternatively, Cartesian coordinates for atomic positions can be used as input. In this case, the adjacency matrix is inferred using a heuristic that searches through all distances to find the bonds (Ångstrøms are used throughout). Cartesian coordinates can be obtained from the Yoshida database,[256] or from a previous quantum theoretical calculation, or directly from basic geometry for the two I_h-fullerenes C_{20} and C_{60}.[100]

Another method to construct a specific fullerene isomer is by the Fowler-Manolopoulos ring-spiral algorithm, for isomers whose faces (or carbon rings) can be unwound in a spiral. 12 *ring spiral pentagon indices* (RSPI) specifying which of the n_f faces along the spiral are pentagons, unambiguously define the fullerene structure, as illustrated in Figure 2.2. From the 12 pentagon indices, the fullerene's dual graph $D_t[G]$ is constructed, *i.e.*, the graph corresponding to G where faces and vertices have switched roles. Details of the construction are given in ref.[32]. We then take the dual to obtain the fullerene graph. This works for almost all fullerenes, except for a very small class of fullerenes which cannot be produced from a ring spiral (abbreviated as NS for the following), the first two of which are NS-T-C_{380} and NS-D_3-C_{384}.

Given a fullerene graph, related fullerenes can be derived from it through local *patch replacements*, such as the Stone-Wales transformation[77] or vertex insertion transformations.[82,83] Non-spiral fullerenes can for example be constructed through vertex insertions from smaller fullerenes which contain a ring spiral (see discussion below).

A 3D structure can be obtained from the fullerene graph using either the 3D Tutte embedding (3D-TE) or the adjacency matrix eigenvector (AME)[32] algorithm followed by force field optimization. The 3D-Tutte embedding maps a planar layout of the fullerene graph, obtained for example by the Tutte embedding,[56] to a sphere in a way that is roughly equidistant, and such that bonds never cross. The angles to the barycentre of each topological level in the planar graph layout are used as longitudes, and the topological depth determines the latitude. In the AME algorithm (or spectral

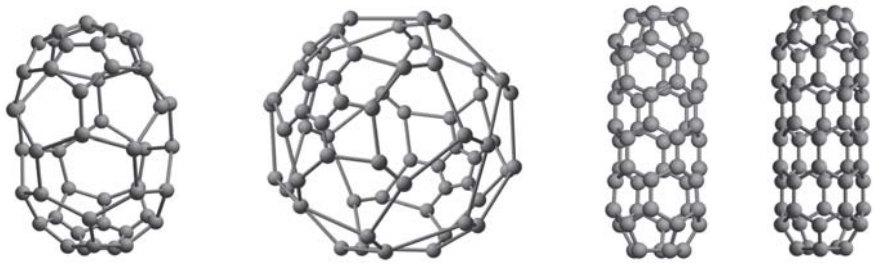


(a) (b)

Figure 2.2 Ring spirals for (a) C_{20} with RSPI $= \{1,2,3,4,5,6,7,8,9,10,11,12\}$, and (b) C_{60} with RSPI $= \{1,7,9,11,13,15,18,20,22,24,26,32\}$.

layout algorithm), the adjacency matrix is diagonalized. From the set of eigenvectors one selects the P-type eigenvectors representing the Cartesian coordinates (X_i, Y_i, Z_i) for all vertices $i = 1, \ldots, n_v$ to construct the Cartesian coordinates of the fullerene.[32] This is, however, not always straightforward, especially for fullerenes that deviate substantially from spherical symmetry, and the Tutte embedding algorithm is perhaps the more reliable solution. Both algorithms have their advantages and disadvantages, *i.e.*, the AME algorithm usually produces an initial geometry closer to the 3D structure than 3D-TE, the latter sets the vertices on a sphere (although other shapes like an ellipsoid, tetrahedron or cylinder with half-sphere caps could be chosen). Figure 2.3 shows the difference between the 3D structures obtained by the two different algorithms in comparison to the 3D structures obtained from a force field and a density functional theory (DFT) treatment.

Program *Fullerene* produces `.xyz` files which can be viewed with standard molecular plotting programs such as *CYLview*[257], *Avogadro*,[258] *Jmol*,[259] *PyMOL*[260] or *VMD*[261]. Files of type `.cc1` (another ASCII file format) contain information about vertex connections, which are useful for other plotting, force field or for quantum chemical programs. Internal coordinates using bond distances, angles and dihedrals are also created.



(a) (b) (c) (d)

Figure 2.3 Structure of D_{5d}-C_{60} carbon nanotube obtained from (a) the AME algorithm, (b) the Tutte embedding, (c) a force-field optimization, and (d) a B3LYP DFT optimization.

2.4 Goldberg-Coxeter transformation

There are several methods to derive higher fullerenes from smaller ones. One such method is the the Goldberg-Coxeter transformation[69] $T^{GC}_{k,l}[C_n]$ that transforms a fullerene C_n into $C_{k^2+kl+l^2}$.[262] The current implementation allows only for leap-frog transformations ($k = l = 1$), LFT, and for Halma transformations ($k \neq 0$ and $l = 0$), HT, of C_n fullerenes, except for $T^{GC}_{k,l}[C_{20}]$, where the RSPIs can be easily obtained as shown by Fowler and Rogers.[263] Successive leap-frog transformations are also implemented, *i.e.*, $(T^{GC}_{1,1})^m[C_n]$, which is useful for the construction of very large fullerenes. Successive leap-frog transformations of C_{50} are shown for example in Figure 2.5.

The general Goldberg-Coxeter transformation $T^{GC}_{k,l}[G_0]$ used here works in the following way:

(*i*) Consider a cubic planar graph G_0 and take its dual $D_t[G_0]$ (faces become vertices and vertices become faces). This transforms the graph G_0 with n_v vertices into a triangulation, *i.e.*, a geodesic sphere with n_v triangles. The dual analogue to the 12 pentagon theorem for fullerene states that if every face of a polyhedron is a triangle and the degree of every vertex is five or six, then the polyhedron has exactly 12 vertices of degree five;

(*ii*) Each triangle T_k ($k = 1, \ldots, n_v$) of the geodesic sphere is tiled with a

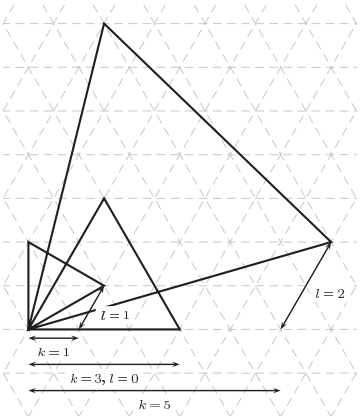


Figure 2.4 Goldberg-Coxeter construction: $T^{GC}_{1,1}$, leap-frog transformation; $T^{GC}_{3,0}$, Halma transformation; $T^{GC}_{5,2}$, general.

set of equilateral triangles t_i. The orientation and relative size (and hence number) of these triangles is given by the Coxeter indices (k, l). Triangles t_i need not be aligned with the edges of T_k, however, as every face T_k is tiled in the same way, complete triangles are obtained when neighboring T_ks are glued together. We end up with a new triangulation, *i.e.*, a larger geodesic sphere $T^{GC}_{k,l} D_t[G_0]$;

(*iii*) Finally we take the dual which yields a new fullerene,

$$G_1 = T^{GC}_{k,l}[G_0] = D_t T^{GC}_{k,l} D_t[G_0] \tag{2.2}$$

If the initial graph G_0 has n_v vertices, the number of vertices of $T^{GC}_{k,l}[G_0]$ is

$$n^{GC}_v = t(k,l) n_v = (k^2 + kl + l^2) n_v \tag{2.3}$$

The value $t(k, l)$ is called the triangulation number. We work with the dual triangulation (geodesic dome)[264] instead of a hexagonal mesh as it is easier to implement. A general $T^{GC}_{k,l}[C_n]$ for all values of k and l is currently in preparation. For the interconversion between the icosahedral GC construction and the RSPIs see Fowler and Rogers.[263]

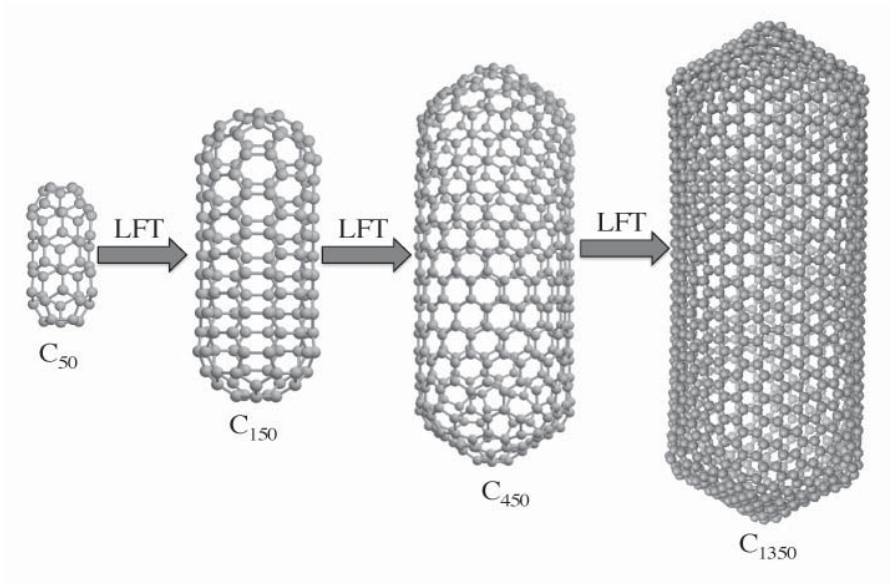


Figure 2.5 Consecutive leap-frog transformations (LFT) starting from D_{5d}-C_{60} nanotube. All structures are force-field optimized.

2.5 Stone Wales transformation and vertex insertions

Brinkmann and Fowler presented a catalogue of possible isomerization transformations and vertex insertions as patch replacements in fullerenes.[74,75] Currently implemented in this program are Stone-Wales transformations,[77] Endo-Kroto 2-vertex insertion,[82] Yoshida-Fowler 4- and 6-vertex insertions,[83] and Brinkmann-Fowler 6-vertex insertion.[75] These are shown schematically in Figure 2.6. The vertex insertions are particularly useful for the construction of the non-spiral fullerenes NS-T-C_{380}[236] or NS-D_3-C_{384} shown in Figure 2.7. NS-T-C_{380} can for example be constructed from T-C_{364} (RSPI = $\{12, 14, 27, 78, 101, 103, 118, 120, 143, 156, 172, 174\}$) by performing four Yoshida-Fowler 4-vertex insertions, and NS-D_3-C_{384} can be constructed from D_3-C_{366} (RSPI = $\{1, 2, 14, 15, 140, 141, 164, 166, 177, 178, 183, 185\}$) by

performing three Brinkmann-Fowler 6-vertex insertions. These two non-spiral fullerenes (and many more) can also be obtained from the House of Graphs database,[35] which can be read by our program.

2.6 Force-field optimizations and vibrational frequencies

The initial fullerene 3D structure coming from the AME or Tutte algorithm is not very accurate and requires a force-field optimization which is done through a Fletcher-Reeves-Polak-Ribiere geometry optimization[265] using analytical gradients. A simple force-field based on the harmonic oscillator approximation (HOFF) is implemented, which considers both bond lengths and angles,[237,238]

$$E_{\mathrm{HOFF}} = \frac{k_p}{2} \sum_{i_p}^{\mathrm{p-edges}} \left(R_{i_p} - R_p\right)^2 + \frac{k_h}{2} \sum_{i_h}^{\mathrm{h-edges}} \left(R_{i_h} - R_h\right)^2 + \frac{f_p}{2} \sum_{j_p}^{60} \left(\theta_{j_p} - \theta_p\right)^2 + \frac{f_h}{2} \sum_{j_h}^{3n_v-60} \left(\theta_{j_h} - \theta_h\right)^2 \quad (2.4)$$

Only bonded pairs of vertices are taken into account. k_p and k_h are the force constants for the two different C–C bonds (set to ≈ 300 kcal Å^{-2}), R_p and R_h the corresponding pentagon and hexagon (adjacent to a hexagon) bond distances (≈ 1.4 Å), and θ_p and θ_h are the corresponding bond angles (108° and 120° respectively). In analogy to the Wu force field, that was developed for I_h-C_{60} only, edges between adjacent pentagons are treated in the same way as edges between a pentagon and a hexagon.

The force-field optimization may lead to distortions of the fullerene to a 3D structure of lower symmetry compared to the original point-group. Moreover, as no dihedral angles are restrained, the molecule might distort away from convexity. In such cases an additional Coulomb repulsive potential can be added (see input instructions) in a pre-optimization step to avoid

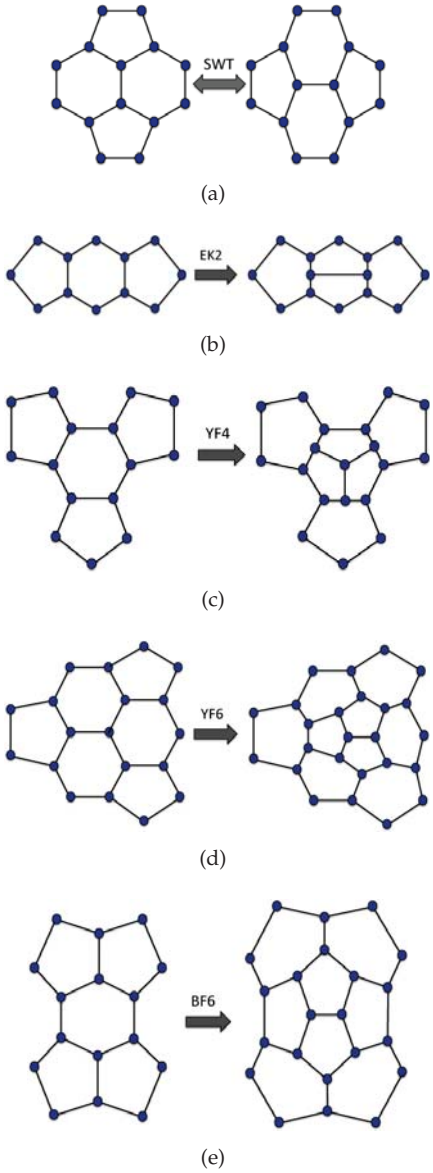
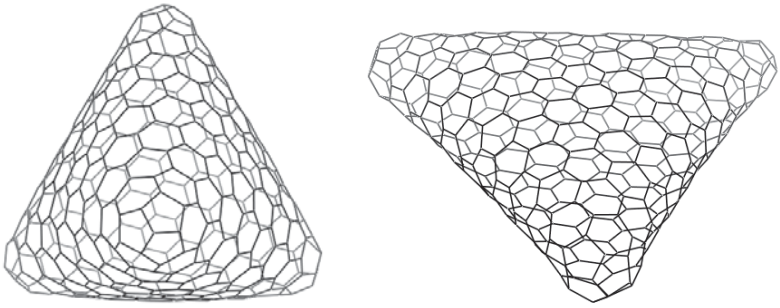


Figure 2.6 (a) Stone-Wales transformations, (b) Endo-Kroto 2-vertex insertion, (c) Yoshida-Fowler 4-vertex insertion, (d) Yoshida-Fowler 6-vertex insertion, and (e) Brinkmann-Fowler 6-vertex insertion.



(a) (b)

Figure 2.7 The two smallest fullerenes which do not contain a ring spiral: (a) NS-T-C_{380}, (b) NS-D_3-C_{384}. Both structures are force-field optimized.

large "dents" or edge crossings in the fullerene structure,

$$E = E_{\mathrm{HOFF}} + E_{\mathrm{Coulomb}} = E_{\mathrm{HOFF}} + \sum_{i}^{n_v} \frac{f_{\mathrm{Coulomb}}}{|\boldsymbol{R}_i - \boldsymbol{R}_0|} \tag{2.5}$$

with $\boldsymbol{R}_i - \boldsymbol{R}_0$ being the distance between vertex i and the barycenter at $\boldsymbol{R}_0$.

A more sophisticated force field using dihedral angles is implemented as well, which we call the extended HO force field. It enhances planarity for areas of connected hexagons and enforces curvature at sites that contain pentagons. The extended HO force field takes three types of bonds (adjacent to 0, 1 or 2 pentagons), two types of angles, and four types of dihedral angles into account. There is one dihedral per atom. Dihedrals θ_{abcd} are defined between one atom a and its three neighbors b, c and d. As one atom is adjacent to three faces (0, 1, 2, or 3 pentagons), there are four different types of dihedrals which differ in their respective zero values and force constants.

The total energy for the extended HO force field is given by

$$E_{\text{extHOFF}} = \frac{f_{\text{pp}}}{2}\sum_{i_{\text{pp}}}^{\text{pp-e}}\left(R_{i_{\text{pp}}} - R_{\text{pp}}\right)^2 + \frac{f_{\text{hp}}}{2}\sum_{i_{\text{hp}}}^{\text{hp-e}}\left(R_{i_{\text{hp}}} - R_{\text{hp}}\right)^2 + \frac{f_{\text{hh}}}{2}\sum_{i_{\text{hh}}}^{\text{hh-e}}\left(R_{i_{\text{hh}}} - R_{\text{hh}}\right)^2 + \frac{f_{\text{p}}}{2}\sum_{j_{\text{p}}}^{60}\left(\theta_{j_p} - \theta_p\right)^2 + \frac{f_{\text{h}}}{2}\sum_{j_{\text{h}}}^{3n_v-60}\left(\theta_{j_h} - \theta_h\right)^2 + \frac{f_{\text{ppp}}}{2}\sum_{k_{\text{ppp}}}^{\text{ppp-v}}\left(\theta_{k_{\text{ppp}}} - \theta_{\text{ppp}}\right)^2 + \frac{f_{\text{hpp}}}{2}\sum_{k_{\text{hpp}}}^{\text{hpp-v}}\left(\theta_{k_{\text{hpp}}} - \theta_{\text{hpp}}\right)^2 + \frac{f_{\text{hhp}}}{2}\sum_{k_{\text{hhp}}}^{\text{hhp-v}}\left(\theta_{k_{\text{hhp}}} - \theta_{\text{hhp}}\right)^2 + \frac{f_{\text{hhh}}}{2}\sum_{k_{\text{hhh}}}^{\text{hhh-v}}\left(\theta_{k_{\text{hhh}}} - \theta_{\text{hhh}}\right)^2 \quad (2.6)$$

where pp-e, hp-e, and hh-e denote the number of edges adjacent to 2, 1, and 0 pentagons, respectively, and ppp-v, hpp-v, hhp-v, and hhh-v the number of vertices between 3, 2, 1, 0 pentagons, respectively. Figure 2.3 shows that the force-field optimized 3D structure for D_{5d}-C_{60} carbon nanotube is quite close to the B3LYP optimized 3D structure.

The second derivative matrix, the Hessian, is produced analytically, and vibrational frequencies can be obtained. The code was tested for the Wu and Weeks parameterization [237,238] with the frequencies provided by Hands *et al.* [239] Note that the force-field parameters in eq. 2.6 have not yet been optimized to known experimental or calculated vibrational frequencies for fullerenes, which is on our to-do list.

2.7 Hamiltonian cycles

The program uses the back-track algorithm developed by Babić [199] to obtain all Hamiltonian cycles (HC) and subsequently, if required, the IUPAC name of the fullerene. Hamiltonian cycles visit every vertex exactly once, returning to the starting point. In a fullerene, a Hamiltonian cycle visits exactly n_v edges,

with $n_v/2$ edges being left out. The number of distinct (non-isomorphic) Hamiltonian cycles has carefully been checked against a simple exhaustive search algorithm for a small number fullerenes up to 110 vertices. Left-right cycles and cyclic vertex permutations count as the same cycle as they are isomorphic to each other. Although counting all Hamiltonian cycles is a #*P*-hard problem, and thus only exponential-time algorithms are known, it is computationally affordable up to about C_{140} for a specific isomer. For larger fullerenes it soon becomes too demanding in computer time. Therefore, for larger fullerenes, the program prints conjectured upper and lower limits for the number of Hamiltonian cycles instead, which are fulfilled for most fullerenes.[266] The existence of Hamiltonian cycles for fullerenes is only conjectured at this stage (part of Barnette's conjecture[244]), but has been verified for all isomers up to C_{316}.[66] Our tight lower bound gives support to the existence of Hamiltonian cycles for all fullerenes.[266]

2.8 Hückel analysis

From the diagonalization of the adjacency matrix

$$A_{ij} = \begin{cases} 1 & \text{if vertices } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

a Hückel analysis can be performed. The eigenvalues yield a good approximation for whether the fullerene has an electronic open- or closed-shell structure.[32] For smaller fullerenes or non-IPR fullerenes, the Hückel analysis may not be very reliable as hybridization with the C(2*s*) orbitals occurs due to non-planarity of the 3D structure (NB: this should not be confused with planarity in graph theoretical terms), and as a consequence the σ–π separation breaks down. The orbital energies are defined as

$$\epsilon_i = \alpha + x_i\beta \tag{2.8}$$

and we adopt $\alpha = -0.21$ au and $\beta = -0.111$ au obtained from the experimental ionization potential (7.58 eV) and excitation energy (3.02 eV) of C_{60} (NB: in C_{60} the electron goes into the second LUMO of t_{1g} symmetry).[267,268] This gives orbital energies for C_{60} in reasonable agreement with DFT Kohn-Sham orbital energies.

The program gives information on the HOMO–LUMO gap. If the highest occupied molecular orbital (HOMO) is degenerate and partially filled, the fullerene may undergo a first-order Jahn-Teller distortion. This can result in a singlet ground state instead of the higher multiplicity state, as for example in the case of C_{20}.[217] Such electronic effects are not captured by our program, and for such systems a proper quantum theoretical calculation is required.

2.9 Stability of fullerenes

Cioslowski *et al.*'s incremental scheme for the calculation of the heat of formation ΔH_f^o for IPR fullerenes[34] is implemented in this program, which uses 25 independent structural descriptors (in the form of patterns in the graph) plus a curvature term,

$$\Delta H_C = \sum_{i=1}^{25} \epsilon_i N_i - 8050.751(n_v - 30.050)^{-1}[\text{kcal/mol}] \qquad (2.9)$$

where N_i stands for the number of specific ring patterns, ϵ_i is the energy contribution to that specific pattern, and the second term in eq. 2.9 is the curvature term, which is repulsive for C_{20} and approaches zero for $N \to \infty$. For example, for C_{60} we just have one pattern in Cioslowski *et al.*'s scheme consisting of a hexagon surrounded by 3 pentagons and 3 hexagons (denoted as pattern 135 in ref.[34]). There are 20 of these, and with $\epsilon_{135} = 44.281$ kcal/mol we obtain $\Delta H_C = 616.814$ kcal/mol for C_{60} (see ref.[34] for details).

Alcami *et al.* devised a similar scheme for general fullerenes.[33] This scheme gives $\Delta H_A = 654.0$ kcal/mol for C_{60}. All these schemes are based on the topology of fullerenes and are (more or less) reliable for smaller to medium sized fullerenes. Alcami *et al.*'s scheme has the correct graphene limit ($n_v \to \infty$), and Cioslowski's scheme performs well for medium sized IPR fullerenes but has an incorrect convergence behavior towards the graphene limit. While these schemes are based on incremental energy contributions to specific face patterns,[33,34] Babić's scheme obtains the π-resonance energy from the matching polynomial of the corresponding fullerene graph.[188,269]

2.10 Topological Indicators

Topological indicators (or indices, also known as connectivity indices) are commonly used in (chemical) graph theory and some have been applied to fullerene theory,[41] *e.g.*, the Wiener index, hyper-Wiener index, Balaban index, Szeged index, Estrada index, mean topological distance, and many more. A topological indicator is a graph invariant natural or real number, i.e., it does not depend on the 2D representation of a graph G. It is however not required that a topological index uniquely defines a graph up to isomorphism, only the converse is true. In the following we introduce the topological indices which are implemented in our program.

The Estrada $E(G)$ index and bipartivity $\beta(G)$ index are obtained from the eigenvalues of the adjacency matrix $A(G)$, $\{x_i, i = 1, ..., n_v\} \in [-3, 3]$, of a fullerene graph G,[145,270]

$$E(G) = \sum_{i=1}^{n_v} e^{x_i} \quad \text{and} \quad \beta(G) = \sum_{i=1}^{n_v} \cosh(x_i)/E(G) \tag{2.10}$$

The Wiener $W(G)$ and hyper-Wiener $WW(G)$ indices are defined as,[130]

$$W(G) = \frac{1}{2} \sum_{\{u,v\} \subseteq V(G)} d(u,v) \quad \text{and} \tag{2.11}$$

$$WW(G) = \frac{1}{2} \sum_{\{u,v\} \subseteq V(G)} \left(d(u,v) + d(u,v)^2 \right) \tag{2.12}$$

Here $V(G)$ is the vertex set of a graph G and $d(u,v)$ is the topological (chemical) distance between the two vertices u and v, *i.e.*, the number of edges in the shortest path connecting them. If we define

$$d_i^{\max} = \max_j \{D_{ij}\} \tag{2.13}$$

we can define the topological diameters D and radius R as[139]

$$D = \max\{d_i^{\max}\} \quad \text{and} \quad R = \min\{d_i^{\max}\} \tag{2.14}$$

We can now define the reverse Wiener index[271] as

$$W_{\text{rev}} = n_v(n_v - 1)D/2 - W \tag{2.15}$$

The Balaban index is defined as[30,142]

$$B(G) = \frac{n_e}{n_e - n_v + 2} \sum_{i<j \in E(G)} \left(W_i W_j\right)^{-1/2}, \tag{2.16}$$

where n_e are the number of edges and W_i is the sum of topological distances between vertex i and all other vertices in the graph,

$$W_i = \sum_{j \in V(G)} d(i,j) \tag{2.17}$$

For fullerenes with equal row sums in the matrix $d(i,j)$, *i.e.*, $\min(W_i) = \max(W_i)$, the following relationship between the Wiener and Balaban indices holds,[139]

$$W(G)B(G) = \frac{9}{4}\frac{n_v^3}{n_v + 4} \tag{2.18}$$

The Szeged index is defined as[272]

$$S_z(G) = \sum_{e \in E(G)} n_{i_j}(e) n_{j_i}(e) \tag{2.19}$$

where $n_{i_j}(e) = |B_{i_j}(e)|$ is the number of vertices that are closer to vertex i than vertex j in an edge $e = (ij) \in E(G)$ (with $E(G)$ being the edge set) and

$$B_{i_j}(e) = \{k | k \in V(G), D_{ik} < D_{jk}\} \tag{2.20}$$

Further mentioned here are two topological efficiency parameters ρ and ρ_E defined by Vukicevic, Ori and co-workers,[135,273]

$$\rho = 2\frac{W(G)}{n_v W_{\min}} \quad \text{with} \quad W_{\min} = \min\{W_i\} \tag{2.21}$$

and

$$\rho_E = \frac{W_{\max}}{W_{\min}} \quad \text{with} \quad W_{\max} = \max\{W_i\} \tag{2.22}$$

The spectral moments $\{\mu_k, k = 0, 1, \dots\}$ are calculated from eigenvalues x_i of the adjacency matrix,

$$\mu_k = \sum_{i=1}^{n_v} x_i^k \tag{2.23}$$

and we have $\mu_0 = n_v$, $\mu_1 = 0$, $\mu_2 = 3n_v$, $\mu_3 = 0$, $\mu_4 = 15n_v$, $\mu_5 = 120$, $\mu_4 = 93n_v - 120$, and $\mu_7 = 1680$. The higher spectral moments depend on the fullerene structure.[274]

One of the first topological indicators introduced to discuss the stability of fullerenes are the neighbor indices for pentagons and hexagons. The program calculates the pentagon neighbor indices,[32,275] hexagon neighbor indices and the pentagon arm indices.[276] Every fullerene isomer can be characterized by a signature of the form $\{i_k|k = 1, \ldots, n\}$. The pentagon indices $(p_i|i = 0, \ldots, 5)$ define the number of pentagons attached to another pentagon, i.e. for IPR fullerenes $p_0 = 12$ and all other pentagon indices are zero. Hexagon indices are similarly defined, i.e. $(h_i|i = 0, \ldots, 5)$, where h_k is the number of hexagons with neighbor index k. In an IPR fullerene every hexagon is adjacent to a minimum of three others and we can restrict the list to (h_3, h_4, h_5, h_6). The pentagon arm indices $(n_i|i = 0, \ldots, 5)$ counts the number of arms of the pentagons. Here an edge E incident to a vertex of a pentagon not belonging to the pentagon is called an arm if, i) both end-vertices of edge E are incident to pentagons, and ii) E shares two neighbor hexagons.[276] This is easily seen to be a Stone-Wales pattern.

We can now define some useful topological invariants.[32] The pentagon index N_p is defined as

$$N_p = \frac{1}{2}\sum_{k=1}^{5} k p_k \quad \text{with} \quad \sum_{k=0}^{5} p_k = 12 \tag{2.24}$$

For IPR fullerenes we have no pentagon attached to another and therefore $p_0 = 12$ and all other $p_i = 0$, and it follows that $N_p = 0$. For IPR fullerenes a more useful index is defined through the hexagon neighbor indices. The standard deviation σ_h of the hexagon neighbor index distribution is defined as

$$\sigma_h = \sqrt{\langle k^2 \rangle - \langle k \rangle^2} \tag{2.25}$$

where

$$\langle k^n \rangle = \frac{\sum_{k=1}^{6} k^n h_k}{\sum_{k=0}^{6} h_k} \quad \text{with} \quad \sum_{k=0}^{6} h_k = \frac{n}{2} - 10 \tag{2.26}$$

Réti and László define the pentagon arm index N_A in a similar way to the

pentagon index,[276]

$$N_A = \frac{1}{2}\sum_{k=1}^{5} k n_k \quad \text{with} \quad \sum_{k=0}^{5} n_k = 12 \tag{2.27}$$

The first and second moments M_1 and M_2 and the variance Var are defined as (real numbers),

$$M_i = \frac{1}{12}\sum_{k=1}^{5} k^i n_k \quad \text{and} \quad \text{Var} = M_2 - M_1^2 \tag{2.28}$$

We now define the Réti-László topological descriptor as Ψ,

$$\Psi = \frac{30 + 6M_1}{1 + 4.5N_p + C(M_1, M_2)} \quad \text{with} \tag{2.29}$$

$$C(M_1, M_2) = \frac{(120M_2)^{1/2}}{(1 + 7M_1)^{1/2}(1 + 0.9\text{Var}^{1.5})} \tag{2.30}$$

If there exists a non-negative integer q for which one of the arm indices $n_q = 12$, the fullerene is called q-balanced. In this case Var $= 0$.[276] If $N_p + N_A = 0$, the fullerene is called strongly isolated. C_{60} contains Stone-Wales patterns and is therefore not strongly isolated. The first leapfrog of C_{60}, C_{180}, and the ones to follow are strongly isolated.

2.11 Volume, Surface Area and Deviation from Spherical Symmetry

It is often useful to determine the volume and surface area of a fullerene, for example to estimate close packing in the solid state or to determine the available space inside a fullerene for possible endohedral encapsulation or for the isoperimetric quotient. We use two measures for the fullerene volume: i) The volume of the fullerene polyhedron P, *i.e.*, with pentagon and hexagon faces. If distortion has made rings non-planar, a triangulation with triangles emanating from ring centers is used; and ii) The volume of the convex hull of the polyhedron. The quotient of the two is used as a

measure for the fullerene's convexity. The volumes V are calculated using the divergence theorem by summing over face normals

$$V(P) = \frac{1}{3}\sum_i \boldsymbol{x}_i \cdot \boldsymbol{n}_i A_i \tag{2.31}$$

where $\boldsymbol{x}_i$ is any point on face i, and A_i is the area of the face. In a similar way, the surface area of the fullerene polyhedron and its convex hull are obtained by summing over the triangles. The orientation of the surface is derived from a planar layout of the fullerene graph, obtained using the Tutte embedding. The convex hull is calculated using the Incremental 3D Convex Hull algorithm.[114]

For I_h-C_{20} and I_h-C_{60} the volumes V and surface areas A can be expressed analytically,

$$V_{C_{20}} = \frac{(15+7\sqrt{5})}{4} R_p^3 \tag{2.32}$$

$$A_{C_{20}} = 3R_p^2\sqrt{5^2+10\sqrt{5}} \tag{2.33}$$

and

$$V_{C_{60}} = \frac{5(3+\sqrt{5})}{12}(2R_p + R_h)^3 - \frac{(5+\sqrt{5})}{2} R_p^3 \tag{2.34}$$

$$A_{C_{60}} = 3R_p^2\sqrt{5^2+10\sqrt{5}} + 5\sqrt{3}\left(R_h^2 + R_p^2 + 4R_pR_h\right) \tag{2.35}$$

The distances R_p and R_h were already defined in eq. 2.4.

It is useful to have a measure of non-sphericity of a fullerene, *i.e.*, by how much a fullerene deviates from an ideal sphere, for example to obtain a single measure for Jahn-Teller distortions.[217] A widely used sphericity measure is the isoperimetric quotient (IPQ) defined as

$$q_{\mathrm{IPQ}} = 36\pi\frac{V^2}{A^3} \quad \text{with} \quad q_{\mathrm{IPQ}} \in [0,1] \tag{2.36}$$

which for an ideal sphere is trivially $q_{\mathrm{IPQ}} = 1$, and for zero volume $q_{\mathrm{IPQ}} = 0$. For C_{20} and C_{60} with equal bond lengths we get $q_{\mathrm{IPQ}}(C_{20}) = 0.75470$ and $q_{\mathrm{IPQ}}(C_{60}) = 0.90317$. We can now define the deviation from a sphere (in %) by

$$D_{\mathrm{IPQ}} = 100\,(1 - q_{\mathrm{IPQ}}) \tag{2.37}$$

The program also determines the sphericity parameter of Diaz-Tendero,[127]

$$q_{\mathrm{SP}} = \sqrt{(a-b)^2 + (a-c)^2 + (b-c)^2}/a \tag{2.38}$$

where $a \geq b \geq c$ are the rotational constants. We deviate from the original definition by dividing the square-root by the rotational constant a to obtain a dimensionless quantity.

Some other useful parameters are determined by the program. The asymmetry parameter of Fowler is defined as,

$$\lambda_{\mathrm{asym}} = \sum_{i=1}^{n_v} \frac{(R_i - R_{\mathrm{av}})^2}{R_{\mathrm{av}}^2} \tag{2.39}$$

where R_i is the radial distance of atom i from the barycenter and R_{av} is the average distance.[129] If all atoms lie on a sphere, we have $\lambda_{\mathrm{asym}} = 0$.

The minimum covering sphere (MCS) of a fullerene is determined using Yildirim's second algorithm.[118,120] Let S be the set of n points $\boldsymbol{p}_i$ ($i = 1, \ldots, n$) in m-dimensional Euclidean space, $\mathbb{R}^m$. The MCS of this set, MCS(S), is a sphere of smallest radius R_{MCS} that encloses the set of points S, and can be expressed as follows,

$$\min_{\boldsymbol{c}_{\mathrm{MCS}}} \max_i \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MCS}}\| \tag{2.40}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\boldsymbol{c}_{\mathrm{MCS}}$ is the center of the MCS. The MCS is uniquely defined and can be expressed as a convex combination of at most $(m+1)$ points (Carathéodory's theorem[277]), hence Yildirim's algorithm stops when at most four points are left over in the iteration process. We changed the first condition in Yildirim's algorithm by choosing R_{max} as the furthest vertex point from the barycenter. In the final statistics there should be 0 points lying outside the sphere and at least 2 points on the sphere. We can now introduce the definition for the MCS distortion parameter D_{MCS} (in % of the smallest bond distance R_{min}),

$$D_{\mathrm{MCS}} = \frac{100}{n_v R_{\mathrm{min}}} \sum_{i=1}^{n_v} (R_{\mathrm{MCS}} - \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MCS}}\|) \tag{2.41}$$

The Van der Waals radius R_{vdW} of carbon (taken as 1.415 Å) can be added to the radius of the MCS, and the volume of an ideal FCC solid is obtained

from ideal sphere packing. The Van der Waals radius is chosen such that for C_{60} the solid-state results of David *et al.* are reproduced.[113] We note that in the case of nonplanar pentagons or hexagons there is no unique definition for the volume of a fullerene, except for the convex hull.

The MCS is biased for the case where few atoms stick out on the surface of a fullerene, and the minimum distance sphere (MDS) may be more appropriate for a measure from spherical distortion,

$$\min_{c_{\mathrm{MDS}} \in \mathrm{CH}(S)} \frac{1}{n_v} \sum_i |R_{\mathrm{MDS}} - \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MDS}}\||$$ (2.42)

with the MDS radius

$$R_{\mathrm{MDS}} = \frac{1}{n_v} \sum_i \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MDS}}\|$$ (2.43)

The restriction of $\boldsymbol{c}_{\mathrm{MDS}}$ to be inside the convex hull is necessary as $\boldsymbol{c}_{\mathrm{MDS}}$ may move out of the fullerene 3D structure in the optimization procedure. In general we have $\boldsymbol{c}_{\mathrm{MDS}} \neq \boldsymbol{c}_{\mathrm{MCP}}$ except for the case where all points lie on the MCS. The MDS may, however, not be uniquely defined, as there could be many (even degenerate) local minima, but for fullerenes it works just fine. This gives perhaps a better definition for the distortion parameter compared to D_{MCS}, as it is not biased to a few points lying outside or inside the fullerene surface. Hence, analogous to the MCS we define a measure for distortion from spherical symmetry through the MDS,

$$D_{\mathrm{MDS}} = \frac{100}{n_v R_{\min}} \sum_{i=1}^{N} |R_{\mathrm{MDS}} - \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MDS}}\||$$ (2.44)

The maximum inner sphere (MIS) is an important measure to estimate whether there is enough space inside the fullerene cage for endohedral encapsulation of atoms or molecules. It is defined as

$$R_{\mathrm{MIS}} = \max_{c_{\mathrm{MIS}} \in \mathrm{CH}(S)} \min_i \|\boldsymbol{p}_i - \boldsymbol{c}_{\mathrm{MIS}}\|$$ (2.45)

The MIS is in general not uniquely defined. For example, consider a fullerene nanotube where many (even degenerate) solutions exist with MISs at different points $\boldsymbol{c}_{\mathrm{MIS}}$ along the tubular axis. For ideal C_{60} (I_h symmetry) the MCS, MDS

and MIS are all identical. For the MIS the Van der Waals radius of carbon can be taken off ($R_{\mathrm{MIS}} - R_{\mathrm{vdW}}$) to determine a measure for the real space available for endohedral encapsulation of atoms or molecules.

2.12 2D Graph representations and Schlegel projections

Fullerenes can be nicely visualized by symmetric planar drawings of their fullerene graphs. Ideally, all vertices and edges should be clearly visible, as well as part of the symmetry elements of the underlying 3D point group. There are a number of different algorithms available trying to achieve this (like the Schlegel projection), *i.e.*, they create $(X_i, Y_i)(i = 1, \ldots, n_v)$ coordinates for a 2D representation of a fullerene graph.[59] For non-spherical fullerenes, like the famous NS-T-C_{380} (see Figure 2.7), or for very large fullerenes, it becomes, however, very difficult to create a good Schlegel projection, and edges may cross in the 2D graph produced. 2D graph representations for C_{60} using a number of different algorithms are shown in Figure 2.8.

In the perspective Schlegel projection (PSP), the 3D representation of the graph is rotated such that a selected vertex or edge of ring (commonly the barycenter of a polygon is chosen) is located at top of the z-axis at some distance below the projection point P, with the z-axis going (not necessarily) through the barycenter of the fullerene. Vertices are then sorted in descending order according to their z-values. The Schlegel projection then yields the projected (X_i, Y_i) coordinates by projecting vertices down on a plane below the fullerene as shown in Figure 2.9. The connections between the points are already written out earlier in the output such that the fullerene graph can be drawn. The corresponding Schlegel diagram is shown in Figure 2.8a.

In the cone Schlegel projection (CSP) the vertices are projected out on an enveloping cone and then down on a plane below the fullerene. The barycenter of the last ring closest to the projection plane should be at the bottom of the fullerene. A scaling factor could be applied for the outer points in the 2D representation of the graph in order to avoid edge crossings. The result of such a projection is shown for C_{60} in Figure 2.8b. This projection

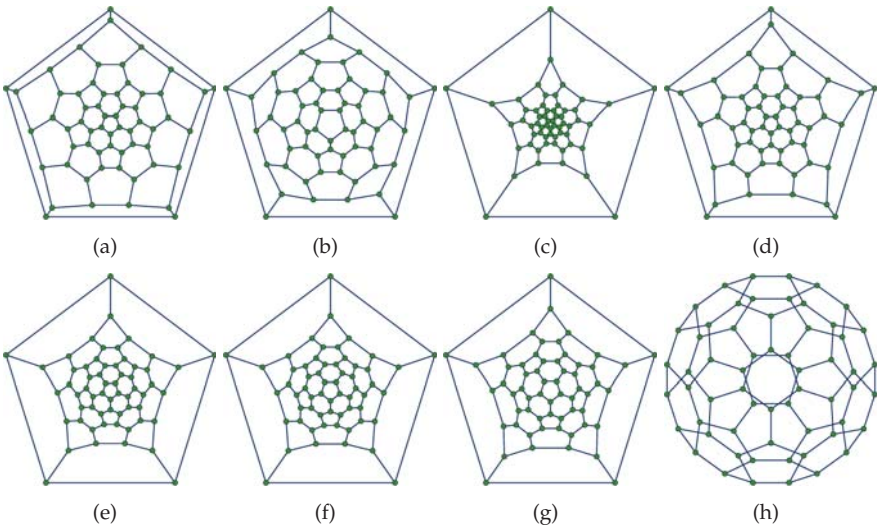


Figure 2.8 2D fullerene structures created using different algorithms for C_{60}. From the left to the right: (a) Perspective Schlegel projection; (b) Cone projection method; (c) Tutte embedding; (d) Tutte embedding plus linear scaling for vertex distances from center of graph; (e) Tutte embedding plus spring embedding optimization; (f) Tutte embedding plus spring and repulsive Coulomb optimization; (g) Pisanski-Plestenjak-Graovac embedding; (h) Kamada-Kawai embedding.

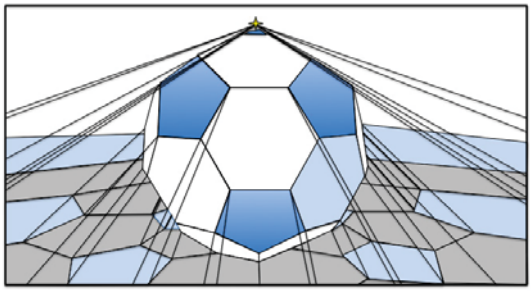


Figure 2.9 Perspective Schlegel projection shown for C_{60}

method is particularly useful for carbon nanotubes where the z-axis of the cone axis is aligned with the main axis of the nanotube.

The result of a Tutte 2D embedding with linear scaling (2D-TE-LS) is shown in Figure 2.8d for C_{60}. Here the Tutte embedding is used,[56] which is guaranteed to be planar and symmetric, but the resulting overcrowding of small polygons near the center of the graph (see Figure 2.8c) is remedied by a linear scaling procedure,

$$\lambda_i = 1 + \frac{f(R_{\min} - R_i)}{R_{\min}} \tag{2.46}$$

i.e., by applying a linear scaling factor λ_i to all vertices v_i ($i = 5(6), \ldots, n_v$), where f is a chosen parameter, $R_{\min}$ is the smallest distance to the vertices of the peripheral taken from the barycenter P of the innermost ring (or, alternatively, the barycenter of the 2D convex hull), and R_i is the distance from that point P to the vertex v_i.

The Pisanski-Plestenjak-Graovac embedding algorithm (PPGA) applied to C_{60} is shown in Figure 2.8g.[278] This is one of the more successful algorithms developed by Pisanski, Plestenjak and Graovac for better visualization of graphs, which is an improvement over usual spring embedders (see for example Figure 2.8e). It has been used by the authors in conjunction with a simulated annealing procedure. Here, the potential between adjacent

vertices is modeled by

$$E_{\mathrm{PPG}} = \sum_{i<j}^{\mathrm{ord}(G)} A_{ij} R_{ij}^2 exp\left\{\alpha \frac{2d_{\max} - d_{iP} - d_{jP}}{d_{\max}}\right\} \tag{2.47}$$

where A_{ij} is the adjacency matrix, R_{ij} the distance between vertex i and j, and α a parameter to be chosen. The distances in eq. 2.47 are defined through the topological distance matrix (also called chemical distance matrix) C, which contains integers C_{ij} for the shortest number of edges required to get from vertex i to vertex j,

$$d_{\max} = \max_{j,j_P} \{C_{ij_P}\} \quad \text{and} \quad d_{ip} = \min_{j_P} \{C_{ij_P}\} \tag{2.48}$$

where j_P is the vertex number belonging to the (5 or 6) peripheral vertices. Instead of a simulated annealing algorithm, we start with a Tutte 2D embedding and perform a minimization procedure for E_{PPG}.

The widely used Kamada-Kawai embedding algorithm[279] shown for C_{60} in Figure 2.8h gives a good 2D view of a 3D structure and therefore has edge crossings (see ref.[279] for details).

2.13 Databases

A compact database is provided for general isomers up to C_{110} and for IPR isomers up to C_{122} including the number of Hamiltonian cycles, and without Hamiltonian cycles up to C_{146} for general and up to C_{192} for IPR isomers. The database contains the RSPIs and other useful information. The RSPI numbering scheme is in canonical order and identical to the one introduced in the book by Fowler and Manolopoulos,[32] that is, each isomer in the book's appendix can be constructed from our database. It is our intention to extend the isomer list beyond C_{146}/C_{192} (without Hamiltonian cycles). The often used Yoshida database[256] has also been added to our database. Finally, files from the *House of Graphs* database can also be used (they can be downloaded from the House of Graphs website).[35] Note that the spiral algorithm cannot compete with the algorithm used in *Buckygen*[66,235] as it goes through all combinations of RSPIs $\{i_k | k = 1, \ldots, 12$ with $i_n < i_m$ for $n < m\}$, and therefore a large number of invalid indices which do not give rise to a valid ring spiral need to be discarded.

2.14 Structure of *Fullerene* and libgraph

This section is aimed at updating and extending section 2.2, with a focus on the underlying library libgraph and the possibilities that arise from it.

In fig 2.1 the sequence of the most important function calls within the main function of *Fullerene* is displayed in an explicit way.

In a more abstract way, all operations in any run of the *Fullerene* program can be summarised by fig. 2.10. Further methods for generating the initial graph, embedding the graph into 3D to obtain the polyhedron, transformations of the graph and the polyhedron, analyses of both, and exporting coordinates may be added in future, without changing the general structure.

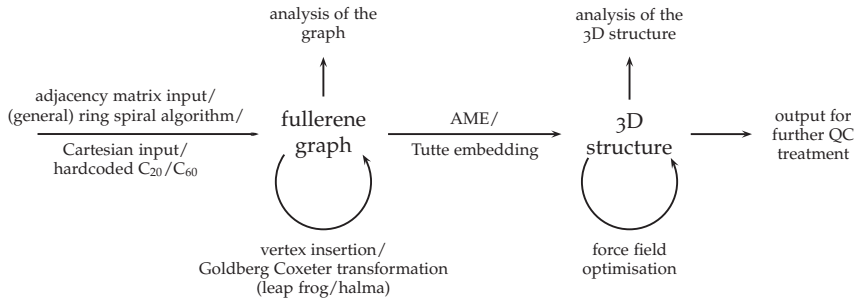


Figure 2.10 Abstract structure of the actions performed by *Fullerene*

This procedure of creating an object, deriving a second object, modifying either, calculating properties of objects, and writing a representation of an object can conveniently be expressed by a class structure in an object oriented programming language.

This structure is reflected by the libgraph library which is written in C++ and serves as a backend for *Fullerene*. The class structure of the most important classes of libgraph is displayed in fig. 2.11; arrows denote inheritance, that is, the class Graph is inherited directly or indirectly by all other classes in this diagram. Several smaller classes that have been omitted for simplicity.

During a run of *Fullerene* objects of type `FullereneGraph` and `Polyhedron` can be created and destroyed through pointers from the Fortran front end.

They are used to represent the graph and the 3D structure of a fullerene as displayed in fig. 2.10.

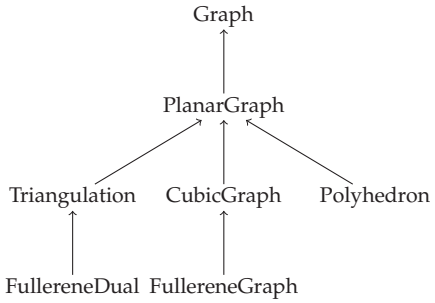


Figure 2.11 Simplified class structure of libgraph

From within the *Fullerene* program the libgraph library is used to provide a number of functions, belonging to different parts of the program run. Here, typically a `FullereneGraph` or `Polyhedron` object is initialised with the Fortran representation of the same object. Pointers to these C++ objects are passed around within a single Fortran function and member functions of the respective classes are called to manipulate the object or to calculate some of its properties.

In the future it is intended to use pointers to libgraph objects more heavily and pass them around within the whole Fortran program, thus avoiding the need to create and destroy objects in different functions and allowing for a much shorter representation of objects in the Fortran code. This will lead to faster and shorter code as well as give the option of treating several fullerene graphs or polyhedra at the same time, e.g. for operating on nested fullerenes.

A second use for libgraph is to directly call its functions from short auxiliary programs written in C++ to solve specific problems that do not fit in the main program, to experiment and to test functionality. There is a collection of such short programs, internally called 'apps' which offer an easy way to optimise arbitrary polyhedra (see the force field described in sec. 4.6), convert between graph descriptions, transform graphs and polyhedra using functions that are not exported to the main program yet,

and perform many other specialised operations. Furthermore, there is a collection of similar programs that are exclusively used for testing and developing new functionality in libgraph.

Not only can libgraph be interfaced by small auxiliary programs, but it would be possible to implement python bindings and create, manipulate and export graphs and polyhedra from short python programs. In the same way, libgraph can be used as a backend to a graphical used interface (GUI). There exists a prototype for a GUI which we intend to integrate into a future release of the *Fullerene* program.

There are a number of features that lend themselves to inclusion into the *Fullerene* program.

One weak point in the generation of realistic 3D structures are the initial 3D coordinates that are optimised using a force field. Both the AME algorithm and the combination of the Tutte embedding of a graph onto a plane and mapping this planar layout onto a sphere may lead to distorted structures when the true structure is large or non-spherical. An algorithm that directly and non-iteratively derives a 3D embedding of a graph is already in work. This embedding produces the correct global shape of the polyhedron, preserves symmetry, and only leads to small local distortions. It is robust even in the case of extremely non-spherical shapes such as nanotubes and should replace the two embedding algorithms that are currently used.

In order to represent nested fullerenes (bucky onions) and fullerenes enclosed in nanotubes (nanotube peapods[280]), the front end could be modified to deal with multiple graphs and polyhedra at the same time. In the backend this would only require few changes, e.g. in the force field, while the representation of each object stays the same.

While nanotubes are a subset of all fullerenes (as long as caps are included), there are more concise and intuitive representations of them, especially if the caps are not the main point of interest. Then, it would be possible to encode a nanotube with only three integers and automatically construct minimal caps to obtain a fullerene graph.

It is possible to extend the scope of the whole program from fullerene graphs, that is polyhedral graphs with pentagons and hexagons only, to polyhedral graphs with faces of size six or smaller, that is all polyhedra with only positive and zero Gaussian curvature. These changes would

mostly affect the Fortran front end of the *Fullerene* program, while libgraph is already capable of representing cubic planar graphs. Extending the functionality to all polyhedral graphs is also possible but requires to modify many more algorithms to deal with negative Gaussian curvature.

3 Non-face spiral fullerenes and general face-spiral algorithms[n]

3.1 Introduction

Fullerenes have exactly 12 pentagons and the location of the pentagons within the network of hexagons uniquely determines the fullerene structure. A convenient way to characterize fullerenes is by using the face-spiral algorithm by Manolopoulos and Fowler (MF).[32,281] This algorithm spirally unwinds (clockwise or anti-clockwise) the faces of a fullerene polyhedron into a string of pentagons and hexagons, and the locations of pentagons in that sequence result in the 12 ring- (or face-) spiral pentagon indices (RSPI) that uniquely determine the fullerene graph G_F up to isomorphism. The RSPI is thus a compact, unique, and canonical representation of fullerene isomers, from which their graphs can easily be constructed.

The success of the MF face-spiral algorithm made it tempting to conclude that every fullerene can be constructed in this manner, because it holds true for the first many million isomers.[32,281] However, this is not so: Manolopoulos and Fowler found the first counterexample in 1993, showing the existence of a T-C_{380} fullerene isomer that cannot be constructed from a face-spiral.[282] We call such fullerenes *non-face-spiral* or simply *non-spiral* (NS) fullerenes in the following. It appears that non-spiral fullerenes are exceedingly rare: the first two non-spiral fullerenes are T-C_{380} and

[n] Sections 3.1 to 3.5 of this chapter have been published as "Structure and properties of the non-face-spiral fullerenes T-C_{380}, D_3-C_{384}, D_3-C_{440}, and D_3-C_{672} and their Halma and Leapfrog transforms" by Wirz et al.[96]. Reprinted with permission from the authors. Copyright 2014 American Chemical Society. The candidate's contribution to the article is quantified on pages 299 ff. Section 3.6 is appended to go into more detail about the general face-spiral algorithm including its implementation.

D_3-C_{384},[83,274,282] which are shown in Figures 3.1 and 3.2. In fact, Brinkmann *et al.* demonstrated that these are the only two NS-fullerenes out of 2,653,606,256,199 fullerenes with 400 atoms or less.[35,66,283] Hence, it is advantageous to identify fullerenes with their canonical face spiral, and to treat the very rare fullerenes that do not have one separately.

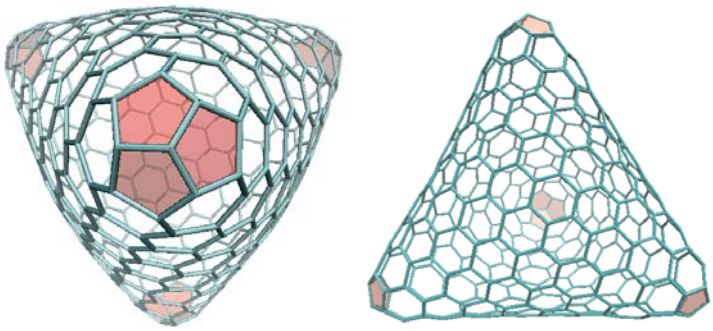


Figure 3.1 Density functional (B3LYP) optimized structure of the NS-fullerene T-C_{380}. View from the top (left) and side (right) of the polyhedron.

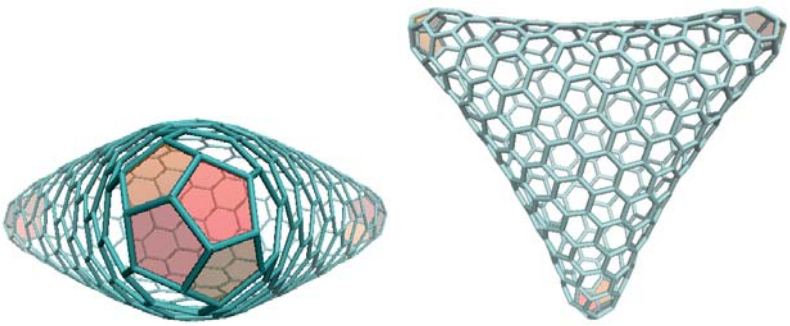


Figure 3.2 Density functional (B3LYP) optimized structure of the NS-fullerene D_3-C_{384}. View from the top (left) and side (right) of the polyhedron.

We recently introduced a comprehensive software package called *Fullerene*, which is a general purpose program for automatically constructing,

transforming, and analyzing fullerenes.[36] The program allows constructing any fullerene (whether or not it has a face spiral) from generalized ring spiral indices by way of a new algorithm, and analyzing topological as well as chemical properties. It produces good approximations to the molecular structures by way of force-field optimization, resulting in geometries that are close to those calculated by a much more demanding quantum theoretical treatment.

The aim of this study was to investigate the structure and topology of the smallest three NS-fullerenes T-C_{380}, D_3-C_{384}, and D_3-C_{440}, as well as the smallest known NS-fullerene with no adjacent pentagons, D_3-C_{672}, using density functional theory (DFT) in order to obtain accurate structural properties. We further aim to study the leapfrog and Halma transforms of T-C_{380}, D_3-C_{384}, D_3-C_{440} and D_3-C_{672}, which are special cases of Goldberg-Coxeter transformations,[68,94] to highlight some important mathematical properties for this class of NS-fullerenes.

3.2 Methods

Our program, *Fullerene*[36], first creates the fullerene graph G_F for each fullerene considered, applying the algorithms as discussed in the next section. The fullerene graph is the undirected, planar, cubic and connected graph describing the C–C bonds, and which uniquely determines a particular fullerene up to inversion. Pairs of enantiomers share the same underlying graph. All four investigated fullerenes as well as the derived structures are chiral but as electronic and geometric properties are independent of inversion we do not distinguish between the enantiomers. To generate the three dimensional molecular structure, the program uses the Tutte embedding algorithm[56] to compute a planar layout of G_F, which is then projected onto a sphere to produce a 3D surface structure without crossing edges. This initial structure is then subjected to a force-field optimization specifically designed for fullerenes as described in detail in Ref.[36].

The force-field optimized structures were taken as input for a refined density functional theory (DFT) optimization, using the gradient corrected Perdew-Burke-Ernzerhof (PBE)[152] functional and the hybrid Becke-Lee-Yang-Parr (B_3LYP)[284,285] functional together with a def2-SVP basis set[153]

and the resolution of identity (RI) approximation.[286] Energies and gradients were obtained with the Turbomole and Gaussian09 program packages.[287,288] Convergence criteria were tightened with respect to the standard values (SCF convergence $< 10^{-10}$ a. u., root mean square force $< 10^{-5}$ a. u. and fine grid for the numerical integration). The optimized structures of T-C_{380} and D_3-C_{384} are shown in Figures 3.1 and 3.2. The next NS-fullerene, D_3-C_{440}, which is similar in shape compared to D_3-C_{384}, is shown in Figure 3.3. The first NS-fullerene with isolated pentagons is D_3-C_{672} and shown in Figure 3.4. We note that the force field implemented in *Fullerene* gives very good initial structures, e. g. the root mean square error in the force-field optimized bond distances is ≈ 0.12 Å compared to the PBE optimized structure.

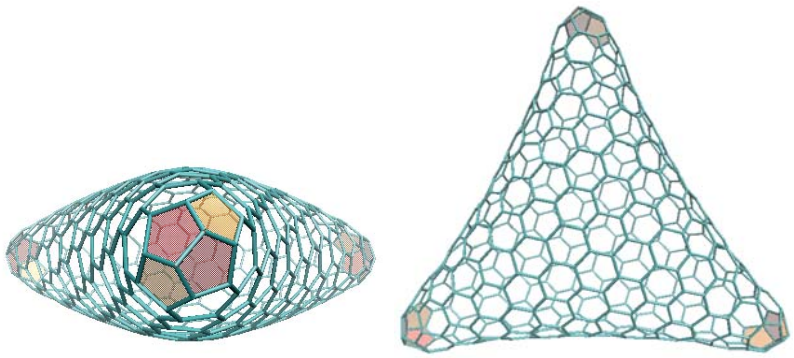


Figure 3.3 Density functional (B3LYP) optimized structure of the NS-fullerene D_3-C_{440}. View from the top (left) and side (right) of the polyhedron.

For each of the four NS-fullerenes, we performed Goldberg-Coxeter (GC) transformations[68,94] $\mathtt{GC}kl(G_F)$ for the special cases $k = l$, and for $l = 0$, in order to investigate whether or not these fullerenes can be unwound into a face spiral. We considered GC-transformed fullerenes up to $k = 7$ for $l = 0$, and to $k = 4$ for $k = l$. These very large structures, comprised of thousands of atoms, were not subjected to DFT optimizations because of the excessive computer time required, but only to force-field optimizations.

In addition, we carried out B3LYP optimizations for the most stable isomer

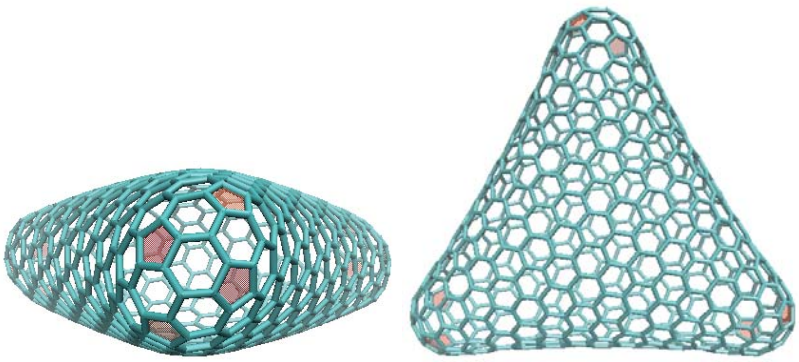


Figure 3.4 Density functional (B3LYP) optimized structure of the NS-fullerene D_3-C_{672}. View from the top (left) and side (right) of the polyhedron.

for a series of C_N structures in the range $20 \leq N \leq 540$ to compare to the stability of the NS-fullerenes. Here we used an augmented correlation consistent PVDZ basis set for $20 \leq N \leq 60$, a correlation consistent PVDZ basis set for $60 \leq N \leq 180$, and a 6-31G basis set[288] for the larger fullerenes with $240 \leq N \leq 540$.

3.3 Results and Discussion

3.3.1 Construction of T-C_{380} from T-C_{364} and D_3-C_{384} from D_3-C_{366} through vertex insertions

The concept of vertex insertions or deletions and formal isomerizations has proven to be useful, as it allows for the step-wise derivation of one fullerene graph from another.[74,75] In particular, fullerene graphs that fail to be constructed using the MF face-spiral algorithm, can be derived from smaller fullerenes that admit a spiral by vertex insertions.

Any patch of hexagons and pentagons can be characterized by its boundary code.[75,289] Any two such patches that are equal in their boundary code can be exchanged for each other. Furthermore, if the symmetry of the boundary code is higher than the symmetry of the patch itself, the patch

can be replaced by a rotated instance of itself. The smallest example of such a rotation is the Stone-Wales transformation.[77] Two patches with the same boundary code must have the same number of pentagons in order to keep the number of pentagons exactly at 12 in the overall fullerene graph, as required by Euler's polyhedron formula. They can, however, differ in their number of hexagons. In patches that contain at most one pentagon, the number of hexagons is uniquely defined by their boundary, i.e., vertex insertions can only be performed by replacing patches that contain at least 2 pentagons.[72,73] Two hexagon-only patches with the same boundary are, however, only guaranteed to be isomorphic if they consist of no more than 24 vertices.[70,72]

In case of a constant number of hexagons a patch replacement is formally an isomerization. Notable examples are the Stone-Wales transformation[77] and the generalized Stone-Wales transformation.[64,85] Formal isomerizations can be represented as the application of 2-switches to a limited domain of the graph. According to Berge's switching theorem,[80,81] every fullerene graph of a given size can be derived from one isomer of the same size by consecutive 2-switches.

In the case of a differing hexagon count, the transformation is called a formal vertex insertion or deletion. The first described examples are the Endo-Kroto 2-vertex insertion,[82] the Yoshida-Fowler 4-vertex insertion,[83] and the Yoshida-Fowler 6-vertex insertion.[83] Brinkmann *et al.* compiled extensive lists of formal isomerizations[74] and formal vertex insertions/deletions.[75]

According to Hasheminezhad *et al.*,[63] every fullerene can be derived from either C_{20} or T_d-C_{28} by applying a sequence of vertex insertions. These vertex insertions (also referred to as 'growth operations') are divided into three classes two of which are of countably infinite size. Hence, as every fullerene can be constructed from C_{20} or T_d-C_{28} (which both admit a spiral), every non-spiralable fullerene can be derived from spiralable fullerenes by applying one or several vertex insertions.

More concretely, the substitution of one patch of vertices for another containing less vertices can be understood as the truncation of a domain of high curvature in the corresponding polyhedron. The curvature (which must always sum to $4\pi = 12 \times \frac{\pi}{3}$ in total) is thereby spread and the distance between some or all pentagons involved increases. Vice versa, the addition

of vertices resembles the capping of one or several faces. As a result, the distance between pentagons in the replaced patch decreases. It is not possible to truncate or cap a domain that contains one or no pentagon. The property of not having a face spiral is a global property of a fullerene graph and cannot be related to a specific subgraph. However, failing spirals typically fail at domains of high curvature, i.e., a subgraph that contains many pentagons in close proximity.

The NS-fullerenes T-C_{380} and D_3-C_{384} can be constructed by subsequent Yoshida-Fowler 4-vertex (YF_4) insertions[83] and Brinkmann-Fowler 6-vertex (BF6) (denoted as G4.14.4.1 $\leftrightarrow$ G4.14.2.2 by Brinkmann *et al.*) insertions,[75] respectively, as shown in eq. 3.1.

$$\begin{aligned} T-C_{364} \xrightarrow{YF_4} C_3-C_{368} \xrightarrow{YF_4} C_2-C_{372} \xrightarrow{YF_4} C_3-C_{376} \xrightarrow{YF_4} T-C_{380} \\ D_3-C_{366} \xrightarrow{BF6} C_2-C_{372} \xrightarrow{BF6} C_2-C_{378} \xrightarrow{BF6} D_3-C_{384} \end{aligned} \tag{3.1}$$

The DFT optimized structures of T-C_{364} and D_3-C_{366}, from which T-C_{380} and D_3-C_{384} are constructed, are shown in Figure 3.5. The corresponding face spiral pentagon indices are listed in Table 3.1. T-C_{364} has no pentagon start for a face spiral, whereas D_3-C_{366} does. From Table 3.2 we also see that D_3-C_{366} admits more face spirals compared to T-C_{364}. All fullerenes leading to D_3-C_{384} by Brinkmann-Fowler 6-vertex insertions have face spirals starting at a pentagon, while the ones leading to T-C_{380} by Yoshida-Fowler 4-vertex insertions have no pentagon start.

3.3.2 Construction of NS-Fullerenes through a Generalized Face Spiral

NS-fullerenes can only practically be found through an extensive search through all possible isomers by using efficient algorithms such as the Brinkmann-Goedgebeur-McKay patch replacement algorithm.[35,66] Once an NS-fullerene has been found it can be completely characterized by a generalized face-spiral algorithm, which we will describe in this section.

The MF face-spiral algorithm was the first algorithm for creating fullerene graphs of a given size.[32,281] Furthermore it offers a short and unambiguous description of fullerene graphs. It works as follows: In the dual of the cubic graph (which is a triangulation of the sphere) three vertices that are

PGS	n_v	RSPI											
	T-C_{364}, C_3-C_{368}, C_2-C_{372}, C_3-C_{376} and GC$kl[T-C_{364}]$												
T	364	12	14	27	78	101	103	118	120	143	156	172	174
C_3	368	12	14	27	78	101	103	118	120	143	173	174	186
C_2	372	12	14	27	78	101	103	119	143	144	175	176	188
C_3	376	13	27	28	80	103	105	121	145	146	177	178	190
T	1092	155	158	187	216	219	232	299	331	361	533	536	547
T	1456	193	211	281	301	317	321	393	397	413	706	722	726
T	3276	459	50	563	704	761	785	814	820	929	1602	1608	1629
T	4368	696	722	728	784	848	854	1270	1302	1426	2153	2159	2166
T	5824	691	821	1098	1168	1240	1248	1540	1620	1723	2810	2881	2906
T	9100	1103	1266	1742	1830	1920	1930	2435	2535	2664	4387	4506	4537
T	9828	1611	1620	1713	1800	1809	1848	2913	3009	3103	4844	4853	4888
T	13104	1611	1807	2534	2640	2748	2760	3534	3654	3809	6314	6493	6530
T	17836	2215	2444	3474	3598	3724	3738	4837	4977	5158	8591	8842	8885
T	17472	2915	2927	3052	3168	3180	3232	5228	5356	5482	8624	8636	8683
	D_3-C_{366}, C_2-C_{372}, C_2-C_{378}, and GC$kl[D_3-C_{366}]$												
D_3	366	1	2	14	15	140	141	164	166	177	178	183	185
C_2	372	1	2	14	15	140	141	164	166	184	185	186	188
C_2	378	1	3	4	5	143	144	167	169	187	188	18	191
D_3	1098	1	7	21	34	429	432	471	475	493	532	536	547
D_3	1464	13	15	95	97	519	521	617	661	703	705	728	732
D_3	3294	9	25	149	152	1228	1231	1444	1450	1580	1583	1637	1643
D_3	4392	1	37	91	150	1746	1752	1876	1918	1926	2043	2156	2190
D_3	5856	111	115	401	532	2063	2067	2451	2726	2808	2812	2910	2918
D_3	9150	41	71	475	480	3396	3401	3996	4006	4387	4392	4547	4557
D_3	13176	195	201	879	1068	4717	4723	5587	6001	6317	6323	6548	6560
D_3	17934	97	141	985	992	6644	6651	7820	7834	8598	8605	8913	8927
D_3	9882	1	79	193	325	3850	3859	4234	4246	4300	4666	4834	4919
D_3	17568	1	153	357	599	6911	6923	7515	7599	7615	8181	8614	8754
	GC$kl[D_3-C_{384}]$												
D_3	1152	1	11	13	24	453	496	498	542	563	572	574	577
D_3	4608	51	108	112	272	1746	1921	2005	2096	2246	2284	2288	2304
D_3	10368	51	186	192	449	4036	4426	4432	4825	5051	5138	5144	5183
	GC$kl[D_3-C_{440}]$												
D_3	1320	14	28	30	70	516	562	564	611	647	656	658	661
	D_3-C_{168} and GC$kl[D_3-C_{672}]$												
D_3	168	4	5	13	14	58	59	74	75	82	83	84	86
D_3	2016	1	41	45	90	711	763	825	849	877	954	958	1008
D_3	8064	1	161	169	351	2893	2997	3345	3393	3449	3815	3823	4030
D_3	18144	1	361	373	784	6547	6703	7561	7633	7717	8584	8596	9068
D_3	32256	1	641	657	1389	11673	11881	13473	13569	13681	15261	15277	16122

Table 3.1 Point group symmetry (PGS), number of vertices n_v, and canonical ring (face) spiral pentagon indices (RSPI) for selected fullerenes. GC$kl[C_n]$ denotes the two-index Goldberg-Coxeter transform of fullerene C_n.

(k,l)	$t(k,l)$	n_v	N_S^{SD}	N_S	N_P	σ_h	$\Delta\epsilon_{\mathrm{HL}}$	W	S_z
				GCkl(T-C$_{364}$)					
(1,0)	1	364	18	216	0	0.7115	0	778866	15215328
(1,1)	2	1092	5	60	0	0.3154	0.485	12186666	439133076
(2,0)	4	1456	25	300	0	0.2767	0	25023132	1044274308
(2,2)	12	4368	5	60	0	0.1638	0.243	390430080	29031693144
(3,0)	9	3276	41	492	0	0.1884	0.281	190158018	12165385368
(3,3)	27	9828	5	60	0	0.1099	0.162	2965469358	334243178292
(4,0)	16	5824	61	732	0	0.1423	0	801528588	69114647592
(4,4)	48	17472	5	60	0	0.0826	0.121	12497408760	606824313414
(5,0)	25	9100	85	1020	0	0.1142	0	2446359618	265400741868
(6,0)	36	13104	113	1356	0	0.0953	0.140	6087715572	795970632660
(7,0)	49	17836	145	1740	0	0.0818	0	13158459978	2013421949184
				GCkl(D_3-C$_{366}$)					
(1,0)	1	366	57	342	6	0.7398	0.235	794568	16091457
(1,1)	2	1098	196	1176	0	0.3788	0.678	12405813	440654361
(2,0)	4	1464	70	420	0	0.3047	0.079	25477143	1046910918
(2,2)	12	4392	266	1596	0	0.1634	0.317	397278072	28568886036
(3,0)	9	3294	175	1050	0	0.1879	0.371	193538835	11997453945
(3,3)	27	9882	902	5412	0	0.1096	0.203	3017221629	326798431890
(4,0)	16	5856	54	324	0	0.1419	0.001	815676747	67618674624
(4,4)	48	17568	639	3834	0	0.0820	0.149	12715128384	1840123932168
(5,0)	25	9150	386	2316	0	0.1139	0	2489397258	258424188237
(6,0)	36	13176	240	1440	0	0.0951	0.174	6194618766	772638775344
(7,0)	49	17934	685	4110	0	0.0816	0	13389275154	1950026430255

Table 3.2 Goldberg-Coxeter indices (k,l), triangulation parameter $t(k,l)$, number of vertices n_v, number of symmetry-distinct (N_S^{SD}) and total (N_S) face-spirals, pentagon neighboring index N_P, hexagon strain parameter σ_h, HOMO–LUMO gap $\Delta\epsilon_{\mathrm{HL}}$ from a Hückel analysis (in eV), Wiener index W, and Szeged index S_z of the Halma and leapfrog transforms of C$_{364}$ and C$_{366}$.

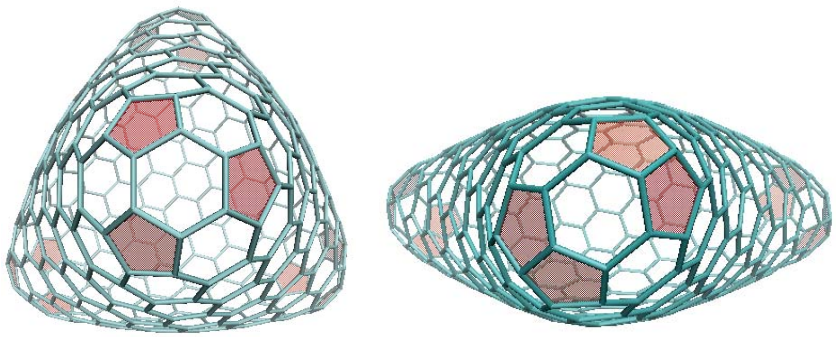


Figure 3.5 Density functional (B3LYP) optimized structures of T-C_{364} (left) and D_3-C_{366} (right) showing the pentagons involved in the Yoshida-Fowler 4-vertex and Brinkmann-Fowler 6-vertex insertions.

mutually connected, i.e., form a face, are chosen as a starting sequence. The remaining vertices are then added one-by-one, such that the next vertex is connected to the previous one and to the vertex that has been added earliest and which has open valencies left. By this procedure a spiral that contains all vertices of the triangulation—corresponding to the faces of the cubic graph—is obtained. In other words, a face spiral in a fullerene is identical to a Hamiltonian spiral path for the fullerene dual. Giving the degree of the vertices (5 or 6 in the case of fullerene graphs) in the order in which they are added to the spiral, uniquely describes the graph. The spiral displayed in Figure 3.6 (a) is denoted as '5555565665655555'. This can be abbreviated by listing the positions of the 12 pentagons only. The aforementioned spiral would then be described by '1, 2, 3, 4, 5, 7, 10, 12, 13, 14, 15, 16'. The graph can unambiguously be reconstructed from this sequence.

The original face-spiral algorithm has the shortcoming that in some fullerene graphs, there are starting sequences that lead to a failing spiral, i.e., there is a step in the spiral generation where the previous vertex and the vertex that has been added earliest and which has open valencies left do not have a common neighbor in the set of remaining vertices. In these cases, the spiral fails (Figure 3.6 (b)). There are fullerene graphs in which *all* starting sequences lead to failing spirals.[236,282] These cannot be generated or described using the original face-spiral algorithm.

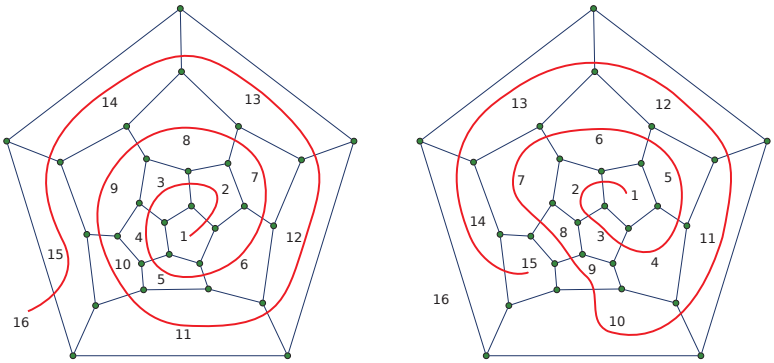


Figure 3.6 Two attempts to construct a spiral in the D_2 isomer of C_{28}: (a) succeeds while (b) misses face 16.

An extended version of the original MF face-spiral algorithm which we implemented recently into the program *Fullerene* eliminates this shortcoming (similar generalized spiral codes have also been proposed by Brinkmann[290] and Fowler and co-workers[289]). During the spiral generation, a walk that traverses the boundary of the subgraph induced by the vertices that have been added to the spiral is maintained. The next vertex to be added is chosen as the vertex in the remaining subgraph that is adjacent to the first and the last vertex in the boundary walk (which is equivalent to the previous criterion). Before adding that vertex, however, it is checked whether the subgraph induced by the set of vertices that have not been added to the spiral yet remains connected after removing it. If this is not the case, a *jump*, i.e., a cyclic shift is performed on the walk. The smallest offset that allows for adding a vertex to the spiral without disconnecting the remaining subgraph is chosen.

By demanding that the remaining subgraph must always remain connected, we ensure that the spiral never gets stuck and hence every fullerene graph can be constructed and described with a general spiral string. As we prove in a forthcoming paper, this algorithm is guaranteed to work for all connected planar cubic graphs.

The general pentagon indices are denoted as $n_1, k_1, \ldots, n_m, k_m; p_1, \ldots, p_{12}$, where n_i, k_i are the positions at which a jump is performed and the respective offset and p_i are the positions of the 12 pentagons. Counting jump- and pentagon-positions starts from 1. If the jump list is empty, we omit the semicolon. The general spiral code describing the spiral in Figure 3.7 (b) is '14, 1; 1, 2, 3, 5, 8, 9, 11, 12, 13, 14, 15, 16', where $n_1, k_1 = 14, 1$ indicates that a cyclic shift of length 1 is performed before adding face 14.

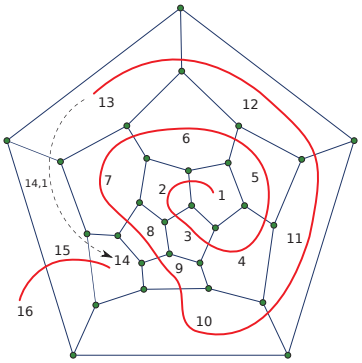


Figure 3.7 The general spiral starting at the same position as the failing spiral in Figure 3.6 (b).

We define the canonical general spiral of a graph, as the shortest of the $6n_v$ spiral codes (where n_v is the number of vertices in the cubic graph). If there is more than one shortest spiral code, the lexicographically smallest is chosen. Eqs. 3.2 to 3.5 list the canonical general spiral codes of the fullerene graphs that are investigated in this paper.

$$C_{380}: \quad 110, 2; 45, 70, 71, 82, 83, 110, 119, 120, 144, 184, 185, 192 \tag{3.2}$$
$$C_{384}: \quad 49, 1; 29, 30, 31, 49, 145, 146, 170, 171, 190, 191, 192, 194 \tag{3.3}$$
$$C_{440}: \quad 62, 1; 39, 40, 41, 62, 170, 171, 197, 198, 218, 219, 220, 222 \tag{3.4}$$
$$C_{672}: \quad 142, 1; 51, 53, 109, 111, 220, 252, 288, 302, 304, 306, 320, 338 \tag{3.5}$$

3.3.3 Electronic Structure, Stability, and Geometry of NS-Fullerenes

The smallest NS-fullerene T-C_{380} has a large HOMO–LUMO gap of $\Delta\epsilon_{HL} = 0.434\,eV$ obtained from a Hückel analysis (Table 3.3). This compares well with the DFT calculations, which give 0.289 eV at the PBE and 0.640 eV at the B3LYP level of theory (the larger gap for the latter functional comes from the exact exchange contribution). D_3-C_{440} has a HOMO–LUMO gap of 0.055 eV, as seen in Table 3.4. In contrast, D_3-C_{384} and D_3-C_{672} have a zero gap, and should undergo a first-order Jahn-Teller distortion into C_2 symmetry (active E-mode). Structurally, these Jahn-Teller distortions are extremely small for such large fullerenes, and not visible to the naked eye. For example, for D_3-C_{384} the optimized distances around the three bonds sharing 2 pentagons where the C_2-axes are situated are 1.4567, 1.4567 and 1.4570 Å (at the optimized B3LYP level of theory). In D_3-symmetry these distances are all equal. However, the DFT calculations show an increased HOMO–LUMO gap of 0.291 eV for D_3-C_{384} using the PBE functional as compared to simple Hückel theory.

Concerning the stability of these fullerenes we can compare to C_{60} by using the following (hypothetical) reaction, which approximates the energy difference per atom:

$$\frac{1}{60}C_{60} \rightarrow \frac{1}{N}C_N + \Delta E_N \tag{3.6}$$

The results together with some other useful information of the optimized DFT structures are shown in Table 3.5. Figure 3.8 shows the stability of the four NS-fullerenes compared to a selection of the most stable fullerenes within a series of isomers from C_{20} to C_{540}. Here ΔE_N is plotted against $N^{-1/2}$ in order to extrapolate to the graphene limit ($N \rightarrow \infty$). We note in passing that in cluster physics a $N^{-1/3}$ behavior is chosen for compact 3D clusters to extrapolate to the solid state. This well-known behavior is derived for example from a liquid-drop model expansion of the free energy.[254,291] For fullerenes the atoms are on the surface of a "sphere", which more realistically points towards a $N^{-1/2}$ behavior. Indeed, this law gives a much better correlation with the calculated ΔE_N values and is therefore applied here.

We first see that the fullerenes separate into two distinct sets, $C_{N<60}$

(k,l)	$t(k,l)$	n_v	N_S^{SD}	N_S	N_P	σ_h	$\Delta\epsilon_{\mathrm{HL}}$	W	S_z
			GCkl(T-C_{380})						
(1,0)	1	380	0	0	12	0.5416	0.434	864762	17592456
(1,1)	3	1140	0	0	0	0.4259	0.428	13513974	493757544
(2,0)	4	1520	0	0	0	0.3250	0.290	27753252	1186112880
(2,2)	12	4560	0	0	0	0.1604	0.206	432893046	32677809552
(3,0)	9	3420	0	0	0	0.1845	0.238	210864450	13744636896
(3,3)	27	10260	0	0	0	0.1076	0.134	3287907642	376364250888
(4,0)	16	6080	0	0	0	0.1393	0.135	888748116	77890894116
(4,4)	48	18240	0	0	0	0.0809	0.099	13856139156	2126356677840
(5,0)	25	9500	0	0	0	0.1118	0.115	2712479754	298658043336
(6,0)	36	13680	0	0	0	0.0933	0.115	6749836440	894840528456
(7,0)	49	18620	0	0	0	0.0800	0.080	14589471486	2261961454188

Table 3.3 Goldberg-Coxeter indices (k,l), triangulation parameter $t(k,l)$, number of vertices n_v, number of symmetry-distinct (N_S^{SD}) and total (N_S) face-spirals, pentagon neighboring index N_P, hexagon strain parameter σ_h, HOMO–LUMO gap $\Delta\epsilon_{\mathrm{HL}}$ from a Hückel analysis (in eV), Wiener index W, and Szeged index S_z of the Halma and leapfrog transforms of C_{380}.

(k,l)	$t(k,l)$	n_v	N_S^{SD}	N_S	N_P	σ_h	$\Delta\epsilon_{\mathrm{HL}}$	W	S_z
				GCkl(D_3-C$_{384}$)					
(1,0)	1	384	0	0	15	0.5191	0	900972	18489834
(1,1)	3	1152	70	420	0	0.4481	0.769	14063379	503477097
(2,0)	4	1536	0	0	0	0.3354	0	28880766	1196989728
(2,2)	12	4608	8	48	0	0.1596	0.350	450300945	32482982520
(3,0)	9	3456	0	0	0	0.1836	0.411	219381558	13686989061
(3,3)	27	10368	178	1068	0	0.1071	0.220	3419854512	371063413764
(4,0)	16	6144	0	0	0	0.1386	0	924571380	77056229226
(4,4)	48	18432	0	0	0	0.0805	0.158	14411798937	2087919325548
(5,0)	25	9600	0	0	0	0.1112	0	2821708806	294297617166
(6,0)	36	13824	0	0	0	0.0928	0.186	7021504974	879481127910
(7,0)	49	18816	0	0	0	0.0796	0	15176486088	777771569091
				GCkl(D_3-C$_{440}$)					
(1,0)	1	440	0	0	15	0.4866	0.055	1267804	27865848
(1,1)	3	1320	19	114	0	0.4196	0.736	19785768	757459407
(2,0)	4	1760	0	0	0	0.3142	0	40628424	1800847608
(2,2)	12	5280	0	0	0	0.1493	0.335	633463320	48809068236
(3,0)	9	3960	0	0	0	0.1718	0.392	308602644	20579946951
(3,3)	27	11880	0	0	0	0.1001	0.209	4810808520	557305601235
(4,0)	16	7040	0	0	0	0.1296	0.004	1300566284	115828515936
(4,4)	48	21120	0	0	0	0.0752	0.150	20273362905	3135159889380
(5,0)	25	11000	0	0	0	0.1040	0.0002	3969179866	442300731228
(6,0)	36	15840	0	0	0	0.0867	0.177	9876814590	1321616670198
(7,0)	49	21560	0	0	0	0.0744	0.002	21347979990	3334263250641
				GCkl(D_3-C$_{672}$)					
(1,0)	1	672	0	0	0	0.4921	0	3632139	100189002
(1,1)	3	2016	194	1164	0	0.2377	0.474	56675358	2734390758
(2,0)	4	2688	0	0	0	0.2073	0	116374455	6513298290
(2,2)	12	8064	463	2778	0	0.1212	0.223	1814461428	177071344866
(3,0)	9	6048	0	0	0	0.1397	0.262	883922250	74585962914
(3,3)	27	18144	1116	6696	0	0.0811	0.145	13779758694	2024937702174
(4,0)	16	10752	0	0	0	0.1052	0	3725140125	420195517416
(4,4)	48	32256	1662	9972	0	0.0609	0	58069545396	11399985523686
(5,0)	25	16800	0	0	0	0.0843	0	11368648473	1605503429328
(6,0)	36	24192	0	0	0	0.0703	0	28289407236	4799169761892
(7,0)	49	32928	0	0	0	0.0603	0	61145300079	12111032215788

Table 3.4 Goldberg-Coxeter indices (k,l), triangulation parameter $t(k,l)$, number of vertices n_v, number of symmetry-distinct (N_S^{SD}) and total (N_S) face-spirals, pentagon neighboring index N_P, hexagon strain parameter σ_h, HOMO–LUMO gap $\Delta\epsilon_{\mathrm{HL}}$ from a Hückel analysis (in eV), Wiener index W, and Szeged index S_z of the Halma and leapfrog transforms of C$_{384}$, C$_{440}$, and C$_{672}$.

Fullerene	DFT	r_{min}	r_{max}	r_{MCS}	ΔE_N	V	A	C_A	q_{IPQ}	D_{MDS}	λ_{asym}
I_h-C_{60}	PBE	1.407	1.457	3.565	0	164.8	150.3	1.0000	0.904	0.0	0.0
	B3LYP	1.398	1.455	3.555	0	163.3	149.4	1.0000	0.904	0.0	0.0
T-C_{364}	PBE	1.398	1.457	11.30	-6.002	2432.2	959.3	1.0157	0.758	15.05	7.449
	B3LYP	1.388	1.454	11.25	-6.351	2408.2	952.7	1.0157	0.758	15.02	7.429
D_3-C_{366}	PBE	1.400	1.463	13.96	-5.155	2119.4	964.5	1.0434	0.566	40.74	26.65
	B3LYP	1.393	1.462	13.90	-5.445	2100.0	958.0	1.0436	0.567	40.52	26.57
T-C_{380}	PBE	1.411	1.462	12.82	-5.430	2487.4	1001.5	1.0260	0.697	17.60	10.84
	B3LYP	1.404	1.461	12.77	-5.712	2462.8	994.7	1.0205	0.697	17.52	10.82
D_3-C_{384}	PBE	1.396	1.464	15.73	-4.724	2183.1	1012.1	1.0635	0.520	44.95	33.62
	B3LYP	1.386	1.462	15.68	-4.978	2162.9	1005.3	1.0632	0.521	44.62	33.53
D_3-C_{440}	PBE	1.407	1.493	16.98	-5.124	2661.8	1161.2	1.0706	0.512	47.72	39.74
	B3LYP	1.387	1.465	17.01	-5.399	2637.7	1153.2	1.0672	0.513	47.40	39.63
D_3-C_{672}	PBE	1.399	1.465	18.89	-6.814	5307.2	1779.0	1.0298	0.566	42.03	49.38
	B3LYP	1.390	1.465	18.82	-7.249	5260.0	1766.9	1.0301	0.567	41.79	49.18

Table 3.5 Properties obtained from DFT optimizations. Smallest (r_{min}) and largest bond distance (r_{max}) in Å, radius of minimum covering sphere r_{MCS} in Å, energetic stability ΔE_N per carbon atom compared to C_{60} in kcal/mol as defined in eq. 3.6, volume V in Å^3 and surface area A in Å^2, convexity parameter C_A, isoperimetric quotient q_{IPQ}, spherical distortion parameter from minimum distance sphere D_{MDS} in %, and Fowler asymmetry parameter λ_{asym}.

and $C_{N\geq 60}$ for the most stable fullerenes, the latter fulfilling the isolated pentagon rule (IPR). This results in two stability lines. Second, for the most stable IPR fullerenes, one obtains the graphene limit at -13 kcal/mol, in reasonable agreement with the estimated value of -9 kcal/mol obtained from heat of formation of graphite and considering the van der Waals interaction between graphene layers.[292,293] Third, the results show that the NS-fullerenes are reasonably stable with a destabilization of less than 2 kcal/mol per C–C bond compared to the most stable IPR fullerenes. The PBE and B3LYP results are very similar.

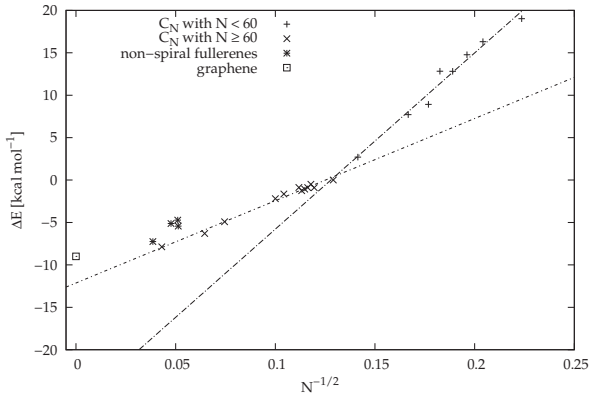


Figure 3.8 Stability of NS-Fullerenes compared to C_{60}, i.e., $E(C_N)/N - E(C_{60})/60$ (in kcal/mol), with E being the total electronic energy. For the two lines the most stable fullerenes have been chosen, i.e., IPR fullerenes for $N > 60$. The stabilities of the IPR fullerenes are extrapolated linearly to the graphene limit.

Table 3.5 contains many other useful properties. We see that T-C_{364} and its Yoshida-Fowler vertex inserted T-C_{380} have very similar volumes and surface areas (trivially the surface area increases from T-C_{364} to T-C_{380}), and so do D_3-C_{366} and D_3-C_{384}. However, a fullerene with a larger number of vertices does not have to be larger in size due to their completely different geometry, as shown by a comparison, for example, between T-C_{364} and D_3-C_{384}. Except for C_{60}, all fullerenes listed in Table 3.5 are slightly non-convex, as shown by the (surface) convexity parameter $C_A = A_{CH}/A$. Here, A is the surface area of the fullerene, and A_{CH} the surface area of its convex hull.

The non-convexity of the fullerenes is clearly visible in Figures 3.1 to 3.5, where we see a slight concave curvature around the hexagons. The last three columns in table 3.5 measure the deviance for each fullerene from the ideal 3D sphere (for a detailed discussion see Ref.[36]).

3.4 The Halma and Leapfrog Transforms of NS-Fullerenes

The Goldberg-Coxeter transformation[68,94] GC*kl* transforms a fullerene C_N to a larger fullerene $C_{N(k^2+kl+l^2)}$ ($t(k,l) = (k^2 + kl + l^2)$ is called the triangulation number) with the same shape and symmetry.[32] Two particularly simple GC-transforms are the *halma* and *leapfrog* transformations, corresponding to $l = 0$ and $k = l$, respectively. We call the $l = 0$ case, which is shown in Figure 3.9 (a), the "Halma" transform due to the construction's similarity to a Halma game board. It is also sometimes called a k-inflation. The leapfrog transformation is illustrated in Figure 3.9 (b).

(a) (b)

Figure 3.9 Two examples for the Goldberg Coxeter transformation: (a) the third Halma transformation GC30 and (b) the third leap frog transformation GC33

We briefly mention some important properties of GC transformations. The GC-transform can be described in terms of multiplication by complex numbers $(k + l\omega)$, where $\omega = e^{\frac{i2\pi}{6}}$, in what is called the *Eisenstein plane*.

Because of this, any number of GC transformations can be composed into a single GC transformation, corresponding to the product of the Eisenstein

numbers, as follows:

$$\mathtt{GC}k_1l_1\mathtt{GC}k_2l_2 = \mathtt{GC}k_1k_2 - l_1l_2l_1k_2 + (k_1 + l_1)l_2 \tag{3.7}$$

$\mathtt{GC}10 = 1$ is the unit transformation. GC transformations commute with each other, i.e.,

$$\mathtt{GC}k_1l_1\mathtt{GC}k_2l_2 = \mathtt{GC}k_2l_2\mathtt{GC}k_1l_1. \tag{3.8}$$

The GC-transform is 12-fold symmetric through the transformations

$$\begin{aligned} &k + l\omega \mapsto (k + l\omega)\omega^j \text{ for } j = 0, \ldots, 5 \\ &k + l\omega \mapsto l + k\omega. \end{aligned} \tag{3.9}$$

From Equation 3.7, we can derive the combined leapfrog and Halma transformations,

$$\mathtt{GC}11\mathtt{GC}11 = (\mathtt{GC}11)^2 = \mathtt{GC}03 = \mathtt{GC}30 \tag{3.10}$$

$$\mathtt{GC}11\mathtt{GC}k0 = \mathtt{GC}kk \tag{3.11}$$

$$\mathtt{GC}k0\mathtt{GC}l0 = \mathtt{GC}kl0 \tag{3.12}$$

Hence we obtain all possible GC transformations of the form $\mathtt{GC}k0$ and $\mathtt{GC}kk$ from such combinations as implemented in *Fullerene*. The two largest Halma transforms of C_{380} and C_{384} studied here are shown in Figures 3.10 and 3.11.

Tables 3.3 and 3.4 list properties of the GC transforms of the four NS fullerenes studied. None of the Halma transforms admit any ring spiral. This leads us to the following conjecture:

Conjecture 1. *Any Halma transform of an NS-fullerene is also an NS-fullerene.*

If this conjecture is true, the class of NS-fullerenes is infinite, even though it is vanishingly small compared to the class of all fullerenes for the range where exhaustive search has been performed, and we expect it to be so for larger N as well.

In Tables 3.3 and 3.4, we see that many of the leapfrogged NS-fullerenes are NS-fullerenes themselves, but that in general they are not. Brinkmann and Fowler have investigated spiralability of the leapfrogs of several million cubic polyhedra and found that these leapfrogs statistically are less likely to have spirals than their parents.[294]

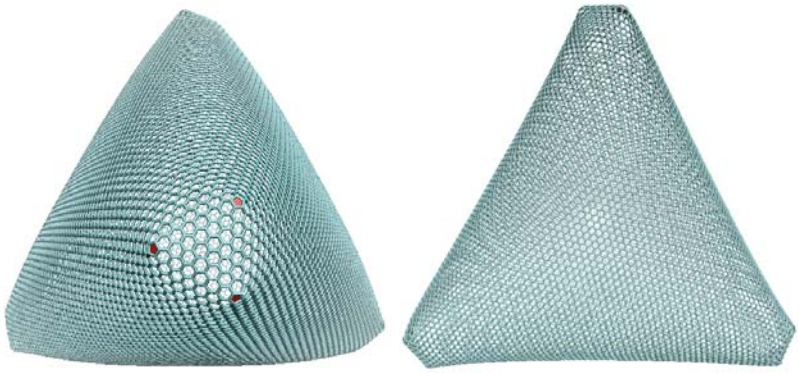


Figure 3.10 The force-field optimized structure of GC70[C_{380}] = C_{18620}. View from the top (left) and side (right) of the polyhedron.

We see from Table 3.4 that a leapfrog of a spiralable fullerene can be either spiralable or non-spiralable: For example, GC11[C_{440}] admits a spiral, but GC11GC11[C_{440}] = GC30[C_{440}] does not. Similarly, the Halma transform of a spiralable fullerene can be either: NS-D_3-C_{672} provides a counterexample to the spiralable fullerenes being closed under the Halma transform, because it can be written as the Halma transform GC20[D_3-C_{168}].

The face spiral pentagon indices for D_3-C_{168} are given in Table 3.1. As expected, the spiral count for the D_3-C_{168} fullerene is low, i.e., we get 11 distinct face spirals out of a total of only 66. NS-D_3-C_{672} is likely to be the smallest Goldberg-Coxeter NS-fullerene. We mention that the D_3-C_{924} = GC11[D_3-C_{308}] may well be the first leapfrog NS-fullerene.[282,294]

For completion and future reference, we also list topological indicators (Wiener and Szeged index) for C_{364}, C_{366}, C_{380}, C_{384}, C_{440}, and C_{672} and their smallest Goldberg-Coxeter transforms for $k = l$ and $l = 0$ or $l = k$ in Tables 3.2, 3.3, and 3.4.

3.5 Conclusions

We investigated the four NS-fullerenes NS-T-C_{380}, NS-D_3-C_{384}, NS-D_3-C_{440}, and NS-D_3-C_{672}, as well as their leapfrog and Halma transforms up to about

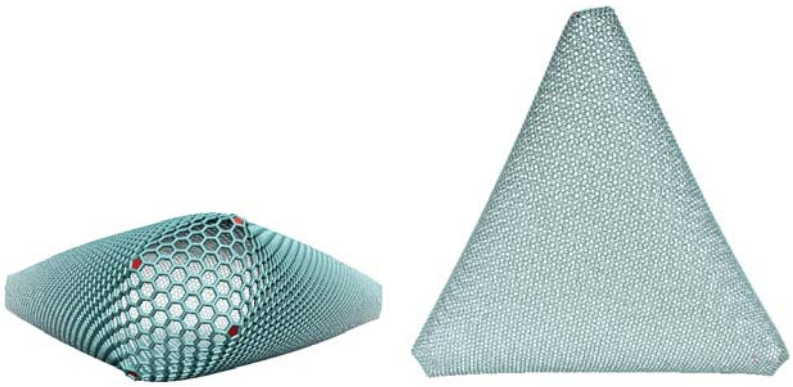


Figure 3.11 The force-field optimized structure of GC70[C_{384}] = C_{18816}. View from the top (left) and side (right) of the polyhedron.

30,000 atoms. The NS-fullerenes were optimized and analyzed quantum chemically at the DFT level of theory, while the larger transformed fullerenes were force-field optimized and analyzed using graph theoretical methods.

NS-*T*-C_{380} has a large HOMO–LUMO gap of 0.434 eV, while C_{440}, C_{384} and C_{672} have small or zero gap, which places them among the semiconductors and conductors, respectively. The predicted Jahn-Teller distortion undergone by the zero-gap fullerenes was found to be so minor as to be undetectable by visual inspection of the structure.

NS-fullerenes typically have clusters of pentagons in close proximity, leading to sharp curvature and reduced stability compared to more spherical fullerenes of equal size, but the investigated NS-fullerenes remain reasonably stable with less than 2 kcal/mol difference per bond to the icosahedral isomers. All four quantum chemically optimized geometries are found to be slightly concave, a feature that is predicted by the force-field optimization, showing that the concavity arises directly from the graph topology rather than the electronic structure.

It was found that the leapfrog transformation may introduce spiralability, while all the Halma transformed fullerenes were unspiralable. We conjecture that the class of NS-fullerenes is closed under the Halma transform. However, the converse is not true: a counterexample is NS-C_{672}, which is the second

Halma transform of C_{168}.

While the vanishingly small proportion of isomers with no spirals makes it difficult to find examples of NS-fullerenes through extensive search, especially for large values of N, new NS-fullerenes can be obtained from existing ones through Halma transformations if our proposed conjecture holds. Furthermore, we conclude that none of the investigated leapfrogs of NS-T-C_{380} are spiralable.

3.6 The generalised face-spiral algorithm

As pointed out in subsection 3.3.2, most fullerene graphs can be encoded with the original face-spiral algorithm by Manolopoulos et al., while in order to encode any polyhedral graph and in particular any fullerene graph, a generalised face-spiral algorithm is required. The algorithm should have the property, that a graph which can be represented by the original spiral algorithm gets encoded by the same expression, that is, generalised representations should be a superset of the original ones. Such generalised face-spiral algorithms have been suggested by Brinkmann[290] and Fowler et al..[289] The problem of getting stuck while creating a spiral from a graph is avoided by defining how to proceed if all adjacent faces have been placed already (Brinkmann) or by avoiding to get stuck by potentially skipping faces (Fowler).

In the program *Fullerene* the algorithm by Fowler et al. has been implemented.

Two functions are required to obtain the spiral string from a polyhedral graph and to construct the graph from a spiral string. Both parts of the algorithm are defined to operate on triangulations and hence consider vertex-spirals (in the triangulation) instead of face-spirals (in the cubic graph); taking the dual is implemented elsewhere and is removed from this algorithm.

The generation of a generalised vertex spiral from a triangulation is declared as `bool Triangulation::get_spiral(const node_t f1, const node_t f2, const node_t f3, vector<int> &spiral, jumplist_t& jumps, bool general) const`, that is, as a member function of the triangulation class, taking the first three vertices of the spiral as arguments. The spiral,

i.e., a vector of face sizes and the required jumps (as a vector of pairs of integers) are retunded through references. `general` is a switch for toggling the general form of the algorithm—switching it off increases the speed of the computation. The return value indicates whether a spiral was found.

The structure of this function is outlined by the following pseudo code:

```
1  bool get_spiral(vertices v1, v2, v3, spiral s, jumplist jl
    %){
2    initial graph ig
3    graph g
4    spiral string s
5    boundary b
6    boundary valencies bv
7    jump state j
8    list of jumps jl
9
10   add first vertex to s
11   remove first vertex from g
12   add first vertex to bv
13
14   add second vertex to s
15   remove second vertex from g
16   update bv
17   add second vertex to bv
18
19   add third vertex to s
20   remove third vertex from g
21   update bv
22   add third vertex to bv
23
24   for (vertices k=3 to N-1){
25     find a vertex k in ig that is connected to first and
        %last vertex in b
26
27     remove k from g
28
29     // jump required?
30     if(g is disconnected)
31       revert removal of k from g
```

```
32        perform cyclic shift by 1 on bv
33        increment j
34        go to beginning of loop
35
36      // end of cyclic shift?
37      if(j != 0 and g is connected)
38        append (k, j) to jl
39        set j to 0
40
41      connect k and first vertex in b
42      while(first element in bv is 0)
43        remove first element from bv
44        connect k and first vertex in b
45
46      connect k and last vertex in b
47      while(last element in bv is 0)
48        remove last element from bv
49        connect k and last vertex in b
50
51      append k to spiral
52      append open valencies of k to bv
53    }// leave loop with one vertex remaining to be added
54
55    if(length of b != valency of the last vertex or
56       not every element in bv is 1)
57      return false
58
59    append the last vertex to spiral
60    return true;
61  }
```

Here, `connect` is a short function that updates the valencies of the two vertices. Removing a vertex from the working copy of the graph, removes not only the vertex itself but all its edges and keeps a backup of the prior state in case the operation has to be reverted. Connecting and removing vertices is implemented as inline void functions, but could elegantly be written as lambda expressions.

The reverse operation is defined as a constructor of the triangulation class. As arguments it requires a spiral string and a list of jumps which may

be empty: `Triangulation(const vector<int>& spiral_string, const jumplist_t& jumps = jumplist_t());`. The spiral string is a vector of N face sizes; any size from 3 upwards is permitted which makes this constructor applicable to any triangulation of the sphere, including ones with negative Gaussian curvature. The jump list is a vector of pairs of two integers, one being the position and the other the length of the respective jump.

This algorithm works as follows:

```
1  Triangulation(spiral string, jumplist){
2    connect the first two vertices
3
4    for (vertex k=3 to N-1){
5      if(a jump is required for k)
6        lookup length l of jump in jumplist
7        perform cyclic shift by l on open valencies
8
9      connect k to the last vertex on boundary
10     connect k to the first vertex on boundary
11
12     while (valencies of the last vertex on the boundary
        %are 0)
13       remove vertices with 0 valency from boundary
14       connect k to the last vertex on boundary
15
16     while (valencies of the first vertex on the boundary
        %are 0)
17       remove vertices with 0 valency from boundary
18       connect k to the first vertex on boundary
19
20     add k to open valencies
21   }
22
23   add edges between the last vertex and each element of
      %the boundary
24
25   construct triangulation from edge set
26 }
```

Here, N is the length of the input spiral, `open valencies` is a vector of integers that holds the valencies of all vertices on the boundary, and `connect` implies that an edge between the two vertices is added to the edge set and the open valencies of both vertices are decremented.

4 Force fields for fullerenes, general polyhedral graphs and triangulations of the sphere[o]

4.1 Introduction

Fullerenes are hollow polyhedral carbon structures consisting of 12 pentagons and F_6 hexagons with $F_6 \geq 0$ and $F_6 \neq 1$.[6,9,32,41] Schleyer and co-workers pointed out that both C_{20}-I_h and C_{60}-I_h are not spherically π aromatic[204] but spherically π anti-aromatic,[203] which for C_{60} explains the large heat of formation (for a more detailed discussion on of fullerene aromaticity see ref.[296]). They concluded that against common belief, fullerenes are not highly stable molecules,[203] showing no "magic" stability for C_{60} compared to other fullerenes (and graphene).

As the pentagons can be distributed in many different ways on a 2D surface (of genus zero),[32] the number of isomers grows rapidly with increasing number of carbon atoms N, i.e., Thurston established a $\mathcal{O}\left(N^9\right)$ polynomial growth for the number of isomers with vertex number N.[20] These isomers come in many different shapes and symmetries,[1] and it is clear that a fast algorithm is required to determine accurate molecular shapes, which then can be refined by a more rigorous quantum theoretical

[o]Sections 4.1 to 4.5 of this chapter have been published as "From Small Fullerenes to the Graphene Limit: A Harmonic Force-Field Method for Fullerenes and a Comparison to Density Functional Calculations for Goldberg-Coxeter Fullerenes up to C_{980}" by Wirz et al.[295] and are reproduced with kind permission from the authors and John Wiley & Sons, Ltd.. The candidate's contribution to the article is itemised on pages 299 ff. Section 4.6 describes a force field for general polyhedral graphs and triangulations of the sphere which is a part of libgraph (see sec. 2.14).

treatment. Such an algorithm should give the experimentaly or computationally predicted bond distances and angles for the 12 pentagons and F_6 hexagons, and the 3D structure should have the right symmetry, shape and Gaussian curvature (which is predominantly positive but may be negative locally in fullerenes).[1]

The fullerene graph $G = (\mathcal{V}, \mathcal{E})$ with vertices $v_i \in V(i = 1, \ldots, N)$ and edges $e_j \in E(j = 1, \ldots, 3N/2)$ contains all the information required to construct a structural carbon framework. A compact short-hand notation for fullerene graphs is given by the Fowler-Manolopoulos face spiral algorithm, which, starting from a selected face, goes through all faces of the fullerene exactly once, i.e., it is identical to a Hamilton spiral path for the dual of the fullerene. Only the locations of the pentagons need to be listed, and if the spiral gets stuck in a cul-de-sac, one can define jumps to remedy the situation. This leads to a generalized spiral algorithm which uniquely defines the fullerene graph. Conversely, from the general face spiral one obtains the graph G (and its vertex adjacency matrix A_{kl}), which can be used to embed the graph in an appropriate 2D surface to obtain an initial 3D structure. Such 2D embedding algorithms are already available, e.g., the Adjacency Matrix Eigenvector (AME) algorithm introduced by Fowler and Manolopoulos,[32,101] or the more recent combination of the Tutte embedding of the graph into a 2D plane with subsequent projection onto the surface of a sphere, which was developed in our group.[1,56]

The next step is to refine the initial crude 3D geometry by a force-field optimization, which should lead to a structure close to the real (experimental) one. It is common for molecular force-fields to distinguish between single and double bonds. This strategy, however, is not practical for fullerenes due to exponential growth of different Kekulé structures with increasing vertex number N. Force fields for fullerenes should therefore be designed avoiding the distinction between single and double bonds. The first force field tailored specifically to fullerenes was developed by Wu et al.[237] It was designed for C_{60}-I_h only using harmonic force field terms between two bond types: bonds adjacent to two hexagons with a bond length of 1.54 Å, and bonds adjacent to a pentagon and a hexagon with a bond length of 1.41 Å. Angles are either part of a pentagon ($\theta_0 = \frac{3}{5}\pi$) or a hexagon ($\theta_0 = \frac{2}{3}\pi$).

A number of extensions and modifications to the Wu force field have been published since.[239,297–301] Except for the force field by Ceulemans et al.,[301]

all others have been exclusively designed for C_{60}-I_h. Here, we introduce a harmonic force-field applicable to all fullerenes as implemented into our program package *Fullerene*, and test the obtained structures for larger fullerenes (obtained from Goldberg-Coxeter transforms of C_{20}) against more accurate density functional (DFT) calculations. As an illustrative example, we use the DFT electronic energies to extrapolate to the graphene limit, and compare our results to periodic DFT calculations for graphene.[302,303] The force-field vibrational frequencies are also used to extrapolate to the zero-point energy limit of graphene.

4.2 Computational Methods

The initial fullerene structures were generated by the program *Fullerene*[P] through a face-spiral algorithm,[36] a Tutte embedding of the graph and subsequent mapping of the planar layout on a sphere followed by a geometry optimization using the general fullerene force-field as introduced in the next chapter. We chose the most stable fullerene structures for each vertex number up to C_{60} (see ref.[304] for details), plus some selected fullerenes up to C_{100}. As candidates for larger fullerenes beyond 100 vertices we chose the Goldberg-Coxeter transforms[32,68,69] of C_{20}, i.e., $GC_{k,l}[C_{20}\text{-}I_h]=C_N$ with $N = 20(k^2 + kl + l^2) \leq 980$. These icosahedral structures were taken as input for refined optimization using density functional theory (DFT). In a first step, the structures were optimized with the Perdew-Burke-Ernzerhof (PBE) functional[152] using the def2-TZVPP basis set for the fullerenes from C_{20} up to C_{60}, and the smaller def2-SVP basis set[153] for the larger fullerenes (from C_{60} onwards) within the Gaussian09 code.[288] Tight geometry convergence criteria and ultra-fine integration grids were applied and the symmetry of the structures was exploited. Starting from these structures, subsequent optimizations for the larger fullerenes were carried out with the Becke-Lee-Yang-Parr hybrid functional (B_3LYP)[284,285] functional as implemented in Gaussian09 and Turbomole (for the optimized structures see supplementary materials).[288,305] This was done to obtain a second value for the graphene limit. We only chose these two commonly applied functionals as some of the

[P]Program *Fullerene* is an open-source code and is freely downloadable at `http://ctcp.massey.ac.nz/index.php?page=fullerenes`

calculations for the larger fullerenes became very computer time intensive. All structures were treated as closed-shell singlet molecules, except for some of the smaller fullerenes as detailed in ref.[304].

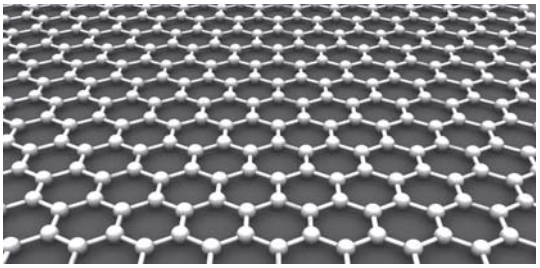


Figure 4.1 The Structure of Graphene

The energy difference between C_{60} and graphene was calculated using the Vienna Ab-Initio Simulation Package (*VASP*),[306] with a plane wave basis set (energy cutoff $E_c = 600\,\text{eV}$) and a 1*s* frozen core PAW dataset.[307,308] Graphene sheets as shown in Figure 4.1, in the graphite arrangement, were kept 10 Å apart, and C_{60} was calculated in a cubic box of size 20 Å. All internal coordinates (and thus, the lattice constants of graphene) were optimized until residual forces were below $1\,\text{meV/Å}$. Here, we chose a variety of functionals as implemented in *VASP*,[306] namely the local density approximation (LDA), the PBE functional as already mentioned,[152] the PBE functional corrected using Grimme's dispersion correction (PBE-D2),[309,310] and the optB88-vdW functional of Langreth and Lundqvist.[311–313]

4.3 A General Harmonic Force Field for Fullerenes

The original program *Fullerene*[36] included a harmonic force field of the form

$$E_{\mathrm{HFF}} = \frac{k_\mathrm{p}}{2} \sum_{i_\mathrm{p}}^{\mathrm{p-edges}} \left(R_{i_\mathrm{p}} - R_\mathrm{p}\right)^2 + \frac{k_\mathrm{h}}{2} \sum_{i_\mathrm{h}}^{\mathrm{h-edges}} \left(R_{i_\mathrm{h}} - R_\mathrm{h}\right)^2 + \frac{f_\mathrm{p}}{2} \sum_{j_\mathrm{p}}^{60} \left(\alpha_{j_\mathrm{p}} - \alpha_\mathrm{p}\right)^2 + \frac{f_\mathrm{h}}{2} \sum_{j_\mathrm{h}}^{3N-60} \left(\alpha_{j_\mathrm{h}} - \alpha_\mathrm{h}\right)^2, \quad (4.1)$$

In this simple force field, introduced by Wu et al. for C_{60}-I_h,[237] bonded pairs of vertices are taken into account. k_p and k_h are the force constants for the two different C–C bonds (set to $\approx 300\,\mathrm{kcal\,Å^{-2}}$), R_p and R_h the corresponding pentagon and hexagon bond distances ($\approx 1.4\,$Å), f_p and f_h are the force constants for the two different bending modes in a pentagon and hexagon with the corresponding bond angles α_p and α_h of 108° and 120°, respectively. As the original Wu force field was developed for C_{60}-I_h only, bonds adjacent to two pentagons were treated in the same way as bonds adjacent to one pentagon and one hexagon in the program *Fullerene*.[36] The geometry can then be optimized using for example a Fletcher-Reeves-Polak-Ribiere geometry optimization with analytical gradients,[265] which is very computer time efficient even for larger fullerenes containing thousands of carbon atoms.[36,96]

The Wu force field is in principle applicable to all fullerenes and usually yields structures that are in good agreement with optimized structures from a more rigorous quantum theoretical treatment such as DFT. The root mean square (rms) deviations between DFT and force-field optimized bond lengths are usually smaller than 0.05 Å.[36] However, the Wu force field suffers from two problems. First, the optimization may converge to local minima, especially if the initial structure is far from the real one. This may result in distortions away from convexity ("dents") or, in the worst case, partially inverted structures with self-intersecting planes. Second, as no dihedral angles are restrained in this force field, the minimum geometry does not

correctly reproduce the convexity or planarity of calculated geometries at DFT level of theory.

The former problem can be alleviated by adding a Coulomb repulsive potential in the initial phase of the geometry optimization,[36]

$$E = E_{\mathrm{HFF}} + E_{\mathrm{Coulomb}} = E_{\mathrm{HFF}} + \sum_{i=1}^{N} \frac{f_{\mathrm{Coulomb}}}{|\boldsymbol{r}_i - \boldsymbol{r}_0|} \tag{4.2}$$

with $\boldsymbol{r}_i - \boldsymbol{r}_0$ being the distance between vertex i and the barycenter at $\boldsymbol{r}_0$. f_{Coulomb} can be chosen as large as necessary to keep the fullerene cage in the right shape. Another more robust as well as more challenging solution to this issue is, to come up with an initial structure that is closer to the global minimum, e.g., by directly deriving the shape of the ideal embedding from the graph.

As a solution to the second problem, we add a dihedral angle term to the Wu force field (the result is denoted as extended harmonic force field in the following, EHFF), which enhances planarity in areas of connected hexagons and enforces convexity in areas of positive Gaussian curvature. In addition to improving the global minimum structure, the dihedral angle term makes the force field more robust with respect to converging to the global minimum.

The extended Wu force field takes three types of bonds (in the ring fusion of 0, 1, or 2 pentagons), two types of angles, and four types of dihedral angles into account. There is one dihedral per atom that is uniquely defined with respect to the three adjacent faces (Fig. 4.2). Dihedrals θ_{abcd} are defined between one atom a and its three neighbors b, c, and d. Note that the connection a–b represents an edge in the graph, while b–c and c–d do not. As one atom is part of three faces (0, 1, 2, or 3 pentagons) there are four different types of dihedrals which differ in their respective zero value and force constant as shown in Fig. 4.2. In the case of mixed adjacent faces, i.e., vertex a being adjacent to one or two pentagons, vertex b is chosen to lie between the two equal faces, a choice that is unique and symmetry preserving. Vertices c and d are placed counter clockwise which leads to positive dihedral angles at convex vertices. This definition requires the least possible number of angle evaluations per vertex compared to for example the usual definition of a torsional motion.

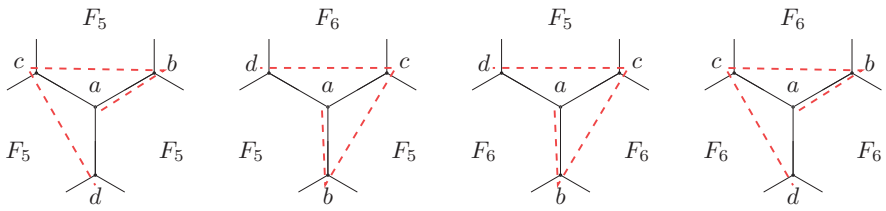


Figure 4.2 Unique and symmetry preserving layout of the four types of dihedral angles θ_{abcd}, depending on the adjacent faces.

The total energy for the EHFF force field is given by

$$\begin{aligned} E_{\mathrm{EHFF}} = {} & \frac{f_{\mathrm{pp}}}{2} \sum_{i_{\mathrm{pp}}}^{\mathrm{pp-e}} \left(R_{i_{\mathrm{pp}}} - R_{\mathrm{pp}} \right)^2 + \frac{f_{\mathrm{hp}}}{2} \sum_{i_{\mathrm{hp}}}^{\mathrm{hp-e}} \left(R_{i_{\mathrm{hp}}} - R_{\mathrm{hp}} \right)^2 \\ & + \frac{f_{\mathrm{hh}}}{2} \sum_{i_{\mathrm{hh}}}^{\mathrm{hh-e}} \left(R_{i_{\mathrm{hh}}} - R_{\mathrm{hh}} \right)^2 + \frac{f_{\mathrm{p}}}{2} \sum_{j_{\mathrm{p}}}^{60} \left(\alpha_{j_{\mathrm{p}}} - \alpha_{\mathrm{p}} \right)^2 \\ & + \frac{f_{\mathrm{h}}}{2} \sum_{j_{\mathrm{h}}}^{3N-60} \left(\alpha_{j_{\mathrm{h}}} - \alpha_{\mathrm{h}} \right)^2 + \frac{f_{\mathrm{ppp}}}{2} \sum_{k_{\mathrm{ppp}}}^{\mathrm{ppp-v}} \left(\theta_{k_{\mathrm{ppp}}} - \theta_{\mathrm{ppp}} \right)^2 \\ & + \frac{f_{\mathrm{hpp}}}{2} \sum_{k_{\mathrm{hpp}}}^{\mathrm{hpp-v}} \left(\theta_{k_{\mathrm{hpp}}} - \theta_{\mathrm{hpp}} \right)^2 + \frac{f_{\mathrm{hhp}}}{2} \sum_{k_{\mathrm{hhp}}}^{\mathrm{hhp-v}} \left(\theta_{k_{\mathrm{hhp}}} - \theta_{\mathrm{hhp}} \right)^2 \\ & + \frac{f_{\mathrm{hhh}}}{2} \sum_{k_{\mathrm{hhh}}}^{\mathrm{hhh-v}} \left(\theta_{k_{\mathrm{hhh}}} - \theta_{\mathrm{hhh}} \right)^2 \end{aligned} \tag{4.3}$$

where pp-e (hp-e, hh-e) denotes the number of edges adjacent to 2 (1, 0) pentagons, and ppp-v (hpp-v, hhp-v, hhh-v) is the number of vertices adjacent to 3 (2, 1, 0) pentagons. If required, the Coulomb repulsive potential in eq. 4.2 can be added to this force field for the initial geometry optimization.

The force-field parameters are set as follows: $R_{\mathrm{hp}} = 1.458\,\text{Å}$ and $R_{\mathrm{hh}} = 1.401\,\text{Å}$ are taken from an electron gas diffraction study of C_{60},[314] which are close to the distances obtained from X-ray diffraction measurements of solid C_{60} ($R_{\mathrm{hp}} = 1.455\,\text{Å}$ and $R_{\mathrm{hh}} = 1.391\,\text{Å}$).[113] $R_{pp} = 1.479\,\text{Å}$ has been estimated from PBE calculations on C_{50} and scaled with the experimental R_{hp} and R_{hh} for C_{60}. The angles in pentagons and hexagons have

been set to $\alpha_p = 108^\circ$ and $\alpha_h = 120^\circ$ under the assumption that faces are planar. The dihedral angles $\theta_{ppp} = 37.38^\circ$, $\theta_{hpp} = 29.20^\circ$, $\theta_{hhp} = 23.49^\circ$ and $\theta_{hhh} = 0.0^\circ$ have been calculated assuming planar faces and the above given bond lengths. It is clear that for many fullerenes the optimized structural parameters can deviate substantially from the ideal ones given in this force-field, for example, two neighboring hexagons in carbon nanotubes are not in one plane. Nevertheless, the structures we obtain are rather accurate as we shall see.

From the least squares fit to B3LYP frequencies of the vibrational spectra of C_{50}-C_1(193), C_{60}-I_h and C_{70}-D_{5h}(8149) (numbering of the isomers stems from lexicographically ordered face-spiral pentagon indices, see ref.[32]), we obtain the following force constants for the force field shown in eq. 4.3 (in N/m or Nm/deg^2): $f_{pp} = 260.0$, $f_{hp} = 390.0$, $f_{hh} = 450.0$, $f_p = 100.0$, $f_h = 100.0$, $f_{ppp} = 35.0$, $f_{pph} = 65.0$, $f_{phh} = 85.0$, $f_{hhh} = 270.0$. The stretching and bending force constants are not too far away from the original force field parameters published by Wu et al.[237] or Ceulemans et al.[300] Within the extended force field, frequencies are obtained through analytical second derivatives. The root mean square (rms) deviation of the fit $\Delta\omega_{rms} = 45.5\,\mathrm{cm}^{-1}$ is relatively small considering the fact that such a simple force-field does not correctly describe the coupling between the stretching, bending and torsional modes in terms of the correct normal modes.

4.4 Results and Discussion

The results of our EHFF and PBE calculations are shown in Tables 4.1 and 4.2. Here, we only give the ideal point group of the graph, which is reproduced by the force field, even though a DFT geometry optimization may reduce the symmetry due to Jahn-Teller distortions (see for example the discussion in refs.[1, 304]). The rms deviations listed for the bond lengths, bond angles and torsion angles show that the extended harmonic force field performs well, especially if we consider that the bond lengths have been fixed to experimental values (note that the rms error for the bond distances can be considerably reduced if we use distances closer to the PBE optimized structure). The rms error is therefore predominantly caused by the fact that the DFT optimized structures have larger bond lengths compared to

experiment. The smallest deviation is observed for C_{60} as we expect from the near ideal spherical geometry of this molecule. Moreover, if we compare the zero-point vibrational energy, which we obtained at the B3LYP level of theory for C_{50}-C_1(193) with 0.1667 eV/atom (this isomer is not listed in Table 4.1), C_{60}-I_h(1812) with 0.1705 eV/atom, and C_{70}-D_{5h}(8149) with 0.1711 eV/atom, we see that these are in excellent agreement with the values obtained from the general force-field with 0.1681 eV/atom, 0.1685 eV/atom, and 0.1726 eV/atom, respectively. The force-field structures of the two largest isolated pentagon rule (IPR) fullerenes C_{980}-I_h and C_{980}-I are shown in Figure 4.3 in comparison with the PBE optimized structures. It is not easy to spot differences between the two structures with the naked eye, except perhaps for the curvature around the pentagons. We conclude that our force-field optimized structures, which are computationally very efficient, are ideal to be used as starting geometries for a more elaborate quantum theoretical treatment.

Tables 4.1 and 4.2 also show the stability of the fullerenes with respect to C_{60} according to the isodesmic reaction $1/60\ E(C_{60}) \rightarrow 1/N\ E(C_N)$ (relative fullerene stability, RFS),

$$\Delta E_{\mathrm{RFS}}(N) = 1/60\ E(C_{60}) - 1/N\ E(C_N) \tag{4.4}$$

where $E(C_N)$ is the total electronic energy obtained from the DFT calculation. Figure 4.4 compares the different RFS energies. It is clear, that for $N \rightarrow \infty$ we approach the graphene limit as the size of planar hexagon sheets increase and the pentagons become less relevant in this limit. An extrapolation to $N \rightarrow \infty$ using all fullerenes shown in Tables 4.1 and 4.2 and a N^{-1} law for $\Delta E_{\mathrm{RFS}}(N)$ (NB: the N^{-1} law comes from the curvature term or strain energy in fullerenes, for details see refs. [34, 52, 192, 302, 315]),[1] gives -8.06 kcal/mol at the PBE level of theory (and -8.42 kcal/mol at the B3LYP level of theory). However, the $\Delta E_{\mathrm{RFS}}(N)$ behavior is not strictly linear with N^{-1}. A linear extrapolation for the $\Delta E_{\mathrm{RFS}}(N)$ values using larger fullerenes with $N \geq 500$ gives $\Delta E_{\mathrm{RFS}} = -8.64$ kcal/mol at the PBE level of theory (-9.22 kcal/mol at the B3LYP level of theory) in much better agreement with the periodic boundary calculations listed in Table 4.3. More importantly, Figure 4.4 supports Schleyer's hypothesis that fullerenes are not highly stable molecules,[203] there is no "magic" stability of C_{60} compared to all the other fullerenes and especially graphene.

N	Isomer	IPG	$R_{\min}^{\mathrm{PBE}}$	$R_{\min}^{\mathrm{EHFF}}$	$R_{\max}^{\mathrm{PBE}}$	$R_{\max}^{\mathrm{EHFF}}$	ΔR_{rms}	$\Delta\alpha_{\mathrm{rms}}$	$\Delta\theta_{\mathrm{rms}}$	A^{PBE}	A^{EHFF}	$\Delta E_{\mathrm{RFS}}^{\mathrm{PBE}}$	$\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$
20	1	I_h	1.409	1.479	1.514	1.479	0.041	1.63	1.88	43.64	45.16	18.148	0.13977
24	1	D_{6d}	1.373	1.460	1.530	1.481	0.054	1.28	1.91	54.66	55.82	15.710	0.14736
26	1	D_{3h}	1.384	1.458	1.540	1.484	0.047	1.69	1.95	59.88	61.24	14.024	0.15020
28	2	T_d	1.430	1.458	1.510	1.482	0.026	1.05	1.57	65.01	66.69	11.735	0.15261
30	3	C_{2v}	1.368	1.406	1.524	1.485	0.034	1.86	3.04	70.16	71.77	10.350	0.15472
32	6	D_3	1.384	1.406	1.503	1.482	0.036	1.48	2.69	75.37	77.08	8.369	0.15654
34	5	C_2	1.389	1.403	1.503	1.483	0.033	1.62	2.67	80.66	82.21	8.029	0.15810
36	15	D_{6h}	1.414	1.401	1.490	1.479	0.030	0.70	1.27	85.87	87.55	7.126	0.15949
38	17	C_2	1.377	1.401	1.491	1.481	0.034	1.25	2.20	91.31	92.71	6.409	0.16072
40	38	D_2	1.377	1.402	1.497	1.480	0.034	1.46	2.61	96.67	97.90	5.749	0.16184
42	45	D_3	1.383	1.402	1.489	1.481	0.030	1.49	2.39	101.8	103.1	5.012	0.16284
44	75	D_2	1.384	1.401	1.481	1.480	0.027	1.36	2.21	107.1	108.3	4.296	0.16368
46	109	C_2	1.395	1.398	1.483	1.481	0.030	1.64	4.30	112.4	113.2	4.066	0.16522
48	171	C_2	1.387	1.399	1.480	1.479	0.027	1.65	4.28	117.6	118.3	3.384	0.16587
50	271	D_{5h}	1.395	1.401	1.468	1.474	0.021	1.01	1.71	122.8	123.9	2.526	0.16591
52	422	C_2	1.389	1.399	1.480	1.479	0.031	1.94	5.57	128.2	128.5	2.512	0.16782
54	540	C_{2v}	1.379	1.398	1.480	1.479	0.024	1.19	3.80	133.4	134.0	1.921	0.16766
56	916	D_2	1.382	1.396	1.469	1.477	0.026	1.54	4.70	138.8	139.0	1.560	0.16877
58	1205	C_{3v}	1.391	1.397	1.467	1.479	0.022	1.48	4.65	144.0	144.2	1.260	0.16911
60	1812	I_h	1.399	1.401	1.452	1.458	0.005	0.06	0.02	149.1	150.1	0.000	0.16851

Table 4.1 Properties of selected fullerenes C_N ($N = 20 + 2n, n \in \mathbb{N}_0 \setminus \{1\}$) up to C_{60} (numbering of the isomers stems from lexicographically ordered face-spiral pentagon indices, see ref.[32]) from the extended harmonic force-field (EHFF) and DFT (PBE) optimizations using a def2-TZVPP basis set for carbon. Ideal point group symmetry IPG, smallest $R_{\min}$ and largest $R_{\max}$ bond distances, root mean square error between the force-field and PBE calculated distances ΔR_{rms}, bond angles $\Delta\alpha_{\mathrm{rms}}$ and torsions between adjacent carbon atoms $\Delta\theta_{\mathrm{rms}}$ (for the torsions a positive value implies a convex arrangement), surface area A in Å^2, the PBE total energy $\Delta E_{\mathrm{RFS}}^{\mathrm{PBE}}$ per carbon atom relative to C_{60} in kcal/mol, and the force-field zero-point vibrational energy contribution $\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$ per carbon atom in eV.

N	Isomer	IPG	$R_{\text{min}}^{\text{PBE}}$	$R_{\text{min}}^{\text{EHFF}}$	$R_{\text{max}}^{\text{PBE}}$	$R_{\text{max}}^{\text{EHFF}}$	ΔR_{rms}	$\Delta\alpha_{\text{rms}}$	$\Delta\theta_{\text{rms}}$	A^{PBE}	A^{EHFF}	$\Delta E_{\text{RFS}}^{\text{PBE}}$	$\Delta E_{\text{ZPV}}^{\text{EHFF}}$
60	1812	I_h	1.407	1.401	1.457	1.458	0.004	0.06	0.03	150.3	150.1	0.000	0.16851
70	8149	D_{5h}	1.402	1.395	1.474	1.461	0.022	2.06	6.38	177.1	174.9	-0.875	0.17262
72	11190	D_{6d}	1.390	1.395	1.470	1.468	0.029	1.69	5.23	182.5	179.8	-0.476	0.17341
74	14246	D_{3h}	1.397	1.397	1.476	1.463	0.026	2.20	7.30	187.9	184.8	-0.913	0.17394
76	19150	D_2	1.395	1.392	1.476	1.466	0.029	2.37	6.95	193.2	189.8	-1.039	0.17456
76	19151	T_d	1.396	1.398	1.477	1.461	0.028	2.28	7.84	193.3	189.8	-0.900	0.17453
78	24105	D_3	1.393	1.392	1.474	1.463	0.032	2.51	7.01	198.5	194.8	-1.222	0.17512
78	24107	C_{2v}	1.385	1.397	1.476	1.464	0.029	2.53	7.88	198.6	194.8	-1.087	0.17509
78	24108	D_{3h}	1.369	1.396	1.472	1.464	0.033	2.26	6.43	198.6	194.7	-1.175	0.17521
78	24109	D_{3h}	1.382	1.398	1.474	1.461	0.029	2.47	8.35	198.6	194.8	-0.897	0.17511
80	31924	I_h	1.424	1.401	1.458	1.457	0.027	2.35	8.80	204.0	199.7	-0.979	0.17562
92	126409	T	1.387	1.395	1.474	1.460	0.037	2.94	8.72	236.1	229.7	-1.681	0.17841
100	285880	C_{2v}	1.376	1.370	1.474	1.468	0.041	3.24	8.08	257.3	249.9	-2.121	0.17991
140	$(2,1)$	I	1.404	1.387	1.461	1.467	0.046	4.31	8.62	364.0	351.0	-3.770	0.18432
180	$(3,0)$	I_h	1.401	1.373	1.457	1.473	0.050	3.89	6.72	470.7	452.8	-4.708	0.18686
240	$(2,2)$	I_h	1.399	1.385	1.452	1.473	0.045	3.64	5.75	630.2	605.3	-5.577	0.18895
260	$(3,1)$	I	1.402	1.384	1.459	1.474	0.045	3.31	5.12	683.5	656.3	-5.475	0.18951
320	$(4,0)$	I_h	1.399	1.385	1.455	1.474	0.043	2.78	4.40	843.1	809.2	-5.964	0.19060
380	$(3,2)$	I	1.399	1.387	1.454	1.472	0.041	2.59	4.12	1002.7	961.9	-6.356	0.19131
420	$(4,1)$	I	1.400	1.388	1.448	1.472	0.039	2.35	3.89	1109.0	1063.9	-6.594	0.19177
500	$(5,0)$	I_h	1.401	1.387	1.447	1.471	0.038	2.07	3.62	1324.4	1267.7	-6.800	0.19238
540	$(3,3)$	I_h	1.400	1.386	1.445	1.470	0.037	2.01	3.52	1428.3	1369.5	-7.019	0.19263
560	$(4,2)$	I	1.401	1.387	1.445	1.470	0.037	1.96	3.45	1481.5	1420.5	-7.002	0.19271
620	$(5,1)$	I	1.401	1.387	1.445	1.469	0.036	1.83	3.35	1641.2	1573.4	-7.152	0.19305
720	$(6,0)$	I_h	1.401	1.386	1.441	1.469	0.035	1.70	3.16	1907.1	1828.2	-7.386	0.19341
740	$(4,3)$	I	1.401	1.385	1.444	1.469	0.035	1.68	3.09	1960.4	1879.1	-7.397	0.19346
780	$(5,2)$	I	1.401	1.385	1.441	1.469	0.035	1.64	3.04	2066.8	1981.0	-7.489	0.19359
860	$(6,1)$	I	1.402	1.384	1.443	1.468	0.035	1.55	2.93	2279.7	2184.9	-7.764	0.19382
960	$(4,4)$	I_h	1.401	1.383	1.440	1.468	0.034	1.49	2.77	2545.7	2439.8	-7.718	0.19401
980	$(5,3)$	I	1.402	1.383	1.443	1.469	0.033	1.48	2.75	2599.0	2490.7	-7.712	0.19405
980	$(7,0)$	I_h	1.401	1.383	1.442	1.468	0.034	1.47	2.77	2599.0	2490.7	-7.709	0.19405
∞	-	-	1.425	1.401	1.425	1.401	0.024	0	0	-	-	-8.863[a]	0.196[b]

Table 4.2 Properties of selected fullerenes C_N for $N \geq 60$ (numbering of the isomers stems from lexicographically ordered face-spiral pentagon indices up to C_{100}, see ref.[32], otherwise the Goldberg-Coxeter symbols (k,l) are given) from the extended harmonic force-field (EHFF) and DFT (PBE/def2-SVP) optimizations. Ideal point group symmetry IPG, smallest R_{min} and largest R_{max} bond distances, root mean square error between the force-field and PBE calculated distances ΔR_{rms}, bond angles $\Delta\alpha_{\text{rms}}$ and torsions between adjacent carbon atoms $\Delta\theta_{\text{rms}}$ (for the torsions a positive value implies a convex arrangement), surface area A in Å^2, the PBE total energy $\Delta E_{\text{RFS}}^{\text{PBE}}$ per carbon atom relative to C_{60} in kcal/mol, and the force-field zero-point vibrational energy contribution $\Delta E_{\text{ZPV}}^{\text{FF}}$ per carbon atom in eV.

[a]See Table 4.3 for graphene data. [b]The ZPV energy contribution of graphene is the extrapolated value from the fullerene data.

Table 4.3 C–C bond distance R_{CC} (in Å), total energies E_T (in eV/atom) for graphene (G) and C_{60}, calculated in a plane wave implementation of various DFT functionals, and the energy difference $\Delta E(\infty)$ (in kcal/mol) between the different allotropes (C_{60}→graphene, per atom).

DFT method	R_{CC}	E_T(G)	E_T(C_{60})	$\Delta E(\infty)$
LDA	1.413	-10.0917	-9.7074	-8.862
PBE	1.425	-9.2271	-8.8530	-8.627
PBE-D2	1.425	-9.2821	-8.9051	-8.694
optB88-vdW	1.423	-7.9515	-7.5702	-8.793
exp.	1.42	-	-	-

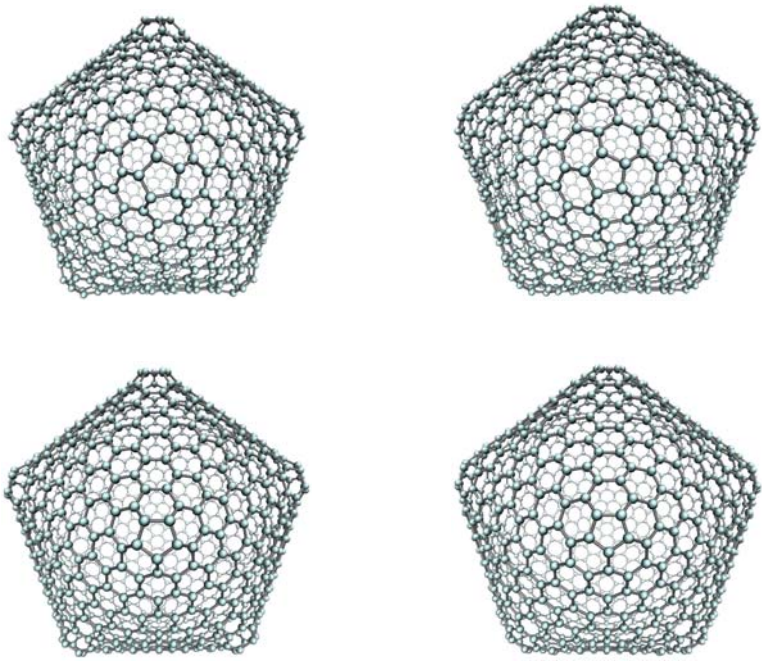


Figure 4.3 Force-field (left) and PBE optimized (right) structures of C_{980}-I (top) and C_{980}-I_h (bottom)

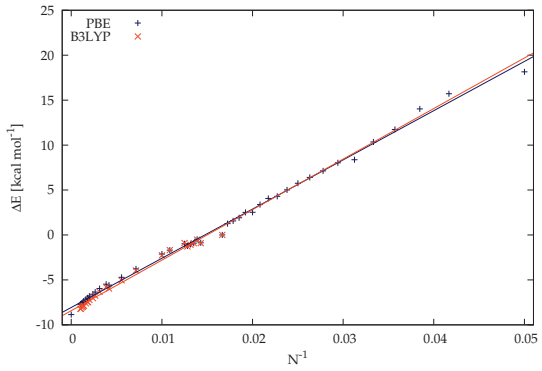


Figure 4.4 $\Delta E^{\mathrm{PBE}}_{\mathrm{RFS}}$ for the fullerenes as listed in Tables 4.1 and 4.2 plotted against N^{-1}. For the PBE functional the graphene limit is shown (see Table 4.3), and for C_{60} onwards B3LYP results are also shown.

The graph representing C_{60} has 90 edges, out of these only $1/3$ are double bonds (giving rise to 12,500 Kekulé structures of which 158 are symmetry distinct).[161–163] Moreover, the heat of formation cannot be simply derived from single and double bond energy increments as there is an additional repulsive curvature term from the 12 pentagons in the fullerene, which is mainly responsible for the difference in stability to graphene.[29] However, this curvature term approaches zero as "spherical" fullerenes with pentagons separated as far as possible from each other grow toward infinity ($N \to \infty$). Such fullerenes can be obtained as Goldberg-Coxeter transformations of C_{20}, i.e., $GC_{k,l}[C_{20}]$. This is the main reason why we used this class of fullerenes (up to C_{980}) to obtain the graphene limit. For a recent interesting discussion on smooth scaling of fullerene properties toward the graphene limit see Lewis et al.[316]

In a recent paper by Radom and co-workers, the heat of formation for C_{60} was calculated accurately to be 602.7 kcal/mol,[151] which translates into 10.0 kcal/mol per carbon atom. This value is already in very good agreement with our PBE or B3LYP graphene limit value (mean value between both levels of theory is 8.9 ± 0.3 kcal/mol which is in the range of values listed in Table 4.3). Taking into account the small Van der Waals interactions between the graphene sheets in graphite (+1.5 kcal/mol)[293] and differences

in zero-point vibrational energy between C_{60} and graphene (-0.63 kcal/mol, see discussion below), and neglecting finite temperature effects, we arrive at a heat of formation for C_{60} of 596 ± 20 kcal/mol in excellent agreement with the value by Radom and co-workers or with the NIST recommended value of 612 ± 24 kcal/mol.[185,317] Note that already in 1996 Dunlap and Boettger discussed the graphene limit for fullerenes by using a local density approximation.[302] In another, Dunlap and Zope obtained an enthalpy of formation of 594 kcal/mol for C_{60} using an Xα approximation (with an α-value of 0.64190).[303]

Turning to the zero-point vibrational energy ΔE_{ZPV}, which we obtain from our force-field calculations, we see from Figure 4.5 that for the larger IPR fullerenes, $\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$ follows an almost perfect linear trend with respect to N^{-1} for C_{60} onwards. In fact, the fullerenes with adjacent pentagons also follow a linear trend, but different to the isolated pentagon fullerenes. From the latter we can extrapolate the harmonic zero-point vibrational energy contribution for graphene as 0.196 eV per carbon atom. This value should be rather accurate as the $\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$ values for C_{50}, C_{60} and C_{70} were in good agreement with the PBE results as already discussed. This value is of the right order as well as can be deduced from calculated phonon dispersion curves and experimental Raman frequencies for graphene.[318] Even though the phonon dispersion curves are known rather accurately,[318–320] we only found one publication using a semi-empirical approach for the zero-point vibrational energy including anharmonicity effects, which gave a value of 0.165 eV per carbon atom somewhat below our extrapolated value of 0.196 eV.[321] Another value (0.181 eV) comes from PBE calculations of graphite,[322] which should be close to the graphene value.

4.5 Summary and Conclusions

We have implemented a general force field as part of the program *Fullerene*,[36] which works for all fullerenes. Structure optimizations show that geometries obtained from this force-field are in excellent agreement with those obtained from density functional theory. This force field is easily extendable to more general polyhedral structures. Isodesmic reaction energies ΔE_{RFS} compared to C_{60}-I_h for Goldberg-Coxeter fullerenes were used to ex-

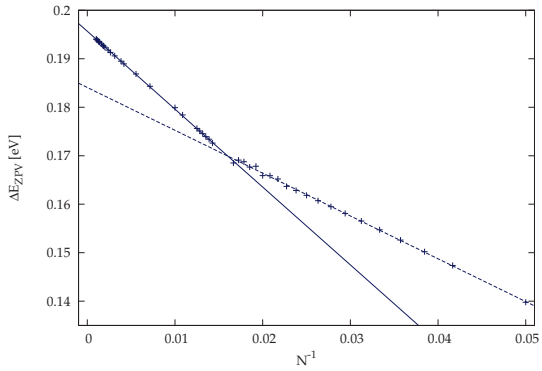


Figure 4.5 $\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$ for the fullerenes as listed in Tables 4.1 and 4.2 plotted against N^{-1}. For the linear fit the data are split in two regions, $N \geq 60$ (solid line) and $N \leq 60$ (dashed line).

trapolate to the graphene limit. The extrapolated ΔE_{RFS} graphene value was used to obtain the heat of formation of C_{60}, which is in very good agreement with the more accurate value published by Radom and co-workers.[151] The force constants of the general force field were adjusted to the vibrational spectrum of three selected fullerenes, which gives rather accurate values for the zero-point vibrational contribution $\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$. Extrapolation to the graphene limit allows to obtain the $\Delta E_{\mathrm{ZPV}}^{\mathrm{EHFF}}$ value for graphene, which is in reasonable agreement with estimates from phonon dispersion calculations.

4.6 A force field for general polyhedral graphs and triangulations of the sphere

The primary force field of the program *Fullerene* is written in Fortran and is limited to embeddings of fullerene graphs. The optimisation is done using the Polak-Ribiere conjugate gradient algorithm[323], that is, evaluating the energy function and its derivatives. The energy function comprises distance, angle and dihedral angle terms. The potential can be chosen between a harmonic and a hyperbolic potential. Hessians can be computed using analytical second derivatives.

In order to optimise 3D embeddings of general polyhedral graphs and

triangulations of the sphere, a second force field has been implemented as a part of libgraph (see section 2.14), or more specifically as a member function of the Polyhedron class.

The optimisation is implemented using the gsl multimin library.[324] Here, again the Polak-Ribiere algorithm is chosen.

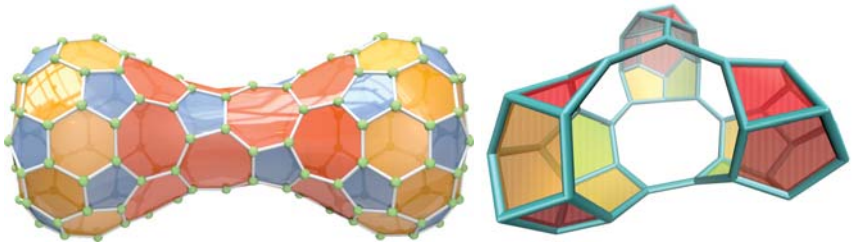
For polyhedral graphs, the energy function comprises distance, angle, and dihedral angle restraints for which the first derivatives are given analytically. For the optimisation of triangulations only distance restrains are employed. Harmonic potentials are used for all force terms.

For distance restraints, individual zero values for each edge length can be input; alternatively a default value is used. Ideal angles α between two edges inside an n-gon are calculated under the assumption, that every face is planar: $\alpha = \pi(1 - \frac{2}{n})$. The ideal dihedral angle at every vertex as defined in fig. 4.2 can be calculated from the three adjacent angles and edge lengths (as three angles and the three edge lengths adjacent to one vertex fully define a tetrahedron). Here, it is assumed that the sum of the three angles is less or equal to 2π, i.e., that the vertex is a singularity with zero or positive Gaussian curvature. For sizes n_1 to n_3 of adjacent faces that translates to $\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3} \geq 0.5$. Otherwise, in the case of a hyperbolic singularity the dihedral angle is biased towards a value of 0. Then, not all adjacent angles can be optimised to their ideal value. Expressing ideal angles and dihedral angles as a function of only edge lengths and face sizes allows for optimising polyhedra with arbitrary face sizes without prior parametrisation.

As long as all singularities with positive Gaussian curvature of the 2D surface are embedded in 3D such that they are convex, there is one well defined optimal embedding up to isomorphism. If convex and concave singularities are permitted, there may be multiple embeddings that represent the 2D surface equally well. This is true not only for discrete but also for continuously curved surfaces.

Here, the first derivatives of all force terms were manually derived, which is infeasible for higher derivatives of non-simple coordinate dependencies like dihedral angles. Analytical expressions for higher derivatives can efficiently by obtained by algorithmic (or automatic) differentiation.[325] This technique makes use of the fact, that in any programming language even complicated functions are expressed as a concatenation of simple mathematical functions. Programs for algorithmic differentiation then apply the

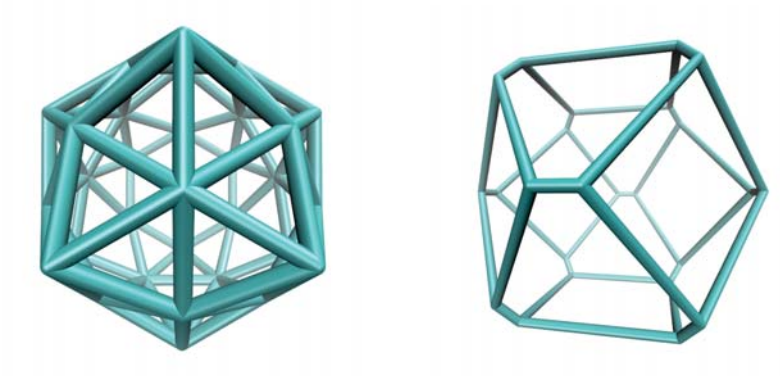
chain rule and return analytical expressions of derivatives (in the respective programming language). Algorithmic differentiation is not to be mistaken for numerical differentiation which suffers from much higher numerical inaccuracy.



(a) (b)

Figure 4.6 (a) C_{86} carbon cage with pentagons, hexagons, and heptagons; (b) an embedding of the Tutte graph into 3D; this is an example for a polyhedral graph with a wide range of face sizes (4, 5, 9, and 10)

Figures 4.6 to 4.7 show examples of structures that were optimised using the libgraph force field.



(a) (b)

Figure 4.7 (a) Optimised 3D embedding of the dual of the icosahedral C_{60} fullerene; (b) Truncated octahedron, optimised with different edge lengths. This structure was used as a precursor for a cage in which C_2 units where inserted into selected bonds.

5 Hamilton Cycles in Fullerene Graphs[q]

5.1 Introduction

A Hamilton cycle (or Hamilton circuit) is a spanning cycle in a graph $G = (\mathcal{V}, \mathcal{E})$ that visits every vertex exactly once. The test for Hamiltonicity is known to be NP-complete,[327–329] while counting Hamilton cycles is $\#P$-complete (in contrast for example to perfect matchings where the counting can be done efficiently through the Pfaffian of the adjacency matrix).[330] Tutte proved in 1956 that every 4-connected planar graph is Hamiltonian.[331] However, while there are a few necessary and sufficient conditions known for Hamilton cycles in general,[332] they are not very useful for cubic planar graphs such as fullerenes.

For cubic graphs with $N = |\mathcal{V}|$ vertices, Eppstein provided an upper bound of $2^{3N/8}$ on the number of Hamilton cycles, and conjectured that the bound could be improved to $2^{N/3}$.[333] He also found a family of graphs that have exactly $2^{N/3}$ Hamilton cycles, thus demonstrating that, if the conjecture is true, the upper bound would be tight. More recently, Heidi Gebauer improved the original upper bound to 1.276^N.[334,335] If one allows for multigraphs, then a cycle that alternates between single and double bonds has exactly $2^{N/2}$ Hamilton cycles. The smallest non-Hamiltonian polyhedral graph is the Herschel graph with 11 vertices,[336] and the smallest non-Hamiltonian cubic polyhedral graph is the Barnette-Bosák-Lederberg graph with 38 vertices (there are 6 non-isomorphic of them).[337,338] Historically, the first counterexample to Tait's conjecture that every 3-regular polyhedron

[q] The following sections 5.1 to 5.5 are being published as "Toward Tight Upper and Lower Bounds for Hamilton Cycles in Fullerene Graphs" by Wirz et al.[326]. The candidate's contribution to the article is itemised on pages 299 ff.

has a Hamilton cycle was the Tutte graph with 46 vertices.[339] Reviews on Hamilton cycles are given by Holton and Aldred[340] and Gould.[332,341]

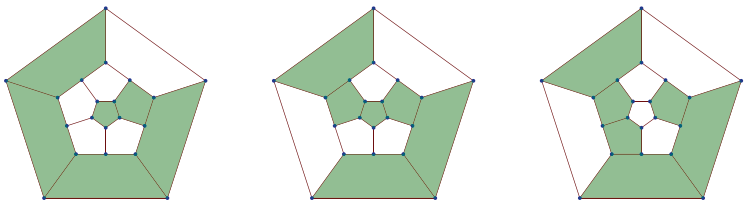
Barnette's famous conjecture states that every 3-connected cubic bipartite planar graph has a Hamilton cycle,[244] and Holton et al. showed that this is true for up to 66 vertices.[342] More important for the discussion here, Barnette's second conjecture states that every 3-connected cubic polyhedral graph with n-gons of $n \leq 6$ (six or fewer edges) is Hamiltonian. Aldred et al. demonstrated that such graphs up to 176 vertices are indeed Hamiltonian,[343] and Brinkmann et al. extended the list to 250 vertices.[344] Goodey proved that the conjecture is true for 3-connected cubic planar graphs that contain only triangles and hexagons.[345] Kardoš presented a computer proof for Bernette's second conjecture in 2014.[346] This proof is constructuve and hence not only guarantees the existence of at least one Hamilton cycle but describes how to construct it in linear time.

Fullerenes belong to the class of 3-connected cubic planar graphs consisting of exactly 12 pentagons and F_6 hexagons ($F_6 \in \mathbb{N} \setminus \{1\}$). It is currently not known whether all fullerene graphs are Hamiltonian. Recent work by Brinkman, Goedgebeur and McKay, who demonstrated the existence of Hamilton cycles for all fullerenes up to 316 vertices by generating all possible non-isomorphic fullerene graphs and testing them for Hamiltonicity, gave strong support to the conjecture that all fullerenes are Hamiltonian.[66] Erman et al. proved that every fullerene graph has a cycle of length at least $\text{int}\{6(N+2)/7\}$, and is therefore almost Hamiltonian.[347] Thomassen proved the existence of Hamilton cycles for torus-shaped fullerenes (ring-shaped nanotubes or torusenes of genus 1) containing hexagons only.[348] Marušić investigated Hamiltonicity in leap-frog fullerenes.[245] Franzblau proved by complete induction that every layered cubic planar graph has a Hamilton cycle.[349] Here we conjecture that the C_{10+10k}-D_{5+} nanotubes with even and odd k (+ sign denotes either h for even k or d for odd k), belonging to the class of layered fullerenes and thus are Hamiltonian, play a special role in providing tight lower and upper bounds for Hamilton cycles for almost all fullerenes. This class of fullerenes are the smallest diameter fullerene nanotubes.[121]

5.2 Method

The program *Fullerene*[36] was used to generate all C_N fullerene graphs for every $N \leq 120$, and to count the number of admitted Hamiltonian cycles for each generated graph. The Hamiltonian cycles of the generated graphs were counted using the back-track algorithm for fullerenes introduced by Babić.[350] Details of the algorithm can be found in ref.[350]. Note that the number of non-isomorphic fullerene graphs with a given number of vertices (called isomers) scales like $\mathcal{O}\left(N^9\right)$,[20] and the backtrack algorithm scales like $\mathcal{O}\left(2^N\right)$.[333] But while the worst case for counting Hamiltonian paths is necessarily exponential, the usual case is so fast that we have been able to count all Hamiltonian paths for all C_N-isomers up to $N = 120$. Counting the Hamiltonian paths for the isomers of C_{120} required about a month on 400 CPUs. For single (5,0) and (6,0) nanotube fullerenes (for the terminology see ref.[351]) we were able to count Hamilton cycles up to 160 vertices before the counting algorithm became too computationally demanding.

For the discussion on the upper bound for the Hamilton cycle count, we also determined the maximum number of perfect matchings for every fullerene isomer. This can be done very efficiently in $\mathcal{O}\left(N^3\right)$ time by way of the Pfaffian of the adjacency matrix, using the Fisher-Kasteleyn-Temperley algorithm.[167]



(A) (B) (C)

Figure 5.1 The three symmetry distinct Hamilton cycles A, B and C (up to 5-fold rotational symmetry and inversion) for C_{20}-I_h with $N_{\mathrm{HC}} = 30$. The Hamilton cycle runs on the boundary of the green shaded area.

5.3 Results and discussion

C_{20} is the smallest fullerene and has exactly 30 Hamilton cycles, of which only 3 (shown in Figure 5.1) are distinct by symmetry operations. Note that any closed non-selfintersecting path (simple closed curve such as a Hamilton cycle) on a genus-0 surface divides the surface in two parts (Jordan curve theorem), which is represented by the two different colors (green and white) for the polygons used in Figure 5.1.

The smallest fullerene nanotubes are of D_{5h} or D_{5d} symmetry because of the pentagon at the top and bottom of the tube. They play a special role for the Hamilton cycle counts as we shall see, and are shown for the two representative candidates C_{70}-D_{5h} and C_{60}-D_{5d} in Figure 5.2 (not to be confused with the famous C_{70}-D_{5h} and C_{60}-I_h buckyballs). These fullerenes belong to the class of zig-zag nanotubes with a chiral vector[351] of $\mathbf{C}_h$ = (5,0) having a cap of six fused pentagons at each end. From basic group theoretical arguments one derives that these two (topological) point groups can only occur at vertex number $N = 10 + 10k$ ($k \in \mathbb{N}$) for fullerenes.[1] In graph theoretical terms, the D_{5h} structure has the top and bottom pentagons aligned in the same way, while in the D_{5d} structure the top pentagon is rotated by $\pi/5$ compared to the bottom pentagon. In fact, the fullerene (5,0) capped carbon nanotubes C_{10+10k} belong to the class of $k+1$-layered cubic planar graphs, $P(m_i), i = 1, ..., k+1$ with $m_1 = 5$ and $m_i = 1$ for all $i > 1$, and thus are Hamiltonian (for details see Franzblau[349]).

An analysis on the number of non-isomorphic Hamilton cycles N_{HC} computed for all fullerene isomers up to a vertex number of $N = 120$ revealed an interesting and perhaps quite unexpected fact: the C_{10+10k}-D_{5+} (5,0) fullerene nanotubes have either the lowest (for $N \geq 70$) or the highest count on Hamilton cycles within the class of $N = 10 + 10k$ fullerene isomers, as the results in Table 5.1 demonstrate. One may therefore be tempted to speculate that these fullerenes also provide good lower and upper bounds for all fullerene isomers of any vertex number. For the upper bound this assumption is incorrect as we shall see.

Table 5.2 lists the number of Hamilton cycles N_{HC} for the C_{10+10k}-D_{5+} (5,0) fullerene nanotubes in comparison with the next largest in diameter C_{12+12k}-D_{6+} (6,0) fullerene nanotubes (shown in Figure 5.3) up to highest computationally feasible vertex numbers. In contrast to the (6,0) fullerene

N	$N_{\mathrm{HC}}^{\mathrm{LB}}$	PG	RSPI	P_1	$N_{\mathrm{HC}}^{\mathrm{UB}}$	PG	RSPI	P_1
20	30	I_h	{1,2,3,4,5,6,7,8,9,10,11,12}	30	30	I_h	{1,2,3,4,5,6,7,8,9,10,11,12}	30
24	34	D_{6d}	{1,2,3,4,5,7,8,10,11,12,13,14}	24	34	D_{6d}	{1,2,3,4,5,7,8,10,11,12,13,14}	24
26	24	D_{3h}	{1,2,3,4,5,7,9,11,12,13,14,15}	21	24	D_{3h}	{1,2,3,4,5,7,9,11,12,13,14,15}	21
28	18	T_d	{1,2,3,5,7,9,10,11,12,13,14,15}	18	43	D_2	{1,2,3,4,5,7,10,12,13,14,15,16}	20
30	20	D_{5h}	{1,2,3,4,5,6,12,13,14,15,16,17}	20	32	C_{2v}	{1,2,3,4,7,10,11,12,13,14,15,16}	17
32	40	C_2	{1,2,3,4,7,10,11,12,14,15,17,18}	16	76	D_2	{1,2,3,4,5,8,12,13,15,16,17,18}	18
34	28	C_2	{1,2,3,4,5,7,13,15,16,17,18,19}	17	66	C_{3v}	{1,2,3,5,7,10,11,14,15,17,18,19}	15
36	42	D_{2d}	{1,2,3,4,7,10,12,15,17,18,19,20}	14	128	D_2	{1,2,3,4,5,8,13,16,17,18,19,20}	18
38	24	D_{3h}^*	{1,2,3,4,5,9,13,17,18,19,20,21}	18	96	C_2	{1,2,3,5,10,11,13,15,16,19,20,21}	12
40	68	C_s	{1,2,3,4,5,12,14,17,19,20,21,22}	14	280	D_{5d}	{1,2,3,4,5,6,17,18,19,20,21,22}	20
42	44	C_2^*	{1,2,3,4,5,7,17,19,20,21,22,23}	16	150	C_{2v}	{1,2,3,4,7,12,15,17,18,20,22,23}	12
44	120	D_{3d}	{1,2,3,4,5,9,18,19,20,22,23,24}	18	327	D_2	{1,2,3,4,5,8,18,19,21,22,23,24}	18
46	76	C_s	{1,2,3,4,5,13,18,19,21,22,23,25}	14	260	C_s	{1,2,3,4,5,14,17,19,20,22,24,25}	13
48	152	D_{2h}	{1,2,3,4,7,10,17,20,23,24,25,26}	14	512	C_1	{1,2,3,4,5,13,19,20,22,23,24,26}	15
50	80	D_{5h}	{1,2,3,4,5,6,22,23,24,25,26,27}	20	410	C_{2v}	{1,2,3,4,5,14,17,20,24,25,26,27}	14
52	262	C_s	{1,2,3,4,5,12,19,22,25,26,27,28}	14	806	D_2	{1,2,3,4,5,8,23,24,25,26,27,28}	18
54	66	D_{3h}^*	{1,2,3,4,14,18,19,22,23,26,27,29}	15	642	C_2	{1,2,3,4,7,12,18,23,26,27,28,29}	12
56	440	C_s	{1,2,3,4,14,19,20,21,24,25,26,29}	9	1746	D_3	{1,2,6,7,17,18,21,22,23,26,27,30}	15
58	173	C_2	{1,2,3,4,5,13,19,27,28,29,30,31}	15	1068	C_s	{1,2,3,4,11,14,17,25,27,28,29,31}	10
60	618	D_{2d}	{1,2,3,4,7,10,25,28,29,30,31,32}	14	3040	D_{5d}	{1,2,3,4,5,6,27,28,29,30,31,32}	20
62	288	C_2	{1,2,3,4,5,13,21,29,30,31,32,33}	14	1802	C_s	{1,2,3,4,11,14,20,26,29,30,31,33}	10
64	1062	C_3	{1,2,3,5,18,22,23,24,27,28,29,32}	12	3340	D_2	{1,2,3,4,5,8,29,30,31,32,33,34}	18
66	197	C_2^*	{1,2,3,4,5,13,23,31,32,33,34,35}	14	3096	C_{2v}	{1,2,3,4,12,19,21,29,31,32,33,35}	10
68	1750	C_1	{1,2,3,4,7,18,23,27,32,34,35,36}	10	6018	D_3	{1,2,3,4,19,23,24,28,29,32,33,36}	15
70	320	D_{5h}	{1,2,3,4,5,6,32,33,34,35,36,37}	20	4818	C_1	{1,2,3,4,11,20,27,30,31,32,34,37}	10
72	2688	C_2	{1,2,3,4,7,21,26,30,33,34,37,38}	12	10428	D_{2h}	{1,2,3,4,20,21,25,26,32,36,37,38}	12
74	1182	C_2	{1,2,3,4,5,13,27,35,36,37,38,39}	14	7832	C_2	{1,2,3,4,14,24,28,29,32,34,36,37}	10
76	4230	T	{1,2,3,19,22,23,26,27,30,31,34,37}	12	15926	D_2	{1,2,3,4,19,20,24,25,37,38,39,40}	12
78	1596	D_{3h}	{1,4,5,12,25,26,30,31,32,36,37,41}	15	12226	C_1	{1,2,3,4,19,20,31,33,34,35,38,40}	8
80	7110	C_s	{1,2,3,4,5,27,29,34,36,38,39,40}	9	35200	D_{5d}	{1,2,3,4,5,6,37,38,39,40,41,42}	20
82	2400	C_{2v}	{1,2,3,19,20,24,25,28,33,38,41,42}	9	20856	C_1	{1,2,3,4,11,20,29,37,39,40,41,43}	10
84	10814	C_2	{1,2,3,4,7,18,27,38,41,42,43,44}	12	39067	C_2	{1,2,3,4,19,20,23,31,40,41,42,44}	12
86	1980	D_3	{1,2,3,4,29,30,34,35,36,40,41,45}	15	33427	C_1	{1,2,3,4,13,22,28,38,41,42,43,45}	10
88	17905	C_1	{1,2,3,4,5,20,29,40,43,44,45,46}	13	76063	C_2	{1,2,3,4,20,21,32,39,42,43,44,46}	12
90	1280	D_{5h}	{1,2,3,4,5,6,42,43,44,45,46,47}	20	51585	C_1	{1,2,3,4,19,20,32,35,40,42,44,46}	8
92	29944	C_2	{1,2,3,5,7,12,38,43,44,46,47,48}	12	117106	C_2	{1,2,3,4,21,22,40,41,44,45,46,48}	12
94	7930	C_2	{1,2,3,19,20,24,28,33,37,42,45,46}	8	90221	C_1	{1,2,3,4,14,32,35,38,39,43,46,47}	10
96	46231	C_2	{1,2,3,4,7,18,33,44,47,48,49,50}	12	209692	D_3	{1,2,3,4,29,34,35,40,41,45,46,50}	15
98	13307	C_1	{1,2,3,4,14,20,36,37,47,48,49,51}	11	156288	C_{2v}	{1,2,3,4,23,32,35,39,40,44,48,49}	10
100	72168	C_2	{1,2,3,5,7,12,44,46,49,50,51,52}	12	417280	D_{5d}	{1,2,3,4,5,6,47,48,49,50,51,52}	20
102	20754	C_s	{1,2,3,4,29,34,38,40,41,45,46,53}	13	249148	C_{2v}	{1,2,3,4,19,22,32,38,49,50,51,53}	10
104	119540	C_2	{1,2,3,5,7,12,45,49,50,51,52,53}	12	686286	D_3^*	{2,8,9,19,35,36,41,42,43,48,49,54}	15
106	40912	C_2	{1,2,3,4,5,13,43,51,52,53,54,55}	14	421194	C_s	{1,2,3,8,29,32,40,42,46,47,50,55}	8
108	184445	C_2	{1,2,3,5,7,12,45,50,52,54,55,56}	12	1104223	D_3	{1,3,4,5,34,35,46,47,52,53,54,56}	15
110	5120	D_{5h}	{1,2,3,4,5,6,52,53,54,55,56,57}	20	743346	C_s	{1,2,3,4,20,31,42,50,53,54,55,57}	10
112	283642	C_2	{1,2,3,5,7,12,51,52,53,55,56,57}	12	1373166	C_2	{1,2,3,4,34,41,42,47,48,49,53,54}	13
114	84329	C_2	{1,2,3,4,5,13,47,55,56,57,58,59}	14	1122238	C_1	{1,2,3,4,30,41,43,48,49,55,57,59}	8
116	463959	C_2	{1,2,3,5,7,12,51,52,56,58,59,60}	12	2874489	D_3	{1,2,3,4,41,42,47,48,49,54,55,60}	15
118	113548	C_2	{1,2,3,4,5,13,49,57,58,59,60,61}	14	2110120	C_1	{1,2,3,4,29,32,42,52,57,58,59,61}	10
120	701267	C_2	{1,2,3,5,7,12,52,57,58,60,61,62}	12	4986880	D_{5d}	{1,2,3,4,5,6,57,58,59,60,61,62}	20

Table 5.1 Exact lower ($N_{\mathrm{HC}}^{\mathrm{LB}}$) and upper bounds ($N_{\mathrm{HC}}^{\mathrm{UB}}$) for the number of distinct Hamilton cycles (HC) in fullerenes up to C_{120}. N is the number of vertices, PG the topological (ideal) point group, RSPI are the canonical face-spiral pentagon indices, $P_1 = \sum_{k=1}^{5} kp_k/2$ is the pentagon signature with with p_k being the pentagon neighbor indices (see ref. [32] for definitions). The fullerene isomers not following the upper or lower bounds defined by the C_{10+10k}-D_{5+} (5,0) nanotubes are indicated by an asterisk behind the symmetry group label.

(a) (b)

Figure 5.2 Two representative polyhedra for (5,0) fullerene nanotubes, (a) C_{70}-D_{5h} and (b) C_{60}-D_{5d}. Pentagon caps are distinguished from the hexagons by shading.

nanotubes, we see a distinct oscillating behavior in N_{HC} between the two different symmetries D_{5h} and D_{5d} (Figure 5.4), thus the (5,0) fullerene nanotubes separate into two different classes depending on their symmetry. We derive and prove analytical formula for the number of Hamilton cycles in D_{5+} (5,0) and D_{6+} (6,0) nanotubes (see Appendix A and B).

Theorem 1. *The k^{th} C_{10+10k}-D_{5h} (5,0) fullerene nanotube (for even k and $k \in \mathbb{N}$) admits exactly*

$$N_{\mathrm{HC}}^{D_{5h}}(k) = 5 \times 2^k \tag{5.1}$$

Hamilton cycles.

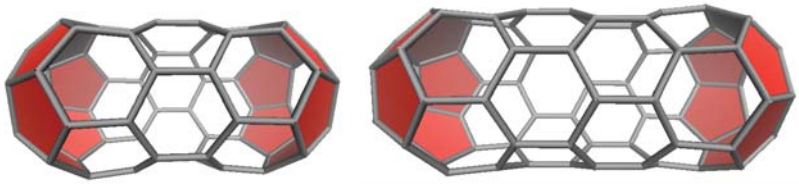
Theorem 2. *The k^{th} C_{10+10k}-D_{5d} (5,0) fullerene nanotube (for odd k and $k \in \mathbb{N}$) admits exactly*

$$N_{\mathrm{HC}}^{D_{5d}}(k) = 5 \times \left[2^k + 4 \cdot 12^{\frac{k-1}{2}}\right] \tag{5.2}$$

Hamilton cycles.

From this we get $N_{\mathrm{HC}}^{D_{5h}}(k) < N_{\mathrm{HC}}^{D_{5d}}(k+1)$ (k even) and $N_{\mathrm{HC}}^{D_{5d}}(k) > N_{\mathrm{HC}}^{D_{5h}}(k+1)$ (k odd), which explains the alternating behavior in the Hamilton cycles for the (5,0) fullerene nanotubes as shown in Figure 5.4. We mention that analytical expressions for the topological Wiener and Szeged indices were found previously for these type of fullerene nanotubes.[133,352]

The two symmetry distinct Hamilton cycles for C_{30}-D_{5h} are shown in Figure 5.5. A comparison to C_{20}-I_h (Figure 5.1), the first "fullerene nanotube" of subgroup D_{5d}, shows that there are more symmetry distinct Hamilton cycles than in C_{30}-D_{5h}. A similar picture arises if we compare the few symmetry distinct Hamilton cycles for C_{50}-D_{5h} (Figure 5.6) with the many



(a) (b)

Figure 5.3 Two representative polyhedra for the (6,0) fullerene nanotubes, (a) C_{60}-D_{6h} and (b) C_{72}-D_{6d}.

for C_{40}-D_{5d} (Figures 5.7–5.9). The prefactor of 5 in the equations of both theorems reflects the five-fold symmetry of the fullerene, i.e., the 5 starting points for Hamilton cycles from the top pentagon at the end of the nanotube. The 2^k law reflects the two possible ways (left-right) one can go in a cubic graph. More details can be found in the appendix.

Upper ($N_{\mathrm{HC}}^{\mathrm{UB}}$) and lower ($N_{\mathrm{HC}}^{\mathrm{LB}}$) bounds for Hamilton cycles of a specific fullerene are listed together with the corresponding face-spiral pentagon indices (FSPIs) and topological point group symmetries in Table 5.1 up to a vertex count of $N = 120$, and separately for isolated pentagon (IPR) fullerenes (non-adjacent pentagons) up to a vertex count of $N = 122$ in Table 5.3. The results are summarized in Figure 5.10. It is immediately apparent that both upper and lower bounds for the number of Hamilton cycles are determined by the non-IPR fullerenes, which leads to the following conjecture,

Conjecture 2. *Lower and upper bounds for Hamilton cycle counts of fullerenes are determined by non-IPR fullerenes.*

From the data listed in Table 5.1 it is apparent that the C_{10+10k}-D_{5h} (5,0) fullerene nano-tubes with even k and face-spiral pentagon indices of FSPI $= (1, 2, 3, 4, 5, 6, F-5, F-4, F-3, F-2, F-1, F)$ (number of faces $F = N/2 + 2$) admit the lowest number of Hamilton cycles within the class of (5,0) nanotubes, leading to the following conjecture,

Conjecture 3. *The number of Hamilton cycles determined by the C_{10+10k}-D_{5h} (5,0) (k even) fullerene nanotubes with $N_{\mathrm{HC}} = 5 \times 2^k$ gives exact lower bounds for this class of fullerenes.*

N	PG	N_{HC}	N	PG	N_{HC}
20	D_{5d}	30	24	D_{6d}	34
30	D_{5h}	20	36	D_{6h}	80
40	D_{5d}	280	48	D_{6d}	328
50	D_{5h}	80	60	D_{6h}	1232
60	D_{5d}	3040	72	D_{6d}	5344
70	D_{5h}	320	84	D_{6h}	23168
80	D_{5d}	35200	96	D_{6d}	103168
90	D_{5h}	1280	108	D_{6h}	460544
100	D_{5d}	417280	120	D_{6d}	2065408
110	D_{5h}	5120	132	D_{6h}	9270272
120	D_{5d}	4986880	144	D_{6d}	41641984
130	D_{5h}	20480	156	D_{6h}	187092992
140	D_{5d}	59760640	168	D_{6d}	840712192
150	D_{5h}	81920			
160	D_{5d}	716800000			

Table 5.2 Number of Hamilton cycles N_{HC} in fullerenes (5,0) C_{10+10k}-D_{5+} and (6,0) C_{12+12k}-D_{6+} ($k \in \mathbb{N}$) fullerene nanotubes as obtained from the back-track algorithm implemented into program *Fullerene*. N is the number of vertices and PG denotes the topological (ideal) point group.

Can we generalize this to all fullerene graphs of any vertex number N? The data in Table 5.1 suggests indeed that we can, if we take care of the very few exceptions which do not follow this lower bound. These exceptions are, C_{38}-D_{3h} with $N_{\mathrm{HC}}^{\mathrm{LB}} = 24$, C_{38}-D_3 with $N_{\mathrm{HC}}^{\mathrm{LB}} = 33$, C_{42}-C_2 with $N_{\mathrm{HC}}^{\mathrm{LB}} = 44$, C_{54}-D_{3h} with $N_{\mathrm{HC}}^{\mathrm{LB}} = 66$, and C_{66}-C_2 with $N_{\mathrm{HC}}^{\mathrm{LB}} = 197$. This leads to our next and most important conjecture,

Conjecture 4. *For every $N \geq 20$, except for the special cases $N = 38$ ($N_{\mathrm{HC}}^{\mathrm{LB}} = 24$), $N = 42$ ($N_{\mathrm{HC}}^{\mathrm{LB}} = 44$), $N = 54$ ($N_{\mathrm{HC}}^{\mathrm{LB}} = 66$), and $N = 66$ ($N_{\mathrm{HC}}^{\mathrm{LB}} = 197$),*

$$N_{\mathrm{HC}}^{\mathrm{LB}}(N) := 5\left(2^{N/10-1}\right) \tag{5.3}$$

is a lower bound to the number of Hamilton cycles admitted by any isomer of C_N. This bound is strict, in the sense that it is attained whenever $N = 10 + 10k$ (with even $k \in \mathbb{N}$) or $N \geq 68$.

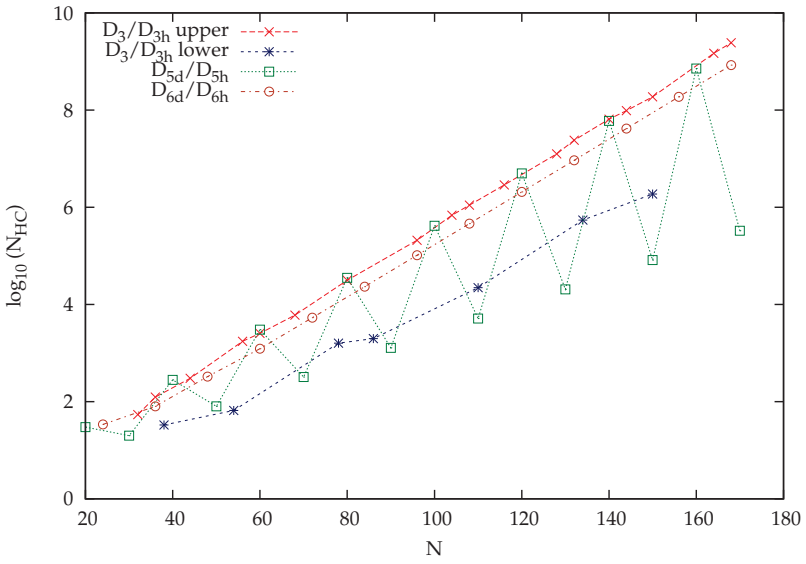
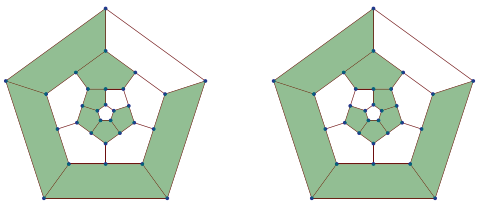


Figure 5.4 Number of Hamilton cycles (logarithmic scale) in (5,0) C_{10+10k}-D_{5+}, and, for comparison, in (6,0) C_{12+12k}-D_{6+} fullerene nanotubes and the D_{3h}/D_3 fullerenes with pentagon neighbor signature $P_1 = 15$ dependent on the vertex number N.

From this lower bound we follow that there are exponentially many Hamilton cycles in fullerenes. The proof for this conjecture will be non-trivial, but if it holds we derive immediately the following important corollary,

Corollary 1. *Every fullerene graph is Hamiltonian.*

The upper bound for the Hamilton cycle count of fullerenes turns out to be far more complicated. Before we discuss our results here, let us mention a nice relationship between Hamilton cycles and perfect matchings in cubic planar graphs. A theorem by Petersen from 1891 states that every cubic bridgeless graph contains a perfect matching.[353] Now, out of the set of the exponentially many perfect matchings for fullerene graphs, there are three perfect matchings which can be assigned to a specific Hamilton cycle (if it exists). Two perfect matchings have all match-edges lying on the Hamilton cycle, hence every Hamilton cycle is a union of two perfect matchings.[354] However, there is no unique mapping from either of these



(A) (B)

Figure 5.5 The two Hamilton paths (up to 10-fold D_{5h} rotational and inversion symmetry) in the D_{5h}-fullerene nanotube C_{30} ($N_{\mathrm{HC}} = 20$).

two perfect matchings back to the generating Hamilton cycle. The third perfect matching consists of all edges that do not lie on the Hamilton cycle. There is a bijective mapping between any Hamilton cycle and this class of perfect matchings. Hence we have the following theorem,

Theorem 3. *The Hamilton cycle count N_{HC} in cubic planar graphs is bounded from above by the number of perfect matchings N_{PM} through*

$$N_{\mathrm{HC}} \leq N_{\mathrm{PM}}. \tag{5.4}$$

Proof. One perfect matching in a cubic planar graph can be obtained as the complement of a Hamilton cycle with respect to the edge set $\mathcal{E}$. Vice versa, the original Hamilton cycle can be obtained as the complement of this perfect matching. Hence, there is a bijective mapping between all Hamilton cycles and a subset of the perfect matchings. This implies that there are fewer or equally many Hamilton cycles than perfect matchings. □

Figure 5.10 shows that the perfect matching count provides a better upper bound for fullerenes than the one proposed by Eppstein ($2^{3N/8}$).[333] In order to get an even tighter upper bound for the Hamilton cycle count, we further analyze the N_{HC} upper bound values for fullerenes shown in Table 5.1. From this we propose,

Conjecture 5. *Among all the isomers for C_{10+10k}, for every odd k, either the fullerene (5,0) nanotube C_{10+10k}-D_{5d} with,*

$$N_{\mathrm{HC}}(N) := 20\left[2^{N/10-3} + 12^{N/20-1}\right] \tag{5.5}$$

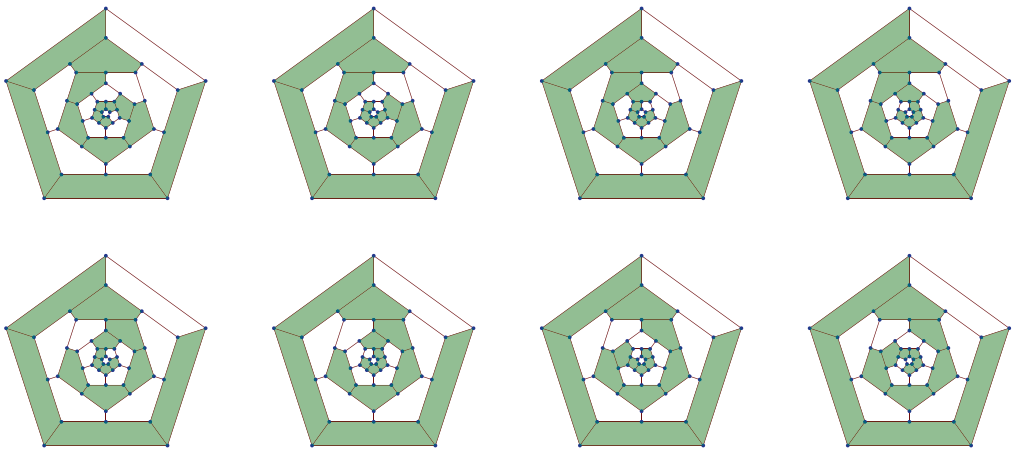


Figure 5.6 The eight Hamilton paths (up to 10-fold D_{5h} symmetry) in D_{5h}-fullerene nanotube C_{50} ($N_{HC} = 80$).

and $N = 10 + 10k \geq 20$, or a possible D_3/D_{3h} fullerene with a trigonal structure and four fused pentagons at each of the three corners, admits the highest number of Hamilton cycles.

Unlike the C_{10+10k}-D_{5d} nanotubes, where we could deduct an analytical formula for the Hamilton cycle count, for the D_3/D_{3h} trigonal structures an analytical formula was not found and may not exist, but requires further investigation. C_{104}-D_3 and C_{132}-D_{3h} shown in Figure 5.11 present two early counterexamples to the assumption that the formula for the Hamilton cycle count given by the C_{20k}-D_{5d} nanotubes present a proper upper bound for fullerenes, with the first counterexample for the C_{20k} nanotube series occurring at $k = 7$ (see figure 5.11). For example, C_{104}-D_3 admits 686,286 ($= 2 \times 3^3 \times 71 \times 179$) Hamilton cycles compared to 685,074 from the formula given in theorem 2, C_{132}-D_{3h} admits 23,932,076=($2^2 \times 7 \times 104,717$) Hamilton cycles compared to 22,126,207 of theorem 2, and C_{140}-D_3 admits 65,031,243 ($3 \times 19 \times 71 \times 16,069$) Hamilton cycles compared to the C_{20k}-D_{5d} nanotube of 59,760,640 (the prime factor decomposition is given in parentheses). The prime factor decomposition shows that the Hamilton cycle counts for the

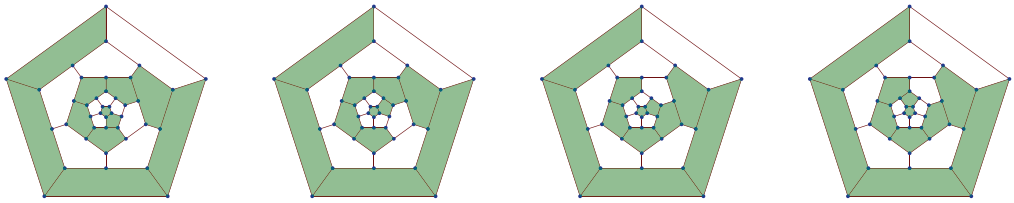


Figure 5.7 The four Hamilton paths of type A (up to 10-fold rotational symmetry and inversion) in the D_{5d}-fullerene nanotube C_{40} ($N^A_{HC} = 40$).

D_3/D_{3h} fullerenes do not have common divisors, and it is very likely that an analytical formula for this class of fullerenes does not exist.

A survey of all C_N-D_3/D_{3h} fullerenes ($N \leq 120$) with high N_{HC} values revealed that they all have the same pentagon neighbor signature of $P_1 = 15$ (with pentagon indices of $p_k = (0,0,6,6,0,0)$) having three sets of four fused pentagons, The D_3/D_{3h} fullerenes with pentagon neighbor signature of $P_1 = 15$ are collected in Table 5.4. A subsequent survey through all fullerene graphs up to $N = 180$ using the House of Graphs database[35] gave in total 18 C_N-D_3/D_{3h} fullerenes with high N_{HC} counts. Out of this set, we have higher Hamilton cycle counts compared to the formula given by the C_{20k}-D_{5d} nanotubes for five fullerenes, which are presented in Figure 5.11. Moreover, four out of the five exceptions have no face spiral pentagon start as the FSPIs in Table 5.4 show.[b] Such fullerenes have been investigated before by Yoshida and Fowler in terms of their pentagon spirals.[355] The upper bound fullerenes between vertex numbers 110 and 118 are shown in Figure 5.12. What is evident is that all of them tend toward structures which separate the pentagons in three sets.

Interestingly, similar to the D_{5h} fullerene nanotubes, there is a set of D_{3h}/D_3 fullerenes with a rather low Hamilton cycle count as well as evident from Figure 5.4. Analyzing the list of upper and lower bound fullerenes in more detail, we see that they all have high pentagon neighbor signatures P_1, thus fusion of pentagons seem to be advantageous for both lower and upper

[b] The C_{168}-D_3 fullerene with no pentagon spiral start has been studied before. The Halma-transform of this fullerene leads to the non-spiral C_{672}-D_3 fullerene.[96]

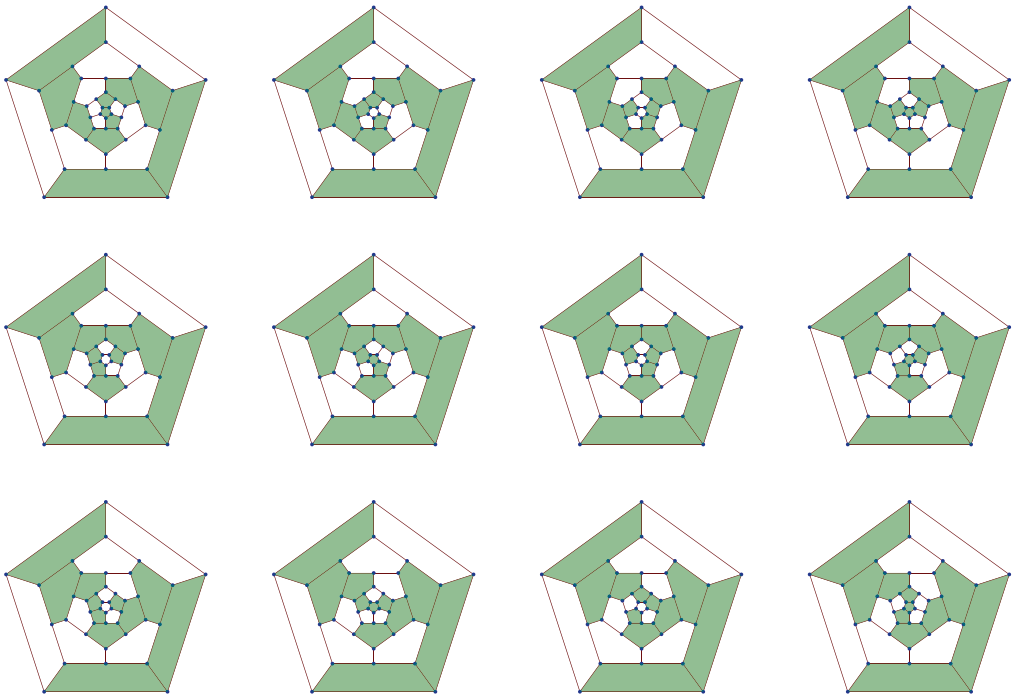


Figure 5.8 The 12 Hamilton paths of type B (up to 10-fold rotational symmetry and inversion) in the D_{5d}-fullerene nanotube C_{40} ($N_{HC}^{B} = 120$).

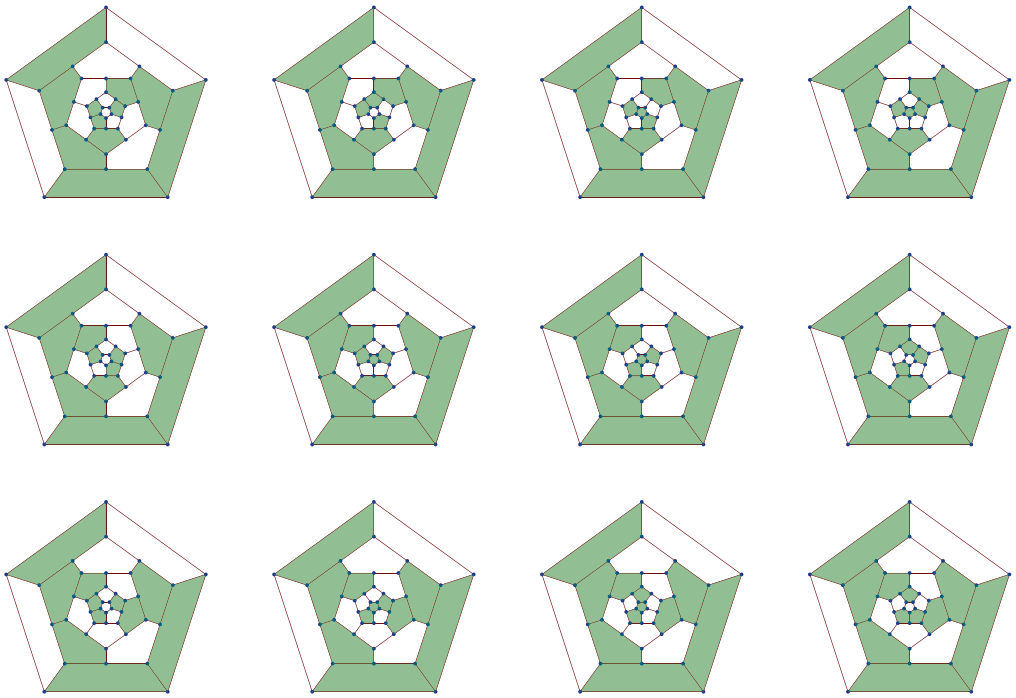


Figure 5.9 The 12 Hamilton paths of type C (up to 10-fold rotational symmetry and inversion) in the D_{5d}-fullerene nanotube C_{40} ($N_{HC}^{C} = 120$).

N	$N_{\mathrm{HC}}^{\mathrm{LB}}$	PG	RSPI	$N_{\mathrm{HC}}^{\mathrm{UB}}$	PG	RSPI
60	1090	I_h	{1,7,9,11,13,15,18,20,22,24,26,32}	1090	I_h	{1,7,9,11,13,15,18,20,22,24,26,32}
70	2790	D_{5h}	{1,7,9,11,13,15,27,29,31,33,35,37}	2790	D_{5h}	{1,7,9,11,13,15,27,29,31,33,35,37}
72	3852	D_{6d}	{1,7,9,11,13,18,22,24,27,34,36,38}	3852	D_{6d}	{1,7,9,11,13,18,22,24,27,34,36,38}
74	4794	D_{3h}	{1,7,9,11,14,23,26,28,30,32,35,38}	4794	D_{3h}	{1,7,9,11,14,23,26,28,30,32,35,38}
76	6078	T_d	{1,7,9,12,14,21,26,28,30,33,35,38}	6643	D_2	{1,7,9,11,13,18,26,31,33,35,37,39}
78	6988	D_3	{1,7,9,11,13,20,25,28,32,34,36,38}	8244	D_{3h}	{1,7,9,11,15,18,22,25,33,37,39,41}
80	9004	C_{2v}	{1,7,9,12,14,20,26,28,32,34,39,42}	10970	D_{5h}	{1,7,10,12,14,19,26,28,32,34,39,42}
82	11226	C_{3v}	{1,7,9,12,14,20,27,32,34,36,38,40}	13614	C_s	{1,7,9,11,13,24,27,29,31,33,36,43}
84	14748	C_s	{1,7,9,12,14,20,27,29,33,35,40,43}	18260	D_{6h}	{1,7,10,13,19,22,25,28,30,34,37,44}
86	17853	D_3	{1,7,10,12,14,19,28,33,35,37,40,45}	21756	C_s	{1,7,9,12,20,24,26,28,34,39,41,44}
88	22661	C_1	{1,7,9,12,14,20,28,34,36,40,42,45}	28652	D_2	{1,7,9,11,13,18,33,35,38,40,42,44}
90	29277	C_2	{1,7,9,12,14,20,28,34,36,39,41,47}	36852	C_{2v}	{1,7,10,13,19,25,28,33,35,38,42,47}
92	36949	C_2	{1,7,9,11,22,24,28,30,37,42,44,47}	47054	C_2	{1,7,9,11,13,24,30,37,40,42,44,46}
94	44730	C_2	{1,7,10,12,14,19,35,37,39,41,43,49}	59118	C_S	{1,7,9,11,14,29,31,33,36,39,44,46}
96	60070	C_1	{1,7,9,12,20,25,27,33,40,45,47,50}	78044	D_{6h}	{1,7,10,14,19,27,33,35,37,43,48,50}
98	71950	C_s	{1,7,12,14,19,22,29,36,42,44,47,49}	95694	C_s	{1,7,9,11,13,29,33,36,38,42,45,48}
100	93986	C_1	{1,7,9,12,25,28,30,34,36,43,49,52}	131690	D_{5d}	{1,7,9,11,13,15,38,40,42,44,46,52}
102	116772	C_{2v}	{1,7,10,12,14,30,34,38,40,43,47,50}	161148	C_s	{1,7,9,13,20,28,34,36,42,47,49,52}
104	149920	C_s	{1,7,9,11,22,26,30,36,43,48,50,53}	207165	C_2	{1,7,10,12,24,27,31,40,44,46,49,52}
106	180243	C_2	{1,7,10,12,25,27,31,34,41,47,52,55}	257746	C_1	{1,7,9,11,23,27,37,41,43,47,53,55}
108	237580	C_s	{1,7,9 11,14,34,38,40,44,52,54,56}	351976	D_2	{1,8,10,13,26,29,31,37,46,48,52,56}
110	244254	D_3	{1,8,10,12,25,27,29,43,45,47,50,57}	426750	D_{5h}	{1,7,9,11,13,15,47,49,51,53,55,57}
112	383218	C_1	{1,7,9,20,23,25,40,42,44,47,49,58}	571622	D_2	{1,7,14,20,22,25,31,41,46,48,51,54}
114	699908	D_3	{1,7,9,11,25,39,44,46,48,51,53,56}	699908	D_3	{1,8,10,16,23,25,42,44,47,49,52,59}
116	630059	C_2	{1,7,9,15,32,34,37,39,44,53,57,60}	1013844	T_h	{1,7,20,23,25,28,32,35,41,44,50,60}
118	723505	C_2	{1,8,10,12,14,28,43,45,48,50,52,61}	1151918	C_s	{1,7,9,11,24,26,42,44,46,52,57,59}
120	1038971	C_1	{1,7,9,12,25,28,44,46,48,52,54,62}	1590875	D_5	{1,7,20,22,25,28,32,43,48,50,53,62}
122	919626	D_3	{1,8,10,23,25,28,30,45,47,52,59,62}	1888558	C_2	{1,7,9,11,25,34,40,44,49,54,59,62}

Table 5.3 Exact lower ($N_{\mathrm{HC}}^{\mathrm{LB}}$) and upper bounds ($N_{\mathrm{HC}}^{\mathrm{UB}}$) for the number of distinct Hamilton cycles (HC) in IPR fullerenes up to C_{122}. For definitions see Table 5.1.

bounds. Figure 5.13 shows that the lower bound structures for the Hamilton cycles of the larger fullerenes have tubular shapes. This perhaps may be the reason why the lower bound is most likely be determined by the Hamilton cycle count of the C_{10+10k}-D_{5h} (5,0) fullerene nanotubes. We note, however, that the number of Hamilton cycles in IPR fullerenes are at the higher end as Figure 5.10 shows.

Let us analyze the Hamilton cycle upper bound for fullerenes in more detail. From the five D_3/D_{3h} fullerenes with high Hamilton cycle counts we derive an approximate Eppstein type upper bound,

$$N_{\mathrm{HC}}^{\mathrm{UB}}(\mathrm{C}_N) < 2^{\frac{N}{5.40}+1.22} \tag{5.6}$$

This is also an upper bound for all fullerenes presented here (see Figure 5.11).

N	N_{HC}	PG	RSPI
32	54	D_3	{1,2,3,5,7,9,10,12,14,16,17,18}
36	124	D_{3h}	{1,2,3,6,10,11,12,14,15,16,18,20}
38	33	D_3	{1,2,3,4,11,12,14,15,16,18,19,21}
44	303	D_3	{1,2,3,4,11,14,15,18,19,21,22,24}
54	66	D_{3h}	{1,2,3,4,14,18,19,22,23,26,27,29}
56	1746	D_3	{1,2,6,7,17,18,21,22,23,26,27,30}
60	2551	D_3	{1,2,3,4,19,20,23,24,25,28,29,32}
68	6018	D_3	{1,2,3,4,19,23,24,28,29,32,33,36}
78	1596	D_{3h}	{1,4,5,12,25,2,30,31,32,36,37,41}
80	25704	D_{3h}	{1,2,3,4,23,28,29,33,34,38,39,42}
80	31380	D_3	{1,3,4,5,23,24,33,34,38,39,40,42}
86	1980	D_3	{1,2,3,4,29,30,34,35,36,40,41,45}
96	209692	D_3	{1,2,3,4,29,34,35,40,41,45,46,50}
104	686286	D_3	{2,8,9,19,35,36,41,42,43,48,49,54}
108	1104223	D_3	{1,3,4,5,34,35,46,47,52,53,54,56}
110	22188	D_{3h}	{1,2,3,4,34,40,41,46,47,52,53,57}
116	2874489	D_3	{1,2,3,4,41,42,47,48,49,54,55,60}
128	12569388	D_3	{1,2,3,4,41,47,48,54,55,60,61,66}
132	23932076	D_{3h}	{14,15,29,30,36,37,51,52,58,59,66,67}
134	545823	D_3	{8,9,20,21,41,42,56,57,63,64,65,69}
140	65031243	D_3	{1,3,4,5,47,48,61,62,68,69,70,72}
144	97034672	D_{3h}	{1,2,3,4,47,54,55,61,62,68,69,74}
150	1870728	D_3	{1,2,3,4,55,56,62,63,64,70,71,77}
164	1475673552	D_3	{14,15,29,30,52,53,68,69,76,77,78,84}
168	2429628676	D_3	{4,5,13,14,58,59,74,75,82,83,84,86}

Table 5.4 The class of fullerenes with D_3 or D_{3h} symmetry and pentagon neighbor signature $P_1 = 15$ having 3 sets of 4 fused pentagons at the corners of the polyhedron up to $N = 180$. For definitions see Table 5.1.

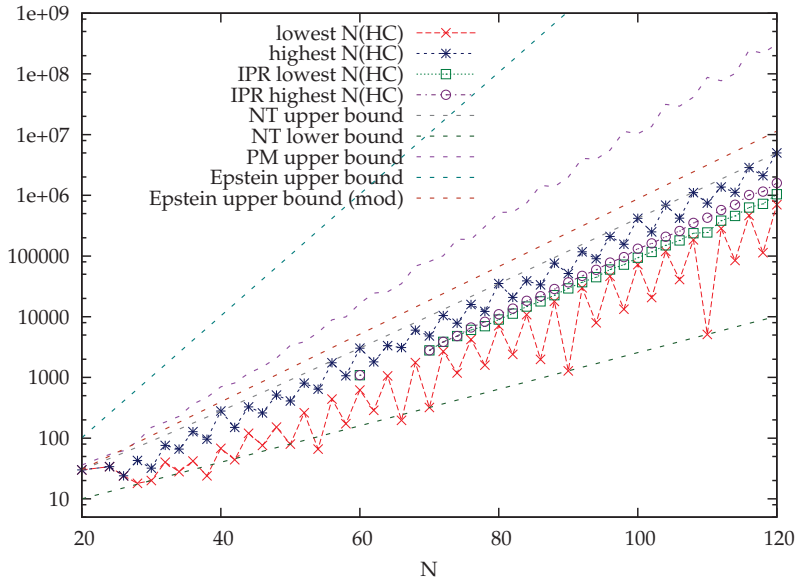


Figure 5.10 Upper and lower bounds for Hamilton cycles in general and IPR fullerenes (shown on a logarithmic scale) as a function of the number of vertices N. Also shown are the Eppstein upper bound ($2^{N/3}$) for cubic graphs, the perfect matching (PM) upper bound, and the modified Eppstein upper bound for fullerene graphs.

We are confident that this is a good upper bound, which demonstrates that the Eppstein upper bound for cubic graphs can be considerably improved for fullerene graphs.

The funnel shape of the lower and upper bound area in Figure 5.10 is easily explained, as the distance between the upper and lower bound increases exponentially with increasing number of vertices,

$$\frac{N_{\mathrm{HC}}^{\mathrm{UB}}(N)}{N_{\mathrm{HC}}^{\mathrm{LB}}(N)} = a_N = 2\left(3^{N/20-1}\right) + 1 \tag{5.7}$$

An interesting feature of the fullerenes is the alternating behavior of the lower and upper bounds for the Hamilton cycles. Small (large) values of lower bounds imply small (large) values of upper bounds in N_{HC}. This is obviously a topological feature of fullerenes, which requires a more detailed

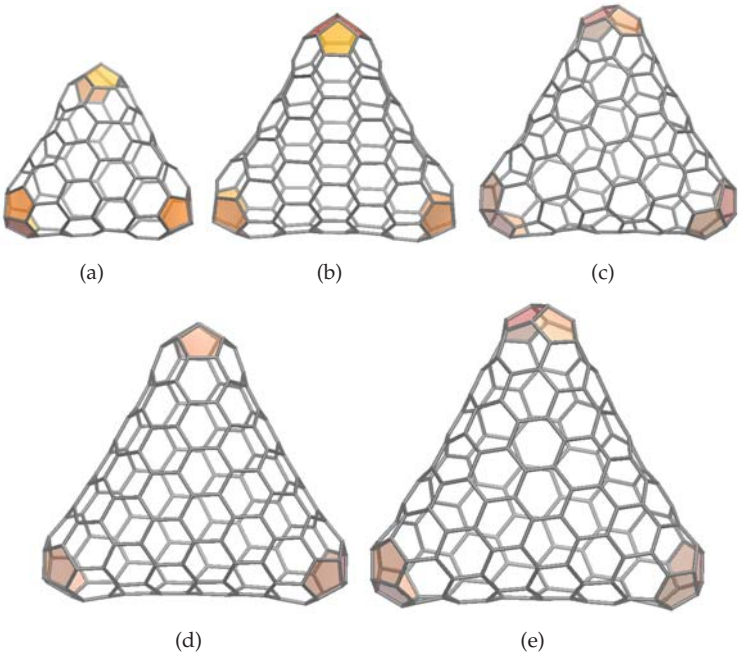


Figure 5.11 The five D_3/D_{3h}-C_N fullerene polyhedra with larger Hamilton cycle count compared to the formula given by the D_{5d} (see theorem 2) nanotubes up to $N = 180$. (a) D_3, $N = 104$, $N_{\mathrm{HC}} = 686,286$; (b) D_{3h}, $N = 132$, $N_{\mathrm{HC}} = 23,932,076$; (c) D_3, $N = 140$, $N_{\mathrm{HC}} = 65,031,243$; (d) D_3, $N = 164$, $N_{\mathrm{HC}} = 1,475,673,552$; and (e) D_3, $N = 168$, $N_{\mathrm{HC}} = 2,429,628,676$.

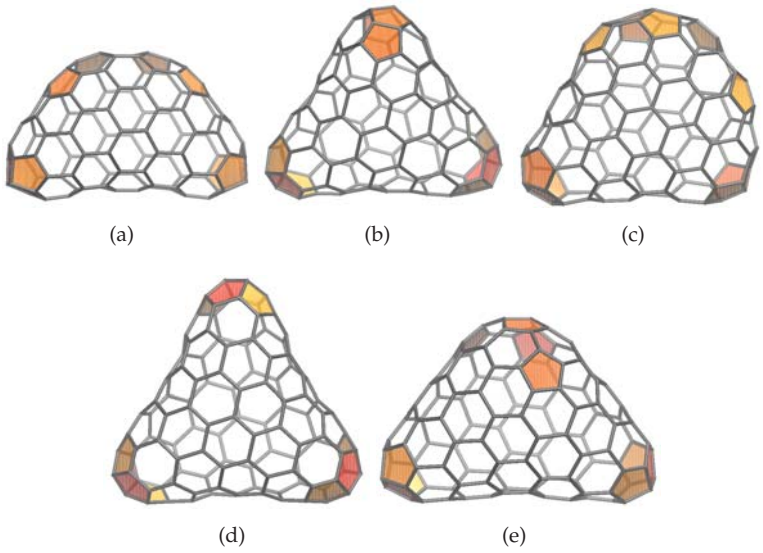


Figure 5.12 The fullerene polyhedra with largest Hamilton cycle count for $N = 110 - 118$. (a) C_{110}-C_s; (b) C_{112}-C_2; (c) C_{114}-C_1; (d) C_{116}-D_3; (e) C_{118}-C_1.

analysis. Nevertheless, we can ask the question how the number of Hamilton cycles is distributed over the fullerene isomers with fixed vertex number N, i.e., are there more fullerene isomers with small N_{HC} count or more with high N_{HC} count?

The relative frequency distributions (histograms) for some of the larger fullerenes are shown in Figure 5.14. We see that for fullerenes with $N = (4n)$ vertices the distributions are positive skewed towards higher Hamilton cycle counts, while for fullerenes with $N = (4n + 2)$ vertices the distributions are slightly left skewed towards lower Hamilton cycle counts in line with the alternating behavior for the upper and lower bounds shown in Figure 5.10. Moreover, the distributions broaden with increasing vertex numbers in line with the funnel shape for the upper and lower bounds in Figure 5.10.

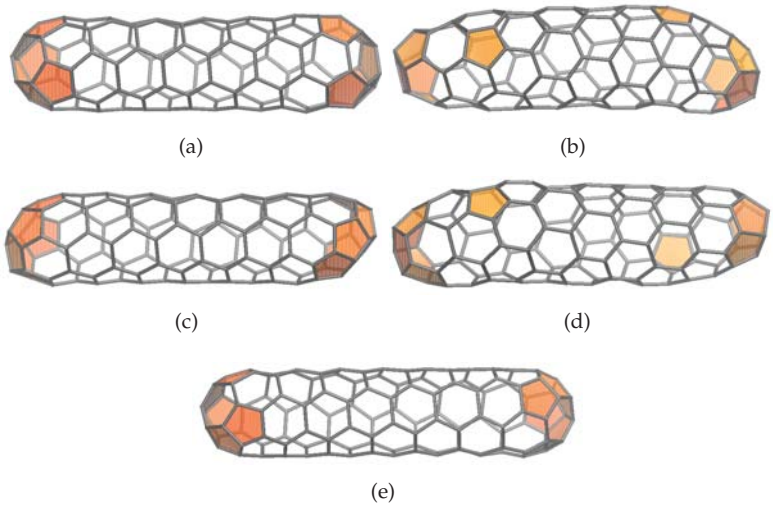


Figure 5.13 The fullerene polyhedra with smallest Hamilton cycle count for $N = 112 - 120$. (a) C_{112}-C_2; (b) C_{114}-C_2; (c) C_{116}-C_2; (d) C_{118}-C_2; (e) C_{120}-C_2.

5.4 Appendix A: The number of Hamilton cycles in C_{10+10k}-D_{5+} (5,0) fullerene nanotubes

Lemma 1. *There is a bijective mapping between all Hamilton cycles in planar graphs and a set of bi-colorings of faces in their planar layouts (up to swapping the two colors). This set comprises all bi-colorings in which a) every vertex is adjacent to at least one shaded and one unshaded face and b) all shaded and all unshaded faces are connected.*

Proof. According to Jordans curve theorem, a closed curve that lies in a surface with genus 0 divides this surface into two connected parts. Conversely dividing this surface into two parts which are both connected defines a closed curve.

A planar graph can be mapped on a surface with genus 0. Hence, choosing a planar layout of a planar graph and shading in some faces (as

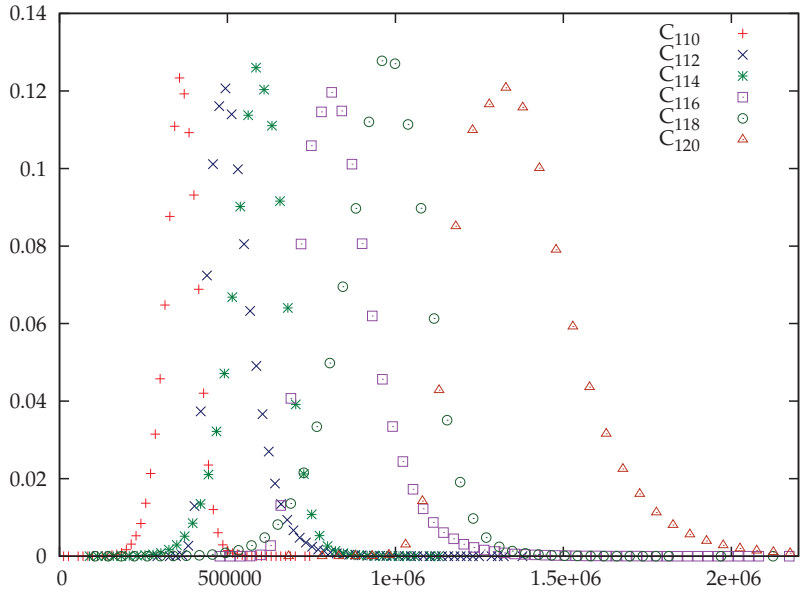


Figure 5.14 Relative frequency distributions for N_{HC} for all fullerenes isomers with a fixed number of vertices ($N = 110 - 120$). For the histograms we chose a width of $\sim 15,000$ ($N = 110$) – $50,000$ ($N = 120$) in N_{HC}.

a means of denoting the two surface parts) such that all shaded and all unshaded faces are connected, defines a single closed curve that follows edges of the graph. If the sets of the shaded or the unshaded faces are disconnected, the border is formed not by one but by several closed curves. The border between the two surface parts does not depend on the choice of which part is shaded and which is not.

Every vertex that is adjacent to at least one shaded and one unshaded face lies on the described closed curve. Hence, in order for all vertices to lie one curve, every vertex needs to by adjacent to at least one of both face types.

That is, in order for the boundary curve to be Hamiltonian, every vertex needs to be adjacent to both face colors and all shaded and all unshaded faces need to be connected. □

Lemma 2. *For shading in a layered graph according to the rules defined in Lemma 1, the legal colorings of each layers depend only on the previous layer and whether the shaded and unshaded faces on the previous layer are connected from the outside, that is, whether in the subgraph that consists of all previous layers all shaded and all unshaded areas are connected or not.*

Proof. As explained in Lemma 1, for a shading of faces in the planar layout of a graph to correspond to a Hamilton cycle, it is required that all shaded and all unshaded faces are connected respectively after all faces are shaded in.

While shading in the graph layer-wise, the sets of shaded or unshaded faces may be disconnected. However, in order to connect them by a consecutive layer, each disconnected group of faces must be adjacent to the central hole where new layers are inserted. That implies, that each layer must comprise shaded and unshaded faces. In order to fully characterize a layer the information of whether all its shaded and unshaded faces are connected by previous layers is required.

Lemma 1 requires that each vertex is adjacent to at least one face of both colors. As all vertices in a layered graph lie on the boundary between one layer and the next, the adjacency of each vertex is fully determined by the two layers and their orientation and does not depend on any other layers. □

In the two types of layered graphs that will be examined, not each layer type is a valid last layer. Following the last full layer there is a single last face that can either be shaded or unshaded. Not for every layer, the requirement of Lemma 1 are fulfilled after terminating with a single face.

We can interpret each possible shading of a single layer together with the information of whether or not its faces are connected from the outside as a vertex in a directed multigraph M. The edges in M are then the options of layers that are accessible from a given layer. If one layer type can be followed by another layer in more than one way, these two vertices are connected by multiple directed edges.

The graph M can be represented by its adjacency matrix $A(H)$, which in case of directed multigraphs may have diagonal entries and needs not be symmetric. Entries A_{ij} in adjacency matrices contain the number of ways to get from one vertex i to a given other vertex j in a single step. If an adjacency

matrix is multiplied with itself k times, the entries $(A^k)_{ij}$ hold information of how many walks of length k there are from vertex i to vertex j.[356]

We can therefore build the adjacency matrix of the graph that describes which layer types may follow each other, raise it to the k^{th} power, and in the row that refers to the starting point, sum over all columns that refer to layers that can be the last layer. This sum is the number of Hamilton cycles that originates from the chosen starting point.

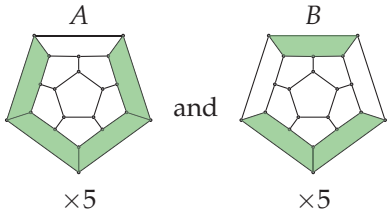
The two different D_{5+} nanotubes are very similar topologically, yet they yield considerably different counts for Hamilton cycles with respect to the number k. Here we prove the following theorem:

Theorem 4. *The D_{5+} fullerene nanotube C_{10+10k} ($k \in \mathbb{N}$) admits exactly*

$$N_{\text{HC}}(k) = 5\left[2^k + 4 \cdot 12^{\frac{k-1}{2}}\chi(k)\right] \tag{5.8}$$

Hamilton cycles with $\chi(k) = 1$ if k is odd and $\chi(k) = 0$ otherwise.

Proof. Consider a C_{10+10k}-D_{5+} fullerene nanotube. Up to five-fold rotation, there are two inequivalent ways to shade the faces in the outer ring. Without loss of generality we choose not to shade the outer face.

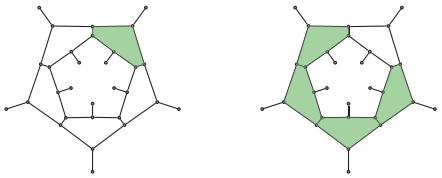


(5.9)

All other choices yield either vertices that are not adjacent to a shaded face or the unshaded faces are separated into two disconnected areas (by shading in all five faces of the outer ring). We will show that the appearance of case B type of Hamilton cycles is responsible for the additional Hamilton cycle count in the even compared to the odd k fullerenes. We now consider the two different cases A and B separately for the possible shadings in the subsequent face layers.

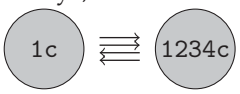
Consider case A (which is denoted as `1234c` in the syntax that is used in the following) as a shading of the first layer. One can easily show that only

two layer types need to be considered for shading of all remaining layers. We denote these two options in a descriptive way as `1c` and `1234c`. The numbers denote the shaded faces in a clockwise or anti-clockwise order (whichever yields smaller sequence of numbers). The `c` and `u` postfix indicates whether the shaded faces of the layer are connected to each other through shaded faces from the previous layers or not.



(5.10)

As `1c` as well as `1234c` can be succeeded by the respective other layer type in exactly two different ways, we obtain the following multigraph:

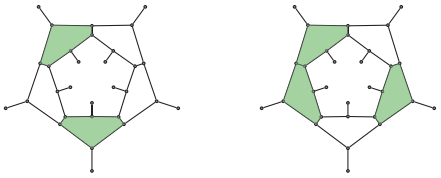


which can be represented by the adjacency matrix $\mathbf{S}$:

$$\mathbf{S} = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}. \tag{5.11}$$

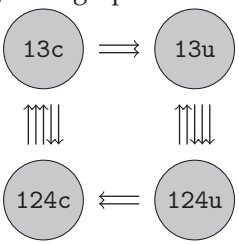
As the starting layer in A is `1234c` and both layers are allowed to be the last layer, the number of Hamilton cycles of C_{10+10k} that originate from A is five times the sum of the upper two elements of the matrix $\mathbf{S}$ raised to the k^{th} power, i.e., $N_{\text{HC}}^{(A)}(k) = 5 \times ((\mathbf{S}^k)_{1,1} + (\mathbf{S}^k)_{1,2})$. Due to the simple structure of the matrix, this is 5×2^k.

If we consider case B as a shading of the first layer, there are four layer types to be considered for shading in the rest of the graph: `13c`, `13u`, `124c`, and `124u`. `124u` is the starting point (the shaded faces are not connected from the outside) and only `13u` or `124c` can be the last layer, with shaded or unshaded last central pentagon face, respectively. In both other cases, shading the central last face uniformly would either yield multiple cycles by disconnecting the shaded or unshaded area, or exclude one vertex from the cycle (making it non-Hamiltonian).



(5.12)

We obtain the following multigraph:



which translates to the adjacency matrix $\mathbf{T}$ (the order of rows and columns corresponds to graphs shown in (5.12)):

$$\mathbf{T} = \begin{pmatrix} 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 3 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 \end{pmatrix}. \tag{5.13}$$

Here the entry point is 124u, i.e., row 4 and the two possible exit points are 13u and 124c, i.e., columns 2 and 3. Five times the sum of the second and third entry in the fourth row of the k^{th} power of the matrix is the number of Hamilton cycles in C_{10+10k} that start from case B:

$$N_{\text{HC}}^{(B)}(k) = 5 \times ((\mathbf{T}^k)_{4,2} + (\mathbf{T}^k)_{4,3}) \tag{5.14}$$

To get a better understanding for the alternating behavior of the sum, we give the second to fourth power of $\mathbf{T}$:

$$\mathbf{T}^2 = \begin{pmatrix} 6 & 0 & 0 & 6 \\ 0 & 6 & 6 & 0 \\ 0 & 6 & 6 & 0 \\ 6 & 0 & 0 & 6 \end{pmatrix}, \mathbf{T}^3 = \begin{pmatrix} 0 & 24 & 24 & 0 \\ 18 & 0 & 0 & 18 \\ 18 & 0 & 0 & 18 \\ 0 & 24 & 24 & 0 \end{pmatrix}, \mathbf{T}^4 = \begin{pmatrix} 72 & 0 & 0 & 72 \\ 0 & 72 & 72 & 0 \\ 0 & 72 & 72 & 0 \\ 72 & 0 & 0 & 72 \end{pmatrix} \tag{5.15}$$

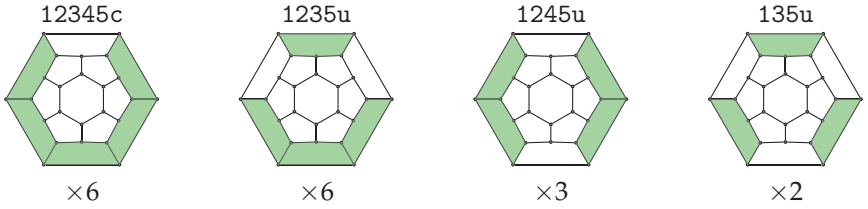
As can be seen from the matrix and its first four powers, $(\mathbf{T}^k)_{4,2}$ and $(\mathbf{T}^k)_{4,3}$ are always equal and alternate between 0 for even k and $2 \cdot 12^{\frac{k-1}{2}}$ each for odd k. The factor of 12 between k and $k+2$ originates from the multiplication of both relevant matrix elements by 3 and 2×2 in alternating steps.

We conclude this proof by summarizing that the overall sum of all Hamilton cycles in D_{5+} (5,0) C_{10+10k} nanotubes is the sum of those starting from cases A and B. As shown above, case A leads to 5×2^k and case B leads to $5 \times 4 \cdot 12^{\frac{k-1}{2}}$ Hamilton cycles for odd k and 0 otherwise. We combine the two cases by defining $\chi(k) = 1$ for odd k and $\chi(k) = 0$ for even k and derive eq.(5.8). □

5.5 Appendix B: The number of Hamilton cycles in C_{12+12k}-D_{6+} (6,0) fullerene nanotubes

We will devise a method for calculating the number of Hamilton cycles in D_{6+} (6,0) fullerene nanotubes in analogy to the above recipe for counting Hamilton cycles in D_{5+} (5,0) fullerene nanotubes.

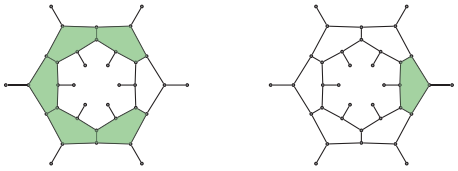
There are four different options for shading in the first layer of the graph, up to sixfold, sixfold, threefold and twofold rotational symmetry respectively (taking C_{24} as a representative example):



(5.16)

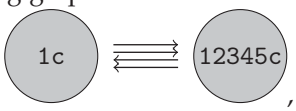
Starting from the first case (12345c) we need to consider only two different layer shadings (in analogy to case A of the previous proof), namely

12345c and 1c:



(5.17)

Both layers may be the last layer and which may follow each other according to the following graph:



,

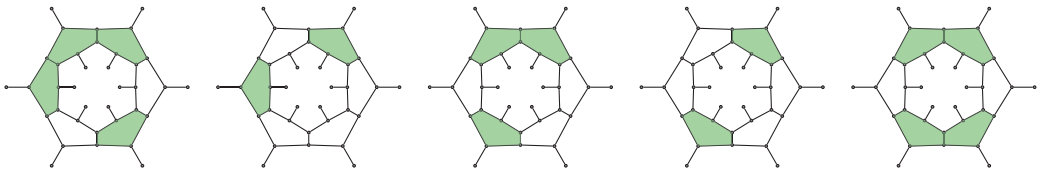
which can be written as the adjacency matrix **U**:

$$\mathbf{U} = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \tag{5.18}$$

The number of Hamilton cycles starting from 12345c is therefore

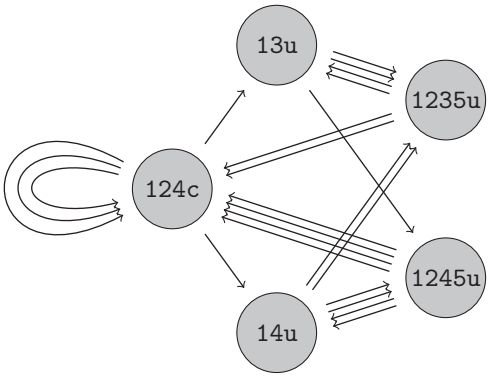
$$N_{\mathrm{HC}}^{(12345c)}(k) = 6 \times ((\mathbf{U}^k)_{1,1} + (\mathbf{U}^k)_{1,2}) = 6 \times 2^k. \tag{5.19}$$

Cases 1235u and 1245u can be treated together because starting from either leads to the same options while shading in the rest of the graph. We need to consider 1235u, 13u, 124c, 14u and 1245u. Only the options 13u and 14u can be the last layer.



(5.20)

We obtain the following graph:



which has the adjacency matrix $\mathbf{V}$:

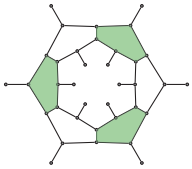
$$\mathbf{V} = \begin{pmatrix} 0 & 2 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 1 & 0 \\ 2 & 0 & 0 & 0 & 2 \\ 0 & 0 & 4 & 2 & 0 \end{pmatrix}. \tag{5.21}$$

Here, the starting points are rows 1 and 5 and the possible final layers are columns 2 and 4. Keeping in mind the degeneracies of the first layer, we get

$$N_{\mathrm{HC}}^{(1235\mathrm{u})}(k) = 6 \times ((\mathbf{V}^k)_{1,2} + (\mathbf{V}^k)_{1,4}) \quad \text{and} \tag{5.22}$$

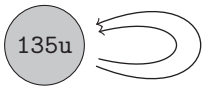
$$N_{\mathrm{HC}}^{(1245\mathrm{u})}(k) = 3 \times ((\mathbf{V}^k)_{5,2} + (\mathbf{V}^k)_{5,4}). \tag{5.23}$$

Shading in the first layer according to 135u, only allows for shading in *all* following layers in the same way.



135u (5.24)

However, there are two possible orientations for each layer.



We skip the adjacency matrix and conclude that there are $N_{\mathrm{HC}}^{(\mathtt{135u})}(k) = 2 \times 2^k$ Hamilton cycles starting from 135u.

In total the number N_{HC} of Hamilton cycles in C_{12+12k} (6,0) nanotubes is given as

$$N_{\mathrm{HC}}(k) = N_{\mathrm{HC}}^{(\mathtt{12345c})}(k) + N_{\mathrm{HC}}^{(\mathtt{1235u})}(k) + N_{\mathrm{HC}}^{(\mathtt{1245u})}(k) + N_{\mathrm{HC}}^{(\mathtt{135u})}(k) \tag{5.25}$$

$$= 6 \times 2^k + 6 \times ((\mathbf{V}^k)_{1,2} + (\mathbf{V}^k)_{1,4}) + 3 \times ((\mathbf{V}^k)_{5,2} + (\mathbf{V}^k)_{5,4}) + 2 \times 2^k, \tag{5.26}$$

and only a matrix version or recursive formula for the number of Hamilton cycles can be derived due to the more complicated structure of matrix $\mathbf{V}$.

6 Gaudienes[r]

6.1 Introduction

Graphene is a hexagonal two-dimensional (2D) lattice connected by trivalent vertices representing sp^2 carbon atoms. This simple network can be modified by introducing different atoms or more complex molecular units for the vertices, or/and by edge extensions and modifications such as adding extra atoms or replacing them by molecular units with ingoing and outgoing bonds lying on straight lines.[358–360] Alternatively, one could also work with its dual, i.e., a triangulated planar surface of hexagonal symmetry containing vertices of degree six, and modify its topology accordingly. There are many different combinations and alterations possible to form such 2D molecular networks, which remain mostly unexplored and could have many interesting electronic properties and applications.[359,360]

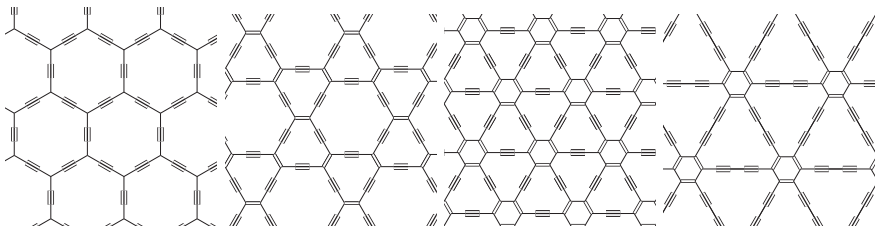


Figure 6.1 The molecular structure of (a) α-graphyne, (b) β-graphyne, (c) γ-graphyne and (d) γ-graphdiyne.

[r]The following sections 6.1 to 6.5 are being published as "Novel hollow all-carbon structures" by Sundholm et al.[357]. The candidate's contribution to the article is itemised on pages 299 ff.

Graphyne shown in Figure 6.1 represents such a class of 2D all-carbon allotropes,[280,361] that is either built from a graphene template or its dual, and some building blocks have already been synthesized by organic chemists.[362–368] Graphynes are one carbon atom thick and consist of sp and sp^2 hybridized carbons in contrast to graphene whose 2D structure consists of six-membered rings of sp^2 hybridized carbons. The α-graphyne structure is formally obtained by inserting –C≡C– units into every carbon-carbon bond of graphene leading to a 2D carbon structure consisting of large hexagonal rings with frustrated three-coordinated carbons in each corner of the 18-membered hexagonal carbon rings. This bond assignment reflects well the symmetry of the structure and approximate bond lengths, however, configurations with only paired electrons are likely to contribute significantly to the electronic structure. The β-graphyne structure consists of similar hexagonal rings as in α-graphyne. However, each of the hexagonal rings is surrounded by six hexadehydro[12]annulene rings yielding an infinite 2D structure with all carbons formally four-valent. The γ-graphyne structure also shown in Figure 6.1 consists of an infinite 2D structure with the six-membered hexagonal carbon (benzoic) rings surrounded by six hexadehydro[12]annulene rings having all carbons four-valent as in β-graphyne. It is easily seen that this structure can be obtained by replacing certain edges in graphene by acetylene units (in graph theoretical terms this is called an edge subdivision), or alternatively by replacing in the dual triangulated network all vertices by benzene units and edges between them by –C≡C– groups. Graphdiynes are derived from the corresponding graphyne structures by replacing the –C≡C– groups with –C≡C–C≡C– units. Graphdiyne films have been synthesized on copper surfaces.[369] The graphyne and graphdiyne structures are shown in Figure 6.1.

The graphene structure consisting of an infinite 2D layer of six-membered rings can be wrapped onto a sphere to give cavernous fullerene-like structures,[370] but with the condition that e.g., 12 six-membered rings are replaced by five-membered ones such that Euler's polyhedral formula is fulfilled.[6] Carbon nanotubes are similarly obtained by rolling up graphene sheets onto a cylinder.[371] Fullerenes, carbon nanotubes, and other carbon-based nanostructures such as graphyne, graphdiyne, graphone, and graphane have been proposed as basic building blocks for a variety of interesting nanotechnological applications.[372–383] As for graphene, hol-

low graphyne and graphdiyne structures can be obtained by wrapping up graphyne sheets, with modifications of course such that Euler's polyhedral formula is fulfilled. Analogously, graphyne-based and graphdiyne-based carbon nanotubes can be obtained by rolling up the corresponding planar sheets. One of the smallest members of this class of molecules is the recently proposed gaudiene molecule, which is an all-carbon molecule consisting of 72 carbon atoms forming a hollow structure of O_h symmetry (Figure 6.2). Gaudiene can be constructed from a truncated octahedron with two thirds of the edges replaced by –C≡C– units.[384] Quantum chemical calculations showed that β-C_{72} is an aromatic molecule with a rather large optical gap. Here, a novel class of hollow carbon structures is proposed. The class of molecules is named gaudienes, because the initial polyhedron was inspired from the work of the Spanish architect Antoni Gaudí.[384]

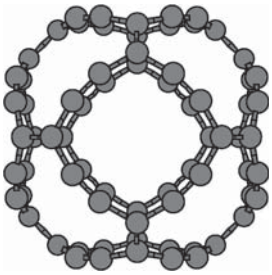


Figure 6.2 The molecular structure of β-C_{72} of O_h symmetry, the smallest β-gaudiene investigated.

6.2 Graph theoretical considerations

Before we start with the discussion of the different gaudienes we briefly analyze their topology and introduce a classification scheme. For the moment we do not distinguish between single, double or triple bonding in the polyhedral graph. Formally, gaudienes belong to the class of convex polyhedra (although a quantum theoretical treatment might end up with a locally non-convex structure).[1] The classification of (non-regular) polytopes is currently an open problem.[385] For the special class of fullerenes

a face-spiral classification scheme has been developed by Manolopoulos et al.,[32,281] which can be generalized for cubic polyhedra.[1,96] Here, we use the fact that simple insertions (deletions) of divalent vertices into edges (edge subdivisions) of a graph G results in a graph G' (which is homeomorphic to G) with a larger vertex set. Note that even though we have a homeomorphism between G and G', G is not a subgraph of G' as the vertex set E is not a subset of E'.

A polyhedral graph G satisfies Euler's polyhedral formula, and by using the handshaking lemma we obtain

$$N - E + F = \sum_{n=2} \left(1 - \frac{n}{2}\right) N_n + \sum_{n=3} F_n = 2 \tag{6.1}$$

where N is the number of vertices in the graph G, E the number of edges, F the number of faces, N_n the number of n-valent vertices, and F_n the number of n-gons. We see that for $n = 2$ the number of divalent vertices is exactly cancelled by the number of extra edges introduced into the graph. In all cases considered here the graph G is cubic, i.e., contain only 3-valent vertices, and Euler's polyhedral formula for the face count becomes[1]

$$3F_3 + 2F_4 + F_5 + \sum_{n\geq 7} (6 - n)F_n = 12 \tag{6.2}$$

Thus, we conclude that one cannot tile a sphere with hexagons only, i.e., one has to introduce extra n-gons (for example 12 pentagons for fullerene graphs; this however does not guarantee that a certain polyhedron exists as for example the fullerene C_{22} does not exist). Introducing extra n-gons can be done in many different ways as we shall see when we discuss the gaudienes in detail.

For the characterization of the various gaudienes we have to describe the cubic graph G or, if G is obtained by some other graph transformation T, the original graph G_0, i.e., $G = T(G_0)$. We list the vertices and faces according to Pederson[385] and describe the subsequent list of vertex insertions into edges of each n-gon separately. For G, or G_0 with no divalent vertices, we use the notation

$$\mathrm{PG}\,[N_3, N_4, N_5, N_6, \ldots; F_3, F_4, F_5, F_6, \ldots] \tag{6.3}$$

where PG is the ideal point group of the polyhedron.[1] For example, the well known fullerene I_h-C_{60} is characterized by

$$I_h\,[60;0,0,12,20] \tag{6.4}$$

This scheme does not uniquely classify the graph, as there are, for example, many non-isomorphic structures for a fullerene with a certain vertex count and symmetry,[1] but becomes useful for the gaudienes introduced here as we shall see.

α-graphynes can be derived from any polyhedral graph by inserting linear –C≡C– units into every bond (see Figure 6.1). This insertion is performed irrespective of the assignment of double bonds in the original structure. We denote the transformed version of a graph G_0 according to this scheme as $G = \alpha(G_0)$. For example, if we chose as the graph G_0 of an α-gaudiene a fullerene with N_0 vertices, denoted as PG$[N_0;0,0,12,(N_0/2-10)]$ and we insert two divalent vertices into all $3N_0/2$ edges, we get a total vertex count of $N_G = 4N_0$, where N_G is the number of vertices in the α-gaudiene. Such an edge subdivision conserves the point group of G_0.

It follows from Thurston's proof, which considers the number of non-isomorphic triangulations of a sphere with all N vertices of valency six or smaller, that the number of nonisomorphic cubic graphs with faces up to hexagons grows as $\mathcal{O}(N^9)$.[20] As every polyhedral graph can be transformed into an α-gaudiene, we can estimate the number of derived α-gaudienes.

β-graphyne analogue carbon cages can be generated starting with any polyhedral (i.e., three connected and planar) graph G_0. First we perform a leapfrog transformation (LF), which is a (1,1) Goldberg-Coxeter transformation.[1] The Goldberg-Coxeter transformation of a polyhedral graph is a polyhedral graph itself, implying that the initial graph can be subjected to any number of consecutive leapfrog transformations. In the obtained graph all faces with sizes $\neq 6$ and a subset of all hexagons are selected, such that each vertex is adjacent to exactly one selected face. Such a selection is guaranteed to exist and to be unique for every polyhedral graph that is obtained through a leapfrog transformation. The edges of all selected faces (i.e., two thirds of all edges), are then replaced by a linear segment with two vertices. The leapfrog transformation triples the number N_0 of vertices of a graph and introduces N_0 extra hexagons,[32] while the described

replacement triples the number of vertices again. Therefore, β-graphyne cages with $N_G = (3 \times 3)N_0$ vertices are accessible from polyhedra with N_0 vertices. We denote this sequence of one leapfrog transformation and the following edge subdivision of G_0 as $\beta(\mathrm{LF}(G_0))$. We denote k consecutive leapfrog transformations of a graph G_0 as $\mathrm{LF}^k(G_0)$.

Equivalently, β-graphynes can be created from the same original graph G_0 by first taking its dual G_0^*. Then, each n-valent vertex in the triangulation is replaced by an n-gon in which into each edge a –C≡C– unit has been inserted. For every pair of vertices connected by an edge in the original graph, the resulting two expanded polygons are connected by an edge.

In a similar fashion, γ-graphyne analogue carbon cages are generated starting with any polyhedral graph G_0. Transformation and selection of faces are performed in the same fashion as above. Then, all edges that are *not* adjacent to any selected face (one third of the edge set), i.e., the complement of the previously picked edges, are replaced by two additional vertices. This replacement scheme doubles the number of vertices, transforming polyhedral graphs with N_0 vertices into $N_G = (3 \times 2)N_0$ vertex cages. Analogously to the previous transformation this is denoted as $\gamma(\mathrm{LF}(G_0))$.

Alternatively, we can start with the dual graph G_0^*, and replace every n-valent vertex by an n-gon. These n-gons are then connected – according to the connectivity of the vertices they originated from – by bonds into which –C≡C– units have been inserted.

α-, β-, and γ-graphdiyne like cages are generated in the same way as the respective graphynes but by replacing edges by linear four-vertex segments instead of two-vertex segments. N_0 vertex polyhedral graphs are then transformed into $N_G = 7N_0$, $N_G = (3 \times 5)N_0$, and $N_G = (3 \times 3)N_0$ vertex cages respectively. Transformations of graphs G_0 to graphdiynes are written with a prepended 'C_4', e.g., C_4-$\alpha(G_0)$.

6.3 Computational methods

Computationally, all structures were generated with a modified version of the program *Fullerene*.[36] Selected small polyhedral graphs were hard coded. These graphs were subjected to leapfrog transformations as detailed in the

previous section. The generation of all the gaudienes considered here are summarized in Table 6.1.

As the force field in program *Fullerene* can only operate on cubic graphs and triangulations, polyhedra were generated at this stage and optimized to bond lengths that correspond to the polyhedron after replacing a subset of bonds by elongated linear segments. After the optimization of the Cartesian coordinates, atoms were inserted and placed in 3D by linear interpolation.

The molecular structures were then further optimized at the density functional theory (DFT) level using the Becke-Perdew generalized gradient approximation (GGA) functional (BP86) in combination with the Karlsruhe split-valence polarization (SVP) basis sets.[386–389] Since calculations at the BP86 level slightly underestimate the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), single-point calculations were performed at the DFT level using Becke's three-parameter functional (B3LYP) together with a SVP basis set.[284,285,389] The DFT calculations were performed with TURBOMOLE version 6.5.[287,390] The Cartesian coordinates of the atomic positions of all studied structures are given as Supporting Information.

6.4 Molecular structures

6.4.1 α-Gaudienes

The optimized structure of the smallest fullerene type α-graphyne based gaudiene (α-C_{80}) is shown in Figure 6.3. The structure belongs to the icosahedral (I_h) point group as does its origin, the dodecahedron represented by the smallest fullerene I_h-C_{20}. Hence, this gaudiene consists of 12 pentagonal 15-membered carbon rings forming a cage. The atoms shared by three rings are formally sp^3 hybridized carbons with only three neighbor carbons leading to a relatively small HOMO-LUMO gap of 0.740 eV at the BP86 level, which can be compared with the HOMO-LUMO gaps of 1.66 eV, 1.45 eV and 1.23 eV for the C_{60}, C_{180} and C_{240} fullerenes calculated at the GGA level.[391] For comparison, DFT calculations using the BP86 functional yield HOMO-LUMO gaps of 1.24 eV and 0.21 eV for β-C_{72} and γ-C_{72}, respectively. At the B3LYP level the HOMO-LUMO gap for α-C_{80} is 1.22 eV.

N	T	G_0	polyhedron(G_0)
80	α	I_h [20; 0, 0, 12]	dodecahedron
72	β(LF)	O_h [8; 0, 6]	cube
216	$\beta(LF^2)$	O_h [8; 0, 6]	cube
648	$\beta(LF^3)$	O_h [8; 0, 6]	cube
120	C_4-β(LF)	O_h [8; 0, 6]	cube
360	C_4-$\beta(LF^2)$	O_h [8; 0, 6]	cube
162	β(LF)	C_{2v} [18; 0, 2, 8, 1]	FS(45556555545)
288	β(LF)	C_2 [32; 0, 0, 12, 6]	fullerene $C_{32}(1)$
288	β(LF)	D_2 [32; 0, 0, 12, 6]	fullerene $C_{32}(2)$
288	β(LF)	D_{3d} [32; 0, 0, 12, 6]	fullerene $C_{32}(3)$
288	β(LF)	C_2 [32; 0, 0, 12, 6]	fullerene $C_{32}(4)$
288	β(LF)	D_{3h} [32; 0, 0, 12, 6]	fullerene $C_{32}(5)$
288	β(LF)	D_3 [32; 0, 0, 12, 6]	fullerene $C_{32}(6)$
72	γ(LF)	D_{6h} [12; 0, 6, 0, 2]	hexagonal prism
648	$\gamma(LF^3)$	D_{6h} [12; 0, 6, 0, 2]	hexagonal prism
108	C_4-γ(LF)	D_{6h} [12; 0, 6, 0, 2]	hexagonal prism
432	$\gamma(LF^2)$	D_{6d} [24; 0, 0, 12, 2]	fullerene $C_{24}(1)$
288	γ(LF)	D_{2h} [48; 0, 0, 12, 14]	fullerene $C_{48}(15)$
288	γ(LF)	C_{2v} [48; 0, 0, 12, 14]	fullerene $C_{48}(17)$
288	γ(LF)	D_{2h} [48; 0, 0, 12, 14]	fullerene $C_{48}(41)$
288	γ(LF)	C_{2v} [48; 0, 0, 12, 14]	fullerene $C_{48}(56)$
288	γ(LF)	C_{2h} [48; 0, 0, 12, 14]	fullerene $C_{48}(80)$
288	γ(LF)	C_{2v} [48; 0, 0, 12, 14]	fullerene $C_{48}(138)$
288	γ(LF)	D_{6d} [48; 0, 0, 12, 14]	fullerene $C_{48}(186)$
288	γ(LF)	D_{6d} [48; 0, 0, 12, 14]	fullerene $C_{48}(189)$

Table 6.1 Investigated gaudienes with N vertices resulting from a transformation T of an origin graph G_0, i.e. $T(G_0)$. A description of the polyhedral graph G_0 is provided as well. For one of the non-regular and less common polyhedron used we give the face spiral (FS) indices n of the n-gons for identification. For the fullerenes (Goldberg polyhedra) we provide the canonical index derived from the list of ring-spiral pentagon indices.[1,32,281]

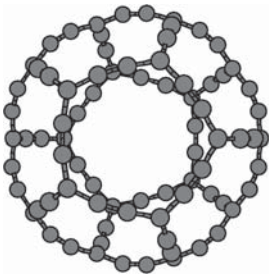


Figure 6.3 The molecular structure of α-C_{80} of I_h symmetry, which is the smallest α-gaudiene investigated.

A very stable molecule with a HOMO-LUMO gap of 5.17 eV is obtained when adding 20 hydrogens to the corners of the α-C_{80} icosahedron. Larger α-gaudienes can easily be constructed by using fullerenes as templates and replacing each carbon-carbon bond with a $-C{\equiv}C-$ moiety. All α-gaudienes have three-coordinated carbons in the intersections between three carbon rings. More stable molecules can be obtained by adding substituents to these three-coordinated carbons.

We should mention that Eq. (6.2) also allows to use other cubic polyhedra. For example, if we chose the truncated octahedron belonging to the Archimedean solids, $O_h\,[24;0,8,0,6]$, containing six squares according to Eq. (6.2), we can expand this graph for all edges in a similar way and we arrive at $O_h{-}C_{96}$, another type of α-gaudiene. Hence the playing field is huge for the construction of new 3D polyhedral molecular structures.

6.4.2 β-Gaudienes

The smallest member of the β-gaudiene class of molecules with four- and six-membered rings involved is the previously proposed β-C_{72} shown in Figure 6.2.[384] It belongs to the O_h point group and can be generated from the simple cube $G_0 = O_h\,[8;0,6]$ or its dual, the octahedron $G_0^* = O_h\,[0,6;8]$. When folding β-graphyne to the hollow C_{72} structure, the edges of eight hexadehydro[12]annulene rings of β-graphyne form six four-sided rings having four carbons on each side with alternating triple and single bonds. Even though hexadehydro[12]annulene is formally antiaromatic,[392–394] β-C_{72} is

aromatic according to the ring-current criterion sustaining a ring current strength of 44.3 nA/T around the molecule, which can be compared to the ring-current strength for benzene of 11.8 nA/T.[384,395] Since all carbons have formally four bonds, β-C_{72} is a stable molecule with a HOMO-LUMO gap of 1.82 eV at the B3LYP/SVP level.

A number of β-gaudiene structures were constructed using small polyhedral graphs as starting points. The smallest one obtained using that procedure was β-C_{162} (C_{2v}), which was obtained using an 18-vertex graph as initial structure, $C_{2v}[18;0,2,8,1]$, belonging to a non-regular polyhedron (see Figure 6.4). This polyhedron does not belong to the common solids and we therefore give the face spiral in Table 6.1. The optimized molecular structure of β-C_{162} (C_{2v}) is shown in Figure 6.5a. It has a rather small HOMO-LUMO gap of 0.48 eV as obtained at the B3LYP level.

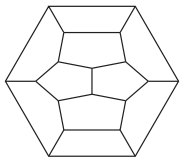
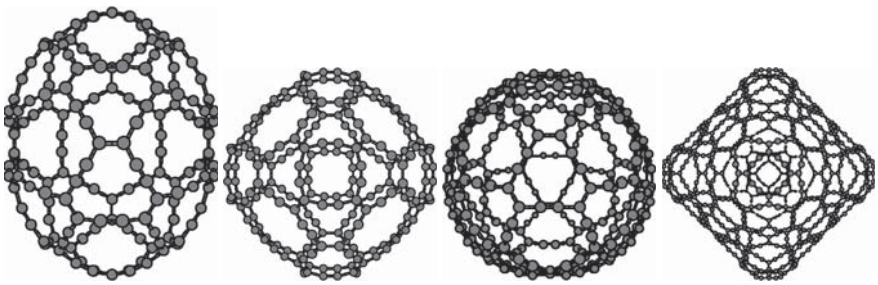


Figure 6.4 Graph of the non-regular $C_{2v}[18;0,2,8,1]$ polyhedron.



(a) (b) (c) (d)

Figure 6.5 The molecular structure of the energetically lowest isomers of β-gaudienes: (a) β-C_{162}, (b) β-C_{216} (O_h), (c) β-C_{288} (D_{3d}), and (d) β-C_{648} (O_h).

Since the topological method ensures that the symmetry of the seed structure is conserved, many different isomers can be constructed when the corresponding fullerene structures are available. The next larger β-gaudiene structures of O_h symmetry are β-C_{216} and β-C_{648}, which have HOMO-LUMO gaps of 0.415 eV and 0.225 eV at the B3LYP level, respectively. These are also shown in Figure 6.5.

A large number of β-gaudienes consisting of 288 carbons can be constructed by using different fullerene isomers such as those of C_{32} as starting structures. The energetically lowest β-C_{288} structure of all investigated isomers (see Table 6.1) belongs to the D_{3d} point group. The energetically lowest β-gaudiene structures obtained in this work are shown in Figure 6.5. The rest of the studied β-C_{288} structures are shown in the Supporting Information. The energetically lowest β-C_{288} isomers are 48.8 kJ/mol (D_3), 51.9 kJ/mol (D_2), 53.2 kJ/mol (C_2), 65.6 kJ/mol (C_2), and 146.6 kJ/mol (D_{3h}) above the D_{3d} structure. The β-C_{288} (O_h) isomer consisting of a β-C_{72} molecule inside β-C_{216} (β-C_{72}@β-C_{216}) is 683 kJ/mol above the lowest β-C_{288} isomer of D_{3d} symmetry. The HOMO-LUMO gaps of the β-C_{288} structures are in the range of 0.27–0.42 eV, with the largest gap for the most stable isomer. The two β-C_{288} structures of C_2 symmetry has negative HOMO-LUMO gaps at the BP86 level.

The largest molecule considered in this study was an onion-like β-C_{72}@β-C_{216}@β-C_{648} cluster of O_h symmetry. The structure is shown in the Supporting Information.

6.4.3 γ-Gaudienes

The smallest γ-gaudiene (γ-C_{72}) studied belongs to the D_{6h} point group and has a HOMO-LUMO gap of 0.519 eV. The optimized molecular structure is shown in Figure 6.6. The benzoic six-membered carbon ring is surrounded by six hexadehydro[12]annulene rings that are fused at the molecular edge. The correct topology of the closed surface is ensured by the six four-membered carbon rings between the hexadehydro[12]annulenes at the edge of the molecule. The next larger γ-gaudiene structure of O_h symmetry is γ-C_{432} (O_h) with a HOMO-LUMO gap of 0.491 eV. The largest γ-gaudiene studied in this work is γ-C_{648}, which has D_{6h} symmetry and a small HOMO-LUMO gap of 0.253 eV at the B3LYP level.

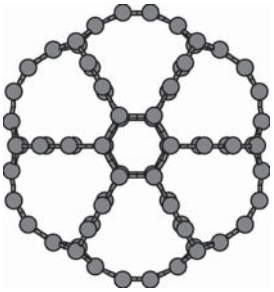


Figure 6.6 The molecular structure of γ-C_{72} of D_{6h} symmetry, which is the smallest γ-gaudiene investigated.

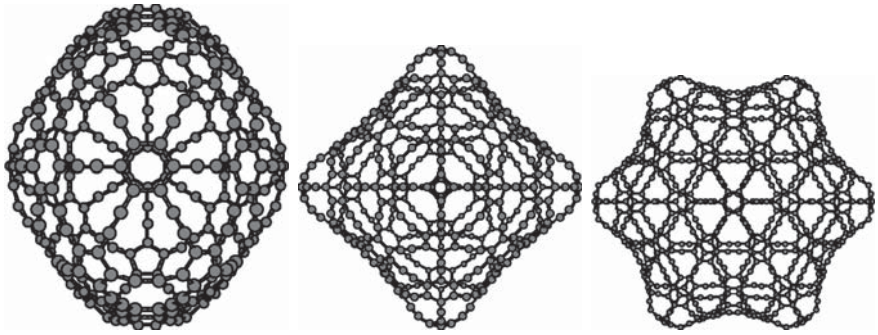
A large number of γ-C_{288} structures were constructed with different fullerene isomers of C_{48} as the origin graph. The energetically lowest γ-C_{288} belongs to the C_{2v} point group. The structure of the energetically lowest γ-C_{288} (C_{2v}) isomer is shown in Figure 6.7. The structures and atomic coordinates of the remaining studied γ-C_{288} clusters are shown in the Supporting Information. The energetically lowest γ-C_{288} isomers are 44.3 kJ/mol (D_{2h}), 47.8 kJ/mol (C_{2h}), 160.0 kJ/mol (D_{6d}), 167.5 kJ/mol (C_{2v}), 287.2 kJ/mol (C_{2v}), 565.5 kJ/mol (D_{2h}), and 873.1 kJ/mol (D_{6d}) above the lowest C_{2v} structure. The calculations show that there are only three low-lying γ-C_{288} isomers among the studied ones.

The HOMO-LUMO gaps of the lowest γ-C_{288} structure is 0.36 eV at the B3LYP level, whereas the HOMO-LUMO gap of the γ-C_{288} are in the range of 0.18–0.47 eV, with the largest gap for γ-C_{288} (D_{2h}). The energetically lower isomer of D_{6d} symmetry has a negative HOMO-LUMO gap at the BP86 level. The molecular structure of the energetically lowest γ-gaudienes obtained in this work are shown in Figure 6.7.

6.4.4 Graphdiyne-based gaudienes

Graphdiyne-based gaudienes are obtained by replacing the –C≡C– units of gaudienes with the twice longer –C≡C–C≡C– (C_4) group.

Even though the obtained molecules have a larger number of triple bonds, the HOMO-LUMO gap is wider than for the corresponding graphyne-based molecules. The HOMO-LUMO gaps calculated at the B3LYP level are 0.877 eV,



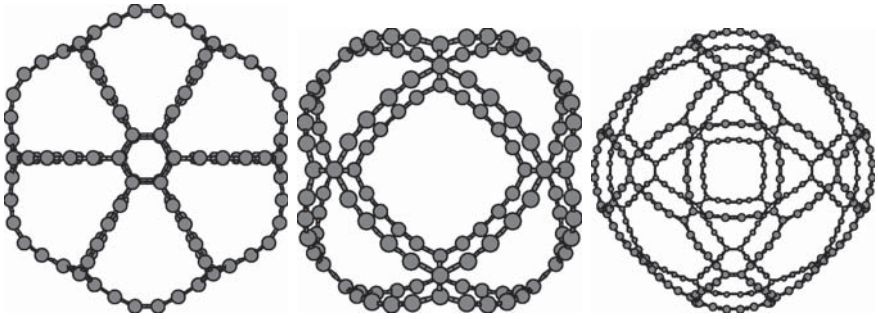
(a) (b) (c)

Figure 6.7 The molecular structure of the energetically lowest isomers of (a) γ-C_{288} (C_{2v}), (b) γ-C_{432} (O_h), and (c) γ-C_{648} (D_{6h}) gaudienes.

1.356 eV, and 1.168 eV for C_4-γ-C_{108}, C_4-β-C_{120}, and C_4-β-C_{360}, respectively. The larger stabilization of the graphdiyne-based gaudienes can be traced back to the molecular properties of dodecahydro[18]annulene, which is an aromatic molecule.[393] The HOMO-LUMO gap of C_4-γ-C_{108} is smaller than for the two other studied C_4-gaudienes due to the significantly larger bond strain of the small four-membered carbon rings at the edge of the molecule. For the C_4-β-gaudienes, the correct topology is obtained by the six 20-membered carbon rings. The molecular structures of the studied C_4-gaudienes are shown in Figure 6.8. The Cartesian coordinates of the atomic positions are given as Supporting Information.

6.5 Discussion and Conclusions

The γ-gaudienes are generally significantly lower in energy than the β-gaudienes for the same number of carbon atoms. The most important exception is β-C_{72}, which is 128.9 kJ/mol below γ-C_{72} due to the ring strain of the four-membered carbon rings of γ-C_{72} and maybe also due to the aromatic stabilization of β-C_{72}.[384] The HOMO-LUMO gap of β-C_{72} of 1.82 eV calculated at the B3LYP level is also much larger than the HOMO-LUMO gap of 0.52 eV for γ-C_{72}. For the small γ-C_{72} cage, the bond strain of the four-



(a) (b) (c)

Figure 6.8 The molecular structure of the investigated graphdiyne-based (C_4) gaudienes: (a) C_4-γ-C_{108} (D_{6h}), (b) C_4-β-C_{120} (O_h), and (c) C_4-β-C_{360} (O_h).

membered carbon rings is not compensated by the energetically favorable structure of γ-graphyne. For the larger gaudienes, the energy compensation from the γ-graphyne structure is much larger than the strain from the four-membered carbon rings leading to a lower energy for the larger γ-gaudienes than for the corresponding β-gaudienes. The lowest isomer of γ-C_{288} is more than 5000 kJ/mol (!) below the energy of the lowest β-C_{288} structure. For larger gaudienes, the energy difference between the γ-gaudiene and β-gaudiene is even bigger. For the largest studied gaudienes, γ-C_{648} is about 11000 kJ/mol below β-C_{648}.

The relative energy per carbon atom is given as a function of $1/N$ in Figure 6.9, where N is the number of carbon atoms. The relative energy depends roughly linearly on $1/N$ leading to the three straight lines in Figure 6.9 showing that γ-gaudiene is the most stable structure and that C_4-gaudienes are the energetically highest ones. Extrapolation to $1/N \to 0$ yields the relative energy of β-graphyne, γ-graphyne, and β-graphdiyne. The calculations shows that β-graphyne is 17 kJ/mol/carbon higher in energy than γ-graphyne and that β-graphdiyne is only 7 kJ/mol/carbon above β-graphyne.

The HOMO-LUMO gaps calculated at the B3LYP level are plotted as a function of $1/N$ in Figure 6.10 yielding three straight lines for three gaudiene classes. Extrapolation of the HOMO-LUMO gap to the limit of $1/N \to 0$ yields

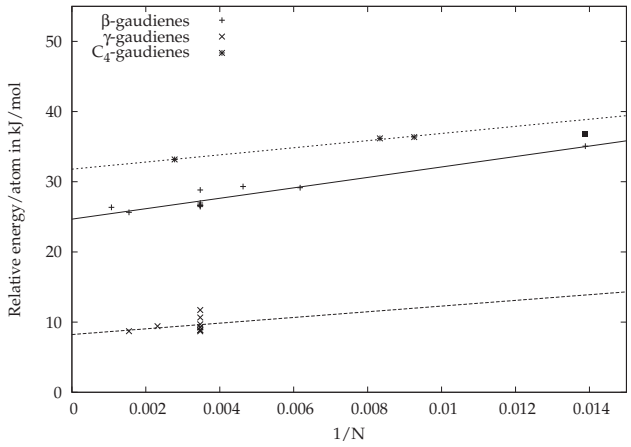


Figure 6.9 The relative energy (in kJ/mol) as a function of $1/N$ of the studied molecular classes. N is the number of atoms. The relative energies are fitted to a line. The outlier at 37 kJ/mol for γ-C_{72} is not considered in the fit.

estimates for the HOMO-LUMO gaps of the β-graphyne, γ-graphyne and β-graphdiyne sheets. The three lines suggest that the HOMO-LUMO gaps for the infinite planar structures are larger than zero. The extrapolated HOMO-LUMO gap of β-graphyne is 0.17 eV. For γ-graphyne we obtain an extrapolated HOMO-LUMO gap of 0.33 eV. The extrapolated HOMO-LUMO gap of 1.25 eV for β-graphdiyne agrees well with the previously reported value of 1.22 eV, which was calculated at the DFT level using a hybrid functional[396] and it is significantly larger than the HOMO-LUMO gap of 0.46 eV calculated at the DFT level using the generalized gradient approximation.[397] Even though the error bars of the present extrapolated values are large due to the few number of points used in the fit, the present calculations suggest that the HOMO-LUMO gap of graphdiyne is larger than 1 eV.

The hollow gaudiene structures might open the avenue to novel materials with interesting properties. For example, triple bonds provide possibilities for functionalizing the molecules with different kinds of substituents that even might couple several gaudiene molecules to polymers or solid-state materials. The gaudiene structures can also be fully or partially saturated with hydrogens without destroying their cage structures. Such molecules are

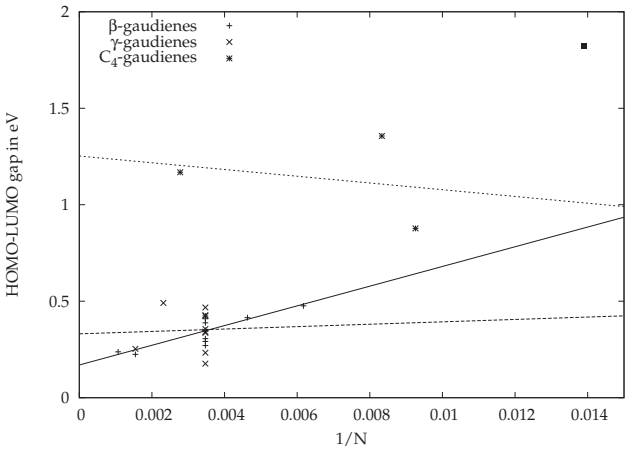


Figure 6.10 The HOMO-LUMO gap (in eV) as a function of $1/N$ of the studied molecular classes. N is the number of atoms. The relative energies are fitted to a line. The outlier at 1.82 eV for β-C_{72} is not considered in the fit.

ordinary saturated hydrocarbons with extraordinary hollow structures. By attaching substituents to the corners of α-gaudienes, molecules with a very large HOMO-LUMO gap are obtained. The optical gap for the C_4-gaudienes is significantly larger than for the two other classes of molecules. Extrapolation to infinitely large clusters suggests that planar graphdiyne has indeed a HOMO-LUMO gap that is significantly larger than zero, which might provide novel possibilities for carbon-based mono-layer structures.[381]

Finally, we point out that there are also other graph theoretical procedures for obtaining cavernous all-carbon structures that have not been discussed here.[398,399] For example, one can insert –C≡C– units into edges running parallel to the main axis of a fullerene nanotube. One can also use other building blocks beside carbon which fit the required topologies, e.g., building units which are used in metal-organic frameworks. This opens up a whole new area of interesting cage structures and materials which could be useful in many interesting applications.

Part II

Biasing molecular dynamics simulations with experimental residual dipolar couplings

The following chapters explore novel ways of restraining molecular dynamics (MD) simulations with experimental residual dipolar couplings (RDCs). RDCs are an observable in nuclear magnetic resonance (NMR) spectroscopy, that holds information on the alignment of interspin vectors relative to the magnetic field and therefore relative to each other. Biasing MD simulations with experimentally obtained observables improves the agreement of simulated and real structures and therefore leads to more accurate values of derived properties of the simulation trajectory.

Chapter 7 is an introduction to MD simulations and NMR spectroscopy with a focus on dipolar couplings and the enhancement of the former with the latter.

Chapter 8 explains in detail two algorithms for fitting alignment tensor components to to experimentally observed RDCs. Their computational cost including the scaling behaviour as well as their numerical accuracy and therefore applicability are assessed.

Chapter 9 considers the parametrisation of alignment by a weight function that expresses the distribution of orientations of the magnetic field relative to the simulated sample. The weight function is expanded in spherical harmonics. Alternatively, alignment is represented as the sum of a finite number of magnetic field vectors. The equations required for restraining are derived. Finally, this approach is compared to the existing model of alignment.

Chapter 10 explores restraining an MD simulation by experimental RDCs while allowing different parts of a molecule to align differently. This novel approach is aimed at curing one of the largest problems in RDC restrained MD simulations, viz the assumption that molecules are rigid.

7 Introduction

7.1 Molecular dynamics

7.1.1 Molecular simulation techniques

The term *molecular modelling* describes a number of computer aided techniques and algorithms that are used to determine properties of molecules with and without time evolution. The most common methods are quantum mechanics (QM), molecular mechanics (MM) and their joint application (QM/MM).

Molecular mechanics describes methods that treat atoms or groups of atoms as classical objects that can be described by Newtonian mechanics. The most prominent examples are Monte Carlo (MC) simulations[400] and molecular dynamics (MD) simulations[401,402]. MC and MD differ in that in MC time does not exist and only energies but no forces are computed while MD explicitly propagates time and in each step energies as well as forces are calculated.

Molecular dynamics is a step wise technique that takes a set of particles where for each particle the 3D coordinates and velocities are known, determines forces on each particle according to a force field, i.e., a set of parameters that define potentials between particles, and calculates new positions and velocities of each particle after a time step Δt. The energy at any given step is the sum of the potential energy, which is given by the force field, and the kinetic energy, which is explicitly known through the velocities, of all particles.

A variation of MD is stochastic dynamics (SD).[403] Here, in addition to the MD procedure, atoms of the solute are subject to a random displacement in each step and velocities are scaled down by a factor to simulate friction. The idea behind SD is, that it allows for simulating a solute molecule without

explicitly treating the solvent, but instead implicitly approximating the effects of the solvent.

MD plays an important role among the molecular modelling techniques as it is the computationally cheapest method that simulates every atom and includes a concept of time. As a result, MD allows simulations of systems of 10^5–10^6 atoms over a period of *ns*–*ms*, carrying out 10^6–10^{12} integration steps.[402,404]

At the same time, the accuracy of MD simulations is lower than semi-empirical or even quantum mechanical methods. Most force fields can only operate on a fixed molecular graph; breaking or forming bonds is not possible. Despite these limitations to electronically simple and non-reacting systems there is a wide range of systems—especially bio-macromolecules—that are successfully modelled using MD. This is particularly true for large systems that are characterised by their dynamic behaviour which requires simulations over long timespans to explore the energy landscape and perform statistical thermodynamics on the trajectory.

7.1.2 Energies and Forces

Force fields divide the total energy of a system and the force on each particle (typically atom) into specific and easy to calculate energy/force terms (eq. 7.2).

$$V_{\text{total}} = V_{\text{bonded}} + V_{\text{nonbonded}} + V_{\text{special}} \tag{7.1}$$

$$= V_{\text{bonds}} + V_{\text{angles}} + \cdots + V_{\text{electrostatic}} + V_{\text{VdW}} + \cdots + V_{\text{XRay}} + V_{\text{NMR}} \tag{7.2}$$

Here, V_{bonded} is the energy of bonded interactions (bonds, angles, dihedrals etc.), $V_{\text{nonbonded}}$ is the energy by physical interactions that do not act through bonds (typical examples are Coulombic interactions and Van der Waals interactions) and V_{special} denotes auxiliary energy terms that are used to impose restraints on a system, e.g. to take experimental data into account. Each of these terms is a sum over interactions between a subset of the simulated particles: Very few particles (e.g. two atoms per bond and four atoms per dihedral) in the case of bonded interactions, or all particles within a chosen radius from a given particle in the case of non-bonded interactions.

Interactions are typically approximated as simple potentials like a harmonic oscillator for bonds (eq. 7.3) and angles or a periodic potential for dihedral angles. More complex potentials can be chosen but increase the computational cost. For example, the bond term is typically expressed as

$$V_{\text{bonds}} = \sum_{i}^{\text{bonds}} \frac{1}{2} k_i (r_i - r_i^{(0)})^2, \tag{7.3}$$

where the index i runs over all bonds, and the values of the force constants k_i and the equilibrium values $r_i^{(0)}$ are specific to the bond type and r_i is the current value of the bond length in the simulated system. The set of these values is fitted to measured or calculated values or vibrational frequencies and is collectively described as the force field.

Forces on atoms i in MD simulations are obtained by deriving the potentials that are given in the force field according to the coordinates of the atoms (eq. 7.4).

$$\boldsymbol{f}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i} \tag{7.4}$$

In general, the forces on one atom in a system are affected by the coordinates of all other atoms in that system. Taking only two-body interactions into account, there are $\frac{1}{2}N(N-1)$ interactions (short range and long range) that contribute, where N is a measure for the system size, i.e., the number of simulated particles. Using a naive approach, the force field would scale $\mathcal{O}(N^2)$. There are algorithms for calculating short range as well as long range interactions that reduce the computational complexity to $\mathcal{O}(N^{\frac{3}{2}})$ (Ewald summation[405]), $\mathcal{O}(N \log(N))$ (particle-particle/particle-mesh summation[406], particle-mesh Ewald summation[407]) or even $\mathcal{O}(N)$ (fast multipole method[408]). The more efficient algorithms only pay off for very large systems.[409] Nevertheless, the calculation of forces is the most expensive step in a molecular modelling simulation (and among those, the non-bonded interactions dominate).

7.1.3 Integration schemes[s]

Given the coordinates $\boldsymbol{r}_t$, the velocity $\boldsymbol{v}_t$ and the mass m of a particle at time t as well as the forces on the particle (which are generally a function of the coordinates of some or all other particles), there are different algorithms to compute its coordinates and velocities $\boldsymbol{r}_{t+\Delta t}$ and $\boldsymbol{v}_{t+\Delta t}$ at time $t + \Delta t$.

A good integration algorithm must fulfil a number of requirements: The energy (or a derived extensive thermodynamic potential, if a thermodynamic ensemble other than the microcanonical is chosen [410]) of the system needs to be conserved, preferably both on a short and on a long time scale. Integration algorithms can be stable with respect to the energy for a small number of cycles but have a long term energy drift. Alternatively, the energy can fluctuate on a short time scale while the energy is conserved on a long time scale. Long term energy conservation is more important than short term energy conservation.

Closely related to the requirement of energy conservation is the demand that the simulation be time reversible, i.e., the trajectories of all particles should be reversed if at one step all velocities are reversed and the simulation continued.

The overall momentum of a simulated system should stay constant. Generally, this means that the sum over the momenta of all particles is zero after each time step. The same requirement exists for the angular momentum.

The previous two points on time reversibility and conservation of (angular) momentum are frequently summarised by demanding algorithms to be area preserving. If one plots the trajectory of any particle in the simulated ensemble in an $\boldsymbol{r}$–$\boldsymbol{v}$ diagram, the area (i.e., 6 dimensional volume) the particle traverses should be constant after a number of initial simulation cycles. For some algorithms the trajectory is increasingly smeared out, which is an indicator for violated time reversibility.

As in any numerical procedure,[t] the integration of the equations of motion is subject to noise. Even worse, all but the smallest molecular

[s] This subsection is predominantly based on ref. [409]

[t] Analytical integration methods cannot be used here because the time evolution of a system that consists of three or more particles that act on each other cannot be expressed in a closed form (three body problem [411]).

systems behave chaotically with the result that two trajectories that differ exponentially drift apart from each other. Small changes in the starting positions or the integration scheme lead to huge differences in the obtained trajectories. We are, however, less interested in the trajectories themselves than in derived properties and the populated regions of the phase space. These thermodynamic properties (such as energy, free energy, temperature and pressure) are relatively insensitive to changes of the trajectory. The obtained thermodynamic properties are only weakly affected by numerical noise. The relative population of different geometries in the phase space of a converged simulation is ideally not affected by diverging trajectories during the simulation or the exact starting point. Strong divergence may, however, lead to qualitatively wrong geometries.

One of the most commonly used step wise integration algorithms is the *Verlet algorithm*. It can easily be derived by adding the Taylor expansions (around t) of $r(t+\Delta t)$ and $r(t-\Delta t)$,

$$r(t \pm \Delta t) = r(t) \pm \frac{\mathrm{d}r(t)}{\mathrm{d}t}\Delta t + \frac{\mathrm{d}^2 r(t)}{\mathrm{d}t^2}\frac{\Delta t^2}{2} \pm \frac{\mathrm{d}^3 r(t)}{\mathrm{d}t^3}\frac{\Delta t^3}{6} + \mathcal{O}(\Delta t^4), \quad (7.5)$$

yielding an expression for $r(t+\Delta t)$ that is accurate to order Δt^4 and does not explicitly depend on $v(r)$,

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + \frac{\mathrm{d}^2 r(t)}{\mathrm{d}t^2}\Delta t^2 + \mathcal{O}(\Delta t^4). \quad (7.6)$$

The difference between the Taylor expansions of $r(t+\Delta t)$ and $r(t-\Delta t)$ yields the velocity at t,

$$r(t+\Delta t) - r(t-\Delta t) = 2\frac{\mathrm{d}r(t)}{\mathrm{d}t}\Delta t + 2\frac{\mathrm{d}^3 r(t)}{\mathrm{d}t^3}\frac{\Delta t^3}{6}, \quad (7.7)$$

i.e.,

$$v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t} + \underbrace{\frac{\mathrm{d}^3 r(t)}{\mathrm{d}t^3}\frac{\Delta t^2}{6}}_{\mathcal{O}(\Delta t^2)}. \quad (7.8)$$

Equivalent to the Verlet scheme is the *Leap Frog algorithm*. It differs from the former in that positions r and velocities v are not given at the

same time but at alternating steps, i.e., the positions are given at $r(t + n\Delta t)$ while the velocity is only computed (and defined) at half integer time steps $v(t + (n + \frac{1}{2})\Delta t)$. Positions and velocities are updated according to

$$v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{\mathrm{d}^2 r(t)}{\mathrm{d}t^2}\Delta t \qquad \text{and} \tag{7.9}$$

$$r(t + \Delta t) = r(t) + v(t + \frac{\Delta t}{2})\Delta t. \tag{7.10}$$

It is of course possible to include more terms of the Taylor expansion resulting in higher precision and therefore allowing for larger time steps. However, this increased short term precision comes at the expense of increased computational cost (due to more complex expressions) and higher memory usage (because coordinates and velocities of previous steps are taken into account and need to be stored). Some of the higher order schemes suffer from an increased long term energy drift.

7.1.4 Constraints

It is often desirable to restrict the flexibility of a system to a subset of its $3N - 6$ degrees of freedom, that is, to introduce constraints. Constraining bond lengths can be used to suppress high frequency vibrations and therefore allow for longer time steps. Constraining a few selected degrees of freedom can be used to accommodate for explicitly known values of these coordinates or to explore the behaviour of a system in case of certain values of internal coordinates.

Constraints can be divided into *holonomic* constraints, *non-holonomic* constraints and those that are neither. Holonomic constraints can be expressed as $f(q, \ldots, t) = 0$, i.e., one can construct a function of coordinates and the time which is required to be zero. Non-holonomic constraints are expressed as a function of coordinates, their derivatives and time, $f(q, \ldots, \dot{q}, \ldots, t) = 0$, whereas the last class of constraints cannot be expressed in either way, but, e.g., as an inequality.

Constraining coordinates in MD simulations requires holonomic constraints.

Lagrangian equations of motion can conveniently be used for finding the exact analytical solution to a constrained problem. The Lagrangian $\mathcal{L}$ is

defined as the difference between the kinetic and the potential energy of a system,

$$\mathcal{L} = \mathcal{K} - \mathcal{U}. \tag{7.11}$$

The Lagrangian equation of motion is given as

$$\frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{q}} = \frac{\partial \mathcal{L}}{\partial q}, \tag{7.12}$$

where q are generalised coordinates.

Given a set of holonomic constraints $\sigma_i(r) = 0$, a constrained Lagrangian $\mathcal{L}'$ can be defined,

$$\mathcal{L}' = \mathcal{L} - \sum_i \lambda_i \sigma_i(r), \tag{7.13}$$

where λ_i are the obtained forces that are required to enforce the chosen constraints.

This method is accurate if applied for differential equations. MD simulations are, however, step wise—using the forces λ_i in one of the above named integration schemes leads to constraints that are only fulfilled within an error margin that is defined by the integration algorithm. This error may be acceptable for the trajectory, but a long term drift of 'constrained' degrees of freedom is not. Furthermore, as soon as more than one constraint is imposed on the system, solving for λ_i requires the diagonalisation of a matrix, which is expensive.

In the case of more than one constraint it is common to employ the SHAKE[412] algorithm. Constrained distances, angles and dihedral angles are formulated as distance constraints of the form $r^2 - d^2 = 0$. In the resulting quadratic equation, the quadratic term is neglected for simplicity. The constraints are then iteratively applied to the system until every constraint is fulfilled within a chosen accuracy. Corrections to a bond distance are carried out in the direction of the bond prior to the first correction step.[412,413] The error due to neglecting the quadratic term is compensated by the iterative procedure. Other algorithms to enforce multiple constraints are RATTLE[414], SETTLE[415], LINCS[416], MSHAKE[417], WIGGLE[418], and SHAPE[419].

Alternatively, in case of single or uncoupled constraints, one can demand that the constraints must be strictly obeyed after each MD step, which leads

to a quadratic equation. This equation has two valid solutions for small displacements—large displacements cannot be corrected.

7.1.5 Restraints

There are experiments that yield structural data regarding the relative position of atoms in a system, such as X-ray diffraction, neutron diffraction and NMR experiments. Such information can be taken into account in MD simulations.

Observables that are back calculated form MD trajectories that are unrestrained with respect to these observable may significantly deviate from the experimentally known values. This implies, that unrestrained simulations are not always a good representation of the real structure and dynamics of a molecule.

Restraining a simulation with respect to observables biases the trajectory to agree better with experimentally known observables, which is desirable for two reasons. First, the force field is fine tuned. Minute deviations in force constants, the shape of potentials, or even missing force terms may lead to relevant differences in derived observables. Restraining the observable is the most direct approach to increase agreement with experiments which is the only way to measure the quality of a simulation. Second, it takes into account interactions that were present in the experiment, but may not be explicitly included in the simulation, like weak interactions with other large molecules or electromagnetic fields. At the same time, the potential energy due to the other force field terms increases; it is therefore important to choose force constants for the restraining terms, that are effective without raising the potential energy to much and without leading to large distortions relative to the unrestrained structure.

Assume there is an observable O_i that depends through a structural parameter s_i (e.g. an angle, a distance, or more complex dependencies) on the coordinates $\boldsymbol{r}_i$ of an atom. One can then introduce restraints by penalising the deviation of back calculated observables O_i from the experimental values O_i^{exp} or by penalising s_i relative to calculated s_i^{exp}. These two approaches are not equivalent in most cases. For the following potentials, s_i and s_i^{exp} will be used to demonstrate potentials, without implying preference over restraining the derived observables.

As experimental data is always subject to inaccuracies, and therefore the 'correct' value of any observable is unknown, constraining a coordinate is not an option. One can, however, restrain a coordinate, that is, introduce an energy term to the force field that penalises divergence from the measured value but does not suppress it entirely. Common dependencies between the energy and the geometric property are harmonic potentials (eq. 7.14), flat-bottom harmonic potentials (eq. 7.15). Physical justifications for these potentials are the assumption that the true potential can be approximated by a harmonic potential for small displacements, and that the location of the minimum is not accurately known and small deviations should therefore not be penalised.

The harmonic potential is expressed as

$$V^{(\text{harmonic})} = \frac{k}{2}(s - s^{\text{exp}})^2, \tag{7.14}$$

where the 'force constant' k is a measure for the accuracy of the experimentally measured s^{exp}. Higher values correspond to a higher accuracy of the measurement and create a steeper function and larger forces and energies.

The flat-bottom potential is defined as

$$V^{(\text{fb})} = \begin{cases} 0 & \text{for } |s - s^{\text{exp}}| \leq s^{(\varepsilon)} \\ \frac{k}{2}(|s - s^{\text{exp}}| - s^{(\varepsilon)})^2 & \text{for } |s - s^{\text{exp}}| \geq s^{(\varepsilon)} \end{cases}, \tag{7.15}$$

where s^{exp} is the measured value and $s^{(\varepsilon)}$ is the maximum difference between current and measured value that is not penalised, and hence a positive value. Both k and $s^{(\varepsilon)}$ are parameters that express the uncertainty of the measurement.

The forces on atoms at $\boldsymbol{r}_i$ due to harmonic and flat-bottom potentials are

$$\boldsymbol{f}_i^{(\text{harmonic})} = -\frac{\partial V^{(\text{harmonic})}}{\partial \boldsymbol{r}_i} = -k(s - s^{\text{exp}})\frac{\partial s}{\partial \boldsymbol{r}_i}, \tag{7.16}$$

and

$$\boldsymbol{f}_i^{(\text{fb})} = -\frac{\partial V^{(\text{fb})}}{\partial \boldsymbol{r}_i} = \begin{cases} \boldsymbol{0} & \text{for } |s - s^{\text{exp}}| \leq s^{(\varepsilon)} \\ -k(|s - s^{\text{exp}}| - s^{(\varepsilon)})\frac{\partial s}{\partial \boldsymbol{r}_i} & \text{for } |s - s^{\text{exp}}| \geq s^{(\varepsilon)} \end{cases}. \tag{7.17}$$

If the instantaneous value of O fluctuates on a time scale shorter than that on which O is measured, then effectively the time-average of O is measured. More specifically, the measured value of O is the convolution of the instantaneous $O(s,t)$ with an exponentially decaying function $g(t) = \frac{1}{\tau}e^{-\frac{t}{\tau}}$,[420]

$$\langle O(s)\rangle_{\text{time}}(t) = (O(s,t) * g(t))(t) \tag{7.18}$$

$$= \frac{1}{\tau(1-e^{-\frac{t}{\tau}})}\int_0^t e^{-\frac{t-t'}{\tau}} O(s,t')\,\mathrm{d}t'. \tag{7.19}$$

Here, τ is a constant that determines the time over which O is averaged. Although the actual time span of a measurement is longer than a whole simulation for many observables, τ is in practice chosen much smaller than the simulation time which allows for approximating the normalisation factor $1/\tau(1-e^{-\frac{t}{\tau}})$ as $\frac{1}{\tau}$ for all but the first frames.

In the limit of very large τ all frames of the simulation are weighted equally.

The harmonic potential using time-averaging is given as

$$V = \frac{k}{2}(\langle O(s)\rangle_{\text{time}} - O^{\text{exp}}(s))^2. \tag{7.20}$$

Other potentials are derived analogously by replacing the instantaneous value of the observable by its time-averaged value.

In practice time averaging leads to slow oscillations of restrained degrees of freedom. This problem can be remedied by using a biquadratic potential, i.e., a product of an instantaneous and a time averaged term.[413]

$$V^{\text{bq}} = \frac{k}{2}\left(O(s) - O^{\text{exp}}(s)\right)^2 \left(\langle O(s)\rangle_{\text{time}} - O^{\text{exp}}(s)\right)^2 \tag{7.21}$$

Deriving a time-averaged potential (e.g. eq. 7.20) according to the coordinates $\boldsymbol{r}_i$ of atom i yields the restraining force (see eq. 7.4):

$$\boldsymbol{f}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i} = -\frac{\partial V}{\partial\langle O(s)\rangle_{\text{time}}}\frac{\partial\langle O(s)\rangle_{\text{time}}}{\partial O(s)}\frac{\partial O(s)}{\partial \boldsymbol{r}_i} \tag{7.22}$$

where

$$\frac{\partial V}{\partial \langle O(s)\rangle_{\text{time}}} = k(\langle O(s)\rangle_{\text{time}} - O(s)^{\text{exp}}) \quad \text{and} \tag{7.23}$$

$$\frac{\partial \langle O(s)\rangle_{\text{time}}}{\partial O(s)} = 1 - e^{-\frac{t-t'}{\tau}}. \tag{7.24}$$

Here, eq. 7.23 depends on the chosen potential only. Eq. 7.24 stems from time-averaging and, as it is a small constant number, it can be absorbed in the force constant and effectively be set to 1. The third term, $\frac{\partial O(s)}{\partial \boldsymbol{r}_i}$, depends on the specific function $O(s)$ and cannot be expressed in a generic way.

In the case of a biquadratic potential, the force on an atom at position $\boldsymbol{r}_i$ is

$$\boldsymbol{f}_i^{\text{bq}} = -\frac{\partial V^{\text{bq}}}{\partial \boldsymbol{r}_i} = -\left(\frac{\partial V^{\text{bq}}}{\partial O(s)} + \frac{\partial V^{\text{bq}}}{\partial \langle O(s)\rangle_{\text{time}}} \frac{\partial \langle O(s)\rangle_{\text{time}}}{\partial O(s)}\right) \frac{\partial O(s)}{\partial \boldsymbol{r}_i}, \tag{7.25}$$

with

$$\frac{\partial V^{\text{bq}}}{\partial O(s)} = k(O(s) - O(s)^{\text{exp}})(\langle O(s)\rangle_{\text{time}} - O(s)^{\text{exp}})^2 \quad \text{and} \tag{7.26}$$

$$\frac{\partial V^{\text{bq}}}{\partial \langle O(s)\rangle_{\text{time}}} = k(O(s) - O(s)^{\text{exp}})^2(\langle O(s)\rangle_{\text{time}} - O(s)^{\text{exp}}). \tag{7.27}$$

The term $\frac{\partial \langle O(s)\rangle_{\text{time}}}{\partial O(s)}$ is specified in eq. 7.24, $\frac{\partial O(s)}{\partial \boldsymbol{r}_i}$ depends on the specific observable.

Not only are measurements of observables O generally slower than fluctuations of O, but as samples usually contain more than one molecule, the measurement averages over a set of molecules in possibly different conformations, thus justifying ensemble averaging, i.e., the simulation of several copies of one molecule and averaging over derived observables.[421]

The ensemble averaged $O(s)$ is given as

$$\langle O(s)\rangle_{\text{traj}} = \frac{1}{N}\sum_{\text{k}}^{N} O_k(\boldsymbol{r}). \tag{7.28}$$

The expressions for the ensemble averaged 'potential' and 'force' are analogous to eqs. 7.20 and 7.22, differing only in the derivative of the average according to the instantaneous value.

$$\frac{\partial \langle O(s) \rangle_{\text{traj}}}{\partial O(s)} = \frac{1}{N} \tag{7.29}$$

7.2 Residual dipolar couplings

Residual dipolar couplings (RDCs) are NMR observables that yield information about the global alignment of inter spin vectors (ISVs), which are typically bonds, relative to each other. The two spins forming an ISV must be in close spatial proximity, while the distance between the ISVs is not restricted. RDCs are therefore sensitive to global alignment of molecular domains, a piece of information that is a valuable complement to the more common local NMR observables like the chemical shift, J-values, and relaxation data due to the nuclear Overhauser effect (NOEs).

7.2.1 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a spectroscopic method in which a sample is irradiated with radio frequency (RF) waves while being exposed to a strong homogeneous magnetic field. In the magnetic field the degeneracy of nuclear spin states is lifted: absorption and emission translate to transitions between different spin states. In modern spectrometers, the spectrum is obtained as a Fourier transform of the consecutive emission. The simplest experiments use a single RF pulse; more involved experiments make use of multiple RF pulses and hence multi-photon processes.

Through the shift, multiplicity, splitting and lineshape of signals NMR spectroscopy reveals information about the local electromagnetic environment of the respective nuclear spin as well as spatial distances between atoms, the local molecular topology, symmetry, and angles and dihedral angles between bonds. As will be explained in subsections 7.2.2 and 7.2.3, NMR spectroscopy can be used to obtain information about data describing the orientation of spin pairs relative to the magnetic field, and therefore global orientation of spin pairs relative to each other.

The energy levels of the combined nuclear spin states of a system are defined by an NMR Hamiltonian $\hat{H}_{\mathrm{NMR}}$ which is the sum of different physical interactions between the local and global magnetic field, single spins and pairs of spins:[422]

$$\hat{H}_{\mathrm{NMR}} = \hat{H}_{\mathrm{Zeeman}} + \hat{H}_{\mathrm{CS}} + \hat{H}_{\mathrm{J}} + \hat{H}_{\mathrm{DD}} + \hat{H}_{\mathrm{Q}} + \cdots \tag{7.30}$$

The terms of eq. 7.30 will be explained in the remainder of this subsection and subsection 7.2.2.

The prevailing term in $\hat{H}_{\mathrm{NMR}}$ is the Zeeman Hamiltonian. It describes the lifting of the degeneracy of the $|s, m_s\rangle$ states in a magnetic field into $2s+1$ states with evenly spaced energy levels. The absolute splitting depends on the magnetic field $\boldsymbol{B}_0$ and the magnetogyric ratio γ of the nucleus. $\hat{\boldsymbol{I}}$ is the nuclear spin operator and $\hat{I}_x$, $\hat{I}_y$, and $\hat{I}_z$ are its components.

$$\hat{H}_{\mathrm{Zeeman}} = -\gamma \boldsymbol{B}_0 \cdot \hat{\boldsymbol{I}} \tag{7.31}$$

Under the assumption that the magnetic field is aligned with the z-axis of the reference frame, eq. 7.31 can be simplified to

$$\hat{H}_{\mathrm{Zeeman}} = -\gamma B_0 \hat{I}_z. \tag{7.32}$$

The second term in eq. 7.30 is the chemical-shift Hamiltonian, which describes the deviation of the local magnetic field at a nucleus from the external field $\boldsymbol{B}_0$. It is a function of the chemical-shift tensor $\boldsymbol{\sigma}$, $\boldsymbol{B}_0$ and the magnetogyric ratio γ:[422,423]

$$\hat{H}_{\mathrm{CS}} = \gamma \boldsymbol{\sigma} \boldsymbol{B}_0 \cdot \hat{\boldsymbol{I}}. \tag{7.33}$$

The sum of the Zeeman and chemical-shift Hamiltonian can be conveniently written as

$$\hat{H}_{\mathrm{Zeeman}} + \hat{H}_{\mathrm{CS}} = -\gamma(\mathbb{1} - \boldsymbol{\sigma})\boldsymbol{B}_0 \cdot \hat{\boldsymbol{I}}. \tag{7.34}$$

$\boldsymbol{\sigma}$ can be transformed into its principal axis frame and with the approximation that it is symmetric, $\boldsymbol{\sigma}$ can be decomposed into two components

$$\boldsymbol{\sigma} = \sigma_{\mathrm{iso}}\mathbb{1} + \begin{bmatrix} \delta_{xx} & 0 & 0 \\ 0 & \delta_{yy} & 0 \\ 0 & 0 & \delta_{zz} \end{bmatrix}, \tag{7.35}$$

and hence $\hat{H}_{\mathrm{CS}}$ can be written as sum of an isotropic ($\hat{H}_{\mathrm{CSI}} = \gamma\sigma_{\mathrm{iso}}\boldsymbol{B}_0 \cdot \hat{\boldsymbol{I}}$) and an anisotropic component ($\hat{H}_{\mathrm{CSA}}$, see chapter 8).[423,424]

$\hat{H}_{\mathrm{J}}$ is the J-coupling or scalar coupling term that describes the indirect dipolar interaction between two nuclear spins i and j mediated by the spin of electrons and therefore perceived as a through-bond interaction.

$$\hat{H}_{\mathrm{J}} = \frac{\gamma_i\gamma_j}{r^3}\hat{\boldsymbol{I}}'_j \cdot \hat{\boldsymbol{I}}_i \tag{7.36}$$

$$= \gamma_i\gamma_j\mathbf{J}\hat{\boldsymbol{I}}_j \cdot \hat{\boldsymbol{I}}_i \tag{7.37}$$

Here, $\mathbf{J}$ is the coupling tensor; multiplication with $\mathbf{J}$ transforms the spin $\hat{\boldsymbol{I}}_j$ into the apparent nuclear spin $\hat{\boldsymbol{I}}'_j$.

In analogy to the chemical-shift tensor $\boldsymbol{\sigma}$, $\mathbf{J}$ can by written as the sum of an isotropic and an anisotropic component, and therefore the Hamiltonian can be decomposed as $\hat{H}_{\mathrm{J}} = \hat{H}_{\mathrm{JI}} + \hat{H}_{\mathrm{JA}}$.[423]

$\hat{H}_{\mathrm{Q}}$ is the quadrupolar interaction Hamiltonian, which describes the interaction of the nuclear electric quadrupole moment Q with the gradient of the surrounding electric field:[422,425]

$$\hat{H}_{\mathrm{Q}} = \sum_{j,k\in\{x,y,z\}} \frac{\partial^2 V}{\partial j \partial k}\hat{\mathbf{Q}}_{jk}, \tag{7.38}$$

where the first factor in the sum is the second spatial derivatives of the electrostatic potential V and the second term is the tensor operator formed by components of the nuclear spin operator $\hat{\boldsymbol{I}}$. Rotating both tensors into their principal axis frame and defining $eq = V_{zz}$ and $\eta = (V_{xx} - V_{yy})/V_{zz}$ yield the following simplified expression:[422]

$$\hat{H}_{\mathrm{Q}} = \frac{e^2qQ}{4I(2I-1)}\left[3\hat{I}_z^2 - \hat{\boldsymbol{I}}(\hat{\boldsymbol{I}}+\mathbf{1}) + \frac{1}{2}\eta(\hat{I}_+^2 + \hat{I}_-^2)\right]. \tag{7.39}$$

Of the above described components of the NMR Hamiltonian, not all commute with the Zeeman Hamiltonian. As a consequence, the observed eigenvalues of $\hat{H}_{\mathrm{NMR}}$ are not eigenvalues of $\hat{H}_{\mathrm{Zeeman}}$ or any of the other components. The eigenstates of $\hat{H}_{\mathrm{NMR}}$ are dominated by a combination of pure spin states for each nucleus (the eigenstates of $\hat{H}_{\mathrm{Zeeman}}$) but other states admix, especially if the energy difference to the admixing state is small due to low magnetic fields.

7.2.2 Dipolar coupling

As is well known from classical magnetostatics, the vector potential $\boldsymbol{A}$ at position $\boldsymbol{r}$ of a magnetic dipole moment $\boldsymbol{\mu}$ is[426,427]

$$\boldsymbol{A} = \frac{\mu_0}{4\pi}\frac{\boldsymbol{\mu}\times\boldsymbol{r}}{r^3}, \tag{7.40}$$

where μ_0 is the vacuum permeability and r is the norm of $\boldsymbol{r}$.

The magnetic field $\boldsymbol{B}$ is the curl of the vector potential $\boldsymbol{A}$:

$$\boldsymbol{B}(\boldsymbol{r}) = \nabla\times\boldsymbol{A} = \frac{\mu_0}{4\pi}\left(\frac{3(\boldsymbol{r}\cdot\boldsymbol{\mu})\boldsymbol{r}}{r^5} - \frac{\boldsymbol{\mu}}{r^3}\right). \tag{7.41}$$

The potential V of a magnetic dipole moment $\boldsymbol{\mu}_1$ in the magnetic field $\boldsymbol{B}$ of a second magnetic dipole moment $\boldsymbol{\mu}_2$ is

$$V = -\boldsymbol{B}(\boldsymbol{r})\cdot\boldsymbol{\mu}_1 = -\frac{\mu_0}{4\pi}\frac{1}{r^3}\left(\frac{3}{r^2}(\boldsymbol{\mu}_1\cdot\boldsymbol{r})(\boldsymbol{\mu}_2\cdot\boldsymbol{r}) - \boldsymbol{\mu}_1\cdot\boldsymbol{\mu}_2\right). \tag{7.42}$$

This classical equation can be transferred into the quantum mechanical picture by replacing the potential and magnetic dipole moment by their respective operators, to obtain the dipolar interaction Hamiltonian:

$$\hat{H}_{\mathrm{DD}} = -\frac{\mu_0}{4\pi}\frac{1}{r^3}\left(\frac{3}{r^2}(\hat{\boldsymbol{\mu}}_1\cdot\boldsymbol{r})(\hat{\boldsymbol{\mu}}_2\cdot\boldsymbol{r}) - \hat{\boldsymbol{\mu}}_1\cdot\hat{\boldsymbol{\mu}}_2\right) \tag{7.43}$$

We can furthermore express the magnetic dipole moment as $\hat{\boldsymbol{\mu}} = \gamma\hat{\boldsymbol{I}}$ and denote the two nuclear spin operators as $\hat{\boldsymbol{I}}$ and $\hat{\boldsymbol{S}}$:

$$\hat{H}_{\mathrm{DD}} = -\frac{\mu_0\gamma_1\gamma_2}{4\pi}\frac{1}{r^3}\left(\frac{3}{r^2}(\hat{\boldsymbol{I}}\cdot\boldsymbol{r})(\hat{\boldsymbol{S}}\cdot\boldsymbol{r}) - \hat{\boldsymbol{I}}\cdot\hat{\boldsymbol{S}}\right) \tag{7.44}$$

This can be written component wise as

$$\hat{H}_{\mathrm{DD}} = -\frac{\mu_0\gamma_1\gamma_2}{4\pi}\frac{1}{r^3}\sum_{i,j\in\{x,y,z\}}\hat{I}_i\left(3\frac{r_i}{r}\frac{r_j}{r} - \delta_{ij}\right)\hat{S}_j, \tag{7.45}$$

or in a much shorter tensor notation:

$$\hat{H}_{\mathrm{DD}} = -\frac{\mu_0\gamma_1\gamma_2}{4\pi}\frac{1}{r^3}\sum_{q=-2}^{2}F_2^q(\theta,\varphi)A_2^{-q}, \tag{7.46}$$

where the individual tensor components are[422,428]

q	$F_2^q(\theta,\varphi)$	A_2^{-q}	
0	$3\cos^2\theta - 1$	$\hat{I}_z\hat{S}_z - \frac{1}{4}(\hat{I}_+\hat{S}_- + \hat{I}_-\hat{S}_+)$	(7.47)
±1	$\sin\theta\cos\theta e^{\mp i\varphi}$	$\frac{3}{2}(\hat{I}_z\hat{S}_\pm \mp \hat{I}_\pm\hat{S}_z)$	(7.48)
±2	$\sin^2\theta e^{\mp 2i\varphi}$	$\frac{3}{4}\hat{I}_\pm\hat{S}_\pm$	(7.49)

Here, $\hat{I}_\pm$ are the raising and lowering operators $\hat{I}_+ = \hat{I}_x + i\hat{I}_y$ and $\hat{I}_- = \hat{I}_x - i\hat{I}_y$.

As already pointed out in subsection 7.2.1, the eigenstate of the Zeeman Hamiltonian (eq. 7.31) is not a pure state of the NMR Hamiltonian (eq. 7.30). In the case of the dipolar interaction Hamiltonian $\hat{H}_{\mathrm{DD}}$ only the first summand of the $m = 0$ term commutes with $\hat{H}_{\mathrm{Zeeman}}$ while the $m = 1$ and $m = 2$ terms contain single and double excitations.

In strong magnetic fields $\boldsymbol{B}_0$ the energy difference between the eigenstate of the Zeeman Hamiltonian and the admixing singly and doubly excited states increases and their contribution becomes negligible. One can therefore approximate $\hat{H}_{\mathrm{DD}}$ by its adiabatic (or secular) part[422,424]

$$\hat{H}_{\mathrm{DD}} = -\frac{\mu_0\gamma_1\gamma_2}{4\pi}\frac{1}{r^3}(3\cos^2\theta - 1)\left(\hat{I}_z\hat{S}_z - \frac{1}{4}(\hat{I}_+\hat{S}_- + \hat{I}_-\hat{S}_+)\right). \tag{7.50}$$

The term $\hat{I}_+\hat{S}_- + \hat{I}_-\hat{S}_+$ (called flip-flop term) admixes a state in which both nuclear spins are flipped: the magnetic spin quantum number of one spin is incremented while the other is decremented. In the special case of spin-$\frac{1}{2}$ nuclei, the whole term is zero for parallel spins and either the first or second half is zero for antiparallel spins.

The remaining non-zero term (in the antiparallel case) is close in energy to the $\hat{I}_z\hat{S}_z$ term only for homonuclear spin systems. The contribution of the flip-flop term is therefore small for heteronuclear systems and can be neglected:[422,424]

$$\hat{H}_{\mathrm{DD}} = -\frac{\mu_0\gamma_1\gamma_2}{4\pi}\frac{1}{r^3}(3\cos^2\theta - 1)\hat{I}_z\hat{S}_z. \tag{7.51}$$

The eigenvalue equation of $\hat{I}_z$ is

$$\hat{I}_z \left| s, m_s \right\rangle = \hbar m_s \left| s, m_s \right\rangle , \tag{7.52}$$

with spin quantum number s and magnetic spin quantum number m_s. Application of $\hat{I}_z$ and $\hat{S}_z$ yields a factor of $\hbar^2 m_s^{(\mathrm{I})} m_s^{(\mathrm{S})}$ which has values of $-\frac{1}{4}\hbar^2$ or $\frac{1}{4}\hbar^2$ depending on whether the two spins have equal or different m_s. The interaction energy between the two dipoles is

$$E_{\mathrm{DD}} = -m_s^{(\mathrm{I})} m_s^{(\mathrm{S})} \hbar^2 \frac{\mu_0 \gamma_1 \gamma_2}{4\pi} \frac{1}{r^3} (3\cos^2\theta - 1). \tag{7.53}$$

A system of two spins has four eigenstates with respect to the Zeeman Hamiltonian, $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$, and $\beta\beta$. Due to the scalar and dipolar interaction, the energies of the two spin triplets are lowered by $E_{\mathrm{J}} + E_{\mathrm{DD}}$ while the two spin singlets are raised by the same energy. Therefore, two peaks are observed in the spectrum: the transition from $\alpha\alpha$ to $\alpha\beta$ and the transition from $\beta\alpha$ to $\beta\beta$ with a splitting of $4(E_{\mathrm{J}} + E_{\mathrm{DD}})$.

NMR spectra can be measured under isotropic conditions which results in the complete cancellation of dipolar splittings or under (weakly) aligned conditions which allows for their measurement (see subsection 7.2.3). J value splitting is (in good approximation) independent of alignment. The isotropically measured splittings between the two peaks is $4E_{\mathrm{J}}$, and the splitting is $4(E_{\mathrm{J}} + E_{\mathrm{DD}})$ when measured under aligned conditions. The difference between these two splittings is[429]

$$D = \frac{4E_{\mathrm{DD}}}{h} \tag{7.54}$$

$$= -\frac{\mu_0 \gamma_1 \gamma_2 \hbar}{8\pi^2} \frac{1}{r^3} (3\cos^2\theta - 1). \tag{7.55}$$

Depending on the alignment of $\boldsymbol{r}$ and the signs of the magnetogyric ratios, D may be positive or negative.

Most RDC theory[429–435] is based on

$$D^{\mathrm{static}} = D^{\mathrm{max}} \frac{3\cos^2\theta - 1}{2}, \tag{7.56}$$

where D is denoted as D^{static} to distinguish the averaged and measured D from the interaction of a statically aligned ISV and D^{max} is a shorthand for

the maximum possible RDC for a given constant inter-spin distance r:

$$D^{\max} = -\frac{\mu_0 \gamma_1 \gamma_2 \hbar}{4\pi^2 r^3} = -\frac{\mu_0 \gamma_1 \gamma_2 h}{8\pi^3 r^3}. \tag{7·57}$$

Nucleus	I	μ/μ_N	$\gamma/10^6\,\mathrm{rad\,s^{-1}T^{-1}}$
^{1}H	$1/2$	2.7928456	267.52196
^{2}H	1	1.21260077	41.662791
^{3}H	$1/2$	5.159714367	285.349779
^{12}C	0	0	0
^{13}C	$1/2$	0.702411	67.2828
^{14}N	1	0.4037607	19.33778
^{15}N	$1/2$	−0.2831892	−27.12621
^{16}O	0	0	0
^{17}O	$5/2$	−1.89380	−36.2808
^{31}P	$1/2$	1.95999	108.394

Table 7.1 Magnetic moments μ/μ_N and magnetogyric ratios γ according to IUPAC[436]

$D^{\max}$ can take positive or negative values, depending on the sign of the magnetogyric ratios (see Tab. 7.1).

Proteins—linear macromolecules that are built from amino acids—are one main field of application of RDCs. Typical RDC values of bonds in and adjacent to the protein backbone lie in the range of $\pm 10^3$–10^5 Hz, e.g., 24.35 kHz for bonded N–H$^{\mathrm{N}}$ RDCs, −4.77 kHz for C$^{\alpha}$–C′ RDCs, and −67.95 kHz for C$^{\alpha}$–H$^{\alpha}$ RDCs.

7.2.3 Averaging and weak alignment

NMR spectroscopy is a slow technique, with the interaction times of chemical shifts, scalar couplings and dipolar couplings ranging from μs to s.[437] Fast structural changes of molecules in the ns range can therefore not be resolved. Furthermore, samples typically contain multiple instances of the same molecule, which are moving independently and therefore have

different conformations. Assuming that the system is ergodic, i.e., that sampling over the time evolution of one molecule is equivalent to sampling over a set of independent molecules, we can use either average to represent the implicitly measured joint average over the ensemble and time.

Averaging over a set of micro states (which is an experimental necessity) loses information; the set of microstates (a time series, ensemble of molecules, or both) uniquely determines the average but the reverse operation is impossible.

Time or ensemble averaged observables that depend on non-observable structural parameters are the average of the instantaneous observable as a function of the structural parameter, and not the observable as a function of the averaged structural parameter (see subsection 7.1.5).

The k^{th} time and ensemble averaged RDC D_k is therefore

$$D_k = -\frac{\mu_0 \gamma_{k_1} \gamma_{k_2} h}{8\pi^3} \left\langle \frac{3\cos^2\theta_k - 1}{2r_k^3} \right\rangle . \tag{7.58}$$

Here, the angle brackets denote weak alignment, i.e., an *implicit* average over the series of orientations of one molecule over time and between different molecules in an ensemble.

The averaged expression in eq. 7.58 can take values between $-1/2\langle r_k^3 \rangle$ and $1/\langle r_k^3 \rangle$. These extreme values correspond to all inter spin vectors lying in the x-y plane, or parallel or antiparallel to the z axis (assuming the magnetic field is aligned with the z-axis), while the term averages to zero under isotropic conditions.

In order to observe dipolar coupling interactions, it is necessary to establish experimental conditions in which molecules weakly deviate from an isotropic distribution of orientations. Absolute values of the averaged term in the order of 10^{-3}–10^{-4} are typically achieved. Higher degrees of alignment are not desirable as they imply an increased line width which may lead to overlapping signals.

Experimentally, low degrees of alignment can be obtained as molecules with a magnetic susceptibility anisotropy align in strong magnetic fields.[438] Higher degrees are realised by addition of liquid crystals which cause the molecules in question to align through either steric or electronic interaction. A large variety of alignment media has been developed[430,439–442]. As will

be elaborated in subsection 7.2.5, structural and dynamic information can be obtained from measuring a sample in a set of alignment media that align the molecule differently and independently.[443–445]

In the following, the term *alignment* is used to describe a distribution of orientations of vectors relative to a coordinate system. A single orientation can be expressed as function of φ and θ. In solution this distribution will be a continuous and differentiable function. *Weak alignment* describes alignment that is close to isotropic, i.e., a function that is close to constant.

Coming from an experimental setup it is natural to express the alignment of a molecule relative to the experimental coordinate system, i.e., the magnetic field which is fixed relative to the coordinate system. One can instead express the alignment of the magnetic field relative to a single static copy of the molecule, a representation that contains the same information. One alignment function is transformed into the other by inversion. However, as the *observed* alignment function is invariant under inversion, the alignment of a molecule with respect to the magnetic field and the alignment of the magnetic field with respect to a molecule are expressed by identical functions. In the following, the alignment of the magnetic field relative to a static molecule will be used.

It is important to stress the difference between weak alignment and time/ensemble averaging in MD simulations. Weak alignment is a non-zero time and ensemble average of the orientation of ISVs relative to the magnetic field *in an experiment*. *In an MD simulation*, however, the magnetic field does not exist as a concept and any interactions that lead to alignment (alignment media, magnetic field) are not in general explicitly simulated. Moreover, usually only one instance of the molecule is simulated. Therefore, the weak alignment is expressed as a set of parameters, rather than an average over molecular structures and their orientation. Time and ensemble averaging, in contrast, explicitly average over molecular structures that have been encountered during an MD simulation. While the experimental weak alignment holds information about the relative orientation of ISVs in one molecule, time and ensemble averaging of the observables that depend on weak alignment add information on structural change and dynamics of the molecule.

7.2.4 Expression of molecular alignment by a tensor

The by far most commonly used parametrisation[u] of the above introduced alignment function is the alignment tensor.[446–448] At this stage, the alignment of the magnetic field relative to a single inter spin vector is described. The extension to sets of ISVs, i.e., molecules or fragments of molecules, including required approximations, will be discussed is the end of this subsection.

One can introduce a coordinate system (molecular frame) relative to which the magnetic field vector $\boldsymbol{B}$ and the k^{th} ISV $\boldsymbol{r}_k$ is given. The angle between $\boldsymbol{B}$ ($\boldsymbol{r}_k$) and the x-axis of the—arbitrarily chosen—molecular frame is β_x (α_x) (see Fig. 8.1); analogous angles are defined for the other axes. $\boldsymbol{B}$ and a unit vector in the direction of $\boldsymbol{r}_k$ are given by

$$\boldsymbol{B} = \begin{pmatrix} \cos\beta_x \\ \cos\beta_y \\ \cos\beta_z \end{pmatrix} \qquad \frac{\boldsymbol{r}_k}{r_k} = \begin{pmatrix} \cos\alpha_x \\ \cos\alpha_y \\ \cos\alpha_z \end{pmatrix}. \tag{7.59}$$

As the length of these vectors is 1, the cosine of the angle θ_k between them is given by

$$\cos\theta_k = \begin{pmatrix} \cos\alpha_x \\ \cos\alpha_y \\ \cos\alpha_z \end{pmatrix} \begin{pmatrix} \cos\beta_x \\ \cos\beta_y \\ \cos\beta_z \end{pmatrix}. \tag{7.60}$$

One can write the averaged angle dependency in equation 7.58 as

$$\left\langle P_2^0(\cos\theta_k)\right\rangle = \left\langle \frac{3\cos\theta_k^2 - 1}{2} \right\rangle \tag{7.61}$$

$$= \tfrac{3}{2}\left\langle \left(\cos\alpha_x\cos\beta_x + \cos\alpha_y\cos\beta_y + \cos\alpha_z\cos\beta_z\right)^2 \right\rangle - \tfrac{1}{2} \tag{7.62}$$

$$\begin{aligned} = \tfrac{3}{2}\Big(&\left\langle \cos^2\beta_x \right\rangle \cos^2\alpha_x + \left\langle \cos^2\beta_y \right\rangle \cos^2\alpha_y + \left\langle \cos^2\beta_z \right\rangle \cos^2\alpha_z \\ &+ 2\left\langle \cos\beta_x\cos\beta_y \right\rangle \cos\alpha_x\cos\alpha_y + 2\left\langle \cos\beta_y\cos\beta_z \right\rangle \cos\alpha_y\cos\alpha_z \\ &+ 2\langle \cos\beta_z\cos\beta_x \rangle \cos\alpha_z\cos\alpha_x \Big) - \tfrac{1}{2}. \end{aligned} \tag{7.63}$$

Only β, the orientation of the molecule with respect to $\boldsymbol{B}$, is averaged while α, the orientation of the inter spin vector with respect to the molecular frame,

[u] A parametrisation is the expression of a function (of any dimensionality) by a finite or infinite set of discrete parameters.

is assumed to be constant. This assumption is accurate as long as we are treating single inter spin vectors only.

Furthermore, in this equation the length r of the ISV has been removed from the averaged term. This approximation is well founded if the ISV represents a bond, while the distance between two spins may be far from constant in other cases.

One can define a tensor $\mathbf{S}$ of rank 2 with elements

$$S_{ij} = \frac{3}{2}\left\langle \cos\beta_i \cos\beta_j \right\rangle - \frac{1}{2}\delta_{ij}, \tag{7.64}$$

which is real, symmetric and traceless and therefore has five independent components. This tensor can be written as a 3×3 matrix, commonly referred to as the *Saupe matrix*.[446,447]

Comparing the following sum

$$\begin{aligned}\textstyle\sum_{i,j\in\{x,y,z\}}^{3} S_{ij}\cos\alpha_i\cos\alpha_j = \tfrac{3}{2}\Big(&\left\langle\cos^2\beta_x\right\rangle\cos^2\alpha_x + \left\langle\cos^2\beta_y\right\rangle\cos^2\alpha_y + \left\langle\cos^2\beta_z\right\rangle\cos^2\alpha_z \\ &+2\left\langle\cos\beta_x\cos\beta_y\right\rangle\cos\alpha_x\cos\alpha_y + 2\left\langle\cos\beta_y\cos\beta_z\right\rangle\cos\alpha_y\cos\alpha_z + 2\langle\cos\beta_z\cos\beta_x\rangle\cos\alpha_z\cos\alpha_x\Big) \\ &-\tfrac{1}{2}\cos^2\alpha_x - \tfrac{1}{2}\cos^2\alpha_y - \tfrac{1}{2}\cos^2\alpha_z\end{aligned} \tag{7.65}$$

to eq. 7.63 while taking into account that $\cos^2\alpha_x + \cos^2\alpha_y + \cos^2\alpha_z = 1$, shows that

$$\left\langle P_2^0(\cos\theta_k)\right\rangle = \sum_{ij}^{3} S_{ij}\cos\alpha_i\cos\alpha_j. \tag{7.66}$$

One can therefore, as an alternative to eq. 7.58, and with the above explained approximation that r is constant, express an RDC as

$$D_k = D_k^{\max}\sum_{ij}^{3} S_{ij}\cos\alpha_i\cos\alpha_j. \tag{7.67}$$

As only five of the tensor's components are independent, one can transform the 3×3 matrix in eq. 7.67 to a vector $\boldsymbol{a}$,

$$D_k = D_k^{\max}\sum_{h}^{5} a_h C_{k,h}, \tag{7.68}$$

where a_h are the five components of the vector and C_h are the respective values of $\cos\alpha_i \cos\alpha_j$. The values of a_h and C_h are given in eq. 8.4 and eq. 8.5 respectively.

The combination of the alignment of the ISV relative to the molecular frame and the alignment of the molecular frame relative to the magnetic field can without approximation be expressed as one averaged and one static degree of freedom. The common choice to average the orientation of the molecular frame relative to $\boldsymbol{B}$ and keep the orientation of the ISV relative to the molecular frame constant is based on the argument that bonds only have limited orientational freedom within a molecule while molecules can rotate freely, limited only by the alignment media and other interactions. Alternatively, the motion of the molecular frame relative to $\boldsymbol{B}$ can be assumed to be fixed while the orientation of the ISV fluctuates within the molecular frame, following the argument that intramolecular motion is faster than the motion of the whole molecule.[421] Both models are mathematically equivalent and lead to identical results.

An even more concise notation than eq. 7.68 is

$$D = D^{\max} \boldsymbol{r}^{\mathsf{T}} \mathbf{S} \boldsymbol{r}, \tag{7.69}$$

that uses only the ISV $\boldsymbol{r}$ and the alignment tensor $\mathbf{S}$. It is referred to as *Yan-Donald tensor notation* by Donald et al.[432], although there are prior applications to RDCs[449] and it is analogous to common expressions in tensor algebra. Eq. 7.69 can easily be derived from eq. 7.67.

The alignment tensor $\mathbf{S}$ can visually be understood as the deviation of an ellipsoid (the actual distribution of orientations) from a sphere (an isotropic distribution of orientations). This deviation is an ellipsoid itself which is centered at the origin and has zero volume. Diagonalising the tensor translates to rotating the ellipsoid (that corresponds to the tensor) in 3D such that the three main axes align with the coordinate system. The main axes of $\mathbf{S}$ may have positive or negative length, however, the sum of the lengths is zero. Eq. 7.69 (as well as eqs. 7.67 and 7.68 of course) evaluates the magnitude of $\mathbf{S}$ in the direction of $\boldsymbol{r}$.

In order to separate the orientation of the alignment ellipsoid and its shape, one can perform a coordinate transformation of the five independent components of $\mathbf{S}$ to the three Euler angles, its axial and its rhombic

component. The Euler angles define the rotation that is required to rotate the ellipsoid in 3D such that its main axis is parallel to the z-axis and the second axis is parallel to the y-axis of the coordinate system. The axial component[429,430]

$$A_a = \frac{3}{2} A_{zz}, \tag{7.70}$$

is a measure for the deviation from isotropic alignment and its rhombic component A_r is a measure for the asymmetry of the alignment

$$A_r = A_{xx} - A_{yy}. \tag{7.71}$$

A_{xx}, A_{yy}, and A_{zz} are the diagonal elements of the diagonalised tensor sorted from smallest to largest absolute value, i.e., A_{zz} is the axis along which the alignment deviates from the isotropic alignment most, A_{yy} and A_{xx} are the two axes with smaller absolute values. The rhombicity of $\mathbf{S}$ is defined as

$$R = A_r / A_a. \tag{7.72}$$

$D_k^{\max}$ can be expressed as a function of the axial and rhombic components in spherical coordinates:[429,430]

$$D_k = \frac{3}{2} D_k^{\max} \left(\cos^2\theta A_{zz} + \sin^2\theta \cos^2\varphi A_{xx} + \sin^2\theta \cos^2\varphi A_{yy} \right) \tag{7.73}$$

$$= D_k^{\max} \left(A_a \frac{3\cos^2\theta - 1}{2} + \frac{3}{4} A_r \sin^2\theta \cos 2\varphi \right). \tag{7.74}$$

Expressing the alignment of a single ISV by an alignment tensor is accurate. The only approximation included in the above equations (e.g., eq. 7.67 and 7.68) is the assumption that r is constant.

If the same $\mathbf{S}$ is used to express the alignment of two or more ISVs, it is implicitly assumed that all vectors have a fixed orientation relative to each other and exhibit the same dynamics. In other words, if all RDCs in a molecule or a molecular fragment are treated by the same alignment tensor then that molecule or fragment is assumed to be rigid. This assumption is generally not well founded; especially for large and flexible molecules like proteins.

7.2.5 Analysing molecular structure and dynamics

Experimental RDCs are used to determine the conformation of large molecules de novo, refine the structure of initial guesses, and to obtain the alignment tensor and hence information on their dynamics.[429,430,433,450]

The conformation of proteins can be constructed from RDC data without prior knowledge in an incremental manner.[429,432,451] Starting with one terminal residue, the dihedral backbone angles φ, θ (and ω) of each residue are calculated up to a low degeneracy. Measuring more than one RDC per residue or measuring the same RDCs in different alignment media reduces the number of angles that are compatible with measured RDCs. It is most common to measure N–H^N RDC but the RDCs of the other two backbone ISVs N–C′ and C^α–C′ as well as C^α–H^α and side chain N–H are measurable as well. One obtains a tree of backbone angles for each residue; the most favorable of the possible conformations can be found using additional NMR data (e.g. NOEs), by comparing to databases of common backbone dihedrals, and by comparing the energies of the structures according to a force field.[432] If fewer RDCs are available, the structure can be reoptimised with a force field after the addition of each residue.

Given *unassigned* RDC data, structural motifs can be identified by comparison with databases of experimental RDCs and secondary structure motifs.[432] Knowledge of the sequence of amino acids is not required as the backbone RDCs are independent of them, and the same folding may occur despite different sequences.

Different methods with varying degrees of approximation are available for calculating the alignment tensor from experimental RDCs with or without Cartesian coordinates of the molecule.

From the unassigned RDCs of a sample and without knowledge of structure, sequence or size of the molecule, the axial and rhombic component can be estimated from the distribution of RDCs.[430,452] This relies on the assumption that sufficiently many RDCs have been measured and that they have no preferential orientation. RDCs between different types of spin pairs can be used but need to be normalised.

Given experimental RDCs and Cartesian coordinates of a molecule, the alignment tensor can be calculated as a best fit using singular value decomposition or the general linear least squares method as will be explained in

chapter 8.

Expressing the alignment of all ISVs in a molecule by a single tensor relies on the assumption that all ISVs have the same dynamics, i.e., their individual alignment tensors all have the same axial and rhombic components. As this assumption is obviously not fulfilled in multidomain molecules, it has been suggested to fit individual tensors to different domains which are expected to be internally rigid but move relative to each other.[453–456]

This approach has been extended to assigning an individual alignment tensor to every ISV in what is called the *model-free approach*.[435,445,457] As input the Cartesian coordinates are required as well as at least five independent sets of experimental RDCs that were measured in different alignment media.

An even more sophisticated algorithm is the *direct interpretation of dipolar couplings* (DIDC).[444,458] Here, individual tensors are calculated for each ISV using at least five independent sets of RDCs but without the molecule's Cartesian coordinates or assignment of peaks.

Both the model-free approach and DIDC use RDCs that are measured in different alignment media. This relies on the assumption that structure and dynamics of the molecule are unaffected by the change of alignment medium.[444]

Using the coordinates of a molecule but *without* experimental RDCs, the alignment tensor can be estimated by sampling the steric and electrostatic interaction of the molecule in different orientations with the alignment medium.[459,460] The most well known implementation of this approach is the program PALES.[460]

Taking only the molecular shape into account, the alignment tensor can be derived from the moment of inertia tensor which can cheaply be calculated from only the Cartesian coordinates.[461] This approximation is justified if the interaction with the alignment medium is dominated by steric interaction.

Once the alignment tensors of individual ISVs or groups of ISVs are known, information on their dynamics can be derived.[462] Approximate molecular structures can be optimised with respect to given alignment tensors as will be expanded in subsection 7.2.7.

7.2.6 Fitting alignment tensor components to experimental RDCs

During the interpretation of experimental results as well as RDC restrained MD simulations, the most common situation is, that assigned experimental RDCs and the Cartesian coordinates of the molecule (or an initial guess thereof) are available.

The alignment tensor that least squares fits the back calculated RDCs to the experimental ones can be found by either singular value decomposition[265,463] (SVD) or the linear least squares method[265,452] (LLS). The former is more commonly applied to RDCs despite the fact that the latter is faster and sufficiently accurate.[464] Both algorithms require a set of five or more independent RDCs to lift the degeneracy of the experimental values with respect to the absolute orientation of the ISV.

A detailed comparison between the two algorithms including timings and and an assessment of the numerical stability in near degenerate situations is provided in the following chapter which has previously been published in Wirz et al.[464].

7.2.7 Restraining MD simulations using RDCs

Restraining the relative orientation of inter spin vectors using experimental RDCs is a special case of the restraining that was introduced in subsection 7.1.5. Employing RDC restraints aids the simulation for the reasons explained previously. They are particularly beneficial as orientational restraints are non-local and can be fulfilled by changing global alignment of molecular fragments relative to each other but–due to their non-local nature–without contradicting the conventional force field terms and hence without raising the potential energy of the system.

In order to restrain atomic coordinates in a simulation it is necessary to define a potential energy function V^{RDC} for the k^{th} RDC and forces $\boldsymbol{f}_i$ on atoms i[421,465,466]

$$V_k^{\mathrm{RDC}} = V_k(D_k, D_k^{\mathrm{exp}}) \tag{7.75}$$

$$\boldsymbol{f}_i = -\frac{\partial V_k^{\mathrm{RDC}}}{\partial \boldsymbol{r}_i} = -\frac{\partial V_k^{\mathrm{RDC}}}{\partial D_k}\frac{\partial D_k}{\partial \boldsymbol{r}_i}, \tag{7.76}$$

that is, the instantaneous RDC interaction D_k and its derivative $\frac{\partial D_k}{\partial \boldsymbol{r}_i}$ according to the coordinates of atom i need to be computed in every MD step.

Calculating D_k requires a representation, i.e., parametrisation, of alignment. As explained in subsection 7.2.4, the parametrisation by an alignment tensor is most commonly used, but any other parametrisation, such as the expansion of the alignment ellipsoid in another basis, can be used as well.

Using the methods explained in the previous subsection, the alignment tensor (in general: the alignment representation) is chosen as the least squares fit of the calculated and measured RDCs. Using other methods for determining the alignment tensor of a molecule mentioned in subsection 7.2.4, e.g., PALES or calculating the alignment tensor based on the moment of inertia tensor is (apart from being inaccurate and inefficient) possible. However, as MD simulations do not explicitly include a magnetic field, i.e., do not have a preferential direction, the obtained tensor would have to be rotated to match the experimental RDCs in a second step, making this procedure even less feasible.

In subsection 7.1.5 the equations are given for restraining, using harmonic potentials (eqs. 7.14 and 7.16) and flat-bottom potentials (eqs. 7.15 and 7.17), which can be combined with time averaging (eqs. 7.20 and 7.22) or ensemble averaging (analogously). These equations are general and apply to restraining using RDCs as well; the only subexpressions that are specific to this problem and were omitted previously are the dependence of the observable $O(\boldsymbol{r})$ (here: $D(\boldsymbol{r})$) on atomic coordinates and its derivative $\frac{\partial O(\boldsymbol{r})}{\partial \boldsymbol{r}_i}$ according to the coordinates, i.e., the restraining force.

$D(\boldsymbol{r})$ is computed most easily in this case using eq. 7.68.

The derivative $\frac{\partial O(\boldsymbol{r})}{\partial \boldsymbol{r}_i}$ can be expanded and computed as

$$\frac{\partial D}{\partial \boldsymbol{r}_i} = \sum_h^5 \frac{\partial D}{\partial C_h} \frac{\partial}{\partial \boldsymbol{r}_i} \frac{C_h}{r_i^3} \qquad \text{with} \tag{7.77}$$

$$\frac{\partial D}{\partial C_h} = D^{\max} a_h r^3 \qquad \text{and} \tag{7.78}$$

$$\frac{\partial}{\partial \boldsymbol{r}_i} \frac{C_h}{r_i^3} = C_h \frac{\partial}{\partial \boldsymbol{r}_i} \frac{1}{r_i^3} + \frac{1}{r_i^3} \frac{\partial C_h}{\partial \boldsymbol{r}_i}. \tag{7.79}$$

The expressions for the five C_h are given in eq. 8.5; the derivatives in eq. 7.79 are easily calculated.

As all restraining potentials refer to internal degrees of freedom, external degrees of freedom, i.e., translation and rotation are unaffected. The sum of all restraining forces on atoms equals zero and the applied torque is zero.

Moltke and Grzesiek[467] have devised a scheme to constrain orientational degrees of freedom using experimental RDCs without the explicit knowledge of the alignment tensor. The accuracy and convergence properties do not significantly suffer from this approximation.[452,467]

In order to analyse a trajectory and assess the deviation between experimental and actual RDCs a quality factor Q is defined as[468]

$$Q = \sqrt{\frac{\sum_k (D_k - D_k^{\text{exp}})^2}{\sum_k (D_k^{\text{exp}})^2}} \tag{7.80}$$

which can be calculated for every frame. Values of Q range from 0 for perfect agreement to ∞. However, only Q values up to 1 are relevant as $Q = 1$ can trivially be achieved by setting all calculated values to 0 by assuming isotropic alignment.

As mentioned above (subsection 7.1.5), measured RDCs are effectively averaged over time and the ensemble of molecules. One can therefore not expect that the calculated RDCs of the simulation agree with the experimental value in every frame. Rather, by the same argument by which time averaged restraints were introduced, Q values should be calculated in an time averaged fashion. The obvious way to define a time averaged Q^{av} is

$$Q^{\text{av}} = \sqrt{\frac{\sum_k (D_k^{\text{av}} - D_k^{\text{exp}})^2}{\sum_k (D_k^{\text{exp}})^2}}, \tag{7.81}$$

although this appears not to have been suggested in literature previously. Here, D_k^{av} is calculated according to eq. 7.19, leading to either one Q^{av} for the whole trajectory or one per frame.

Almost arbitrarily small Q values can be achieved by choosing too large force constants k (see for example eq. 7.14) which cause distortions in the overall structure away from minima as defined by the remaining terms of the force field. It is therefore necessary to validate the overall structure. A common measure for structural change is the root mean square deviation

(RMSD) of all atoms from a reference structure (e.g. an X-ray structure) which—for simplicity—is approximated by the RMSD of all C^{α} atoms.

Q values as well as the RMSD of a trajectory are typically not calculated during the simulation but in a post-processing step and only for a small number of frames based on the stored trajectory.

8 Fitting alignment tensor components to experimental RDCs, CSAs and RQCs[v]

The measurement of residual dipolar couplings[429,433,469] (RDCs) by solution state nuclear magnetic resonance (NMR) spectroscopy gives information about the alignment of the vectors linking pairs of nuclei with non-zero spin relative to the magnetic field. The measured RDC value D_k is a function of $\langle \cos^2(\theta(\boldsymbol{r}_k, \boldsymbol{r}_H)) \rangle$, where $\theta(\boldsymbol{r}_k, \boldsymbol{r}_H)$ is the angle between the inter-spin vector $\boldsymbol{r}_k$ and the magnetic field director $\boldsymbol{r}_H$. Specifically, D_k is given as

$$D_k(\boldsymbol{r}_k, \boldsymbol{r}_H) = -\frac{\gamma_{k_1}\gamma_{k_2}\mu_0 h}{8\pi^3} \left\langle \frac{3\cos^2(\theta(\boldsymbol{r}_k, \boldsymbol{r}_H)) - 1}{2r_k^3} \right\rangle, \tag{8.1}$$

where γ_{k_1} and γ_{k_2} are the gyromagnetic ratios of the two spins, μ_0 is the magnetic permittivity of vacuum, h is Planck's constant, and r_k is the distance between the two spins. The averaging in eq. 8.1 encompasses ensemble as well as time averaging.

Chemical shift anisotropies (CSAs)[470,471] and residual quadrupolar couplings (RQCs)[467,470] are other types of NMR observables that yield information about the global alignment of atoms and their local environment in a molecule relative to a magnetic field and, therefore, indirectly relative to each other. The measured anisotropic contribution δ_k^{an} of the local environment to the chemical shift δ_k of a nucleus and the residual quadrupolar interaction of nuclei of angular momentum $I > \frac{1}{2}$ with their electronic environment

[v] This chapter has been published in "Fitting alignment tensor components to experimental RDCs, CSAs and RQCs" by Wirz et al.[464] and is reproduced with kind permission from Springer Science and Business Media. The candidate's contribution to the article is quantified on pages 299 ff.

have the same angle dependence as RDCs, and can therefore be expressed by the same alignment tensor.[467,470]

8.1 Theory

8.1.1 Description of alignment

In order to extract structural information from RDCs it is necessary to introduce a parametrization of the averaged angle dependence in eq. 8.1. This is achieved by defining first, a coordinate system (molecular frame), and then, relative to this, the orientation of the magnetic field vector $\boldsymbol{r}_H$ and the inter-spin vector $\boldsymbol{r}_k$ (Figure 8.1). The alignment of the molecule relative to the magnetic field can be then described using an alignment tensor (Saupe matrix[447]), leading to the concise formula

$$D_k(\boldsymbol{r}_k, \mathbf{S}) = D_k^{\max} \sum_{i,j} S_{ij} \Delta_{kij}(\boldsymbol{r}_k), \qquad i,j \in x,y,z\,, \tag{8.2}$$

where $D_k^{\max} = -\frac{\gamma_{k_1}\gamma_{k_2}\mu_0 h}{8\pi^3}\frac{1}{r_k^3}$ is the maximum possible RDC for a pair of spins k_1 and k_2 at a given distance r_k, $S_{ij} = \frac{3}{2}\left\langle \cos\beta_i \cos\beta_j \right\rangle - \frac{1}{2}\delta_{ij}$ is the $(i,j)^{\text{th}}$ component of the alignment tensor and $\Delta_{kij} = \cos\alpha_{ki}\cos\alpha_{kj}$ expresses the (static) orientation of the k^{th} inter spin vector relative to the molecular frame.

CSAs and RQCs allow for an equivalent parametrization in terms of an alignment tensor $\mathbf{S}$; the respective equations can be found in, for example, refs. [467] and [463].

Equation 8.2 is based on three approximations: First, the length of the inter spin vector $\boldsymbol{r}_k$, which is generally between a bonded pair of atoms, is assumed to be constant. Second, the orientation of $\boldsymbol{r}_k$ in the molecular frame is treated as fixed, i.e., all $\boldsymbol{r}_k$ that are represented by the same $\mathbf{S}$ have a fixed alignment relative to each other, which is commonly summarized as assuming the molecule to be rigid. Third, it is assumed that neither the length nor the alignment of $\boldsymbol{r}_k$ in the molecular frame affects the alignment of the molecular frame with the magnetic field. An alternative, which leads to analogous equations, is to treat the orientation of the molecular frame

with respect to the magnetic field as constant and the alignment of the inter spin vector within the molecule as variable.[421]

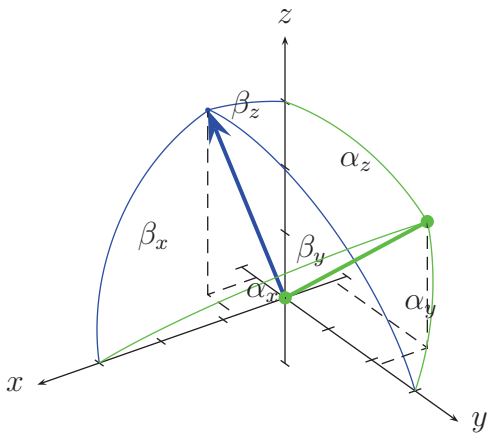


Figure 8.1 The inter-spin vector $\boldsymbol{r}_k$ (green) and magnetic field vector $\boldsymbol{r}_H$ (blue) with respect to the molecular frame; angles between the inter spin vector and the axes of the molecular frame are denoted as $\alpha_{\{x,y,z\}}$; angles between the magnetic field vector and the axes of the molecular frame are denoted as $\beta_{\{x,y,z\}}$.

The alignment tensor $\mathbf{S}$ that describes the weak alignment of the molecular frame in a magnetic field is traceless and symmetric and can therefore be reduced to five independent components, which allows eq. 8.2 to be simplified to

$$D_k(\boldsymbol{r}_k, \boldsymbol{a}) = D_k^{\max} \sum_h^5 a_h C_{k,h}(\boldsymbol{r}_k), \tag{8.3}$$

with

$$\begin{aligned} a_1 &= \tfrac{3}{2}\langle \cos^2 \beta_x \rangle - \tfrac{1}{2}, \\ a_2 &= \tfrac{3}{2}\langle \cos^2 \beta_y \rangle - \tfrac{1}{2}, \\ a_3 &= \tfrac{3}{2}\langle \cos \beta_x \cos \beta_y \rangle, \\ a_4 &= \tfrac{3}{2}\langle \cos \beta_x \cos \beta_z \rangle, \\ a_5 &= \tfrac{3}{2}\langle \cos \beta_y \cos \beta_z \rangle, \end{aligned} \tag{8.4}$$

and

$$\begin{aligned} C_{k,1}^{\mathrm{RDC}} &= \cos^2\alpha_x - \cos^2\alpha_z, \\ C_{k,2}^{\mathrm{RDC}} &= \cos^2\alpha_y - \cos^2\alpha_z, \\ C_{k,3}^{\mathrm{RDC}} &= 2\cos\alpha_x\cos\alpha_y, \\ C_{k,4}^{\mathrm{RDC}} &= 2\cos\alpha_x\cos\alpha_z, \\ C_{k,5}^{\mathrm{RDC}} &= 2\cos\alpha_y\cos\alpha_z. \end{aligned} \tag{8.5}$$

Analogous expansions of the CSA and RQC can be derived from their respective Hamiltonians[467,470] to obtain $\boldsymbol{C}_k^{\mathrm{CSA}}$ and $\boldsymbol{C}_k^{\mathrm{RQC}}$. Dividing by the respective numerical interaction constant provides the reduced interaction and allows for fitting to an alignment tensor.

In the following, we limit ourselves to describing RDCs for simplicity. However, all sums over RDC interactions can be extended to mixed sums over RDC, CSA and/or RQC interactions with the appropriate substitutions of D_k, $C_{k,h}^{\mathrm{RDC}}$, etc, to obtain a common alignment representation, analogous to the joint evaluation of RDCs and CSAs by ref. [463].

8.1.2 Efficient determination of the optimal alignment tensor

The parametrization described above leads to the following question: Given a molecular topology and a number of experimentally measured RDCs, how can the five independent components of the alignment tensor be determined that best describe the molecular orientation, and therefore also best reproduce the experimental results?

Solving this problem is relevant for both assessing ensembles of molecular structures to ascertain how well they agree with experimental data, and biasing simulations to agree with experimental values, because in either case, fitting of the alignment tensor $\mathbf{S}$ is required either for each structure of the ensemble or at each cycle (e.g. integration time step, in the case of molecular dynamics (MD) simulations).

Ref. [463] proposed an elegant solution, in which an equation $\mathbf{A}\boldsymbol{a} = \boldsymbol{b}$ is constructed, where $\mathbf{A}$ is an $n \times 5$ matrix that holds the five $C_{k,h}$ (eq. 8.5) for each interaction, $\boldsymbol{b}$ is a vector of size n that holds the reduced experimentally measured RDCs, $\boldsymbol{a}$ ($\boldsymbol{x}$ in ref. [463]) is a vector of size 5 that holds

the five independent components of the alignment tensor (eq. 8.4) and n is the number of experimentally observed interactions. Diagonalizing $\mathbf{A}$ by singular value decomposition (SVD) into $\mathbf{U}$, $\mathbf{\Sigma}$, and $\mathbf{V}^\intercal$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal$, inverting $\mathbf{\Sigma}$, and solving for $\boldsymbol{a}$ gives the exact result for $n = 5$ and the least squares fit for more than five independent components. SVD is numerically very stable in cases where $\mathbf{\Sigma}$ has near-zero eigenvalues and even allows for handling underdetermined systems, a property that makes this method very reliable and suitable for systems with few, possibly dependent, interactions, such as parallel inter spin vectors.

For the SVD of a matrix $\mathbf{A}$ of size $n \times m$ there are two algorithms available: the Golub-Reinsch algorithm, which has a computational cost of $4n^2m + 8nm^2 + 9m^3$ flops, and the R-SVD algorithm, which costs $4n^2m + 22m^3$ flops if all three resulting matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$ are required,[472] which is the case for the specific task of solving a system of equations by SVD.[463] For fitting to RDCs, n is the number of interactions (≥ 5) and m is the number of independent alignment tensor components ($= 5$). As m is constant in these applications, the time complexity of both algorithms is $\mathcal{O}(n^2)$, that is, they scale quadratically with the number of interactions.

In systems with a large number of observables ($n \gg 5$) that can be expressed in terms of an alignment tensor, underdetermined matrices $\mathbf{A}$ that lead to zero and near-zero eigenvalues of $\mathbf{\Sigma}$ are a negligible risk, but speed becomes relevant, particularly when the fitting needs to be done for every structure in a large ensemble or long trajectory, or at every (2 fs) integration step in a long (μs – ms) MD simulation.

A faster algorithm for obtaining the alignment tensor is linear least squares[265,473] (LLS), which scales linearly with the number of RDCs. LLS can be derived by defining a potential

$$V(\boldsymbol{D}; \boldsymbol{D}^{\text{exp}}, \boldsymbol{D}^{\text{max}}, \boldsymbol{w}) = \sum_{k}^{\text{RDC}} V_k(D_k; D_k^{\text{exp}}, D_k^{\text{max}}, w_k), \tag{8.6}$$

that penalizes the deviation of calculated from experimental RDCs. Individual weight factors w_k specific to each interaction allow, for instance, measurements made with greater confidence to be assigned higher weight. In order to give equal weight to different types of interactions, i.e., RDCs, CSAs and RQCs, as well as RDCs between different types of spins, the potential can be formulated in terms of reduced interactions, e.g., for RDCs,

$D_k^{\text{red,calc}} = D_k / D_k^{\max}$ and $D_k^{\text{red,exp}} = D_k^{\text{exp}} / D_k^{\max}$. The harmonic potentials V_k are defined as:

$$V_k = \sum_k^{\text{RDC}} \frac{1}{2} w_k \left(D_k^{\text{red,calc}}(\boldsymbol{r}_k, \boldsymbol{a}) - D_k^{\text{red,exp}} \right)^2 . \tag{8.7}$$

The use of reduced RDCs differentiates Equation 8.7 from the similarly defined restraining potential by ref. [421] in that it minimizes $||\boldsymbol{D}^{\text{red,calc}} - \boldsymbol{D}^{\text{red,exp}}||$ instead of $||\boldsymbol{D}^{\text{calc}} - \boldsymbol{D}^{\text{exp}}||$.

The $\boldsymbol{a}$ that minimizes $||\boldsymbol{D}^{\text{red,calc}} - \boldsymbol{D}^{\text{red,exp}}||$ can then be found by deriving V (eq. 8.7) according to $\boldsymbol{a}$,

$$\boldsymbol{0} = \frac{\partial V}{\partial \boldsymbol{a}} = \sum_k^{\text{RDC}} w_k \left(\boldsymbol{a} \cdot \boldsymbol{C}_k(\boldsymbol{r}_k) \right) \boldsymbol{C}_k(\boldsymbol{r}_k) - \sum_k^{\text{RDC}} w_k D_k^{\text{red,exp}} \boldsymbol{C}_k(\boldsymbol{r}_k), \tag{8.8}$$

which can be rewritten as a system of five linear equations $\mathbf{A}\boldsymbol{a} = \boldsymbol{f}$, where

$$A_{h,h'} = \sum_k^{\text{RDC}} w_k C_{k,h} C_{k,h'} \quad \text{and} \tag{8.9}$$

$$f_{h'} = \sum_k^{\text{RDC}} w_k D_k^{\text{red,exp}} C_{k,h'}, \tag{8.10}$$

and solved by standard techniques such as lower upper (LU) decomposition[265] to yield $\boldsymbol{a}$.

The time limiting step of this computation is the summation over all interactions to obtain the 25 elements of $\mathbf{A}$ and the five elements of $\boldsymbol{f}$, which takes a time that is proportional to the number of interactions n. The matrix diagonalization takes constant time because $\mathbf{A}$ has constant size. Therefore, this algorithm scales as $\mathcal{O}(n)$, that is, linearly with the number of interactions.

8.2 Results

8.2.1 Calculation time

To explore how these two algorithms perform in practise, we fitted an alignment tensor to differently sized sets of RDCs measured for human

ubiquitin in uncharged bicelles[474] using each algorithm and measured the calculation time (Figure 8.2). The smallest two sets consist of the 68 N–H^N RDCs and a subset thereof of size 30. The 66 C^α–H^α, 64 C^α–C′ and 66 C′–N RDCs were added incrementally to generate sets of increasing size.

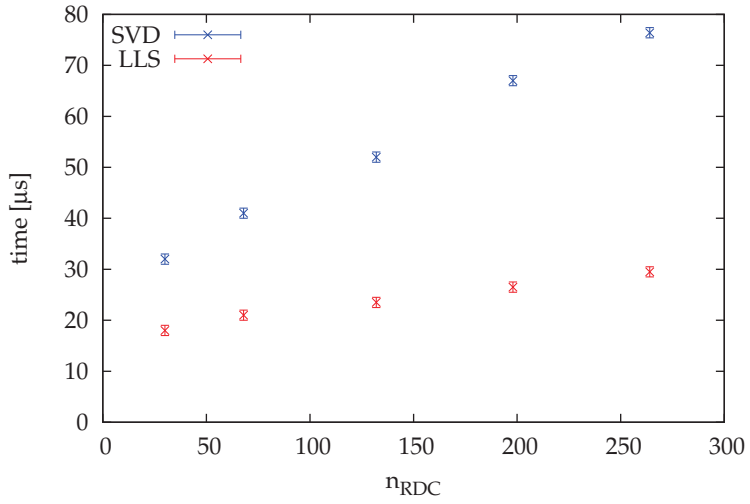


Figure 8.2 Time required to obtain a least squares fitted alignment tensor using the SVD or LLS algorithm as a function of the number of RDCs.

The calculation times shown in Figure 8.2 are for the single function that takes an $n \times 5$ matrix (as an array) of $C_{k,h}$, as well as weights and experimental values of the RDCs, and returns the five independent alignment tensor components. The functions for least squares fitting using SVD or LLS are part of the GROMOS++ suite of programs for preparing and analyzing biomolecular simulations.[475] The SVD and matrix diagonalization are implemented using GSL[476], which uses the Golub-Reinsch SVD algorithm and solves the system of linear equations via LU decomposition. Times were measured using `gettimeofday()`, which has a resolution of 1 μs, on a single core of a 2.67 GHz Intel/Nehalem processor.

As can be seen in Figure 8.2, the LLS algorithm vastly improves the speed of the tensor fitting procedure. However, the quadratic behavior of the SVD algorithm[472,476] is not observed for the numbers of RDCs examined here

because the prefactor of the linear term dominates for low numbers of RDCs. Furthermore, there is a constant term in both algorithms which contributes strongly for small n but becomes less relevant for larger n.

8.2.2 Stability

We also sought to explore how robust each algorithm is, as LLS is known to be less stable in the case of near-singular $\mathbf{A}$.[265] We constructed a case with only five RDCs, in which two of the corresponding inter-nuclear vectors are manipulated so as to become increasingly parallel. We find that for two vectors that differ by an angle of at least 10^{-5} rad, the calculated RDCs are equal to the input on at least the first 7 (11) significant digits using LLS (SVD). When the two vectors differ by an angle of 10^{-7} rad, the calculated RDCs remain equal to the input on the first 3 digits for LLS and the first 10 digits for SVD. Finally, when two of the inter-nuclear vectors differ by an angle of just 10^{-9} rad, some of the LLS back-calculated RDCs differ from the input values on the first digit (but the relative difference is below 10 %) while SVD still yields values that are correct on the first 8 significant digits.

In the extreme case of inter spin vectors that are parallel to within numerical noise, the matrix diagonalization in LLS may fail. A simple means of allowing a RDC-restrained MD simulation to proceed is to selectively catch the GSL error code and omit the tensor update for that MD step.

We emphasize that the situation described here is rather implausible: not only is it unlikely that any two inter-nuclear vectors in a real molecule will be parallel to such high precision, but most experimentally-measured datasets contain far more than five RDCs.

Other than for the cases described above where one or both of the fitting algorithms become unstable, the calculated RDCs as well as the fitted tensor components are identical up to more than 10 significant figures, which is larger than the precision with which RDCs are typically measured experimentally, thus the results of using either algorithm can be treated as equal.

8.2.3 Availability

The LLS algorithm for obtaining the least squares fitted alignment tensor has been implemented in the GROMOS biomolecular simulation software[477–479] and the related GROMOS++ suite of analysis programs[475] (program `fit_rdc`), all of which are written in C++.

8.3 Conclusions

We compare two least squares fitting algorithms for obtaining the alignment tensor that describes the best fit between a molecular structure and a set of experimental RDC, CSA or RQC values. The SVD fit algorithm is to compute the five independent alignment tensor components $\boldsymbol{a}$ in a time that scales as $\mathcal{O}(n^2)$, whereas the LLS algorithm scales as $\mathcal{O}(n)$. While the calculation time required for SVD is dominated by the prefactor of the linear term, and thus does not scale as predicted for small n, LLS is still significantly faster for small n. Importantly, we find LLS to also be rather robust even in the limiting case of five RDCs and near-parallel inter nuclear vectors, so that it is unlikely to fail under more realistic conditions. We therefore recommend the use of LLS during MD simulations, as the faster computation speed of this algorithm will be advantageous for restraining simulations to fit with large sets of NMR observables that may combine RDCs, CSAs and RQCs in the longer simulations that are now possible due to improved parallelization of code and increasing high performance computing resources.

9 Expression of molecular alignment by a weight function

The following chapter explores the representation of alignment by a weight function, where the value of the function expresses the relative abundance of ISVs having a given orientation relative to the magnetic field or vice versa. This weight function is expanded in two different bases (sections 9.1 and 9.2). For both representations potential energy functions and forces are derived which can be used for restraining MD simulations (sections 9.3 and 9.4). Applicability and usefulness of representing alignment by a weight function as well as the differences to the representation by an alignment tensor are assessed (section 9.6). Finally, the implementation of the above described methods and integration of that code into the GROMOS program for biomolecular simulations is explained (section 9.5).

When calculating RDCs and collectively expressing the weak alignment of molecules in solution by an alignment tensor, it is implicitly assumed that the molecules are rigid. This assumption is particularly problematic for flexible molecules like proteins as their dynamics are crucial for their function. These novel alignment representations and their application to restraining simulations are aimed at avoiding this implicit assumption and hence allowing for more faithful MD simulations.

9.1 Expansion in magnetic field vectors

The RDC interaction of a single spin pair that is statically aligned with respect to the magnetic field is given by eq. 7.56. Assume that multiple instances of the ISV are oriented differently relative to a magnetic field; one

can then express D_k by an explicit average over those statically aligned ISV. Equivalently, D_k can be expressed as an average of the interactions of one instance of one ISV with multiple magnetic field vectors (MFV) that may point in different directions, which is an equivalent but mathematically more convenient model.

Eq. 7.56 can be rewritten in a way that makes working with explicit magnetic field vectors $\boldsymbol{r}_h$ more natural:

$$D_k^{\text{static}} = D_k^{\max} \frac{1}{2} \left(3 \left(\frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k r_h} \right)^2 - 1 \right). \tag{9.1}$$

The average of eq. 9.1 over a set of N_H magnetic field vectors $r_H(\theta_h, \varphi_h)$ is

$$D_k = \frac{1}{N_H} \sum_h^{N_H} D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta_h, \varphi_h)). \tag{9.2}$$

Eq. 9.2 can be rewritten using $f(x_0) = \int f(x)\delta(x - x_0)\,\mathrm{d}x$ which leads to

$$D_k = \frac{1}{N_H} \sum_h^{N_H} \int^{2\pi} \int^{\pi} D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi))\delta(\theta - \theta_h, \varphi - \varphi_h) \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi, \tag{9.3}$$

that is, $D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta_h, \varphi_h))$ is written as the convolution of $D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi))$ and a δ function.

One can now express the normalised sum over δ functions by a distribution $g(\theta, \varphi)$, that serves as a general weight function that obviously does not impose any restrictions on the alignment of a single ISV.

$$D_k(\boldsymbol{r}_k, g) = \int^{2\pi} \int^{\pi} D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi)) g(\theta, \varphi) \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi. \tag{9.4}$$

The integral of g on the unit sphere is normalised to 1. In the limit of very large N_H, g can be treated as a smooth function of θ and φ. For typical NMR experiments, the assumption of large N_H is justified as the number of magnetic field vectors represents the (large) number of instances of this ISV in the sample, and as the weight function g represents the physical alignment of N_H inter spin vectors in space, a smooth function is appropriate.

Eq. 9.2 can be interpreted as an expansion of the weight function $g(\theta, \varphi)$ in the base of δ functions, which is orthonormalised and has infinite dimension.

A set of MFVs uniquely determines the resulting RDC interaction. The reverse is not true: there are sets of MFV that lead to the same RDC. For instance, equivalent sets of MFV are trivially generated by adding a second MFV to every existing one, thus halving their weight; vectors can be permuted arbitrarily; infinitely more equivalent sets can be created by continuously moving MVF on paths of equivalent points.

As will become clear in the following subsection, the minimal number of MFV that are sufficient to express general alignment is 5, if they are independent. Lower numbers can suffice to express alignment with certain constraints, down to a single vector that expresses static alignment. Thus, as each MFV requires at least two parameters θ and φ (in spherical coordinates the radius may be omitted), this representation is less efficient than the tensor representation as 10 rather than 5 parameters characterise the alignment representation.

If the same alignment representation is to be used for different ISV, one has to assume that they have the same dynamics, which is to say they are rigidly connected. In other words, molecular fragments and molecules in which all ISV share the same alignment representation are assumed to be rigid. This limitation is the same that underlies the tensor representation of alignment.

9.2 Expansion in spherical harmonics

The alignment of ISVs in a magnetic field may alternatively be parametrised by expanding the weight function $g(\theta, \varphi)$ in the basis of real spherical harmonic functions (RSH) $Y^{(\mathrm{r})}_{l,m}(\theta, \varphi)$.[480] The weight function $g(\theta, \varphi)$ (eq. 9.4) is then defined by the expansion coefficients $\boldsymbol{c}$:

$$g(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{l,m} Y^{(\mathrm{r})}_{l,m}(\theta, \varphi). \tag{9.5}$$

This expansion does not impose any restrictions on $g(\theta, \varphi)$, as the set of RSH forms a complete and orthonormalised basis for real $f(\theta, \varphi)$. D_k can therefore be rewritten as

$$D_k = \int_{}^{2\pi} \int_{}^{\pi} D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi)) \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{l,m} Y_{l,m}^{(\text{r})}(\theta, \varphi) \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi. \tag{9.6}$$

Furthermore, one can expand $D_k^{\text{static}}(\theta, \varphi)$ in terms of real spherical harmonics:

$$D_k = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} c_{l,m} c'_{l',m'} \int_{}^{2\pi} \int_{}^{\pi} Y_{l',m'}^{(\text{r})}(\theta, \varphi) Y_{l,m}^{(\text{r})}(\theta, \varphi) \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi. \tag{9.7}$$

Of all $c'_{l',m'}$, however, only the coefficients for $l' = 2$ may be non-zero in eq. 9.7, because $D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi))$ with $\boldsymbol{r}_k = (0, 0, r_k)$ or $\boldsymbol{r}_k = (0, 0, -r_k)$ is equal to $Y_{2,0}^{(\text{r})}(\theta, \varphi)$ up to a prefactor (compare eq. 7.56), and any real spherical harmonic function $Y_{l,m}^{(\text{r})}(\theta, \varphi)$ that has been rotated in Cartesian space can be expanded using only $Y_{l,m}^{(\text{r})}(\theta, \varphi)$ with the same l.[481,482] In other words, in the transformation matrix that rotates a function in Cartesian space by acting on its expansion coefficients, all entries with different l are zero, i.e., the transformation matrix is block diagonal sparse. These transformation matrices that rotate RSH are the equivalent of the more well known Wigner-D-matrices that rotate complex spherical harmonic functions.

As RSH form an orthogonal basis, only the coefficients $c_{2,m}$ need to be considered for the calculation of D_k. Eq. 9.6 can therefore be simplified to

$$D_k = \int_{}^{2\pi} \int_{}^{\pi} D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi)) \sum_{m=-2}^{2} c_{2,m} Y_{2,m}^{(\text{r})}(\theta, \varphi) \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi, \tag{9.8}$$

that is, the weight function g can be encoded with only the five coefficients $c_{2,m}$ without making any approximations about D_k. While $g(\theta, \varphi)$ expanded using only second order RSH functions may differ from the physical distribution of orientations, the difference cannot be observed through dipolar interactions. Any contributions from RSH with orders different from 2 cannot be observed, as the observed value is the convolution of the

exact $g(\theta, \varphi)$ and D_k^{static}. D_k^{static} a rotated version of $Y_{2,0}^{(\text{r})}(\theta, \varphi)$ and hence orthogonal to all RSH except those with $l = 2$.

For the same reason, five independent MFV are sufficient (and necessary) to express alignment without any approximation (see subsection 9.1).

Equation 9.8 is an expansion of $D_k(\boldsymbol{r}_k, g)$ in a five dimensional basis,

$$\frac{D_k(\boldsymbol{r}_k, g)}{D_k^{\max}} = \sum_{m=-2}^{2} c_{2,m} X_{k,m}(\boldsymbol{r}_k) = \boldsymbol{c}\boldsymbol{X}_k(\boldsymbol{r}_k), \tag{9.9}$$

where the basis functions $X_{k,m}$ are analytically given as

$$X_{k,m}(\boldsymbol{r}_k) = \int^{2\pi}\int^{\pi} \frac{D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta, \varphi))}{D_k^{\max}} Y_{2,m}^{(\text{r})}(\theta, \varphi) \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi, \tag{9.10}$$

and can be simplified to

$$\begin{aligned}
X_{k,1}(\boldsymbol{r}_k) &= 2\sqrt{\frac{3\pi}{5}} \sin^2\theta \cos\varphi \sin\varphi / r_k^2 &&= 2\sqrt{\frac{3\pi}{5}} xy / r_k^2, \\
X_{k,2}(\boldsymbol{r}_k) &= 2\sqrt{\frac{3\pi}{5}} \sin\theta \sin\varphi \cos\theta / r_k^2 &&= 2\sqrt{\frac{3\pi}{5}} yz / r_k^2, \\
X_{k,3}(\boldsymbol{r}_k) &= \sqrt{\frac{\pi}{5}} (3\cos^2\theta - 1) / r_k^2 &&= \sqrt{\frac{\pi}{5}} (3z^2 - 1) / r_k^2, \\
X_{k,4}(\boldsymbol{r}_k) &= 2\sqrt{\frac{3\pi}{5}} \sin\theta \cos\varphi \cos\theta / r_k^2 &&= 2\sqrt{\frac{3\pi}{5}} xz / r_k^2, \\
X_{k,5}(\boldsymbol{r}_k) &= \sqrt{\frac{3\pi}{5}} (\sin^2\theta \cos^2\varphi - \sin^2\theta \sin^2\varphi) / r_k^2 &&= \sqrt{\frac{3\pi}{5}} (x^2 - y^2) / r_k^2.
\end{aligned} \tag{9.11}$$

This basis is orthogonal but not normalised; the norm of each of the five basis functions is $\frac{4\pi}{5}$.

By the same argument that applied to MFV and the tensor representation, assigning multiple ISV the same alignment representation, i.e., the same expansion coefficients $\boldsymbol{c}$, assumes that they have the same dynamics. That is to say, molecular fragments with more than one ISV with the same alignment are considered rigid. Complete flexibility is only given when each ISV has an individual alignment representation.

Given a set of Cartesian molecular coordinates of at least five ISV and the corresponding measured RDCs, the coefficients $\boldsymbol{c}$ can be fitted to agree with the experiment in the same way that alignment tensor components can be determined (see subsection 7.2.6). That is, the penalty function (as a function of the expansion coefficients) has *one* minimum that can be computed analytically.

9.3 Restraining in terms of magnetic field vectors

The general strategy for restraining RDCs is, as explained in subsections 7.1.5 and 7.2.7, to choose a parametrisation of the alignment, update the representation in each MD step taking into account the current and possibly previous molecular coordinates, and determine the forces on molecular atoms. Forces are determined by devising a potential energy function (see subsection 7.1.5) that penalises the deviation of calculated (using atomic coordinates and measured RDCs) from experimental RDCs and deriving this function according to the atomic coordinates (see eq. 7.76).

In the case of MFV there is (even for a given number of vectors) no unique way of placing them to minimise e.g. a harmonic penalty function. In particular, the optimal set of vectors cannot be calculated in closed form as is possible for tensors components. Therefore, it is necessary to store the state, i.e., the coordinates of all MFV, between MD steps and propagate a representation through the simulation.

For updating the MFV three methods were considered:

1. minimising the penalty function with the MFV coordinates as variables and fixed atomic coordinates,
2. performing one MD step in which MFVs move in the 'field' created by atoms per MD step of the simulation, and
3. performing one stochastic dynamics (SD) step on the MFVs in an analogue fashion per MD step of the simulation.

For all three of these updating methods it is required to define 'forces' on the MFV due to the potential energy function. For that, MFV are treated

as 'magnetic field particles' that are situated on the surface of a unit sphere, that is, the lengths of the MFV are constant.

The force on one 'magnetic field particle' with coordinates $\boldsymbol{r}_h$ due to all RDC interactions k is

$$\boldsymbol{f}_h = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial \boldsymbol{r}_h} = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial D_k} \frac{\partial D_k}{\partial D_{k,h}^{\mathrm{static}}} \frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_h}, \tag{9.12}$$

where $\frac{\partial V_k^{\mathrm{RDC}}}{\partial D_k}$ depends on the chosen potential energy function (see e.g. eqs. 7.16 and 7.17). The second factor equates to $\frac{1}{N_H}$ as can easily be seen from eq. 9.2. The last term, the derivative of $D_{k,h}^{\mathrm{static}}$ (eq. 9.1), i.e., the RDC of an ISV $\boldsymbol{r}_k$ which is statically aligned with a magnetic field vector $\boldsymbol{r}_h$ is

$$\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_h} = 3D_{k,h}^{\max} \frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k r_h} \frac{1}{r_h} \left(\frac{\boldsymbol{r}_k}{r_k} - \frac{\boldsymbol{r}_h (\boldsymbol{r}_k \cdot \boldsymbol{r}_h)}{r_k r_h^2} \right), \tag{9.13}$$

and can be simplified, using the fact that r_h is constant:

$$\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_h} = 3D_{k,h}^{\max} \frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k r_h} \frac{\boldsymbol{r}_k}{r_k r_h}, \tag{9.14}$$

and, more specifically, that it has length 1:

$$\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_h} = 3D_{k,h}^{\max} \frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k} \frac{\boldsymbol{r}_k}{r_k}. \tag{9.15}$$

The force on magnetic field particles is always in the direction of the ISV.

If the MFV are to be updated by minimising V^{RDC}, at each step of the simulation the MFV coordinates are optimised starting at the final coordinates of the previous step using a numerical optimisation algorithm until convergence of V^{RDC} is achieved up to a chosen threshold. It is required to provide an initial guess for the MFV coordinates which is used in the first step. There is no guarantee that the optimisation finds the global minimum.

Performing MD or SD with the magnetic field particles serves to escape local minima. The 'particles' are assigned a 'mass' which may differ between individual particles and a 'velocity'. The velocities need to be initialised in

the beginning of the simulation such that they follow a Boltzmann distribution that matches a chosen temperature.

At each MD step of the simulation, *one* MD step or SD step is performed on the magnetic field particles using the stored masses, positions, and velocities (see subsection 7.1.3). In a second substep constraints are enforced, viz. the length of all MFV is reset to 1. As these constraints are not coupled, this can be done in a single SHAKE step; it is not necessary to resolve the constraints iteratively as is done e.g. for bond lengths (see subsection 7.1.4). The new state, i.e., coordinates, velocities and—in the case of SD—four random values, is stored for the next simulation cycle.

Once the alignment parameters are updated using either of the three above described methods, the forces $\boldsymbol{f}_i$ on atoms i that are part of ISVs $\boldsymbol{r}_k$ are calculated by deriving eq. 9.2 according to the coordinates $\boldsymbol{r}_i$ of the respective atom i:

$$\boldsymbol{f}_i = -\sum_k \sum_h^{N_H} \frac{\partial V_k^{\mathrm{RDC}}}{\partial \boldsymbol{r}_i} = -\sum_k \sum_h^{N_H} \frac{\partial V_k^{\mathrm{RDC}}}{\partial D_k} \frac{\partial D_k}{\partial D_{k,h}^{\mathrm{static}}} \frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_k} \frac{\partial \boldsymbol{r}_k}{\partial \boldsymbol{r}_i} \tag{9.16}$$

Analogously to eq. 9.12, the first factor in eq. 9.16 depends on the employed penalty function, and the second factor is $\frac{1}{N_H}$. The third factor is

$$\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_k} = 3D_k^{\max} \left[\frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k^2 r_h^2} \boldsymbol{r}_h - \frac{1}{2} \frac{\boldsymbol{r}_k}{r_k^2} \left(5 \left(\frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k r_h} \right) - 1 \right) \right], \tag{9.17}$$

and the fourth factor is

$$\frac{\partial \boldsymbol{r}_k}{\partial \boldsymbol{r}_i} = \delta_{ik_2} \mathbb{1} - \delta_{ik_1} \mathbb{1} \tag{9.18}$$

and distinguishes between the two atoms k_1 and k_2 that form the ISV $\boldsymbol{r}_k = \boldsymbol{r}_{k_2} - \boldsymbol{r}_{k_1}$. Eq. 9.17 can be simplified under the assumption that r_k is constant.

$$\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_k} = 3D_k^{\max} \frac{\boldsymbol{r}_k \cdot \boldsymbol{r}_h}{r_k^2 r_h^2} \boldsymbol{r}_h \tag{9.19}$$

This is a reasonable assumption in in the case of MD simulations, where it is common to constrain the lengths of bonds that can be described by a potential with a large force constant and therefore vibrate with high frequencies and low amplitudes.

Time and ensemble averaging can easily be introduced by expressing the potential energy function V^{RDC} as a function of the averaged $\langle D_k \rangle_{\mathrm{time/ensemble}}$ instead of the instantaneous D_k. The forces $\boldsymbol{f}_h^{\mathrm{av}}$ on the magnetic field particles h and the forces $\boldsymbol{f}_i^{\mathrm{av}}$ on atoms i due to averaged RDC interactions are then

$$\boldsymbol{f}_h^{\mathrm{av}} = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial \langle D_k \rangle} \frac{\partial \langle D_k \rangle}{\partial D_k} \frac{\partial D_k}{\partial D_{k,h}^{\mathrm{static}}} \frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_h} \tag{9.20}$$

and

$$\boldsymbol{f}_i^{\mathrm{av}} = -\sum_k \sum_h^{N_H} \frac{\partial V_k^{\mathrm{RDC}}}{\partial \langle D_k \rangle} \frac{\partial \langle D_k \rangle}{\partial D_k} \frac{\partial D_k}{\partial D_{k,h}^{\mathrm{static}}} \frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_k} \frac{\partial \boldsymbol{r}_k}{\partial \boldsymbol{r}_i}. \tag{9.21}$$

All these factors have been given above: $\frac{\partial V_k^{\mathrm{RDC}}}{\partial \langle D_k \rangle}$ in eq. 7.23 for an harmonic potential, although other potentials may be used; $\frac{\partial \langle D_k \rangle}{\partial D_k}$ in eq. 7.24 or eq. 7.29 for time and ensemble averaging respectively; $\frac{\partial D_k}{\partial D_{k,h}^{\mathrm{static}}}$ is $\frac{1}{N_H}$ as mentioned above; $\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_h}$ in eq. 9.13; $\frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{r}_k}$ in eq. 9.17; and $\frac{\partial \boldsymbol{r}_k}{\partial \boldsymbol{r}_i}$ in eq. 9.18. As a side effect of this redundancy, time and ensemble averaged forces may be computed at almost no extra cost.

9.4 Restraining in terms of spherical harmonics

Restraining MD simulations using RDCs that are expressed by a weight function expanded in spherical harmonics (see section 9.2) works analogously to restraining using the MFV representation. However, as pointed out in section 9.2, the penalty function V^{RDC} as a function of the five expansion coefficients $\boldsymbol{c}$ and with the atom coordinates and experimental RDCs as parameters has one single minimum that can be found analytically. Hence, there is no chance of choosing the coefficients $\boldsymbol{c}$ such that they correspond to a local minimum of V^{RDC} and we can limit ourselves to determining them such that V^{RDC} is minimised. Performing MD or SD on $\boldsymbol{c}$ in this case would yield coefficients that are further away from the global minimum and lead to a worse fit; this may, however, be of interest for comparison with the

previously introduced MFV representation, and in the following all required formulas will be given.

In each MD step of the simulation, the coefficients $\boldsymbol{c}$ can be obtained by minimisation using either of the two algorithms explained in subsection 7.2.6, or by performing an MD or SD step as explained above. For the latter two options, the 'forces' on the coefficients are required which can be written as one five dimensional vector $\boldsymbol{f}_c$.

$$\boldsymbol{f}_c = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial \boldsymbol{c}} = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial D_k} \frac{\partial D_k}{\partial \boldsymbol{c}} \tag{9.22}$$

Here, $\frac{\partial V_k^{\mathrm{RDC}}}{\partial D_k}$ depends on the chosen potential, and $\frac{\partial D_k}{\partial \boldsymbol{c}}$ is (compare eq. 9.8)

$$\frac{\partial D_k}{\partial \boldsymbol{c}} = \int^{2\pi} \int^{\pi} D_k^{\mathrm{static}}(\boldsymbol{r}_k, \boldsymbol{r}_H(\theta,\varphi)) \boldsymbol{Y}_2^{(\mathrm{r})}(\theta,\varphi) \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi, \tag{9.23}$$

where $\boldsymbol{Y}_2^{(\mathrm{r})}$ is the vector of second order real spherical harmonics. Simplified components of $\frac{\partial D_k}{\partial \boldsymbol{c}}$ can be derived from eq. 9.11 by multiplication with $D_k^{\max}$.

Given the forces $\boldsymbol{f}_c$, the coefficients $\boldsymbol{c}$ can be updated as described above. The forces $\boldsymbol{f}_i$ on atoms i are then computed using

$$\boldsymbol{f}_i = -\sum_k \frac{\partial V^{\mathrm{RDC}}}{\partial \boldsymbol{r}_i} = -\sum_k \frac{\partial V^{\mathrm{RDC}}}{\partial D_k} \frac{\partial D_k}{\partial D_k^{\mathrm{static}}} \frac{\partial D_k^{\mathrm{static}}}{\partial \boldsymbol{r}_k} \frac{\partial \boldsymbol{r}_k}{\partial \boldsymbol{r}_i}, \tag{9.24}$$

of which the first factor depends on the chosen potential, the second is $\frac{\partial D_k}{\partial D_k^{\mathrm{static}}} = 2\sqrt{\frac{\pi}{5}}$ (compare eq. 9.8), factor-3 is given in eq. 9.19 and the fourth factor has already been mentioned in eq. 9.18.

In complete analogy to eqs. 9.20 and 9.21, restraining can be extended to include time or ensemble averaging with

$$\boldsymbol{f}_c^{\mathrm{av}} = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial \langle D_k \rangle} \frac{\partial \langle D_k \rangle}{\partial D_k} \frac{\partial D_k}{\partial D_{k,h}^{\mathrm{static}}} \frac{\partial D_{k,h}^{\mathrm{static}}}{\partial \boldsymbol{c}} \tag{9.25}$$

and

$$\boldsymbol{f}_i^{\mathrm{av}} = -\sum_k \frac{\partial V_k^{\mathrm{RDC}}}{\partial \langle D_k \rangle} \frac{\partial \langle D_k \rangle}{\partial D_k} \frac{\partial D_k}{\partial D_k^{\mathrm{static}}} \frac{\partial D_k^{\mathrm{static}}}{\partial \boldsymbol{r}_k} \frac{\partial \boldsymbol{r}_k}{\partial \boldsymbol{r}_i}. \tag{9.26}$$

All derivatives have been specified above.

9.5 Implementation

In the context of this work, a force term was implemented and integrated into the GROMOS program for biomolecular simulation.[483] The implementation allows for using all three above described parametrisations of alignment, i.e., representations of the magnetic field, in combination with the three above described methods of updating the alignment parameters between MD steps of the simulation. Restraining can be done with respect to instantaneous RDCs, time averaged RDCs (eq 7.20), or using a biquadratic potential (eq 7.21). As potentials users can choose between flat-bottom potentials (eq. 7.15) and their limiting case of harmonic potentials (eq. 7.14). All of the above choices can be combined arbitrarily, thus offering large flexibility.

In GROMOS, the sequence of actions to be performed in each MD step is stored in an object of type `Algorithm_Sequence`, which is defined as `class Algorithm_Sequence: public vector<Algorithm *>`. `Algorithm` is an abstract base class that is inherited by each algorithm. Important examples for algorithms are the calculation of forces, integration schemes, multiple constraint solvers, thermostats, and barostats. During the initialisation of a simulation a single object of type `Algorithm_Sequence` is populated according the user's choices and used throughout the simulation.

One special algorithm is the type `Forcefield` which is defined as `class Forcefield: public vector<Interaction *> public Algorithm`, i.e., it is an `Algorithm` as well as a vector of pointers to `Interaction`. `Interaction` itself is a virtual base class that is inherited by every interaction that is part of the forcefield as for example the potentials describing bonds, angles, dihedral as well as nonbonded interactions and special interactions (see subsection 7.1.2).

One of these interactions is the `RDC_Restraint_Interaction`, defined as `class RDC_Restraint_Interaction: public Interaction` which has been written as a part of this work and will be explained in more detail in the remainder of this section.

Each interaction class, i.e., each class that inherits `Interaction`, defines a constructor that, apart from constructing the object, sets the name to a preset

value, and two functions `int init(...)` and `int calculate_interactions(...)`. `init` is called exactly once while the simulation is set up; `calculate_interactions` is called once in every MD cycle; this is where the interactions are implemented. In addition to these, the Interaction base class implements a function for printing the cumulated time of a timer, which is used for measuring the time spent in `calculate_interactions`.

The following code summarises the header file of the `RDC_Restraint_Interaction` class. Here, the name of the interaction is hard coded as 'RDCRestraint'. The `init` function has no functionality beyond printing the chosen restraining mode.

```
1  class RDC_Restraint_Interaction: public Interaction{
2  public:
3    RDC_Restraint_Interaction(): Interaction("RDCRestraint")
     % {}
4
5    virtual int init(topology::Topology &topo,
6             configuration::Configuration &conf,
7             simulation::Simulation &sim,
8             std::ostream &os = std::cout,
9             bool quiet = false);
10
11   virtual int calculate_interactions(topology::Topology &
     %topo,
12                      configuration::Configuration &conf,
13                      simulation::Simulation &sim);
14 };
```

In `calculate_interactions(...)` one out of three functions is called depending on the chosen magnetic field representation. Each of the functions contains a second switch statement to distinguish between the modes of updating the magnetic field representation.

As infrastructure, there are global structures that store data that is either constant throughout the simulation or affects more than one interaction type. The RDC interaction term reads from the global list of atomic coordinates which is stored in the `Configuration` class, and writes to the global energies (specifically the special energies) and forces on atoms.

RDC specific data is divided into three subsets each of which is stored

in a `struct` of which there is either a single instance or exactly one vector of instances. In `Simulation`, constant data that specifies the simulation mode is stored. In `Topology`, there is a vector in which for each RDC the constant parameters (i.e., the two atoms, weight, experimental value, and magnetogyric ratios) are stored. In `Configuration` the continually updated state of the magnetic field representation and the RDCs is stored, including current and averaged RDCs, positions, masses and velocities (if applicable to the updating method). Of these RDC specific structures, only the one located in `Configuration` is written to.

Each of the nine combinations follows the same general scheme:

1. the 'forces' on the magnetic field representation are computed,
2. the representation is updated, and
3. the 'forces' on atoms according to the updated magnetic field are calculated. In this step, the energies and forces due to the restraining potential are written to the global structure storing them. It is important for the overall program performance to write to globally shared variables as infrequently as possible. This is especially important in parallelised programs.

As the forces on the magnetic field representation are required when doing MD as well as SD, separate functions `void _calculate_forces_vectors_MF(...)`, `void _calculate_forces_tensor_T(...)`, and `void _calculate_forces_clm_SH(...)` were written for all magnetic field representations.

Furthermore, for every restraining type the forces on atoms need to be computed for which three functions `void _calculate_forces_atoms_MF(...)`, `void _calculate_forces_atoms_T(...)`, and `void _calculate_forces_atoms_SH(...)` were implemented.

In the cases of the tensor and spherical harmonic representations, there is a separate function to calculate the optimal tensor components and spherical harmonic expansion coefficients, implementing the LLS algorithm explained in chapter 8.

For the optimisation of the magnetic field vector positions from a given starting configuration, the conjugate Fletcher Reeves algorithm is used as implemented by the Gnu Scientific Library (GSL).[476] The GSL multidimensional

minimisation framework[324] implements different algorithms to minimise arbitrary functions of more than one variable. For this, three additional functions need to be provided that return a) the function value, b) the function's first derivatives, and c) both. There are optimisation algorithms available that do not require derivatives, at the cost of slower convergence. As the MFVs have constant lengths their positions can be optimised in spherical coordinates, which reduces the number of variables from $3N$ to $2N$.

Eq. 9.8 suggests that numerical integration is required for the representation of the magnetic field by a weight function expanded in spherical harmonics. Numerical integration of functions of θ and φ can be done using the quadrature scheme devised by Lebedev.[484,485] For that, the function is evaluated at a number of tabulated points on the sphere, and multiplied with tabulated weights; these results are then summed to obtain the integral. The number and position of points (collectively called a grid) depend on the function to integrate. There are grids available in literature, that range from order 3 to order 131. A grid of order n is guaranteed to exactly integrate spherical harmonics of up to order n. Eq. 9.8 integrates over the product of two functions which are both second order spherical harmonics (or a sum thereof), therefore, a grid of order 5 and with 14 points suffices to integrate the product exactly. As proven by eq. 9.11, however, the integrals can be precalculated analytically and therefore be computed much more efficiently.

9.6 Equivalence of spherical harmonics and tensor representations

The representation of alignment by an alignment tensor (subsection 7.2.4 and by a weight function that is expanded in real spherical harmonics (section 9.2) are completely equivalent as will be shown in the following.

In the alignment tensor representation, the k^{th} RDC $D_k = D_k(\boldsymbol{r}_k, \boldsymbol{a})$ of a structure can in expressed as (eq. 7.68)

$$\frac{D_k(\boldsymbol{r}_k, \boldsymbol{a})}{D_k^{\max}} = \sum_h^5 a_h C_{k,h}(\boldsymbol{r}_k) \tag{9.27}$$

$$= \boldsymbol{a}^{\mathsf{T}} \boldsymbol{C}_k(\boldsymbol{r}_k), \tag{9.28}$$

where $\boldsymbol{a}$ are coefficients and the basis functions $\boldsymbol{C}_k(\boldsymbol{r}_k)$ in Cartesian and spherical coordinates are (eq. 8.5)

$$\begin{aligned}
C_{k,1}(\boldsymbol{r}_k) &= (x^2 - z^2)/r_k^2 &&= (\sin^2\theta\cos^2\varphi - \cos^2\theta)/r_k^2\\
C_{k,2}(\boldsymbol{r}_k) &= (y^2 - z^2)/r_k^2 &&= (\sin^2\theta\sin^2\varphi - \cos^2\theta)/r_k^2\\
C_{k,3}(\boldsymbol{r}_k) &= 2xy/r_k^2 &&= 2\sin^2\theta\cos\varphi\sin\varphi/r_k^2\\
C_{k,4}(\boldsymbol{r}_k) &= 2xz/r_k^2 &&= 2\sin\theta\cos\varphi\cos\theta/r_k^2\\
C_{k,5}(\boldsymbol{r}_k) &= 2yz/r_k^2 &&= 2\sin\theta\cos\theta\sin\varphi/r_k^2.
\end{aligned}$$

These basis functions are not normalised. ($\int\int \boldsymbol{C}_k\boldsymbol{C}_k^\mathsf{T}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi = \frac{16\pi}{15}\mathbb{1}$)

In the spherical harmonic representation, the k^{th} RDC $D_k = D_k(\boldsymbol{r}_k, g)$ of a structure can be expressed as (eq. 9.9)

$$\frac{D_k(\boldsymbol{r}_k, g)}{D_k^{\max}} = \int_{}^{2\pi}\int_{}^{\pi} \frac{D_k^{\text{static}}(\boldsymbol{r}_k, \boldsymbol{r}_h(\theta,\varphi))}{D_k^{\max}} g(\theta,\varphi)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi \tag{9.29}$$

$$= \sum_{m=-2}^{2} c_m X_{k,m}(\boldsymbol{r}_k) \tag{9.30}$$

$$= \boldsymbol{c}^\mathsf{T}\boldsymbol{X}_k(\boldsymbol{r}_k), \tag{9.31}$$

where $\boldsymbol{c}$ are coefficients and the basis functions $\boldsymbol{X}(r_k(\theta,\varphi))$ in Cartesian and spherical coordinates are (eq. 9.11)

$$\begin{aligned}
X_{k,1}(\boldsymbol{r}_k) &= 2\sqrt{\frac{3\pi}{5}}xy/r_k^2 &&= 2\sqrt{\frac{3\pi}{5}}\sin^2\theta\cos\varphi\sin\varphi/r_k^2\\
X_{k,2}(\boldsymbol{r}_k) &= 2\sqrt{\frac{3\pi}{5}}yz/r_k^2 &&= 2\sqrt{\frac{3\pi}{5}}\sin\theta\cos\theta\sin\varphi/r_k^2\\
X_{k,3}(\boldsymbol{r}_k) &= \sqrt{\frac{\pi}{5}}(3z^2 - 1)/r_k^2 &&= \sqrt{\frac{\pi}{5}}(3\cos^2\theta - 1)/r_k^2\\
X_{k,4}(\boldsymbol{r}_k) &= 2\sqrt{\frac{3\pi}{5}}xz/r_k^2 &&= 2\sqrt{\frac{3\pi}{5}}\sin\theta\cos\varphi\cos\theta/r_k^2\\
X_{k,5}(\boldsymbol{r}_k) &= \sqrt{\frac{3\pi}{5}}(x^2 - y^2)/r_k^2 &&= \sqrt{\frac{3\pi}{5}}(\sin^2\theta\cos^2\varphi - \sin^2\theta\sin^2\varphi)/r_k^2.
\end{aligned}$$

These basis functions are not normalised either. ($\int\int \boldsymbol{X}_k \boldsymbol{X}_k^\intercal \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi = \frac{16\pi^2}{25}\mathbb{1}$)

As the RDC D_k is, of course, independent of the alignment representation, it is

$$\boldsymbol{a}^\intercal \boldsymbol{C}_k = \boldsymbol{c}^\intercal \boldsymbol{X}_k, \tag{9.32}$$

which can be solved for the coefficients $\boldsymbol{c}$ and $\boldsymbol{a}$.

$$\boldsymbol{a}^\intercal \mathbf{R} = \boldsymbol{c}^\intercal \qquad\qquad \boldsymbol{a}^\intercal = \boldsymbol{c}^\intercal \mathbf{R}^{-1} \tag{9.33}$$

That is, one can perform a change of basis from the tensor representation (eq. 9.28) to the spherical harmonic representation (eq. 9.31) by multiplying the coefficients with the 5×5 transformation matrix $\mathbf{R}$. The components R_{ij} are given by

$$R_{ij} = \frac{\int^{2\pi}\int^{\pi} C_{k,i}(\theta,\varphi) X_{k,j}^\intercal(\theta,\varphi) \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\varphi}{\|X_k\|^2} \tag{9.34}$$

with

$$\mathbf{R} = \begin{pmatrix} 0 & 0 & -\frac{1}{2}\sqrt{\frac{5}{\pi}} & 0 & \frac{1}{2}\sqrt{\frac{5}{3\pi}} \\ 0 & 0 & -\frac{1}{2}\sqrt{\frac{5}{\pi}} & 0 & -\frac{1}{2}\sqrt{\frac{5}{3\pi}} \\ \sqrt{\frac{5}{3\pi}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sqrt{\frac{5}{3\pi}} & 0 \\ 0 & \sqrt{\frac{5}{3\pi}} & 0 & 0 & 0 \end{pmatrix}. \tag{9.35}$$

The reverse operation is done by multiplication with $\mathbf{R}^{-1}$.

9.7 Conclusion

As shown in sections 9.3 and 9.4 both alignment representations considered in this chapter suffer from the same shortcoming as the commonly used alignment tensor representation, that is, they implicitly assume all ISV that share the same alignment representation to be statically aligned. This can be generalised to a stronger statement: it is not possible to represent alignment without assuming that all ISV that are represented by the same alignment

parametrisation have the same dynamics and hence have a static relative orientation. This problem can be avoided in the analysis on NMR experiments by using multiple alignment media which gives enough information to individually express the alignment of each ISV. However, in restraining MD simulations there is no option to obtain additional information to lift the degeneracy of RDCs other than combining them into groups of at least five which neglects some of the internal degrees of freedom of the molecule. This is a restriction that arises from the number of free parameters that are required to express alignment (with respect to RDCs); it cannot be overcome by a different choice of the representation.

It is an obvious next step to explore ways of subdividing molecules into individually aligned blocks such that a maximum part of the overall internal flexibility is recovered and the side effects of implied rigidity are minimised. This will be undertaken in chapter 10.

The computational effort of representing alignment by a set of magnetic field vectors is significantly higher while the expansion of a weight function in RSH can be implemented entirely equivalent to (and equally efficient as) the alignment tensor representation. As all three representations are equally well suited to describe alignment, there is no advantage of either of the above introduced representations over the well known alignment tensor.

The option of restraining RDCs using different combinations of potentials and time averaging options has been implemented as part of GROMOS and hence made available to the biomolecular simulation community.

10 Alignment Blocks

One of the biggest challenges in interpreting NMR spectra with the help of RDCs as well as in restraining MD simulations with experimental RDCs is the implied static relative alignment of all ISVs that share the same alignment representation. That is, a molecule or molecular domain defined by the set of ISVs it incorporates is assumed to be rigid.

In the analysis of the dynamics of ISVs and in structure determination, it is common practice to subdivide molecules into subunits whose alignment is treated independently. In de novo structure determination (see subsection 7.2.5), these units are chosen to be as small as possible, as smaller building blocks grant higher flexibility. Blocks of two or three residues (depending on the residue type and hence the number of RDCs per residue) are common for data measured in one alignment medium.[429,451] In the determination of dynamics, the focus lies on the relative motion of larger known structural domains, as for example in the two-domain protein calmodulin[486], the three-domain factor H[487] or the dimeric HIV-1 protease[488] which has two segments in each monomer. Here each molecular domain is assigned one alignment block.

As the alignment tensor which is the most concise alignment description is defined by five independent parameters, there are at least five independent experimental RDCs required to lift the degeneracy of ISV orientations that lead to the same measured value.[430] This restriction is universal and does not depend on the chosen alignment parametrisation. It is not possible to devise a different representation or any other means of lowering the number of required experimental RDCs.

The minimal number of ISVs per domain can be reduced to 1 by measuring RDCs in at least five alignment media, thus yielding five or more independent RDCs per ISV.[435,444,457,458] Mixed approaches with more than one ISV per domain and more than one alignment medium are possible under the condition that five or more independent experimental RDCs are available for

each molecular domain that shares the same alignment, i.e., the product of the number of alignment media and the number of ISVs in a given domain needs to be five or greater.

In this chapter the concept of subdividing molecules into individually aligned blocks will be applied to the restraining of MD simulations. Restraining each block with an alignment parametrisation that is specifically fitted to this (presumably rigid) domain should lead to alignment representations that reflect the dynamics of each subdomain more realistically and through restraining lead to molecular structures that are closer to the experimental structures.

10.1 Restraining using alignment blocks

Subdividing a molecule into alignment blocks and restraining them independently is mathematically and computationally easily derived from the previously explained restraining (see subsection 7.2.7).

Eqs. 7.75 and 7.76 define in a general way the restraining potential V_k^{RDC} of a single RDC k and the 'force' $\boldsymbol{f}_i$ on a single atom i. These terms are translated to the potential $V^{(\mathrm{tot})}$ and forces $\boldsymbol{f}_i^{(\mathrm{tot})}$ in the case of multiple alignment blocks by summing over all B alignment blocks b and the RDCs k in each group:

$$V^{(\mathrm{tot})} = \sum_{b}^{B} \sum_{k}^{\mathrm{RDC}_b} V_k^{\mathrm{RDC}} \tag{10.1}$$

$$\boldsymbol{f}_i^{(\mathrm{tot})} = -\sum_{b}^{B} \sum_{k}^{\mathrm{RDC}_b} \frac{\partial V_k^{\mathrm{RDC}}}{\partial \boldsymbol{r}_i}. \tag{10.2}$$

It is possible to choose alignment blocks to overlap arbitrarily. The expected restraining forces on each ISV should not depend on the number of blocks the respective RDC is part of. Therefore, the weights of each RDC in each group should be adjusted such that the sum over all weights of one RDC in different groups is 1 (or any chosen value).

In the following, the tensor representation of alignment will be used—the other above introduced representations are equivalent and lead to the same results.

It seems desirable to choose small alignment blocks in order to allow for maximum flexibility. However, at least five RDCs per block are required to lift the degeneracy. Choosing blocks of exactly size five is possible but ineffective because an alignment tensor can be fitted to any five RDCs without deviation of the calculated from the experimental RDCs and hence the resulting forces on atoms always vanish.

While restraining individual blocks of a molecule can be expected to improve these structures internally, any information on relative alignment of molecular domains in the experiment is lost. That is, orientational restraints impose restraints only on ISV that are treated by the same alignment representation.

However, although the relative alignment of separate blocks is unrestrained, their fitted alignment tensors may be analysed and compared. Similar axial and rhombic components (eqs. 7.70 and 7.71) of consecutive blocks indicate a rigid connection between them; differing axial and rhombic components suggest independent motion in the experiment. Comparing the Euler angles that diagonalise both tensors holds only information on the orientation of both blocks in the simulation (which can be obtained more easily by other means) and reveals no information on their relative experimental alignment.

The implementation of RDC based orientational restraining described in section 9.5 has been extended by the option to group RDCs into blocks and represent their alignment independently. The choice of blocks is very general: there is no constraint on the number, size (other than being larger that 5), or overlap of defined blocks. Blocks do not have to consist of RDCs in consecutive residues. Not every RDC has to be included in a group, but as omitted RDCs are entirely unrestrained problems may occur in subsequent analysis, therefore this is not recommended.

10.2 Application to $^2F1^3F1$ from human fibronectin (2CKU)

Fibronectin is an extracellular protein that plays an important role in wound healing and other processes. It forms a chain of modules, of which by far

the most abundant modules are F1, F2 and F3. The secondary structure of each of these modules is based on β-sheets in the complete absence of α-helices.[489–491] While the structure of individual modules is well understood, the relative orientation of the modules is an open question. It is assumed that the functionality of fibronectin crucially depends on the alignment and dynamics of different modules.[491]

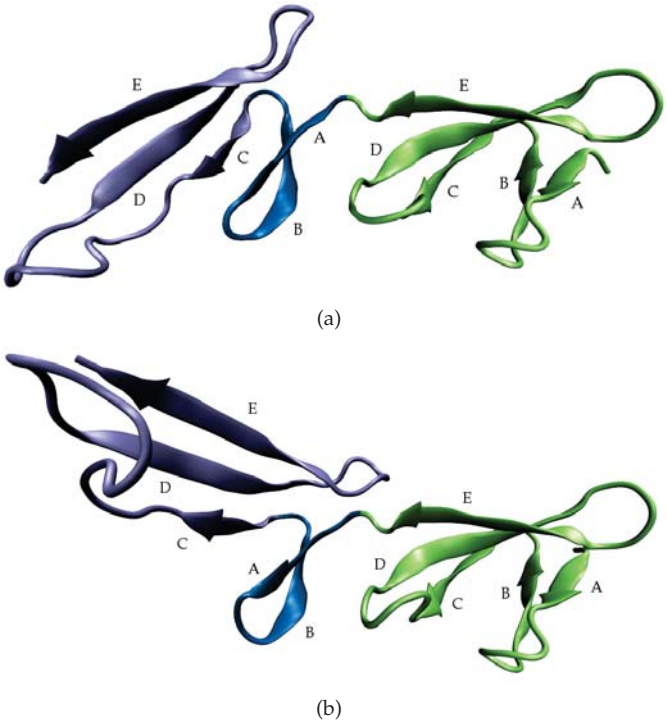


Figure 10.1 Two NMR structures (frames 9 (a) and 8 (b)) of the ^{2}F1^{3}F1 module pair of human fibronectin from 2CKU. Module 1 (residues 1–45) in green, module 2 (residues 46–57 and 58–89) in blue and purple respectively. (a) linear alignment of the two modules, (b) angled alignment.

The F1 module consists of 5 β-strands (see fig. 10.1) which are labelled from A to E. A and B as well as C, D and E form a β-sheet. The first β-sheet

folds over the second, partly due to the interaction of two aromatic residues in strands B and C.

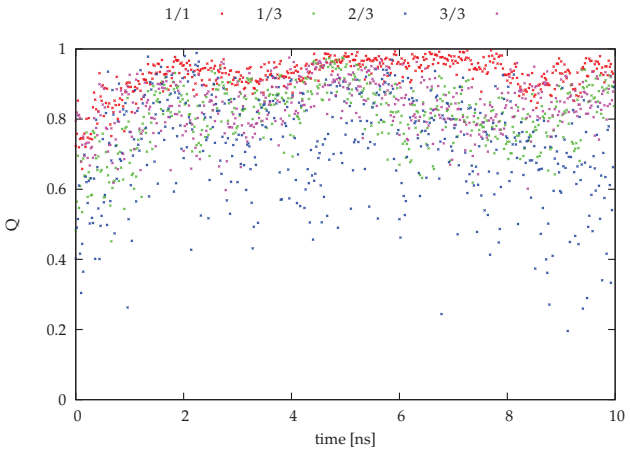
There is an NMR structure of the ^{2}F1^{3}F1 module pair of human fibronectin available (2CKU[492]). The reported structure comprises 15 frames which represent the structural variability, specifically the angle between the two modules. Furthermore, in the second module ^{3}F1 (residues 46–89 of the pair) the alignment of the two β-sheets (residues 46–57 and 58–89) is subject to variation (compare fig. 10.1 (a) and (b)). This structure has been refined using experimental RDCs and a single alignment tensor.

In order to explore the effect of restraining an MD simulation with different numbers of alignment blocks on the structure and on the accuracy with which experimental RDCs are reproduced, the described fibronectin module pair has been simulated for 10 ns.

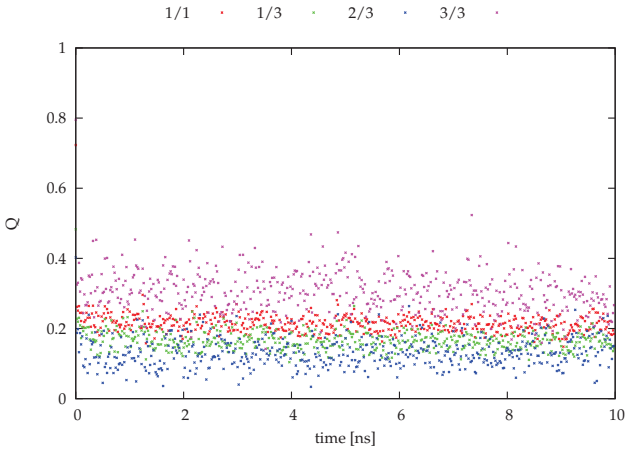
The initial coordinates were taken from the first frame of the NMR model structure deposited in the PBD. The simulation of the protein in a box of water was carried out with the GROMOS program for molecular simulation using the 54a7[493] force field at a temperature of 298 K. Three simulations were performed: unrestrained with respect to RDCs or restrained with experimental backbone N–H^N RDCs,[492] where either the whole molecule was represented by the same alignment representation or three blocks (residues 1–45, 46–57, and 58–89, see fig. 10.1) were represented individually. These blocks were chosen using prior knowledge of the protein's secondary structure. For the restraints harmonic potentials with a force constant of $3\,\mathrm{kJ\,ps^{-2}}\,N_A$ were used. Smaller values of the force constant lead to smaller effects while the simulation becomes unstable between 10 and $15\,\mathrm{kJ\,ps^{-2}}\,N_A$.

In the following the results of the two restrained simulations will be compared with the unrestrained simulation as reference.

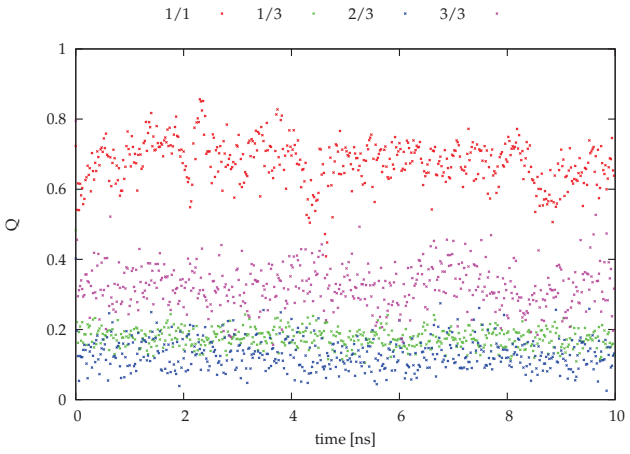
Figures 10.2 (a) to 10.2 (c) show the Q values (eq. 7.80) of the three simulation modes. In each case Q values were calculated based on one alignment block (red) and the three subblocks that are used for restraining (green, blue, and purple). Restraining the protein as a single block leads to good Q values based on a single block as well as each of the three subblocks. Restraining the three subblocks leads to equally low individual Q values but the quality of the fit of the whole molecule is worse (0.6–0.7 instead of 0.2–0.3). This indicates that the three substructures fit the experiment



(a)



(b)



(c)

Figure 10.2 Q values of a simulation of fibronectin, (a) unrestrained, (b) restrained using one alignment block, and (c) restrained using three alignment blocks and with respect to one (red [RDCs 1–59, residues 1–89]) and three (green [RDCs 1–29, residues 1–45], blue [RDCs 30–39, residues 46–57], purple [RDCs 40–59, residues 58–89]) alignment blocks

equally well in both simulation modes. The different levels of agreement between experiment and simulation of the three subblocks can be explained by the fragments sizes and the lower number of residues that are part of β-strands in module ^{3}F1 (see fig. 10.3 (c)). Restraining with only one tensor limits the conformational space to those structures that are rigid and match the experiment while individual restraining does not take into account the relative alignment of the substructures and hence allows the secondary structure elements to move relative to each other. Hence, in the simulation restrained with 3 blocks, experimental and backcalculated RDCs match poorly when the RDCs are backcalculated using a single alignment block. However, this does not imply, that the simulated structure disagrees with the experiment.

Either restraining mode is a vast improvement over the unrestrained simulation (fig. 10.2 (a)) in terms of the agreement with the experimental data.

Figures 10.3 (a) to 10.3 (c) display the occurrence of secondary structure elements during the three simulations. This analysis is based on hydrogen bonds. In each figure, the 15 reported NMR structures are plotted at negative times.

As described above, the secondary structure of the F1 module is defined by 2 β-sheets which are formed by two and three β-strands. Both restrained simulations reproduce strands A to E of the first module and C to E of the second module throughout the simulation. Strands A and B of the second module only persist in some frames, although the structure still forms the characteristic loop between the A and B strands even while the strands themselves are not detected.

In both simulations, restraining increases the number of residues that are part of β-sheets relative to the unrestrained simulation. Comparison with the reported NMR structures suggests, that this is a better representation of the real structure.

For each simulation the root mean square deviation (RMSD) between the backbone C^α coordinates of the first and the current frame was plotted (figs. 10.4 (a) to 10.4 (c)). Specifically, for each trajectory seven combinations of fitting to one set of atoms and calculating the RMSD for a (possibly different) set of atoms were performed. In red, all residues were taken into account for the fit and the RMSD, in green, blue, purple only one of the three

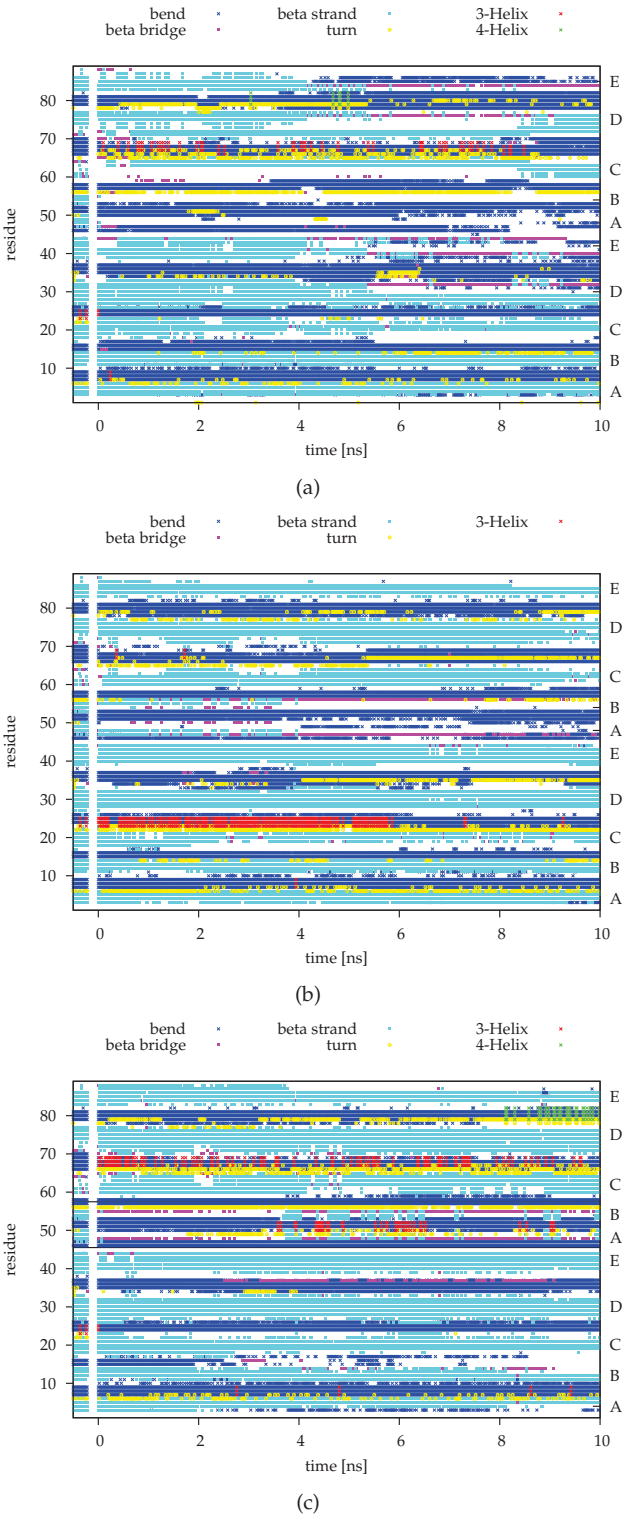


Figure 10.3 Secondary structure motifs during (a) an unrestrained simulation, (b) a simulation restrained with one and (c) three alignment blocks. The secondary structure motifs of the NMR structures reported in 2CKU [492] are plotted at negative times.

blocks was taken into account for fitting to the first frame and the RMSD is calculated with respect to all residues, and in cyan, yellow, black only one block is used for fitting and the RMSD. The first four of these deviations probe predominantly for conformational changes between the chosen blocks while the last three probe for fluctuations within the blocks.

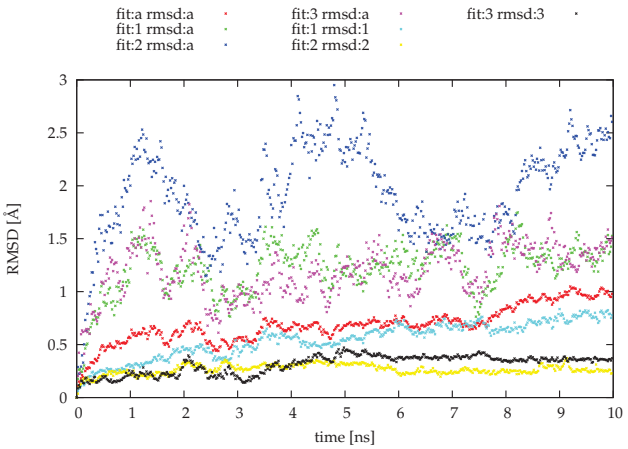
Analogously to the behaviour of Q values, the three individual blocks maintain their structure equally well in the two restrained simulations. Differences between the three blocks can be explained by their different sizes: smaller fragments lead to lower deviation for purely geometric reasons. The main difference between the restrained simulations is the lower deviation when the simulation is restrained with one alignment block and the RMSD of the whole protein is calculated, indicating a lower flexibility.

Again, the structural fluctuations are significantly higher in the unrestrained simulation.

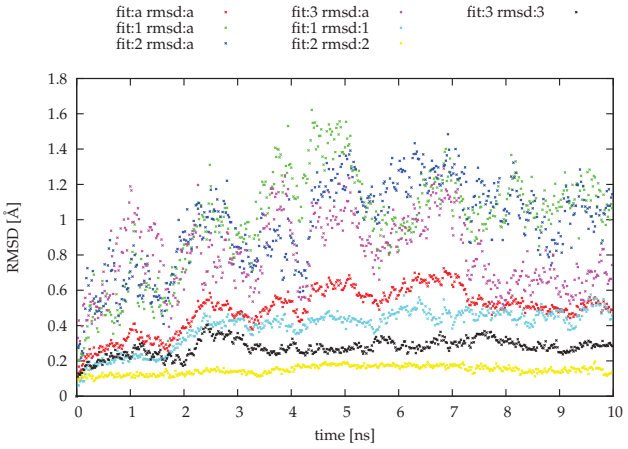
As is shown in the above plots of Q values, restraining three blocks of fibronectin individually does not improve their internal structure. The relative alignment of different block is unrestrained, and hence deviates from the initial structure. Although the measured RDCs agree with the structures in both restrained simulations equally well, the varying angles between the β-sheets in the set of reported NMR structures suggests some inter-domain flexibility which is reproduced more faithfully when secondary structure elements are assigned individual alignment representations.

10.3 Conclusions and outlook

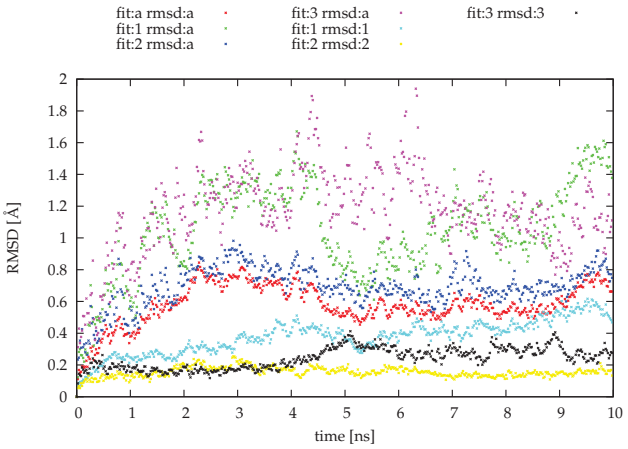
The force term for restraining simulations with experimental RDCs (see sec. 9.5) has been extended by the option to divide the system into arbitrary subunits and restrain them individually. The fibronectin example in sec. 10.2 shows that alignment block implementation is functional. However, this example is not an ideal showcase as the internal structure of the three blocks is not improved relative to the single tensor simulation. It is therefore necessary to explore further which systems benefit from subdivision into alignment blocks. As the measure for agreement between experiment and simulation are Q values, time averaged Q values (eq. 7.81) have to be used.



(a)



(b)



(c)

Figure 10.4 RMSD between the first frame and current frames during (a) an unrestrained simulation, (b) a simulation restrained with one and (c) three alignment blocks.

Furthermore, it has to be investigated what block sizes, subdivision schemes (equal size vs variable size), possible overlaps, and force constants lead to the best results. A compromise between too large alignment blocks that suppress flexibility and too small blocks that discard information on relative orientation and lead to large fluctuations has to be found. In future work, the implications of the above listed variables on derived thermodynamic properties, structural flexibility as well as the fit quality parameter Q should be assessed.

Currently the subdivision of a system into alignment blocks is chosen manually. This implies that prior knowledge of the system is required which adds a source of errors. It would be desirable to make this a black-box procedure and assign alignment blocks automatically. This could be achieved by making use of the fact that tensors fitted to rigidly connected subblocks of a system have the same rhombic and axial components. Hence, calculating the rhombic and axial component of tensors fitted to overlapping blocks of minimal size yields a dependence of tensor dimensions on protein residues based on which alignment blocks can be assigned during simulation setup. It is to be considered to reassign blocks in the course of a simulation based on the same technique.

References

[1] Schwerdtfeger P, Wirz LN, Avery J. The topology of fullerenes. *WIREs Comp Mol Sci* 2015, 5:96–145. doi:10.1002/wcms.1207.

[2] Ōsawa E. Superaromaticity. *Kagaku (Chemistry)* 1970, 25:854–863.

[3] Ōsawa E, Kroto HW, Fowler PW, Wasserman E. The evolution of the football structure for the C_{60} molecule: A retrospective [and discussion]. *Phil Trans R Soc Lond A* 1993, 343:1–8. doi:10.1098/rsta.1993.0035.

[4] Bochvar DA, Gal'pern EG. Electronic structure of the molecules C_{20} and C_{60}. *Proc Acad Sci SSSR* 1973, 209:239–241.

[5] Stankevich I, Nikerov M, Bochvar D. Structural chemistry of crystalline carbon: geometry, stability, electronic spectrum. *Russ Chem Rev* 1984, 53:640–655. doi:10.1070/RC1984v053n07ABEH003084.

[6] Kroto HW, Heath JR, O'Brien SC, Curl RF, Smalley RE. C_{60}: Buckminsterfullerene. *Nature* 1985, 318:162–163. doi:10.1038/318162a0.

[7] Kroto HW. C_{60}: Buckminsterfullerene, the celestial sphere that fell to earth. *Angew Chem Int Ed* 1992, 31:111–129. doi:10.1002/anie.199201113.

[8] Hare JP, Kroto HW. A postbuckminsterfullerene view of carbon in the galaxy. *Acc Chem Res* 1992, 25:106–112. doi:10.1021/ar00015a002.

[9] Kroto H. Symmetry, space, stars and C_{60}. *Rev Mod Phys* 1997, 69:703–722. doi:10.1103/RevModPhys.69.703.

[10] Krätschmer W, Lamb LD, Fostiropoulos K, Huffman DR. Solid C_{60}: a new form of carbon. *Nature* 1990, 347:354–358. doi:10.1038/347354a0.

[11] Mattauch J, Ewald H, Hahn O, Strassmann F. Hat ein Caesium-Isotop langer Halbwertszeit existiert? Ein Beitrag zur Deutung ungewöhnlicher Linien in der Massenspektrographie. *Z Physik* 1943, 120:598–617.

[12] Iijima S. Direct observation of the tetrahedral bonding in graphitized carbon black by high resolution electron microscopy. *J Cryst growth* 1980, 50:675–683. doi:10.1016/0022-0248(80)90013-5.

[13] Buseck P, Tsipursky S, Hettich R. Fullerenes from the geological environment. *Science* 1992, 257:215–217.

[14] Cami J, Bernard-Salas J, Peeters E, Malek SE. Detection of c60 and c70 in a young planetary nebula. *Science* 2010, 329:1180–1182.

[15] Berné O, Tielens AGGM. Formation of buckminsterfullerene (C_{60}) in interstellar space. *Proc Natl Acad Sci* 2012, 109:401–406. doi:10.1073/pnas.1114207108.

[16] Schultz HP. Topological organic chemistry. polyhedranes and prismanes. *J Org Chem* 1965, 30:1361–1364. doi:10.1021/jo01016a005.

[17] Prinzbach H, Weiler A, Landenberger P, Wahl F, Worth J, Scott LT, Gelmont M, Olevano D, v Issendorff B. Gas-phase production and photoelectron spectroscopy of the smallest fullerene C_{20}. *Nature* 2000, 407:60–63. doi:10.1038/35024037.

[18] Sutton D. *Platonic and Archimedean Solids*. Wooden Books, New York, 2002.

[19] Kroto HW. The stability of the fullerenes C_n, with n = 24, 28, 32, 36, 50, 60 and 70. *Nature* 1987, 329:529–531. doi:10.1038/329529a0.

[20] Thurston WP. Shapes of polyhedra and triangulations of the sphere. *Geometry & Topology Monographs* 1998, 1:511–549.

[21] Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, Zhang P. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 2013, 497:643–646. doi:10.1038/nature12162.

[22] Tagmatarchis N, Okada K, Tomiyama T, Yoshida T, Kobayashi Y, Shinohara H. A catalytic synthesis and structural characterization of a new [84]fullerene isomer. *Chem Commun* 2001, 1366–1367. doi:10.1039/B103679N.

[23] Popov AA, Yang S, Dunsch L. Endohedral fullerenes. *Chem Rev* 2013, 113:5989–6113. doi:10.1021/cr300297r.

[24] Yamada M, Akasaka T, Nagase S. Carbene additions to fullerenes. *Chem Rev* 2013, 113:7209–7264. doi:10.1021/cr3004955.

[25] Franco JU, Hammons JC, Rios D, Olmstead MM. New tetraazaannulene hosts for fullerenes. *Inorg Chem* 2010, 49:5120–5125. doi:10.1021/ic1002513.

[26] York APE. Inorganic fullerenes, onions, and tubes. *J Chem Educ* 2004, 81:673. doi:10.1021/ed081p673.

[27] Scott LT, Boorum MM, McMahon BJ, Hagen S, Mack J, Blank J, Wegner H, de Meijere A. A rational chemical synthesis of C_{60}. *Science* 2002, 295:1500–1503. doi:10.1126/science.1068427.

[28] Kabdulov M, Jansen M, Amsharov KY. Bottom-up C_{60} fullerene construction from a fluorinated $C_{60}H_{21}F_9$ precursor by laser-induced tandem cyclization. *Chem Eur J* 2013, 19:17262–17266. doi:10.1002/chem.201303838.

[29] Cioslowski J. *Electronic Structure Calculations on Fullerenes and Their Derivatives*. Wooden Books, New York, 1995.

[30] Balaban AT, Liu X, Klein DJ, Babić D, Schmalz TG, Seitz WA, Randic M. Graph invariants for fullerenes. *J Chem Inf Comput Sci* 1995, 35:396–404. doi: 10.1021/ci00025a007.

[31] Zhu HY, Klein DJ. Graph-geometric invariants for molecular structures. *J Chem Inf Comput Sci* 1996, 36:1067–1075. doi:10.1021/ci960025a.

[32] Fowler PW, Manolopoulos DE. *An Atlas of Fullerenes*. Dover Publications Inc., Mineola, New York, 2nd ed., 2006.

[33] Alcamí M, Sánchez G, Díaz-Tendero S, Wang Y, Martín F. Structural patterns in fullerenes showing adjacent pentagons: C_{20} to C_{72}. *J Nanosci Nanotechnol* 2007, 7:1329–1338. doi:10.1166/jnn.2007.311.

[34] Cioslowski J, Rao N, Moncrieff D. Standard enthalpies of formation of fullerenes and their dependence on structural motifs. *J Am Chem Soc* 2000, 122:8265–8270. doi:10.1021/ja001109+.

[35] Brinkmann G, Goedgebeur J, Mélot H, Coolsaet K. House of graphs: a database of interesting graphs, available at http://hog.grinvin.org. *Discrete Appl Math* 2013, 161:311–314.

[36] Schwerdtfeger P, Wirz L, Avery J. Program fullerene: A software package for constructing and analyzing structures of regular fullerenes. *J Comput Chem* 2013, 34:1508–1526. doi:10.1002/jcc.23278.

[37] Fowler PW. Fullerene stability and structure. *Contemp Phys* 1996, 37:235–247. doi:10.1080/00107519608217530.

[38] Baldridge KK, Siegel JS. Of graphs and graphenes: Molecular design and chemical studies of aromatic compounds. *Angew Chem Int Ed* 2013, 52:5436–5438. doi:10.1002/anie.201300625.

[39] Trinajstic N. *Chemical Graph Theory*. CRC, Boca Raton, FL, 1983.

[40] Kotschik D. The topology and combinatorics of soccer balls. *Am Sci* 2006, 94:350–357. doi:10.1511/2006.60.1001.

[41] Cataldo F, Graovac A, Ori O. *The mathematics and topology of fullerenes*. Springer, Berlin, 2011.

[42] Dresselhaus MS, Dresselhaus G, Eklund PC. *Science of Fullerenes and Carbon Nanotubes*. Academic Press, New York, 1995.

[43] Curl RF. Dawn of the fullerenes: experiment and conjecture. *Rev Mod Phys* 1997, 69:691–702. doi:10.1103/RevModPhys.69.691.

[44] Kadish KM, Ruoff RS (eds.). *Fullerenes: Chemistry, Physics, and Technology*. Wiley-Interscience, New York, 2000.

[45] Hirsch A, Brettreich M, Wudl F. *Fullerenes: Chemistry and Reactions*. Wiley-VCH, Weinheim, 2005.

[46] Rodríguez-Fortea A, Irle S, Poblet JM. Fullerenes: formation, stability, and reactivity. *WIREs Comp Mol Sci* 2011, 1:350–367. doi:10.1002/wcms.21.

[47] Sheka EF. *Fullerenes: Nanochemistry, Nanomagnetism, Nanomedicine, Nanophotonics*. CRC Press, Taylor & Francis, Boca Radon, 2011.

[48] Darwish AD. Fullerenes. *Annu Rep Prog Chem, Sect A: Inorg Chem* 2012, 108:464–477. doi:10.1039/C2IC90017C.

[49] Whitney H. Non-separable and planar graphs. *Trans Amer Math Soc* 1932, 34:339–362. doi:10.1090/S0002-9947-1932-1501641-2.

[50] Beineke LW, Wilson RJ (eds.). *Selected Topics in Graph Theory*. Academic Press, London, New York, San Francisco, 1978.

[51] Steinitz E. Polyeder und Raumeinteilungen. In Klein F, Meyer W (eds.), *Encyclopädie der mathematischen Wissenschaften*, vol. 3 (Geometries), 1–139. B.G. Teubner Verlag, Leipzig, 1922.

[52] Schmalz TG, Seitz WA, Klein DJ, Hite GE. Elemental carbon cages. *J Am Chem Soc* 1988, 110:1113–1127. doi:10.1021/ja00212a020.

[53] Grünbaum G, Motzkin TS. The number of hexagons and the simplicity of geodesics on certain polyhedra. *Can J Math* 1963, 15:744–751. doi:10.4153/CJM-1963-071-3.

[54] Killblane C, Gao Y, Shao N, Zeng XC. Search for lowest-energy nonclassical fullerenes III: C_{22}. *J Phys Chem A* 2009, 113:8839–8844. doi:10.1021/jp9016745.

[55] Janežič D, Miličević A, Nikolić S, Trinajstić N. In Gutman I (ed.), *Graph Theoretical Matrices in Chemistry*, vol. 3 of *Mathematical Chemistry Monographs*. University of Kragujevac, Kragujevac, 2007.

[56] Tutte WT. How to Draw a Graph. *Proc Lond Math Soc* 1963, 13:743–767.

[57] Brandenburg FJ, Himsolt M, Rohrer C. An experimental comparison of force-directed and randomized graph drawing algorithms. In Brandenburg FJ (ed.), *Graph Drawing*, vol. 1027 of *Lecture Notes in Computer Science*, 76–87. Springer Berlin Heidelberg, 1996. doi:10.1007/BFb0021792.

[58] Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Software: Practice and Experience* 1991, 21:1129–1164. doi:10.1002/spe.4380211102.

[59] Plestenjak B. An algorithm for drawing Schlegel diagrams. *see http://www-lpfmfuniljsi/plestenjak/Papers/NICEGRpdf* 1996, 1–10.

[60] Liu X, Klein DJ, Schmalz TG, Seitz WA. Generation of carbon-cage polyhedra. *J Comput Chem* 1991, 12:1252–1259. doi:10.1002/jcc.540121013.

[61] Brinkmann G, Dress AW. A constructive enumeration of fullerenes. *Journal of Algorithms* 1997, 23:345–358. doi:10.1006/jagm.1996.0806.

[62] Brinkmann G, Franceus D, Fowler PW, Graver JE. Growing fullerenes from seed: Growth transformations of fullerene polyhedra. *Chem Phys Lett* 2006, 428:386–393. doi:10.1016/j.cplett.2006.07.040.

[63] Hasheminezhad M, Fleischner H, McKay BD. A universal set of growth operations for fullerenes. *Chem Phys Lett* 2008, 464:118–121. doi:10.1016/j.cplett.2008.09.005.

[64] Balaban AT, Schmalz TG, Zhu H, Klein DJ. Generalizations of the Stone-Wales rearrangement for cage compounds, including fullerenes. *J Mol Struct (THEOCHEM)* 1996, 363:291–301. doi:10.1016/0166-1280(95)04448-5.

[65] Graver JE, Graves CM, Graves SJ. Fullerene patches II. *Ars Math Contemp* 2014, 7:405–421.

[66] Brinkmann G, Goedgebeur J, McKay BD. The generation of fullerenes. *J Chem Inf Model* 2012, 52:2910–2918. doi:10.1021/ci3003107.

[67] Brinkmann G, Goedgebeur J, Van Cleemput N. The history of the generation of cubic graphs. *J Chem Inf Model* 2013, 5:67–89.

[68] Goldberg M. A class of multi-symmetric polyhedra. *Tohoku Math J* 1937, 43:104–108.

[69] Coxeter HSM. Virus macromolecules and geodesic domes. In Butcher JC (ed.), *A Spectrum of mathematics - Essays Presented to H. G. Forder*, 98–107. Oxford University press, 1971.

[70] Guo X, Hansen P, Zheng M. Boundary uniqueness of fusenes. *Discrete Appl Math* 2002, 118:209–222. doi:10.1016/S0166-218X(01)00180-9.

[71] Graver JE. The (m,k)-patch boundary code problem. *MATCH Commun Math Comput Chem* 2003, 48:189–196.

[72] Brinkmann G, Graver JE, Justus C. Numbers of faces in disordered patches. *J Math Chem* 2009, 45:263–278. doi:10.1007/s10910-008-9403-6.

[73] Graver JE, Graves CM. Fullerene patches I. *Ars Math Contemp* 2010, 3:109–120.

[74] Brinkmann G, Fowler PW, Justus C. A catalogue of isomerization transformations of fullerene polyhedra. *J Chem Inf Comput Sci* 2003, 43:917–927. doi:10.1021/ci020069l.

[75] Brinkmann G, Fowler PW. A catalogue of growth transformations of fullerene polyhedra. *J Chem Inf Comput Sci* 2003, 43:1837–1843. doi:10.1021/ci030017b.

[76] Bornhöft J, Brinkmann G, Greinus J. Pentagon–hexagon-patches with short boundaries. *Eur J Combin* 2003, 24:517–529. doi:10.1016/S0195-6698(03)00034-9.

[77] Stone AJ, Wales DJ. Theoretical studies of icosahedral C_{60} and some related species. *Chem Phys Lett* 1986, 128:501–503. doi:10.1016/0009-2614(86)80661-3.

[78] Babić D, Bassoli S, Casartelli M, Cataldo F, Graovac A, Ori O, York B. Generalized Stone-Wales transformations. *Mol Simulat* 1995, 14:395–401. doi: 10.1080/08927029508022032.

[79] Astakhova TY, Vinogradov GA. New isomerization operations for fullerene graphs. *J Mol Struct (THEOCHEM)* 1998, 430:259–268. doi:10.1016/S0166-1280(98)90253-6.

[80] Ori O, Putz MV, Gutman I, Schwerdtfeger P. Generalized Stone-Wales transformations for fullerene graphs derived from Berge's switching theorem. In Gutman I, Pokric B, Vukicevic D (eds.), *Ante Graovac – Life and Works*, vol. 16, 259–272. Mathematical Chemistry Monographs, 2014.

[81] Berge C. *Graphs and Hypergraphs*. Elsevier, New York, 1973.

[82] Endo M, Kroto HW. Formation of carbon nanofibers. *J Phys Chem* 1992, 96:6941–6944. doi:10.1021/j100196a017.

[83] Yoshida M, Fowler PW. Systematic relationships between fullerenes without spirals. *Chem Phys Lett* 1997, 278:256–261. doi:10.1016/S0009-2614(97)00980-9.

[84] Podlivaev A, Openov L. Stone-Wales transformation paths in fullerene C_{60}. *J Exp Theoret Phys Lett* 2005, 81:533–537. doi:10.1134/1.1996764.

[85] Ori O, Cataldo F, Putz MV. Topological anisotropy of Stone-Wales waves in graphenic fragments. *Int J Mol Sci* 2011, 12:7934–7949. doi: 10.3390/ijms12117934.

[86] Saha B, Irle S, Morokuma K. Hot giant fullerenes eject and capture C_2 molecules: QM/MD simulations with constant density. *J Phys Chem C* 2011, 115:22707–22716. doi:10.1021/jp203614e.

[87] Murry RL, Strout DL, Odom GK, Scuseria GE. Role of sp^3 carbon and 7-membered rings in fullerene annealing and fragmentation. *Nature* 1993, 366:665–667. doi:10.1038/366665a0.

[88] Walsh TR, Wales DJ. Relaxation dynamics of C_{60}. *J Chem Phys* 1998, 109:6691–6700. doi:10.1063/1.477319.

[89] Dunk PW, Kaiser NK, Hendrickson CL, Quinn JP, Ewels CP, Nakanishi Y, Sasaki Y, Shinohara H, Marshall AG, Kroto HW. Closed network growth of fullerenes. *Nature Commun* 2012, 3:855. doi:10.1038/ncomms1853.

[90] Irle S, Zheng G, Wang Z, Morokuma K. The C_{60} formation puzzle "solved": QM/MD simulations reveal the shrinking hot giant road of the dynamic fullerene self-assembly mechanism. *J Phys Chem B* 2006, 110:14531–14545. doi:10.1021/jp061173z.

[91] Tan YZ, Chen RT, Liao ZJ, Li J, Zhu F, Lu X, Xie SY, Li J, Huang RB, Zheng LS. Carbon arc production of heptagon-containing fullerene[68]. *Nature Commun* 2011, 2:420. doi:10.1038/ncomms1431.

[92] Austin SJ, Fowler PW, Manolopoulos DE, Zerbetto F. The Stone Wales map for C_{60}. *Chem Phys Lett* 1995, 235:146–151. doi:10.1016/0009-2614(95)00082-F.

[93] Budyka MF, Zyubina TS, Ryabenko AG, Muradyan VE, Esipov SE, Cherepanova NI. Is C_2 cluster ingested by fullerene C_{60}? *Chem Phys Lett* 2002, 354:93–99. doi:10.1016/S0009-2614(02)00113-6.

[94] Dutour M, Deza M. Goldberg-Coxeter construction for 3- and 4-valent plane graphs. *Electron J Comb* 2004, 11:1–49.

[95] Avery JE. Goldberg-Coxeter transformation for cubic polyhedral graphs: A practical algorithm. *unpublished* 2014.

[96] Wirz LN, Avery JE, Schwerdtfeger P. Structure and properties of the non-face-spiral fullerenes T-C_{380}, D_3-C_{384}, D_3-C_{440} and D_3-C_{672} and their halma and leapfrog transforms. *J Chem Inf Model* 2014, 54:121–130. doi:10.1021/ci4005578.

[97] Avery JE, Knudsen MT, Wirz LN. Minimal geodesics and shortest paths on planar cubic graph topologies. *to be published* 2014.

[98] Bakowies D, Bühl M, Thiel W. Can large fullerenes be spherical? *J Am Chem Soc* 1995, 117:10113–10118. doi:10.1021/ja00145a025.

[99] Bakowies D, Bühl M, Thiel W. A density functional study on the shape of C_{180} and C_{240} fullerenes. *Chem Phys Lett* 1995, 247:491–493. doi:10.1016/S0009-2614(95)01222-2.

[100] Senn P. Computation of the cartesian coordinates of buckminsterfullerene. *J Chem Educ* 1995, 72:302–303. doi:10.1021/ed072p302.

[101] Manolopoulos DE, Fowler PW. Molecular graphs, point groups, and fullerenes. *J Chem Phys* 1992, 96:7603–7614. doi:10.1063/1.462413.

[102] Cvetković D, Fowler P, Rowlinson P, Stevanović D. Constructing fullerene graphs from their eigenvalues and angles. *Linear Algebra and its Applications* 2002, 356:37–56. doi:10.1016/S0024-3795(02)00280-X.

[103] Pisanski T, Shawe-Taylor J. Characterizing graph drawing with eigenvectors. *J Chem Inf Comput Sci* 2000, 40:567–571. doi:10.1021/ci9900938.

[104] Fowler P, Pisanski T, Shawe-Taylor J. Graph drawing. In Tamassia R, Tollis I (eds.), *Lecture Notes in Computer Science*, vol. 894, 282–285. Springer, Berlin, 1995.

[105] László I, Graovac A, Pisanski T, Plavšić D. Graph drawing with eigenvectors. In Putz MV (ed.), *Carbon Bonding and Structures*, vol. 5 of *Carbon Materials: Chemistry and Physics*, 95–115. Springer Netherlands, 2011. doi:10.1007/978-94-007-1733-6_6.

[106] Mani P. Automorphismen von polyedrischen Graphen. *Math Ann* 1971, 192:279–303. doi:10.1007/BF02075357.

[107] Deza M. The symmetries of cubic polyhedral graphs with face size no larger than 6. *MATCH Commun Math Comput Chem* 2009, 61:589–602.

[108] Wirz LN, Schwerdtfeger P, Avery JE. A generalised face-spiral algorithm for cubic polyhedral graphs. *unpublished* 2014.

[109] Balasubramanian K. Applications of combinatorics and graph theory to spectroscopy and quantum chemistry. *Chem Rev* 1985, 85:599–618. doi:10.1021/cr00070a005.

[110] Faghani M. Symmetry of a toroidal fullerene. *Optoelectronics and Advanced Materials - Rapid Communications* 2010, 4:1844–1846.

[111] Fujita S. Application of coset representations to the construction of symmetry adapted functions. *Theoret Chim Acta* 1990, 78:45–63. doi:10.1007/BF01112352.

[112] King RB, Diudea MV. The chirality of icosahedral fullerenes: a comparison of the tripling (leapfrog), quadrupling (chamfering), and septupling (capra) transformations. *J Math Chem* 2006, 39:597–604. doi:10.1007/s10910-005-9048-7.

[113] David WIF, Ibberson RM, Matthewman JC, Prassides K, Dennis TJS, Hare JP, Kroto HW, Taylor R, Walton DRM. Crystal structure and bonding of ordered C_{60}. *Nature* 1991, 353:147–149. doi:10.1038/353147a0.

[114] Devadoss SL, O'Rourke J. *Discrete and Computational Geometry*. Princeton University Press, Princeton, 2011.

[115] Pisanski T, Kaufman M, Bokal D, Kirby EC, Graovac A. Isoperimetric quotient for fullerenes and other polyhedral cages. *J Chem Inf Comput Sci* 1997, 37:1028–1032. doi:10.1021/ci970228e.

[116] Sylvester JJ. A question in the geometry of situation. *Q J Pure Appl Math* 1857, 1:79.

[117] Elzinga DJ, Hearn DW. The minimum covering sphere problem. *Manag Sci* 1972, 19:96–104.

[118] Hopp TH, Reeve CP. An algorithm for computing the minimum covering sphere in any dimension. In COMMERCE UDO (ed.), *NIST IR5381*, NIST, 1–8. NIST, 1996.

[119] Nielsen F, Nock R. Approximating smallest enclosing balls. In Laganá A, Gavrilova M, Kumar V, Mun Y, Tan C, Gervasi O (eds.), *Computational Science and Its Applications – ICCSA 2004*, vol. 3045 of *Lecture Notes in Computer Science*, 147–157. Springer Berlin Heidelberg, 2004. doi:10.1007/978-3-540-24767-8_16.

[120] Yıldırım E. Two algorithms for the minimum enclosing ball problem. *SIAM J Optimiz* 2008, 19:1368–1391. doi:10.1137/070690419.

[121] Fowler P. Cylindrical fullerenes: The smallest nanotubes? *J Phys Chem Solids* 1993, 54:1825–1833. doi:10.1016/0022-3697(93)90295-3.

[122] Kumar P, Yıldırım E. Minimum-volume enclosing ellipsoids and core sets. *J Optim Theory Appl* 2005, 126:1–21. doi:10.1007/s10957-005-2653-6.

[123] Kumar P, Yıldırım E. Computing minimum-volume enclosing axis-aligned ellipsoids. *J Optim Theory Appl* 2008, 136:211–228. doi:10.1007/s10957-007-9295-9.

[124] Sun P, Freund RM. Computation of minimum-volume covering ellipsoids. *Oper Res* 2004, 52:690–706. doi:10.1287/opre.1040.0115.

[125] Todd MJ, Yıldırım EA. On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Appl Math* 2007, 155:1731–1744. doi:10.1016/j.dam.2007.02.013.

[126] Nasu K, Taketsugu T, Nakano T, Nagashima U, Hosoya H. Stability of small fullerenes C_n (n=36,40 and 60): A topological and molecular orbital approach. *Theoret Chim Acta* 1995, 90:75–86. doi:10.1007/BF01113841.

[127] Díaz-Tendero S, Martín F, Alcamí M. Structure and electronic properties of fullerenes C_{52}^{q+}: Is C_{52}^{2+} an exception to the pentagon adjacency penalty rule? *ChemPhysChem* 2005, 6:92–100. doi:10.1002/cphc.200400273.

[128] Díaz-Tendero S, Alcamí M, Martín F. Fullerene C_{50}: Sphericity takes over, not strain. *Chem Phys Lett* 2005, 407:153–158. doi:10.1016/j.cplett.2005.03.065.

[129] Fowler PW, Heine T, Zerbetto F. Competition between even and odd fullerenes: C_{118}, C_{119}, and C_{120}. *J Phys Chem A* 2000, 104:9625–9629. doi:10.1021/jp0019815.

[130] Wiener H. Structural determination of paraffin boiling points. *J Am Chem Soc* 1947, 69:17–20. doi:10.1021/ja01193a005.

[131] Ori O, D'Mello M. A topological study of the structure of the C_{76} fullerene. *Chem Phys Lett* 1992, 197:49–54. doi:10.1016/0009-2614(92)86020-I.

[132] Ori O, Cataldo F, Vukičević D, Graovac A. Wiener way to dimensionality. *Iranian J Math Chem* 2010, 1:5–15.

[133] Graovac A, Ori O, Faghani M, Ashrafi A. Distance property of fullerenes. *Iranian J Math Chem* 2011, 2:99–107.

[134] Koorepazan-Moftakhar F, Ashrafi AR, Ori O, Putz MV. Sphericality of some classes of fullerenes measured by topology. In *Fullerenes: Chemistry, Natural Sources and Technological Applications*, NIST. Nova Publ., New York, in press., 2014.

[135] Vukičević D, Cataldo F, Ori O, Graovac A. Topological efficiency of C_{66} fullerene. *Chem Phys Lett* 2011, 501:442–445. doi:10.1016/j.cplett.2010.11.055.

[136] Ori O, Cataldo F, Graovac A. Topological ranking of C_{28} fullerenes reactivity. *Fuller Nanotub Carbon Nanostruct* 2009, 17:308–323. doi:10.1080/15363830902782332.

[137] Wang CR, Kai T, Tomiyama T, Yoshida T, Kobayashi Y, Nishibori E, Takata M, Sakata M, Shinohara H. Materials science: C_{66} fullerene encaging a scandium dimer. *Nature* 2000, 408:426–427. doi:10.1038/35044195.

[138] Katritzky AR, Karelson M, Petrukhin R. The CODESSA PRO project, 2005.

[139] Fowler PW, Caporossi G, Hansen P. Distance matrices, Wiener indices, and related invariants of fullerenes. *J Phys Chem A* 2001, 105:6232–6242. doi: 10.1021/jp0104379.

[140] Hosoya H. On some counting polynomials in chemistry. *Discrete Appl Math* 1988, 19:239–257. doi:10.1016/0166-218X(88)90017-0.

[141] Gutman I. A formula for the Wiener number of trees and its extension to graphs containing cycles. *Graph Theory Notes of New York* 1994, 27:9–15.

[142] Balaban AT. Topological indices based on topological distances in molecular graphs. *Pure Appl Chem* 1983, 55:199–206. doi:10.1351/pac198855020199.

[143] Balaban AT. Chemical graphs: Looking back and glimpsing ahead. *J Chem Inf Comput Sci* 1995, 35:339–350. doi:10.1021/ci000025a001.

[144] Zhang H, Balasubramanian K. Spectral moments of fullerene cages. *Mol Phys* 1993, 79:727–745. doi:10.1080/00268979300101581.

[145] Estrada E. Characterization of 3D molecular structure. *Chem Phys Lett* 2000, 319:713–718. doi:10.1016/S0009-2614(00)00158-5.

[146] de la Peña JA, Gutman I, Rada J. Estimating the Estrada index. *Linear Algebra and its Applications* 2007, 427:70–76. doi:10.1016/j.laa.2007.06.020.

[147] Diudea MV, Gutman I, Jantschi L. *Molecular Topolgy*. Nova Science, New York, 2001.

[148] Austin SJ, Fowler PW, Manolopoulos DE, Orlandi G, Zerbetto F. Structural motifs and the stability of fullerenes. *J Phys Chem* 1995, 99:8076–8081. doi: 10.1021/j100020a035.

[149] Ju Y, Liang H, Zhang J, Bai F. A note on Fowler-Manolopoulos predictor of fullerene stability. *MATCH Commun Math Comput Chem* 2010, 64:419–424.

[150] Stevanović D. Remarks on Fowler-Manolopoulos predictor of fullerene stability. *MATCH Commun Math Comput Chem* 2011, 66:285–292.

[151] Karton A, Chan B, Raghavachari K, Radom L. Evaluation of the heats of formation of corannulene and C_{60} by means of high-level theoretical procedures. *J Phys Chem A* 2013, 117:1834–1842. doi:10.1021/jp312585r.

[152] Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys Rev Lett* 1996, 77:3865–3868. doi:10.1103/PhysRevLett.77.3865.

[153] Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys Chem Chem Phys* 2005, 7:3297–3305. doi:10.1039/B508541A.

[154] Iranmanesh A, Alizadeh Y, Mirzaie S. Computing Wiener polynomial, Wiener index and Hyper Wiener index of C_{80} fullerene by GAP program. *Fuller Nanotub Carbon Nanostruct* 2009, 17:560–566. doi:10.1080/15363830903133204.

[155] Charlier JC, Eklund P, Zhu J, Ferrari A. Electron and phonon properties of graphene: Their relationship with carbon nanotubes. In Jorio A, G D, Dresselhaus M (eds.), *Carbon Nanotubes: Advanced Topics in the Synthesis, Structure, Properties and Applications*. Springer, Berlin, 2008.

[156] Khamatgalimov AR, Kovalenko VI. Electronic structure and stability of C_{80} fullerene IPR isomers. *Fuller Nanotub Carbon Nanostruct* 2011, 19:599–604. doi:10.1080/1536383X.2010.504951.

[157] Xu L, Cai W, Shao X. Prediction of low-energy isomers of large fullerenes from C_{132} to C_{160}. *J Phys Chem A* 2006, 110:9247–9253. doi:10.1021/jp057181h.

[158] Shao N, Gao Y, Yoo S, An W, Zeng XC. Search for lowest-energy fullerenes: C_{98} to C_{110}. *J Phys Chem A* 2006, 110:7672–7676. doi:10.1021/jp0624092.

[159] Shao N, Gao Y, Zeng XC. Search for lowest-energy fullerenes 2: C_{38} to C_{80} and C_{112} to C_{120}. *J Phys Chem C* 2007, 111:17671–17677. doi:10.1021/jp0701082.

[160] Klein DJ, Balaban AT. Clarology for conjugated carbon nano-structures: Molecules, polymers, graphene, defected graphene, fractal benzenoids, fullerenes, nano-tubes, nano-cones, nano-tori, etc. *Open Org Chem J* 2011, 5:27–61. doi:10.2174/1874364101105010027.

[161] Klein DJ, Schmalz TG, Hite GE, Seitz WA. Resonance in C_{60} buckminsterfullerene. *J Am Chem Soc* 1986, 108:1301–1302. doi:10.1021/ja00266a032.

[162] Vukičević D, Kroto HW, Randić M. Atlas of Kekulé valence structures of buckminsterfullerene. *Croat Chem Acta* 2005, 78:223–234.

[163] Randić M, Kroto HW, Vukičević D. Numerical Kekulé structures of fullerenes and partitioning of π-electrons to pentagonal and hexagonal rings. *J Chem Inf Model* 2007, 47:897–904. doi:10.1021/ci600484u.

[164] Austin S, Fowler P, Hansen P, Monolopoulos D, Zheng M. Fullerene isomers of C_{60}. Kekulé counts versus stability. *Chem Phys Lett* 1994, 228:478–484. doi:10.1016/0009-2614(94)00965-1.

[165] Esperet L, Kardoš F, King AD, Král' D, Norine S. Exponentially many perfect matchings in cubic graphs. *Adv Math* 2011, 227:1646–1664. doi:10.1016/j.aim.2011.03.015.

[166] Kardoš F, Král' D, Miškuf J, Sereni JS. Fullerene graphs have exponentially many perfect matchings. *J Math Chem* 2009, 46:443–447. doi:10.1007/s10910-008-9471-7.

[167] Lovász L, Plummer MD. *Matching Theory*. Elsevier, Amsterdam, 1986.

[168] Pauling L, Brockway LO, Beach J. The dependence of interatomic distance on single bond-double bond resonance. *J Am Chem Soc* 1935, 57:2705–2709. doi:10.1021/ja01315a105.

[169] Narita S, Morikawa T, Shibuya T. Linear relationship between the bond lengths and the Pauling bond orders in fullerene molecules. *J Mol Struct (THEOCHEM)* 2000, 532:37–40. doi:10.1016/S0166-1280(00)00563-7.

[170] Rogers KM, Fowler PW. Leapfrog fullerenes, Hückel bond order and Kekulé structures. *J Chem Soc, Perkin Trans 2* 2001, 18–22. doi:10.1039/B007520P.

[171] Gross L, Mohn F, Moll N, Schuler B, Criado A, Guitián E, Peña D, Gourdon A, Meyer G. Bond-order discrimination by atomic force microscopy. *Science* 2012, 337:1326–1329. doi:10.1126/science.1225621.

[172] Graver JE, Hartung EJ, Souid AY. Clar and Fries numbers for benzenoids. *J Math Chem* 2013, 51:1981–1989. doi:10.1007/s10910-013-0193-0.

[173] Bernáth A, Kovács ER. NP-hardness of the Clar number in general plane graphs. www.cs.elte.hu/egres/qp/egresqp-11-07.ps, 2011.

[174] Ye D, Zhang H. Extremal fullerene graphs with the maximum Clar number. *Discrete Appl Math* 2009, 157:3152 – 3173. doi:10.1016/j.dam.2009.06.007.

[175] Zhang H, Ye D, Liu Y. A combination of Clar number and Kekulé count as an indicator of relative stability of fullerene isomers of C_{60}. *J Math Chem* 2010, 48:733–740. doi:10.1007/s10910-010-9706-2.

[176] Fowler PW. Localised models and leapfrog structures of fullerenes. *J Chem Soc, Perkin Trans 2* 1992, 145–146. doi:10.1039/P29920000145.

[177] Došlić T. Leapfrog fullerenes have many perfect matchings. *J Math Chem* 2008, 44:1–4. doi:10.1007/s10910-007-9287-x.

[178] Hartung E. Fullerenes with complete Clar structure. *Discrete Appl Math* 2013, 161:2952–2957. doi:10.1016/j.dam.2013.06.009.

[179] Fowler PW, Austin SJ, Dunning OJ, Dias JR. Symmetry properties of the leapfrog transformation for fullerenes and benzenoids. *Chem Phys Lett* 1994, 224:123–130. doi:10.1016/0009-2614(94)00525-7.

[180] Zhang H, Ye D. An upper bound for the Clar number of fullerene graphs. *J Math Chem* 2007, 41:123–133. doi:10.1007/s10910-006-9061-5.

[181] Herndon WC. Resonance energies of aromatic hydrocarbons. quantitative test of resonance theory. *J Am Chem Soc* 1973, 95:2404–2406. doi:10.1021/ja00788a073.

[182] Herndon WC, Ellzey J M Lawrence. Resonance theory. V. Resonance energies of benzenoid and nonbenzenoid pi systems. *J Am Chem Soc* 1974, 96:6631–6642. doi:10.1021/ja00828a015.

[183] Liu Y, O'Brien S, Zhang Q, Heath J, Tittel F, Curl R, Kroto H, Smalley R. Negative carbon cluster ion beams: New evidence for the special nature of C_{60}. *Chem Phys Lett* 1986, 126:215–217. doi:10.1016/S0009-2614(86)80042-2.

[184] Slanina Z, Uhlík F, Zhao X, Adamowicz L, Nagase S. Relative stabilities of C_{74} isomers. *Fuller Nanotub Carbon Nanostruct* 2007, 15:195–205. doi:10.1080/15363830701236423.

[185] Beckhaus HD, Verevkin S, Rüchardt C, Diederich F, Thilgen C, Meer HUT, Mohn H, Müller W. C_{70} ist stabiler als C_{60}: Experimentelle Bestimmung der Bildungswärme von C_{70}. *Angew Chem* 1994, 33:996–998. doi:10.1002/ange.19941060916.

[186] Wang Z, Ke X, Zhu Z, Zhu F, Ruan M, Chen H, Huang R, Zheng L. A new carbon solid made of the world's smallest caged fullerene C_{20}. *Phys Lett A* 2001, 280:351–356. doi:10.1016/S0375-9601(00)00847-1.

[187] Lin F, Sorensen ES, Kallin C, Berlinsky AJ. C_{20}, the smallest fullerene. In Sattler KD (ed.), *Handbook of Nanophysics: Clusters and Fullerenes*, Handbook of Nanophysics, chap. 29. CRC Press, Boca Raton, USA, 2010.

[188] Babić D, Brinkmann G, Dress A. Topological resonance energy of fullerenes. *J Chem Inf Comput Sci* 1997, 37:920–923. doi:10.1021/ci9700283.

[189] Bühl M, Hirsch A. Spherical aromaticity of fullerenes. *Chem Rev* 2001, 101:1153–1184. doi:10.1021/cr990332q.

[190] Haddon R, Brus L, Raghavachari K. Electronic structure and bonding in icosahedral C_{60}. *Chem Phys Lett* 1986, 125:459–464. doi:10.1016/0009-2614(86)87079-8.

[191] Haddon R, Brus L, Raghavachari K. Rehybridization and π-orbital alignment: the key to the existence of spheroidal carbon clusters. *Chem Phys Lett* 1986, 131:165–169. doi:10.1016/0009-2614(86)80538-3.

[192] Bakowies D, Thiel W. MNDO study of large carbon clusters. *J Am Chem Soc* 1991, 113:3704–3714. doi:10.1021/ja00010a012.

[193] Fowler PW, Hansen P, Stevanović D. A note on the smallest eigenvalue of fullerenes. *MATCH Commun Math Comput Chem* 2003, 48:37–48.

[194] Došlić T. The smallest eigenvalue of fullerene graphs – closing the gap. *MATCH Commun Math Comput Chem* 2013, 70:73–78.

[195] Randić M, Trinajstić N, Živković T. Molecular graphs having identical spectra. *J Chem Soc, Faraday Trans 2* 1976, 72:244–256. doi:10.1039/F29767200244.

[196] Aihara J. A new definition of Dewar-type resonance energies. *J Am Chem Soc* 1976, 98:2750–2758. doi:10.1021/ja00426a013.

[197] Gutman I, Milun M, Trinajstic N. Graph theory and molecular orbitals. 19. Nonparametric resonance energies of arbitrary conjugated systems. *J Am Chem Soc* 1977, 99:1692–1704. doi:10.1021/ja00448a002.

[198] Balasubramanian K. Exhaustive generation and analytical expressions of matching polynomials of fullerenes C_{20}-C_{50}. *J Chem Inf Comput Sci* 1994, 34:421–427. doi:10.1021/ci00018a032.

[199] Babić D, Balaban AT, Klein DJ. Nomenclature and coding of fullerenes. *J Chem Inf Comput Sci* 1995, 35:515–526. doi:10.1021/ci00025a020.

[200] Salvador JM, Hernandez A, Beltran A, Duran R, Mactutis A. Fast partial-differential synthesis of the matching polynomial of C_{72-100}. *J Chem Inf Comput Sci* 1998, 38:1105–1110. doi:10.1021/ci9800155.

[201] Randić M. Novel insight into Clar's aromatic π-sextets. *Chem Phys Lett* 2014, 601:1–5. doi:10.1016/j.cplett.2014.03.073.

[202] Lu X, Chen Z. Curved π-conjugation, aromaticity, and the related chemistry of small fullerenes ($<C_{60}$) and single-walled carbon nanotubes. *Chem Rev* 2005, 105:3643–3696. doi:10.1021/cr030093d.

[203] Chen Z, Wu JI, Corminboeuf C, Bohmann J, Lu X, Hirsch A, Schleyer PvR. Is C_{60} buckminsterfullerene aromatic? *Phys Chem Chem Phys* 2012, 14:14886–14891. doi:10.1039/C2CP42146A.

[204] Hirsch A, Ruoff Z, Jiao H. Spherical aromaticity in I_h symmetrical fullerenes: The $2(n+1)2$ rule. *Angew Chem Int Ed* 2000, 39:3915–3917. doi:10.1002/1521-3773(20001103)39:21<3915::AID-ANIE3915>3.0.CO;2-O.

[205] Fedorov AS, Fedorov DA, Kuzubov AA, Avramov PV, Nishimura Y, Irle S, Witek HA. Relative isomer abundance of fullerenes and carbon nanotubes correlates with kinetic stability. *Phys Rev Lett* 2011, 107:175506. doi:10.1103/PhysRevLett.107.175506.

[206] W Fowler P. Fullerene graphs with more negative than positive eigenvalues: The exceptions that prove the rule of electron deficiency? *J Chem Soc, Faraday Trans* 1997, 93:1–3. doi:10.1039/A605413G.

[207] Manolopoulos DE, Woodall DR, Fowler PW. Electronic stability of fullerenes: eigenvalue theorems for leapfrog carbon clusters. *J Chem Soc, Faraday Trans* 1992, 88:2427–2435. doi:10.1039/FT9928802427.

[208] Fowler PW, Austin SJ. Closed-shell carbon frameworks: Leapfrog fullerenes and decorated spheriphane hydrocarbons. *J Chem Inf Comput Sci* 1994, 34:264–269. doi:10.1021/ci00018a006.

[209] Deza M, Fowler PW, Rassat A, Rogers KM. Fullerenes as tilings of surfaces. *J Chem Inf Comput Sci* 2000, 40:550–558. doi:10.1021/ci990066h.

[210] Fowler P, Rogers K, Fajtlowicz S, Hansen P, Caporossi G. Facts and conjectures about fullerene graphs: Leapfrog, cylinder and Ramanujan fullerenes. In Betten A, Kohnert A, Laue R, Wassermann A (eds.), *Algebraic Combinatorics and Applications*, 134–146. Springer Berlin Heidelberg, 2001. doi:10.1007/978-3-642-59448-9_10.

[211] Bakowies D, Gelessus A, Thiel W. Quantum-chemical study of C_{78} fullerene isomers. *Chem Phys Lett* 1992, 197:324–329. doi:10.1016/0009-2614(92)85777-8.

[212] Bakowies D, Kolb M, Thiel W, Richard S, Ahlrichs R, Kappes MM. Quantum-chemical study of C_{84} fullerene isomers. *Chem Phys Lett* 1992, 200:411–417. doi:10.1016/0009-2614(92)87013-F.

[213] Chen Z, Thiel W. Performance of semiempirical methods in fullerene chemistry: relative energies and nucleus-independent chemical shifts. *Chem Phys Lett* 2003, 367:15–25. doi:10.1016/S0009-2614(02)01660-3.

[214] Noël Y, De La Pierre M, Zicovich-Wilson CM, Orlando R, Dovesi R. Structural, electronic and energetic properties of giant icosahedral fullerenes up to C_{6000}: insights from an ab initio hybrid DFT study. *Phys Chem Chem Phys* 2014, 16:13390–13401. doi:10.1039/C4CP01442A.

[215] Lu X, Chen Z, Thiel W, Schleyer PvR, Huang R, Zheng L. Properties of fullerene[50] and D_{5h} decachlorofullerene[50]: A computational study. *J Am Chem Soc* 2004, 126:14871–14878. doi:10.1021/ja046725a.

[216] Fowler PW, Sandall JPB. Predictions of special signatures of fullerenes. second-order Jahn-Teller effects on the structures of C_{44}, C_{56}, C_{68} and C_{92}. *J Chem Soc, Perkin Trans 2* 1994, 1917–1921. doi:10.1039/P29940001917.

[217] Paulus B. Electronic and structural properties of the cage-like molecules C_{20} to C_{36}. *Phys Chem Chem Phys* 2003, 5:3364–3367. doi:10.1039/B304539K.

[218] Grossman JC, Mitas L, Raghavachari K. Structure and stability of molecular carbon: Importance of electron correlation. *Phys Rev Lett* 1995, 75:3870–3873. doi:10.1103/PhysRevLett.75.3870.

[219] Zhang BL, Wang CZ, Ho KM, Xu CH, Chan CT. The geometry of small fullerene cages: C_{20} to C_{70}. *J Chem Phys* 1992, 97:5007–5011. doi:10.1063/1.463854.

[220] Chen Z, Jiao H, Bühl M, Hirsch A, Thiel W. Theoretical investigation into structures and magnetic properties of smaller fullerenes and their heteroanalogues. *Theoret Chem Acc* 2001, 106:352–363. doi:10.1007/s002140100284.

[221] Chen Z, Heine T, Jiao H, Hirsch A, Thiel W, Schleyer PvR. Theoretical studies on the smallest fullerene: from monomer to oligomers and solid states. *Chem Eur J* 2004, 10:963–970. doi:10.1002/chem.200305538.

[222] Gerhardt P, Löffler S, Homann KH. Polyhedral carbon ions in hydrocarbon flames. *Chem Phys Lett* 1987, 137:306–310. doi:10.1016/0009-2614(87)80889-8.

[223] Zhang QL, O'Brien SC, Heath JR, Liu Y, Curl RF, Kroto HW, Smalley RE. Reactivity of large carbon clusters: Spheroidal carbon shells and their possible relevance to the formation and morphology of soot. *J Phys Chem* 1986, 90:525–528. doi:10.1021/j100276a001.

[224] Smalley RE. Self-assembly of the fullerenes. *Acc Chem Res* 1992, 25:98–105. doi:10.1021/ar00015a001.

[225] Wakabayashi T, Achiba Y. A model for the C_{60} and C_{70} growth mechanism. *Chem Phys Lett* 1992, 190:465–468. doi:10.1016/0009-2614(92)85174-9.

[226] Wakabayashi T, Kikuchi K, Shiromaru H, Suzuki S, Achiba Y. Ring-stacking considerations on higher fullerene growth. *Z Phys D* 1993, 26:258–260. doi:10.1007/BF01425683.

[227] von Helden G, Gotts NG, Bowers MT. Experimental evidence for the formation of fullerenes by collisional heating of carbon rings in the gas phase. *Nature* 1993, 363:60–63. doi:10.1038/363060a0.

[228] George S Hammond VJK (ed.). *Fullerenes: Synthesis, Properties, and Chemistry of Large Carbon Clusters*. ACS Symposium Series. ACD, 1992.

[229] Ōsawa E. Formation mechanism of C_{60} under nonequilibrium and irreversible conditions – an annotation. *Fuller Nanotub Carbon Nanostruct* 2012, 20:299–309. doi:10.1080/1536383X.2012.655104.

[230] Sasaki K, Wakasaki T, Matsui S, Kadota K. Distributions of C_2 and C_3 radical densities in laser-ablation carbon plumes measured by laser-induced fluorescence imaging spectroscopy. *J Appl Phys* 2002, 91:4033–4039. doi:10.1063/1.1455151.

[231] Curl RF, Lee MK, Scuseria GE. C_{60} buckminsterfullerene high yields unraveled. *J Phys Chem A* 2008, 112:11951–11955. doi:10.1021/jp806951v.

[232] Zhang J, Bowles FL, Bearden DW, Ray WK, Fuhrer T, Ye Y, Dixon C, Harich K, Helm RF, Olmstead MM, Balch AL, Dorn HC. A missing link in the transformation from asymmetric to symmetric metallofullerene cages implies a top-down fullerene formation mechanism. *Nature Chem* 2013, 5:880–885. doi:10.1038/NCHEM.1748.

[233] Langa De La Puente F, Nierengarten J. Fullerenes: Principles and applications, 2007.

[234] Kroto HW, Fischer JE, Cox D. *The fullerenes*. Newnes, 1993.

[235] Brinkmann G, Goedgebeur J, McKay BD. *Buckygen* – a program for the efficient generation of all nonisomorphic fullerenes. http://caagt.ugent.be/buckygen/.

[236] Brinkmann G, Goedgebeur J, McKay BD. The smallest fullerene without a spiral. *Chem Phys Lett* 2012, 522:54–55. doi:10.1016/j.cplett.2011.11.056.

[237] Wu Z, Jelski DA, George TF. Vibrational motions of buckminsterfullerene. *Chem Phys Lett* 1987, 137:291–294. doi:10.1016/0009-2614(87)80221-X.

[238] Weeks DE, Harter WG. Rotation-vibration spectra of icosahedral molecules. II. Icosahedral symmetry, vibrational eigenfrequencles, and normal modes of buckminsterfullerene. *J Chem Phys* 1989, 90:4744–4771. doi:10.1063/1.456571.

[239] Hands ID, Dunn JL, Bates CA. A complete nearest-neighbor force field model for C_{60}. *J Chem Phys* 2004, 120:6912–6923. doi:10.1063/1.1683105.

[240] Graver JE. Kekulé structures and the face independence number of a fullerene. *Eur J Combin* 2007, 28:1115–1130. doi:10.1016/j.ejc.2006.03.003.

[241] Randić M, Vukičević D. Kekulé structures of fullerene C_{70}. *Croat Chem Acta* 2006, 79:471–481.

[242] Papadimitriou CH, Steiglitz K. *Combinatorial Optimization: Algorithms and Complexity*. Dover Books on Computer Science, New York, 1998.

[243] Ciesielski A, Krygowski TM, Cyrański MK. *Symmetry* 2010, 2:1390–1400.

[244] Barnette D. In Tutte WT (ed.), *Recent Progress in Combinatorics*, 343. Academic Press, New York, 1969.

[245] Marušić D. Hamilton cycles and paths in fullerenes. *J Chem Inf Model* 2007, 47:732–736.

[246] Robinson RW, Wormald NC. Almost all regular graphs are hamiltonian. *Random Struct Alg* 1994, 5:363–374. doi:10.1002/rsa.3240050209.

[247] Gray JJ. *The Hilbert Challenge*. Oxford University Press, Oxford, 2000.

[248] Fernando Langa JFN. *Fullerenes - Principles and Applications*. Royal Society of Chemistry, Cambridge, 2007.

[249] Brinkmann G, McKay BD. Construction of planar triangulations with minimum degree 5. *Discrete Math* 2005, 301:147–163. doi:10.1016/j.disc.2005.06.019.

[250] Brinkmann G. *fullgen* – a program for generating nonisomorphic fullerenes,. http://cs.anu.edu.au/∼bdm/plantri/.

[251] Brinkmann G, Friedrichs OD, Lisken S, Peeters A, Cleemput NV. Cage – a virtual environment for studying some special classes of plane graphs – an update. *MATCH Commun Math Comput Chem* 2010, 63:533–552.

[252] Pisanski T. *Vega* – a Mathematica based system for manipulating graphs. http://vega.ijp.si/.

[253] Myrvold W, Bultena B, Daugherty S, Debroni B, Girn S, Minchenko M, Woodcock J, Fowler PW. *MATCH Commun Math Comput Chem,* 2007, 58:403–422.

[254] Tonner R, Frenking G, Lein M, Schwerdtfeger P. Packed to the rafters: Filling up C_{60} with rare gas atoms. *ChemPhysChem* 2011, 12:2081–2084. doi: 10.1002/cphc.201100360.

[255] Richeson DS. *Euler's Gem: The polyhedron formula and the birth of topology*. Princeton University Press, 2012.

[256] Yoshida M. VRML gallery of fullerenes. http://www.jcrystal.com/steffenweber/gallery/Fullerenes/Fullerenes.html, 1991.

[257] Legault CY. *CYLview*, Université de Sherbrook, Canada. http://www.cylview.org.

[258] *Avogadro*, an advanced molecule editor and visualizer designed for cross-platform use in computational chemistry, molecular modeling, bioinformatics, materials science, and related areas,. http://avogadro.openmolecules.net.

[259] *JMOL*: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org.

[260] *PyMOL*, a user-sponsored molecular visualization system on an open-source foundation. http://www.pymol.org/.

[261] *VMD*, visual molecular dynamics. http://www.ks.uiuc.edu/Research/vmd/.

[262] Graver JE, Monachino YA. Highly symmetric fullerenes and nanotubes. *Symmetry: Culture and Science* 2008, 19:317–340.

[263] Fowler PW, Rogers KM. Spiral codes and Goldberg representations of icosahedral fullerenes and octahedral analogues. *J Chem Inf Comput Sci* 2001, 41:108–111. doi:10.1021/ci9901486.

[264] Fowler PW, Redmond DB. Symmetry aspects of bonding in carbon clusters: the leapfrog transformation. *Theor Chim Acta* 1992, 83:367–375.

[265] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, Cambridge New York Port Chester Melbourne Sydney, 2nd ed., 1997.

[266] Avery J, Babić D, Schwerdtfeger P. *n/a* .

[267] Bauernschmitt R, Ahlrichs R, Hennrich FH, Kappes MM. *J Am Chem Soc* 1998, 120:5052–5059.

[268] de Vries J, Steger H, Kamke B, Menzel C, Weisser B, Kamke W, Hertel I. *Chem Phys Lett* 1992, 188:159–162.

[269] Babić D, Ori O. Matching polynomial and topological resonance energy of C_{70}. *Chem Phys Lett* 1995, 234:240–244. doi:10.1016/0009-2614(95)00035-3.

[270] Došlić T. Bipartivity of fullerene graphs and fullerene stability. *Chem Phys Lett* 2005, 412:336–340. doi:10.1016/j.cplett.2005.07.013.

[271] Balaban AT, Mills D, Ivanciuc O, Basak S. Reverse Wiener indices. *Croat Chem Acta* 2000, 73:923–941.

[272] Heydari A, Taeri B. Szeged index of nanotubes. *Eur J Combin* 2009, 30:1134–1141. doi:10.1016/j.ejc.2008.09.009.

[273] Ori O. private communications, 2012.

[274] Fowler PW. Systematics of fullerenes and related clusters. *Phil Trans R Soc Lond* 1993, 343:39–52.

[275] Albertazzi E, Domene C, Fowler PW, Heine T, Seifert G, Van Alsenoy C, Zerbetto F. Pentagon adjacency as a determinant of fullerene stability. *Phys Chem Chem Phys* 1999, 1:2913–2918.

[276] Réti T, László I. On the combinatorial characterization of fullerene graphs. *Acta Polytech Hung* 2009, 6:85–93.

[277] Carathéodory C. Über den Variabilitätsbereich der fourier'schen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico di Palermo* 1911, 32:193–217. doi:10.1007/BF03014795.

[278] Pisanski T, Plestenjak B, Graovac A. NiceGraph program and its application in chemistry. *Croat Chem Acta* 1995, 68:283–292.

[279] Kamada T, Kawai S. An algorithm for drawing general undirected graphs. *Inf Process Lett* 1989, 31:7–15.

[280] Georgakilas V, Perman JA, Tucek J, Zboril R. Broad family of carbon nanoallotropes: Classification, chemistry, and applications of fullerenes, carbon dots, nanotubes, graphene, nanodiamonds, and combined superstructures. *Chem Rev* 2015, 115:4744–4822. doi:10.1021/cr500304f.

[281] Manolopoulos DE, May JC, Down SE. Theoretical studies of the fullerenes: C_{34} to C_{70}. *Chem Phys Lett* 1991, 181:105–111. doi:10.1016/0009-2614(91)90340-F.

[282] Manolopoulos DE, Fowler PW. A fullerene without a spiral. *Chem Phys Lett* 1993, 204:1–7. doi:10.1016/0009-2614(93)85597-H.

[283] Brinkmann G, Goedgebeur J, Mélot H, Coolsaet K. House of graphs: a database of interesting graphs, 2013.

[284] Becke AD. Density-functional thermochemistry. iii. the role of exact exchange. *J Chem Phys* 1993, 98:5648–5652.

[285] Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 1988, 37:785–789.

[286] Weigend F. Accurate coulomb-fitting basis sets for h to rn. *Phys Chem Chem Phys* 2006, 8:1057–1065.

[287] Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C. Electronic Structure Calculations on Workstation Computers: The Program System TURBOMOLE. *Chem Phys Letters* 1989, 162:165–169.

[288] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J,

Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ. Gaussian09 Revision D.01, 2009.

[289] Fowler PW, Graovac A, Žerovnik J, Pisanski T. *A generalized ring spiral algorithm for coding fullerenes and other cubic polyhedra*. Preprint series / Institute of Mathematics, Physics and Mechanics, Department of Mathematics, University of Ljubljana. Inst., 1998.

[290] Brinkmann G. Problems and scope of spiral algorithms and spiral codes for polyhedral cages. *Chem Phys Lett* 1997, 272:193–198. doi:10.1016/S0009-2614(97)88009-8.

[291] Brack M. The physics of simple metal clusters: self-consistent jellium model and semiclassical approaches. *Rev Mod Phys* 1993, 65:677–732. doi:10.1103/RevModPhys.65.677.

[292] Beckhaus HD, Rüchardt C, Kao M, Diederich F, Foote CS. The stability of buckminsterfullerene (C_{60}): Experimental determination of the heat of formation. *Angew Chem Int Ed* 1992, 31:3–64.

[293] Dappe YJ, Basanta MA, Flores F, Ortega J. Weak chemical interaction and van der waals forces between graphene layers: A combined density functional and intermolecular perturbation theory approach. *Phys Rev B* 2006, 74:205434–1–9. doi:10.1103/PhysRevB.74.205434.

[294] Brinkmann G, Fowler PW. Spiral coding of leapfrog polyhedra. *J Chem Inf Comput Sci* 1998, 38:463–468. doi:10.1021/ci970107e.

[295] Wirz LN, Tonner R, Hermann A, Sure R, Schwerdtfeger P. From small fullerenes to the graphene limit: A harmonic force-field method for fullerenes and a comparison to density functional calculations for Goldberg–Coxeter fullerenes up to C_{980}. *J Comput Chem* 2015. doi:10.1002/jcc.23894.

[296] Garcia-Borràs M, Osuna S, Luis JM, Swart M, Solà M. The role of aromaticity in determining the molecular structure and reactivity of (endohedral metallo)fullerenes. *Chem Soc Rev* 2014, 43:5089–5105. doi:10.1039/C4CS00040D.

[297] Cyvin SJ, Brendsdal E, Cyvin BN, Brunvoll J. Molecular vibrations of footballene. *Chem Phys Lett* 1988, 143:377–380. doi:10.1016/0009-2614(88)87050-7.

[298] Jishi RA, Mirie RM, Dresselhaus MS. Force-constant model for the vibrational modes in C_{60}. *Phys Rev B* 1992, 45:13685–13689. doi: 10.1103/PhysRevB.45.13685.

[299] Feldman JL, Broughton JQ, Boyer LL, Reich DE, Kluge MD. Intramolecular-force-constant model for C_{60}. *Phys Rev B* 1992, 46:12731–12736. doi: 10.1103/PhysRevB.46.12731.

[300] Ceulemans A, Fowler PW, Vos I. C_{60} vibrates as a hollow sphere. *J Chem Phys* 1994, 100:5491–5500. doi:10.1063/1.467167.

[301] Ceulemans A, Titeca BC, Chibotaru LF, Vos I, Fowler PW. Complete bond force field for trivalent and deltahedral cages: Group theory and applications to cubane, closo-dodecahedrane, and buckminsterfullerene. *J Phys Chem A* 2001, 105:8284–8295. doi:10.1021/jp0036792.

[302] Dunlap BI, Boettger JC. Local-density-functional study of the fullerenes, graphene and graphite. *J Phys B: At Mol Opt Phys* 1996, 29:4907.

[303] Dunlap BI, Zope RR. Efficient quantum-chemical geometry optimization and the structure of large icosahedral fullerenes. *Chem Phys Lett* 2006, 422:451–454. doi:10.1016/j.cplett.2006.02.100.

[304] Sure R, Tonner R, Schwerdtfeger P. A systematic study of rare gas atoms encapsulated in small fullerenes using dispersion corrected density functional theory. *J Comput Chem* 2015, 36:88–96. doi:10.1002/jcc.23787.

[305] Ahlrichs R, Armbruster MK, Bär M, H-P Baron RB, Crawford N, Deglmann P, Ehrig M, Eichkorn K, Elliott S, Furche F, Haase F, Häser M, Hättig C, Hellweg A, Horn H, Huber C, Huniar U, Kattannek M, Kölmel C, Kollwitz M, May K, Nava P, Ochsenfeld C, Öhm H, Patzelt H, Rappoport D, Rubner O, Schäfer A, Schneider U, Sierka M, Treutler O, Unterreiner B, von Arnim M, Weigend F, Weis P, Weiss H, 2012.

[306] Kresse G, Furthmüller J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys Rev B* 1996, 54:11169–11186.

[307] Blöchl PE. Projector augmented-wave method. *Phys Rev B* 1994, 50:17953–17979.

[308] Kresse G, Joubert D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys Rev B* 1999, 59:1758.

[309] Grimme S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J Comput Chem* 2006, 27:1787–1799.

[310] Risthaus T, Grimme S. Benchmarking of london dispersion-accounting density functional theory methods on very large molecular complexes. *J Chem Theory Comput* 2013, 9:1580–1592.

[311] Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI. Van der waals density functional for general geometries. *Phys Rev Lett* 2004, 92:246401. doi: 10.1103/PhysRevLett.92.246401.

[312] Klimeš J, Bowler DR, Michaelides A. Chemical accuracy for the van der waals density functional. *J Phys: Condens Matter* 2010, 22:022201.

[313] Lee K, Murray ED, Kong L, Lundqvist BI, Langreth DC. Higher-accuracy van der waals density functional. *Phys Rev B* 2010, 82:081101. doi: 10.1103/PhysRevB.82.081101.

[314] Hedberg K, Hedberg L, Bethune DS, Brown CA, Dorn HC, Johnson RD, de Vries M. Bond lengths in free molecules of buckminsterfullerene, C_{60}, from gas-phase electron diffraction. *Science* 1991, 254:410–412. doi: 10.1126/science.254.5030.410.

[315] Barnard AS, Russo SP, Snook IK. Size dependent phase stability of carbon nanoparticles: Nanodiamond versus fullerenes. *J Chem Phys* 2003, 118:5094–5097. doi:10.1063/1.1545450.

[316] Lewis GR, Bunting WE, Zope RR, Dunlap BI, Ellenbogen JC. Smooth scaling of valence electronic properties in fullerenes: From one carbon atom, to C_{60}, to graphene. *Phys Rev A* 2013, 87:052515. doi:10.1103/PhysRevA.87.052515.

[317] Afeefy HY, Liebman JF, Stein SE. *Neutral Thermochemical Data. In NIST Chemistry WebBook, NIST Standard Reference Database, Number 69; Linstrom, P. J., Mallard, W. G., Eds.* National Institute of Standards and Technology: Gaithersburg, MD; see http://webbook.nist.gov (accessed January 1, 2015)., 2006.

[318] Dubay O, Kresse G. Accurate density functional calculations for the phonon dispersion relations of graphite layer and carbon nanotubes. *Phys Rev B* 2003, 67:035401. doi:10.1103/PhysRevB.67.035401.

[319] Mounet N, Marzari N. First-principles determination of the structural, vibrational and thermodynamic properties of diamond, graphite, and derivatives. *Phys Rev B* 2005, 71:205214. doi:10.1103/PhysRevB.71.205214.

[320] Lazzeri M, Attaccalite C, Wirtz L, Mauri F. Impact of the electron-electron correlation on phonon dispersion: Failure of LDA and GGA DFT functionals in graphene and graphite. *Phys Rev B* 2008, 78:081406. doi: 10.1103/PhysRevB.78.081406.

[321] Locht IL. *The effect of temperature on the phonon dispersion relation in graphene*. Master thesis, Radboud University of Nijmegen, Nijmegen, Netherlands, 2012.

[322] Grochala W. Diamond: Electronic ground state of carbon at temperatures approaching 0 K. *Angew Chem Int Ed* 2014, 53:3680–3683. doi: 10.1002/anie.201400131.

[323] Polak E. *Computational Methods in Optimization*, vol. 77. New York, 1971.

[324] http://www.gnu.org/software/gsl/manual/html_node/Multidimensional-Minimization.html.

[325] Griewank A, Walther A. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2nd ed., 2008.

[326] Wirz LN, Babić D, Avery J, Schwerdtfeger P. Toward tight upper and lower bounds for Hamilton cycles in fullerene graphs. *to be submitted* 2015.

[327] Karp RM. Reducibility among combinatorial problems. In Miller RE, Thatcher JW (eds.), *Complexity of Computer Computations*, 85–103. Plenum Press, 1972.

[328] Rubin F. A search procedure for Hamilton paths and circuits. *J ACM* 1974, 21:576–580. doi:10.1145/321850.321854.

[329] Garey MR, Johnson DS. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Oxford University Press, New York, 1983.

[330] Liśkiewicz M, Ogihara M, Toda S. The complexity of counting self-avoiding walks in subgraphs of two-dimensional grids and hypercubes. *Theoretical Computer Science* 2003, 304:129–156.

[331] Tutte WT. A theorem on planar graphs. *Trans Amer Math Soc* 1956, 82:99–116.

[332] Gould RJ. Advances on the Hamiltonian problem – a survey. *Graphs Combinat* 2003, 19:7–52. doi:10.1007/s00373-002-0492-x.

[333] Eppstein D. The traveling salesman problem for cubic graphs. *J Graph Algorit Applic* 2007, 11:61–81.

[334] Gebauer H. On the number of hamilton cycles in bounded degree graphs. *Proc 5th Workshop on Analytic Algorithmics and Combinatorics (ANALCO)* 2008.

[335] Gebauer H. Enumerating all hamilton cycles and bounding the number of hamilton cycles in 3-regular graphs. *Electronic J Combinat* 2011, 18:P132.

[336] Dillencourt MB. Polyhedra of small order and their Hamiltonian properties. *J Comb Theory B* 1996, 66:87–122. doi:10.1006/jctb.1996.0008.

[337] Lederberg J. Dendral-64: A system for computer construction, enumeration and notation of organic molecules as tree structures and cyclic graphs. part ii. topology of cyclic graphs. Tech. rep., Interim Report to the National Aeronautics and Space Administration Grant NsG 81-60, 1965.

[338] Holton D, McKay B. The smallest non-hamiltonian 3-connected cubic planar graphs have 38 vertices. *J Comb Theory B* 1988, 45:305–319. doi:10.1016/0095-8956(88)90075-5.

[339] Tutte WT. On Hamiltonian circuits. *J London Math Soc* 1946, s1-21:98–101. doi:10.1112/jlms/s1-21.2.98.

[340] Holton D, Aldred RE. Planar graphs, regular graphs, bipartite graphs and Hamiltonicity. *Aust J Combinat* 1999, 20:111–132.

[341] Gould RJ. Updating the hamiltonian problem - a survey. *J Graph Theory* 1991, 15:121–157. doi:10.1002/jgt.3190150204.

[342] Holton D, Manvel B, McKay B. Hamiltonian cycles in cubic 3-connected bipartite planar graphs. *J Comb Theory B* 1985, 38:279–297. doi:10.1016/0095-8956(85)90072-3.

[343] Aldred R, Bau S, Holton D, McKay B. Nonhamiltonian 3-connected cubic planar graphs. *SIAM J Discrete Math* 2000, 13:25–32. doi:10.1137/S0895480198348665.

[344] Brinkmann G, McKay BD, von Nathusius U. Backtrack search and look-ahead for the construction of planar cubic graphs with restricted face sizes. *MATCH Commun Math Comput Chem* 2003, 48:163–177.

[345] Goodey PR. A class of Hamiltonian polytopes. *J Graph Theor* 1977, 1:181–185. doi:10.1002/jgt.3190010213.

[346] Kardoš F. A computer-assisted proof of a Barnette's conjecture: Not only fullerene graphs are Hamiltonian. *ArXiv e-prints* 2014.

[347] Erman R, Kardoš F, Miškuf J. Long cycles in fullerene graphs. *J Math Chem* 2009, 46:1103–1111. doi:10.1007/s10910-008-9495-z.

[348] Thomassen C. Tilings of the torus and the Klein bottle and vertex transitive graphs on a fixed surface. *Trans Am Math Soc* 1991, 323:605–635.

[349] Franzblau D. Construction of hamiltonian cycles in layered cubic planar graphs. *Graphs and Combinatorics* 2002, 18:259–270. doi:10.1007/s003730200019.

[350] Babić D. Nomenclature and coding of fullerenes. *J Chem Inf Comput Sci* 1995, 35:515–526.

[351] Dresselhaus MS, Eklund PC. Phonons in carbon nanotubes. *Adv Phys* 2000, 49:705–814.

[352] Faghani M, Ashrafi AR. The topological study of an infinite family of fullerenes with 10n carbon atoms. *Fuller Nanotub Carbon Nanostruct* 2013, 21:561–567. doi:10.1080/1536383X.2011.643424.

[353] Petersen J. Die theorie der regulären graphs. *Acta Mathematica* 1891, 15:193–220. doi:10.1007/BF02392606.

[354] Frieze A. On the number of perfect matchings and hamilton cycles in ϵ-regular non-bipartite graphs. *Electronic J Combinat* 2000, 7:2.

[355] Yoshida M, Fowler PW. Dihedral fullerenes of threefold symmetry with and without face spirals. *J Chem Soc, Faraday Trans* 1997, 93:3289–3294. doi: 10.1039/A702351K.

[356] Schwenk AJ, Wilson RJ. Eigenvalues of graphs. In Beineke LW, Wilson RJ (eds.), *Selected Topics in Graph Theory*. Academic Press, London New York SanFrancisco, 1978.

[357] Sundholm D, Wirz LN, Schwerdtfeger P. Novel hollow all-carbon structures. *Nanoscale* 2015, 7:15886–15894. doi:10.1039/C5NR04370K.

[358] Ramakrishna Matte HSS, Gomathi A, Manna AK, Late DJ, Datta R, Pati SK, Rao CNR. MoS_2 and WS_2 analogues of graphene. *Angew Chem Int Ed* 2010, 122:4153–4156. doi:10.1002/ange.201000009.

[359] Ivanovskii AL. Graphene-based and graphene-like materials. *Russ Chem Rev* 2012, 81:571–605. doi:10.1070/RC2012v081n07ABEH004302.

[360] Xu M, Liang T, Shi M, Chen H. Graphene-like two-dimensional materials. *Chem Rev* 2013, 113:3766–3798. doi:10.1021/cr300263a.

[361] Baughman RH, Eckhardt H, Kertesz M. Structure-property predictions for new planar forms of carbon: Layered phases containing sp^2 and sp atoms. *J Chem Phys* 1987, 87:6687–6699. doi:10.1063/1.453405.

[362] Diederich F. Carbon scaffolding: building acetylenic all-carbon and carbon-rich compounds. *Nature* 1994, 369:199–207. doi:10.1038/369199a0.

[363] Haley MM, Brand SC, Pak JJ. Carbon networks based on dehydrobenzoannulenes: Synthesis of graphdiyne substructures. *Angew Chem Int Ed* 1997, 36:836–838. doi:10.1002/anie.199708361.

[364] Bunz UHF, Rubin Y, Tobe Y. Polyethynylated cyclic π-systems: scaffoldings for novel two and three-dimensional carbon networks. *Chem Soc Rev* 1999, 28:107–119. doi:10.1039/A708900G.

[365] Spitler EL, Johnson CA, Haley MM. Renaissance of annulene chemistry. *Chem Rev* 2006, 106:5344–5386. doi:10.1021/cr050541c.

[366] Haley MM. Synthesis and properties of annulenic subunits of graphyne and graphdiyne nanoarchitectures. *Pure Appl Chem* 2008, 80:519–532. doi:10.1351/pac200880030519.

[367] Diederich F, Kivala M. All-carbon scaffolds by rational design. *Adv Mater* 2010, 22:803–812. doi:10.1002/adma.200902623.

[368] Tahara K, Yamamoto Y, Gross DE, Kozuma H, Arikuma Y, Ohta K, Koizumi Y, Gao Y, Shimizu Y, Seki S, Kamada K, Moore JS, Tobe Y. Syntheses and properties of graphyne fragments: Trigonally expanded dehydrobenzo[12]annulenes. *Chem Eur J* 2013, 19:11251–11260. doi:10.1002/chem.201300838.

[369] Li G, Li Y, Liu H, Guo Y, Li Y, Zhu D. Architecture of graphdiyne nanoscale films. *Chem Commun* 2010, 46:3256–3258. doi:10.1039/B922733D.

[370] Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA. Electric field effect in atomically thin carbon films. *Science* 2004, 306:666–669. doi:10.1126/science.1102896.

[371] Iijima S. Helical microtubules of graphitic carbon. *Nature* 1991, 354:56–58. doi:10.1038/354056a0.

[372] Ihara S, Itoh S. Helically coiled and toroidal cage forms of graphitic carbon. *Carbon* 1995, 33:931–939. doi:10.1016/0008-6223(95)00022-6.

[373] Meyyappan M, Srivastava D. Carbon Nanotubes. In Goddard III WA, Brenner DW, Lyshevski SE, Iafrate GJ (eds.), *Handbook of Nanoscience, Engineering, and Technology*, chap. 18, 1–26. CRC Press, Boca Raton Fl, 2003.

[374] Dinadayalane TC, Leszczynski J. Remarkable diversity of carbon–carbon bonds: structures and properties of fullerenes, carbon nanotubes, and graphene. *Struct Chem* 2010, 21:1155–1169. doi:10.1007/s11224-010-9670-2.

[375] Enyashin AN, Ivanovskii AL. Graphene allotropes. *Phys Stat Sol B* 2011, 248:1879–1883. doi:10.1002/pssb.201046583.

[376] Mas-Ballesté R, Gómez-Navarro C, Gómez-Herrero J, Zamora F. 2D materials: to graphene and beyond. *Nanoscale* 2011, 3:20–30. doi:10.1039/conr00323a.

[377] Malko D, Neiss C, Viñes F, Görling A. Competition for graphene: Graphynes with direction-dependent dirac cones. *Phys Rev Letters* 2012, 108:086804. doi:10.1103/PhysRevLett.108.086804.

[378] Zhang S, Zhang Y, Huang S, Wang C. Theoretical investigations of sp-sp^2 hybridized zero-dimensional fullerenynes. *Nanoscale* 2012, 4:2839–2842. doi:10.1039/C2NR30299C.

[379] Zoppi L, Siegel JS, Baldridge KK. Electron transport and optical properties of curved aromatics. *WIREs Comput Mol Sci* 2013, 3:1–12. doi:10.1002/wcms.1107.

[380] Zheng JJ, Zhao X, Zhao Y, Gao X. Two-dimensional carbon compounds derived from graphyne with chemical properties superior to those of graphene. *Sci Rep* 2013, 3:1271. doi:10.1038/srep01271.

[381] Peng Q, Dearden AK, Crean J, Han L, Liu S, Wen X, De S. New materials graphyne, graphdiyne, graphone, and graphane: review of properties, synthesis, and application in nanotechnology. *Nanotechnol Sci Appl* 2014, 7:1–29. doi:10.2147/NSA.S40324.

[382] Xi J, Wang D, Shuai Z. Electronic properties and charge carrier mobilities of graphynes and graphdiynes from first principles. *WIREs Comput Mol Sci* 2015, 5:215–227. doi:10.1002/wcms.1213.

[383] Li Z, Smeu M, Rives A, Maraval V, Chauvin R, Ratner MA, Borguet E. Towards graphyne molecular electronics. *Nature Commun* 2015, 6:6321. doi:10.1038/ncomms7321.

[384] Sundholm D. C_{72}: gaudiene, a hollow and aromatic all-carbon molecule. *Phys Chem Chem Phys* 2013, 15:9025–9028. doi:10.1039/C3CP51042E.

[385] Pedersen JJ. The challenge of classifying polyhedra. *California Math* 1977, 2:21–28.

[386] Vosko SH, Wilk L, Nusair M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin-Density Calculations - a Critical Analysis. *Can J Phys* 1980, 58:1200–1211.

[387] Perdew JP. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys Rev B* 1986, 33:8822–8824.

[388] Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 1988, 38:3098–3100.

[389] Schäfer A, Horn H, Ahlrichs R. Fully Optimized Contracted Gaussian-Basis Sets for Atoms Li to Kr. *J Chem Phys* 1992, 97:2571–2577.

[390] TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[391] Casella G, Bagno A, Saielli G. Spectroscopic signatures of the carbon bucky-onions $C_{60}@C_{180}$ and $C_{60}@C_{240}$: a dispersion-corrected DFT study. *Phys Chem Chem Phys* 2013, 15:18030–18038. doi:10.1039/C3CP53273A.

[392] Pople JA, Untch KG. Induced paramagnetic ring currents. *J Am Chem Soc* 1966, 88:4811–4815. doi:10.1021/ja00973a009.

[393] Jusélius J, Sundholm D. The aromaticity and antiaromaticity of dehydroannulenes. *Phys Chem Chem Phys* 2001, 3:2433–2437.

[394] Jusélius J, Sundholm D. Polycyclic Antiaromatic Hydrocarbon. *Phys Chem Chem Phys* 2008, 10:6630–6634.

[395] Fliegl H, Sundholm D, Taubert S, Jusélius J, Klopper W. Magnetically Induced Current Densities in Aromatic, Antiaromatic, Homoaromatic, and Nonaromatic Hydrocarbons. *J Phys Chem A* 2009, 113:8668–8676. doi: 10.1021/jp9029776.

[396] Jiao Y, Du A, Hankel M, Zhu Z, Rudolph V, Smith SC. Graphdiyne: a versatile nanomaterial for electronics and hydrogen purification. *Chem Commun* 2011, 47:11843–11845. doi:10.1039/C1CC15129K.

[397] Long M, Tang L, Wang D, Li Y, Shuai Z. Electronic structure and carrier mobility in graphdiyne sheet and nanoribbons: Theoretical predictions. *ACS Nano* 2011, 5:2593–2600. doi:10.1021/nn102472s.

[398] Sebastiani D, Parker MA. Polyhedral phenylacetylenes: The interplay of aromaticity and antiaromaticity in convex graphyne substructures. *Symmetry* 2009, 1:226–239. doi:10.3390/sym1020226.

[399] Ivanovskii A. Graphynes and graphdyines. *Prog Solid State Chem* 2013, 41:1–19. doi:10.1016/j.progsolidstchem.2012.12.001.

[400] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953, 21:1087–1092. doi:10.1063/1.1699114.

[401] Alder BJ, Wainwright TE. Phase transition for a hard sphere system. *J Chem Phys* 1957, 27:1208–1209. doi:10.1063/1.1743957.

[402] Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *nature structural biology* 2002, 9:646–652.

[403] van Gunsteren WF, Berendsen HJC, Rullmann JAC. Stochastic dynamics for molecules with constraints: Brownian dynamics of n-alkanes. *Mol Phys* 1981, 44:69–95. doi:10.1080/00268978100102291.

[404] Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman J, Ho CR, Ierardi DJ, Kolossváry I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, , Wang SC. Anton, a special-purpose machine for molecular dynamics simulation. *Comm ACM* 2008, 51:91–97. doi:10.1145/1364782.1364802.

[405] Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann Phys* 1921, 369:253–287. doi:10.1002/andp.19213690304.

[406] Eastwood JW, Hockney RW. Shaping the force law in two-dimensional particle-mesh models. *J Comput Phys* 1974, 16:342–359. doi: http://dx.doi.org/10.1016/0021-9991(74)90044-8.

[407] Darden T, York D, Pedersen L. Particle mesh Ewald: An $n\log(n)$ method for Ewald sums in large systems. *J Chem Phys* 1993, 98:10089–10092. doi: 10.1063/1.464397.

[408] Greengard L, Rokhlin V. A fast algorithm for particle simulations. *J Comput Phys* 1987, 73:325–348. doi:10.1016/0021-9991(87)90140-9.

[409] Frenkel D, Smit B. *Understanding Molecular Simulation: From algorithms to applications*. Academic Press, 2nd ed., 2006.

[410] Hünenberger PH. Thermostat algorithms for molecular dynamics simulations. *Adv Polym Sci* 2005, 173:105–149. doi:10.1007/b99427.

[411] Randall M. A webcomic of romance, sarcasm, math, and language. http://xkcd.com/613/.

[412] Ryckaert JP, Ciccotti G, Berendsen HJ. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977, 23:327–341. doi:10.1016/0021-9991(77)90098-5.

[413] van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, , Tironi IG. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.

[414] Andersen HC. RATTLE: A "velocity" version of the SHAKE algorithm for molecular dynamics calculations. *J Comput Phys* 1983, 52:24–34. doi:10.1016/0021-9991(83)90014-1.

[415] Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 1992, 13:952–962. doi:10.1002/jcc.540130805.

[416] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 1997, 18:1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[417] Kräutler V, van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* 2001, 22:501–508. doi:10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.

[418] Lee SH, Palmo K, Krimm S. WIGGLE: A new constrained molecular dynamics algorithm in Cartesian coordinates. *J Comput Phys* 2005, 210:171–182. doi:10.1016/j.jcp.2005.04.006.

[419] Tao P, Wu X, Brooks BR. Maintain rigid structures in Verlet based Cartesian molecular dynamics simulations. *J Comput Phys* 2012, 137:134110:1–9.

[420] Torda AE, Brunne RM, Huber T, Kessler H, van Gunsteren WF. Structure refinement using time-averaged j-coupling constant restraints. *J Biomol NMR* 1993, 3:55–66.

[421] Hess B, Scheek RM. Orientation restraints in molecular dynamics simulations using time and ensemble averaging. *J Magn Reson* 2003, 164:19–27. doi: 10.1016/S1090-7807(03)00178-2.

[422] Abragam A. *The principles of nuclear magnetic resonance*. Clarendon Press, Oxford, 1961.

[423] Smith SA, Palke WE, Gerig JT. The Hamiltonians of NMR, part I. *Concepts Magn Reson* 1992, 4:107–144.

[424] Smith SA, Palke WE, Gerig JT. The Hamiltonians of NMR, part II. *Concepts Magn Reson* 1992, 4:181–204.

[425] Suits BH. Nuclear quadrupole resonance spectroscopy. In Vij DR (ed.), *Handbook of Applied Solid State Spectroscopy*. Springer, 2006.

[426] Feynman RP, Leighton RB, Sands ML. *The Feynman lectures on physics—Mainly electromagnetism and matter*. Addison-Wesley, 1964.

[427] Landau LD, Lifschitz EM. *The classical theory of fields*, vol. 2 of *Course of theoretical physics*. Butterworth Heinemann.

[428] Hu W, Wang L. Residual dipolar couplings: Measurements and applications to biomolecular studies. vol. 58 of *Annu. Rep. NMR Spectro.*, 231–303. Academic Press, 2006. doi:10.1016/S0066-4103(05)58005-0.

[429] Prestegard JH, Bougault CM, Kishore AI. Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* 2004, 104:3519–3540. doi: 10.1021/cr030419i.

[430] Bax A, Kontaxis G, Tjandra N. Dipolar couplings in macromolecular structure determination. *Meth Enzymol* 2001, 339:127–150.

[431] Bax A, Grishaev A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr Opin Struct Biol* 2005, 15:563–570. doi:10.1016/j.sbi.2005.08.006.

[432] Donald BD, Martin J. Automated NMR assignment and protein structure determination using sparse diplor coupling constraints. *Prog Nucl Magn Reson Spectrosc* 2009, 55:101–127.

[433] Blackledge M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog Nucl Magn Reson Spectrosc* 2005, 46:23–61. doi:10.1016/j.pnmrs.2004.11.002.

[434] Palmer AG. NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 2004.

[435] Tolman JR, Ruan K. NMR residual dipolar couplings as probes of biomolecular dynamics. *J Am Chem Soc* 2006.

[436] Harris RK, Becker ED, de Menezes SMC, Goodfellow R, Granger P. NMR nomenclature. nuclear spin properties and conventions for chemical shifts. *Pure Appl Chem* 2001, 73:1795–1818.

[437] Bryant RG. The NMR time scale. *J Chem Educ* 1983, 60:933–935. doi: 10.1021/ed060p933.

[438] Zweckstetter M. Residual dipolar couplings. EMBO Summer School, Il Ciocco, Italy, 2004.

[439] Prestegard JH, Kishore AI. Partial alignment of biomolecules: an aid to NMR characterization. *Curr Opin Chem Biol* 2001, 5:584–590. doi:10.1016/S1367-5931(00)00247-7.

[440] Tolman JR, al Hashimi HM. NMR studies of biomolecular dynamics and structural plasticity using residual dipolar couplings. In *Annu. Rep. NMR Spectro.*, vol. 51 of *Annu. Rep. NMR Spectro.*, 105 – 166. Academic Press, 2003. doi:10.1016/S0066-4103(03)51003-1.

[441] Fleming K, Matthews S. Media for studies of partially aligned states. In *Protein NMR Techniques*, chap. 6, 79–88. Springer, 2004. doi:10.1385/1-59259-809-9:079.

[442] Ruan K, Tolman JR. Composite alignment media for the measurement of independent sets of NMR residual dipolar couplings. *J Am Chem Soc* 2005, 127:15032–15033. doi:10.1021/ja055520e.

[443] Al-Hashimi H, Valafar H, Terrell M, Zartler E, Eidsness M, Prestegard J. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J Magn Reson* 2000, 143:402–406. doi:10.1006/jmre.2000.2049.

[444] Tolman JR. A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy. *J Am Chem Soc* 2002.

[445] Peti W, Meiler J, Brüschweiler R, Griesinger C. Model-free analysis of protein backbone motion from residual dipolar couplings. *J Am Chem Soc* 2002, 124:5822–5833. doi:10.1021/ja011883c.

[446] Saupe A, Englert G. High-resolution nuclear magnetic resonance spectra of orientated molecules. *Phys Rev Lett* 1963, 11:462–464. doi:10.1103/PhysRevLett.11.462.

[447] Saupe A. Kernresonanzen in kristallinen Flüssigkeiten und in kristallinflüssigen Lösungen. *Z Naturforschg* 1964, 19 a:161–171.

[448] Tjandra N, Bax A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *SCIENCE* 1997, 278:1111–1114.

[449] Miao X, Mukhopadhyay R, Valafar H. Estimation of relative order tensors, and reconstruction of vectors in space using unassigned RDC data and its application. *J Magn Reson* 2008.

[450] Prestegard JH. New techniques in structural NMR – anisotropic interactions. *Nat Struct Biol* 1998.

[451] Tian F, Valafar H, Prestegard JH. A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. *J Am Chem Soc* 2001, 123:11791–11796. doi:10.1021/ja011806h.

[452] Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S. An easy way to include weak alignment constraints into NMR structure calculation. *J Biomol NMR* 2001.

[453] Weaver JL, Prestegard JH. Nuclear magnetic resonance structural and ligand binding studies of BLBC, a two-domain fragment of barley lectin. *Biochemistry* 1998, 37:116–128. doi:10.1021/bi971619p.

[454] Fischer MWF, Losonczi JA, Weaver JL, Prestegard JH. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* 1999, 38:9013–9022. doi:10.1021/bi9905213.

[455] Skrynnikov NR, Goto NK, Yang D, Choy WY, Tolman JR, Mueller GA, Kay LE. Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with β-cyclodextrin. *J Mol Biol* 2000, 295:1265–1273. doi:10.1006/jmbi.1999.3430.

[456] Tian F, Al-Hashimi HM, Craighead JL, Prestegard JH. Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. *J Am Chem Soc* 2001, 123:485–492. doi:10.1021/ja002900l.

[457] Meiler J, Prompers JJ, Peti W, Griesinger C, Bruschweiler R. Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J Am Chem Soc* 2001, 123:6098–6107. doi:10.1021/ja010002z.

[458] Briggman KB, Tolman JR. De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. *J Am Chem Soc* 2003, 125:10164–10165. doi:10.1021/ja035904+.

[459] Ferrarini A. Modeling of macromolecular alignment in nematic virus suspensions. application to the prediction of NMR residual dipolar couplings. *J Phys Chem B* 2003, 107:7923–7931. doi:10.1021/jp034366k.

[460] Zweckstetter M. NMR: Prediction of molecular alignment from structure using the PALES software. *Nature Protocols* 2008.

[461] Azurmendi HF, Bush CA. Tracking alignment from the moment of inertia tensor (TRAMITE) of biomolecules in neutral dilute liquid crystal solutions. *J Am Chem Soc* 2002, 124:2426–2427.

[462] Fredriksson K, Louhivuori M, Permi P, Annila A. On the interpretation of residual dipolar couplings as reporters of molecular dynamics. *J Am Chem Soc* 2004, 126:12646–12650. doi:10.1021/ja048287d.

[463] Losonczi JA, Andrec M, Fischer MW, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 1999, 138:334–342. doi:10.1006/jmre.1999.1754.

[464] Wirz LN, Allison JR. Fitting alignment tensor components to experimental RDCs, CSAs and RQCs. *J Biomol NMR* 2015, 62:25–29. doi:10.1007/s10858-015-9907-x.

[465] Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol* 1997, 4:732–738. doi: 10.1038/nsb0997-732.

[466] Bayer P, Varani L, Varani G. Refinement of the structure of protein-RNA complexes by residual dipolar coupling analysis. *J Biomol NMR* 1999, 14:149–155. doi:10.1023/A:1008360331296.

[467] Moltke S, Grzesiek S. Structural constraints from residual tensorial couplings in high resolution NMR without an explicit term for the alignment tensor. *J Biomol NMR* 1999, 15:77–82. doi:10.1023/A:1008309630377.

[468] Cornilescu G, Marquardt JL, Ottiger M, Bax A. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc* 1998, 120:6836–6837. doi:10.1021/ja9812610.

[469] Lipsitz RS, Tjandra N. Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct* 2004. doi: 10.1146/annurev.biophys.33.110502.140306.

[470] Mehring M. *Principles of High Resolution NMR in Solids*. Springer Verlag, Berlin Heidelberg New York, 2nd ed., 1983. doi:10.1007/978-3-642-68756-3.

[471] Mason J. Conventions for the reporting of nuclear magnetic shielding (or shift) tensors suggested by participants in the NATO ARW on NMR shielding constants at the university of maryland, college park, july 1992. *Solid State Nucl Magn Reson* 1993, 2:285–288. doi:10.1016/0926-2040(93)90010-K.

[472] Golub GH, van Loan CF. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4th ed., 2012.

[473] Sass HJ, Cordier F, Hoffmann A, Rogowski M, Cousin A, Omichinski JG, Löwen H, Grzesiek S. Purple membrane induced alignment of biological macromolecules in the magnetic field. *J Am Chem Soc* 1999, 121:2047–2055. doi:10.1021/ja983887w.

[474] Ottiger M, Bax A. Determination of relative N–H^{N} , N–C′, C^{α}–C′, and C^{α}–H^{α} effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J Am Chem Soc* 1998, 120:12334–12341. doi:10.1021/ja9826791.

[475] Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K, Oostenbrink C, Schmid N, Steiner D, Wang D, van Gunsteren WF. The GROMOS++ software for the analysis of biomolecular simulation trajectories. *J Chem Theory Comput* 2011, 7:3379–3390. doi:10.1021/ct2003622.

[476] GNU scientific library reference manual. http://www.gnu.org/software/gsl/.

[477] Schmid N, Allison JR, Dolenc J, Eichenberger AP, Kunz APE, van Gunsteren WF. Biomolecular structure refinement using the GROMOS simulation software. *J Biomol NMR* 2011, 51:265–281. doi:10.1007/s10858-011-9534-0.

[478] Schmid N, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF. Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. *Comput Phys Commun* 2012, 183:890–903. doi:10.1016/j.cpc.2011.12.014.

[479] Kunz APE, Allison JR, Geerke DP, Horta BAC, Hünenberger PH, Riniker S, Schmid N, van Gunsteren WF. New functionalities in the GROMOS biomolecular simulation software. *J Comput Chem* 2012, 33:340–353. doi:10.1002/jcc.21954.

[480] Su Z, Coppens P. Rotation of real spherical harmonics. *Acta Crystallogr A* 1994, 50:636–643. doi:10.1107/S0108767394003077.

[481] Ivanic J, Ruedenberg K. Rotation matrices for real spherical harmonics. direct determination by recursion. *J Phys Chem* 1996, 100:6342–6347. doi:10.1021/jp953350u.

[482] Green R. Spherical harmonic lighting: The gritty details. In *Game Developers' Conference*. 2003 .

[483] GROMOS software for biomolecular simulation. http://www.gromos.net.

[484] Lebedev VI. Values of the nodes and weights of quadrature formulas of Gauss–Markov type for a sphere from the ninth to seventeenth order of accuracy that are invariant with respect to an octahedron group with inversion. *Zh Vychisl Mat Mat Fiz* 1975, 15:48–54.

[485] Lebedev VI. Quadratures on the sphere. *Zh Vychisl Mat Mat Fiz* 1976, 16:293–306.

[486] Bertini I, Bianco CD, Gelis I, Katsaros N, Luchinat C, Parigi G, Peana M, Provenzani A, Zoroddu MA. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc Natl Acad Sci* 2004, 101:6841–6846. doi:10.1073/pnas.0308641101.

[487] Maciejewski M, Tjandra N, Barlow PN. Estimation of interdomain flexibility of N-terminus of Factor H using residual dipolar couplings. *Biochemistry* 2011. doi:10.1021/bi200575b.

[488] Roche J, Louis JM, Bax A. Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. *ChemBioChem* 2015, 16:214–218. doi:10.1002/cbic.201402585.

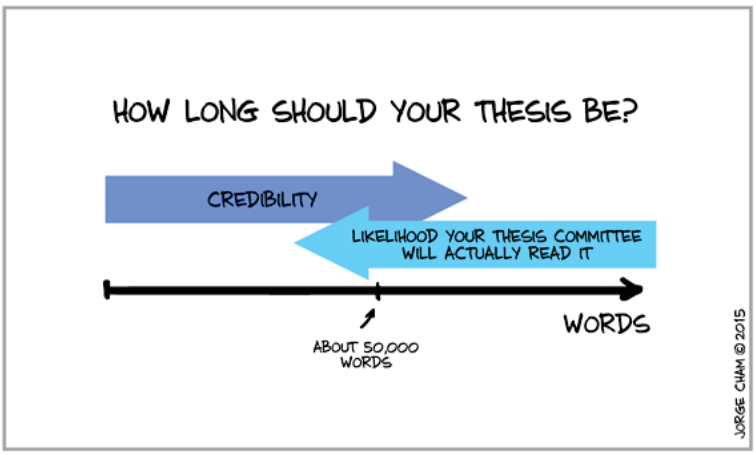
[489] Potts JR, Campbell ID. Fibroneetin structure and assembly. *Curr Opin Cell Biol* 1994, 6:648–655.

[490] Williams MJ, Phan I, Harvey TS, Rostagno A, Gold LI, Campbell I. Solution structure of a pair of fibronectin type 1 modules with fibrin binding activity. *J Mol Biol* 1994, 235:1302–1311.

[491] Potts JR, Campbell ID. Structure and function of fibronectin modules. *Matrix Biol* 1996, 15:313–320.

[492] Rudiño-Piñera E, Ravelli RB, Sheldrick GM, Nanao MH, Korostelev VV, Werner JM, Schwarz-Linek U, Potts JR, Garman EF. The solution and crystal structures of a module pair from the staphylococcus aureus-binding site of human fibronectin—a tale with a twist. *J Biomol NMR* 2007, 368:833–844. doi: 10.1016/j.jmb.2007.02.061.

[493] Poger D, Van Gunsteren WF, Mark AE. A new force field for simulating phosphatidylcholine bilayers. *J Comput Chem* 2010, 31:1117–1125. doi: 10.1002/jcc.21396.




```
> find -name \*.tex -print0 | xargs -0 wc -w | tail -n 1
> 97574 total
> texcount -inc -total master.tex | grep 'Words in text'
> Words in text: 53805
```

Tell me if you—against all odds—reached this page.

Statements of Contribution to Doctoral Thesis Containing Publications

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Schwerdtfeger P, Wirz LN, Avery J. The topology of fullerenes. WIREs Comp Mol Sci 2015, 5:96–145. doi:10.1002/wcms.1207.

In which Chapter is the Published Work: 1

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 1/3

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  Major contribution to sections 1.4 (graph generation), 1.5 (local transformations [patch replacements], but not to global transformations [Goldberg-Coxeter transformation]), 1.14 (gas phase formation), and the omitted sections on the face spiral algorithm and the forcefield. Minor contributions to sections 1.1 (graph theory), 1.6 (geometry), 1.11 (perfect matchings), and 1.12 (stability).

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:45:19 +12'00'

Candidate's Signature

2015/07/15
Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:07:33 +02'00'

Principal Supervisor's signature

17/07/2015
Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Schwerdtfeger P, Wirz L, Avery J. Program fullerene: A software package for constructing and analyzing structures of regular fullerenes. J Comput Chem 2013, 34:1508–1526. doi:10.1002/jcc.23278.

In which Chapter is the Published Work: 2

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate: 1/3

  and / or
- Describe the contribution that the candidate has made to the Published Work:

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:39:34 +12'00'

Candidate's Signature

2015/07/15
Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:06:35 +02'00'

Principal Supervisor's signature

17/07/2015
Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Wirz LN, Tonner R, Avery J, Schwerdtfeger P. Structure and properties of the nonface-spiral fullerenes T-C380, D3-C384, D3-C440, and D3-C672 and their halma and leapfrog transforms. J Chem Inf Model 2014, 54:121–130. doi:10.1021/ci4005578.

In which Chapter is the Published Work: 3

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 2/3
  and / or
- Describe the contribution that the candidate has made to the Published Work:

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:41:22 +12'00'

Candidate's Signature

2015/07/15
Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:07:14 +02'00'

Principal Supervisor's signature

17/07/2015
Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Wirz LN, Tonner R, Hermann A, Sure R, Schwerdtfeger P. From small fullerenes to the graphene limit: A harmonic force-field method for fullerenes and a comparison to density functional calculations for Goldberg–Coxeter fullerenes up to C980. J Comput Chem 2015. doi:10.1002/jcc.23894.

In which Chapter is the Published Work: 4

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  Programming and fitting of the force field

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:49:14 +12'00'

Candidate's Signature

2015/07/15

Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:08:16 +02'00'

Principal Supervisor's signature

17/07/2015

Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Wirz LN, Babic D, Avery J, Schwerdtfeger P. Toward tight upper and lower bounds for Hamilton cycles in fullerene graphs 2015.

In which Chapter is the Published Work: 5

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

Proofs for upper bound of the number of Hamilton cycles (using perfect matchings) and of the exact number of HC in two types of nanotubes

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:53:03 +12'00'

Candidate's Signature

2015/07/15
Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:08:36 +02'00'

Principal Supervisor's signature

17/07/2015
Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Sundholm D, Wirz LN, Schwerdtfeger P. Novel hollow all-carbon structures 2015.

In which Chapter is the Published Work: 6

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

Division and programming of algorithms for transforming cubic graphs into gaudiene-graphs, creation of structures

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:54:31 +12'00'

Candidate's Signature

2015/07/15
Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:08:59 +02'00'

Principal Supervisor's signature

17/07/2015
Date

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Lukas Wirz

Name/Title of Principal Supervisor: Prof. Peter Schwerdtfeger

Name of Published Research Output and full reference:

Wirz LN, Allison JR. Fitting alignment tensor components to experimental RDCs, CSAs and RQCs. J Biomol NMR 2015, 62:25–29. doi:10.1007/s10858-015-9907-x.

In which Chapter is the Published Work: 8

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 3/4
  and / or
- Describe the contribution that the candidate has made to the Published Work:

Lukas Wirz
Digitally signed by Lukas Wirz
DN: cn=Lukas Wirz, o, ou=ctcp, email=l.wirz@massey.ac.nz, c=NZ
Date: 2015.07.15 16:47:20 +12'00'

Candidate's Signature

2015/07/15
Date

Peter Schwerdtfeger
Digitally signed by Peter Schwerdtfeger
DN: cn=Peter Schwerdtfeger, o=Massey University, ou=New Zealand Institute for Advanced Study, email=p.a.schwerdtfeger@massey.ac.nz, c=NZ
Date: 2015.07.17 09:07:55 +02'00'

Principal Supervisor's signature

17/07/2015
Date