Cover page

Exam information
NFYK10020E - Physics Thesis 60 ECTS, Niels Bohr Institute - Kontrakt:114918 (Benjamin Heuser)

Handed in by
Benjamin Heuser
krn362@alumni.ku.dk

Exam administrators
Eksamensteam, tel 35 33 64 57
eksamen@science.ku.dk

Assessors
James Emil Avery
Examiner
j.avery@nbi.ku.dk

Ilya Solovyev
Co-examiner
ilia@sdu.dk

Kurt Valentin Mikkelsen
Internal co-examinator
kmi@chem.ku.dk

Hand-in information

Titel, engelsk: Folding Carbon, Computational Study of the Auto-Assembly of Existing Fullerene Precursor Molecules, Towards Fast Automated Quality Assessment of Precursors for Arbitrary Fullerene Structures

Tro og love-erklæring: Yes

UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

MSc in Physics

Folding Carbon

Computational Study of the Auto-Assembly of Existing Fullerene Precursor Molecules, Towards Fast Automated Quality Assessment of Precursors for Arbitrary Fullerene Structures

Benjamin Heuser

Supervised by Assoc. Prof., Ph.D. James E. Avery and Professor, Ph.D., Dr.Scient. Kurt V. Mikkelsen

July 17, 2020

Benjamin Heuser

Folding Carbon

MSc in Physics, July 17, 2020

Supervisors: Assoc. Prof., Ph.D. James E. Avery and Professor, Ph.D., Dr.Scient. Kurt V. Mikkelsen

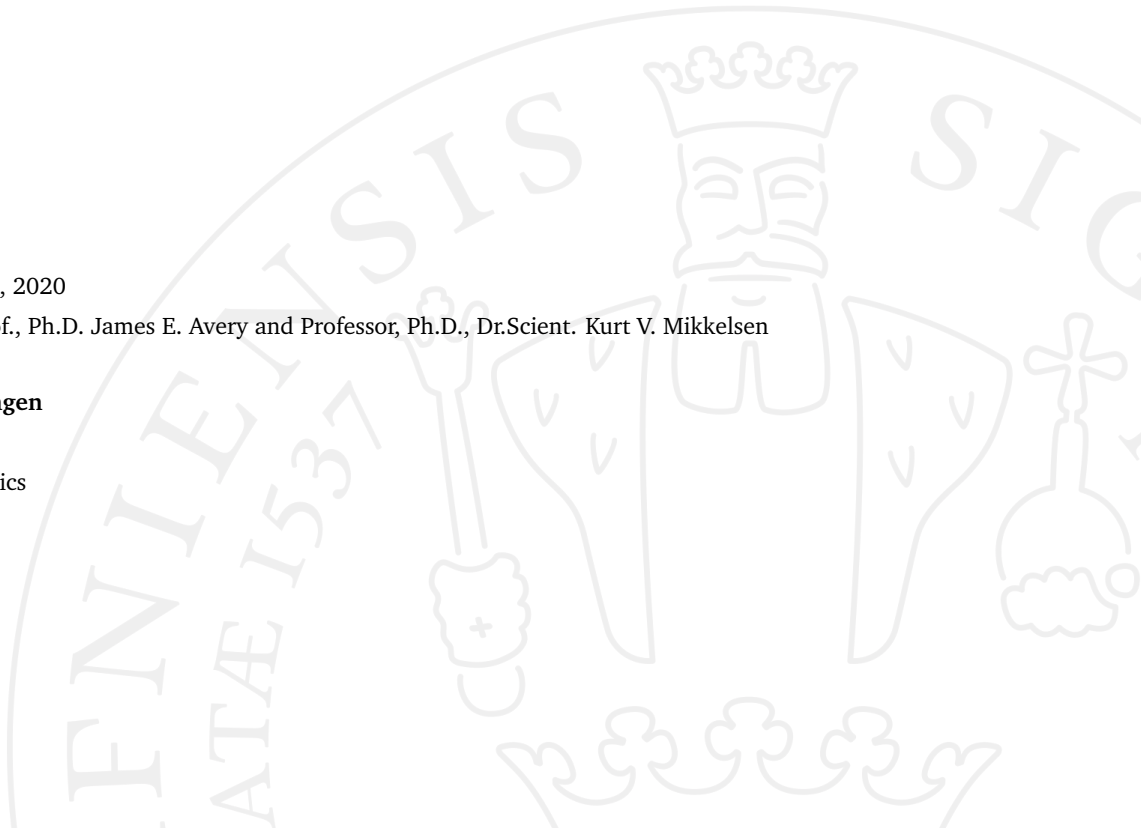
University of Copenhagen

Niels Bohr Institute

Masters Degree in Physics

Blegdamsvej 17

2100 Copenhagen N



Abstract

Fullerenes are a class of polyhedral all-carbon molecules with a closed surface consisting of pentagonal and hexagonal rings. Theory predicts infinitely many stable fullerenes, of which only a handful have yet been produced and analysed. The chemical and electronic properties of fullerenes are of major interest to the scientific community and their applications, predicted and existing are manifold. There are rational synthesis paths for a few fullerenes available, which make use of auto-assembling planar polyarenes. Would it be possible to generalise this concept to any fullerene isomer? The description of fullerenes as mathematical graphs permit the fast automated construction of all possible precursors molecules. The number of possible unfoldings for a given fullerene is already in the order 10^{38} for the C_{60} and growing rapidly. This makes them inaccessible by computational means. To select a specific precursor molecule, rules have to be found to predict what a good precursor should look like. There are too many molecules to test them by any means, theoretical or experimental. Therefore, a fundamental understanding of the mechanism of existing auto-assembly is required to generalise this concept.
This thesis builds a software framework to systematically analyse the reaction mechanism of existing precursor auto-assembly. The precursor geometry was interpolated from the planar to the closed fullerene geometry to decrease the distance of the bond-forming carbon atoms. Each step, a restricted quantum mechanical geometry optimisation was performed to determine at which point a cyclo-dehydrogenation would be energetically favourable. The key periphery atoms were placed either on the inside or the outside of the forming geodesic carbon cage. While no cyclo-dehydrogenation could be achieved for an outside placement of the periphery atoms, the inside placement led to a cyclo-dehydrogenation.
All precursor geometries obtained by quantum mechanical optimisation have been analysed for their bond lengths, bond angles and face planarity to justify the application of future force field methods in the simulation of fullerene precursor auto-assembly. All analysed quantities were found to justify a harmonic approximation for the associated potentials at all points along the studied auto-assembly path.
A thorough comparison of the description of geometric and energetic quantities in the halogen assisted dehydrogenation in the $C_2H_4 + C_2H_3F \longrightarrow C_4H_6 + HF$ reaction was performed. In this, the quality of the three different functionals B3LYP, CAM-B3LYP and M062X with and without Grimme (GD3) dispersion was assessed in comparison to Coupled-Cluster (CC) calculations. While the geometric description of all three functionals is in very good agreement with the CC results for all analysed quantities, the difference found in the energetic accuracy of the functionals was significant.

Contents

1 Introduction **1**

1.1 Project Context . . . 3
1.2 Software . . . 4

2 Theory **5**

2.1 Fullerenes . . . 5
2.1.1 Geometric Properties . . . 5
2.1.2 Chemical Bonds within Fullerenes . . . 6
2.2 Synthesis of Fullerenes . . . 7
2.2.1 Rational Synthesis from Flat Precursor Molecules . . . 7
2.2.2 Flash Vacuum Pyrolysis . . . 8
2.3 Describing Precursors as a Rigid Bodies with Hinges . . . 9
2.4 Graph-theoretical Background . . . 10
2.4.1 Dual Graphs . . . 12
2.4.2 Fullerene Unfolding and their Graphs . . . 12
2.4.3 Minimum Spanning Tree . . . 13
2.5 Rigid Body Vertex Construction . . . 14
2.5.1 Planar Vertex Coordinates . . . 14
2.5.2 Vertex Coordinates in Three Dimensional Space . . . 17
2.5.3 Avoiding Double Rotations . . . 19
2.6 Quantum Mechanical Background . . . 20
2.6.1 From one to Many Particles . . . 20
2.6.2 Hartree Fock Theory . . . 22
2.6.3 Coupled-Cluster Theory . . . 26
2.6.4 Density Functional Theory . . . 29
2.6.5 Basis Sets . . . 33
2.6.6 Dispersion . . . 34
2.7 Force Field Methods . . . 35
2.7.1 Force Field Energy in Molecular Mechanics . . . 36
2.7.2 Gradient of the Energy . . . 41
2.7.3 Obtaining Force Constants . . . 42
2.7.4 Validity of the Potential Approximations . . . 44

3 Method **47**

3.1 Obtaining Relevant Graph Information . . . 47
3.2 Comparison of Different Functionals . . . 49
3.2.1 Comparison of Geometric Properties . . . 51

3.2.2 Comparison of Reaction Enthalpy . . . 52
3.3 DFT Study of Optimal C_{60}-Precursor Geometry . . . 52
3.4 Closing the C_{60} Precursor with Atoms Placed Inside of the Bowl . . . 54
3.5 Finding the Optimal Parameter Values along a Path . . . 56

4 Results and Discussion **57**
4.1 Functional Comparison for a Fluorine Assisted Dehydrogenation of an sp^2 Hybridised Carbon Atom . . . 57
4.1.1 Comparison of the Optimal Geometries for CC and DFT . . . 58
4.1.2 Comparison of the Reaction Enthapy . . . 65
4.2 Geometries Along an Auto Assembly Path for the C_{60} Precursor Molecule . . . 67
4.2.1 Closing Precursor Molecule at the First Hinge . . . 67
4.2.2 Dehydrogenation Along the Assembly Path . . . 75
4.2.3 Placing the Halogen and Hydrogen Atom on the Inside . . . 76
4.3 Summary . . . 80

5 Conclusion and Outlook **83**

Appendices **85**

A Appendix A **87**
A.1 Geometric Quantities Along the Closing Path with Chlorine in the Precursor Molecule's key positions . . . 87
A.2 Geometric Quantities Along the Closing Path for the Inside Fluorine Path . . . 88

A Bibliography **89**

1 Introduction

Fullerenes are closed-surface, polyhedral molecules formed solely by carbon atoms. The first fullerene to be discovered was the C_{60} fullerene by Kroto et al. in 1985 [1]. This particular fullerene consists of exactly 60 carbon atoms in 20 hexagonal and 12 pentagonal faces, arranged in the same way as the modern-day football. It is also referred to as the Buckminsterfullerene or Buckyball, named after the famous architect Richard Buckminster Fuller due to its similarity to his geodesic dome constructions. Fullerenes come in many shapes and forms and have intrigued scientist all over the world with possible applications ranging from biomedical research through hydrogen storage to solar cell development. Their geometric properties, with only pentagonal and hexagonal faces, as well as their distinctive chemical bond structure makes fullerene molecules very well suited for systematic analysis based on computational methods and graph theory.

The elegance of this description of fullerene molecules as mathematical graphs offers a large variety of possibilities for systematic computational studies of fullerenes, and a rich field of study has grown around it [2]. Many properties of fullerene molecules can be determined directly from the bond information, which can be obtained with incredible efficiency for the full isomer spaces [3–5]. The geometrical structure and possible unfoldings can likewise be obtained algorithmically from the bond graph, and it is possible to solve wave equations, which could predict molecular properties, on the fullerene surface by describing the polyhedral surface as a manifold [6]. All of these possibilities are currently being explored at the University of Copenhagen (UCPH) by an interdisciplinary research team combining methods from theoretical physics, computational and mathematical chemistry and computer science into a new intermediate sub-field.

The number of potentially stable fullerenes is theoretically unlimited and other spherical fullerene molecules like C_{70} and C_{78} (constituted by 70 and 78 carbon atoms) have been produced and studied. However, the current means of fullerene production are extremely limited and rely upon procedures like laser ablation, or resistive heating of graphite. Although those methods have produced kilogram quantities of the C_{60} buckyball and small quantities of other fullerenes, they have major disadvantages. They are non-selective and thus, the production of larger fullerene molecules happen merely "accidental". Therefore, the possibility of producing a specific isomer of a fullerene is near to impossible.

A first solution to this problem was introduced by Scott et al., who constructed a rational chemical synthesis path for the icosahedral C_{60} fullerene in 2002 from an auto-assembling planar precursor molecule [7]. The precursor molecule contained all carbon atoms as well as 75 of the 90 carbon-carbon bonds and chlorine atoms in some key positions. When exposed to flash vacuum pyrolysis, the precursor molecule was designed to selectively auto-assemble to the Buckminsterfullerene. This breakthrough in fullerene synthesis was soon followed by the discovery of selective precursor molecules for isomers of the C_{78} and C_{84} fullerene [8, 9]. The original C_{60} precursor molecule has been altered by exchanging

the chlorine with fluorine and placing them at strategic positions within the precursor, by Kabdulov et al. in 2010, which improved the yield of the reaction significantly [10]. Otero et al. used the fully hydrogenated C_{60} precursor on a platinum (111) surface and reached yields nearly $100\ \%$ [11].

The huge success of rational synthesis from auto-assembling precursors, in combination with the very high regularity of fullerene molecules, raises the hope that this procedure of a rational synthesis can be generalised and automated to any fullerene isomer. A selective rational synthesis for any fullerene would offer researchers access to an entirely new class of molecules in large quantities. The electronic and chemical properties of the few fullerenes, which can be synthesised today provide a glimpse into the possible applications and further investigation of this type of molecules is of major interest. A fundamental understanding of the auto-assembly process would make it possible to suggest precursor molecules directly from the bond information of any fullerene. If the systematics of the reaction were to be understood, the precursor molecules with the highest probability to auto-assemble could be suggested by an algorithm. The problem is, that for any fullerene there exist astronomically many possible precursor molecules, only growing with the size of the fullerene. How could one know which precursor to synthesise?

The number of possible unfoldings is too large to be processable even with largest computers and fastest algorithms. Therefore, rules have to be derived that define a good precursor molecule. This subset could then be analysed by fast automated algorithms to select the best precursor candidates. The number of those is still going to be in the hundreds. As quantum chemical calculations are not feasible even for hundreds of possible precursor molecules a faster method like a Molecular Dynamic Simulation (MD) is required. These methods treat the quantum mechanical system by the reduction to classical Force Fields (FF) to simplify and speed up calculations. To determine how the respective FF should be parameterised to best describe the auto-assembly of a precursor, the geometries and energies of a precursor molecule during the auto-assembly reaction have to be studied in detail. There are no methods to experimentally study these energies and geometries. Thus, these quantities should be studied theoretically with the most accurate theoretical descriptions available.

The most accurate calculations of molecular geometries and energies are achieved by Coupled Cluster (CC) methods, but due to the extremely high computational cost they are only feasible for very small system sizes way below fullerene atom numbers. Density Functional Theory (DFT) offers a method which can be performed on bigger systems and still yield an acceptable result. However, DFT is a very tricky business where the choice of functional can heavily influence the result for reasons which are almost unpredictable a priori and any results obtained should be treated with caution and the choice of functional justified.

This thesis investigated the auto-assembly process of the well established icosahedral C_{60} precursor molecule for two different halogens along a suspected reaction path. The geometries, as well as the energies of the two precursor molecules, are analysed by a DFT method. An additional study is performed for an alternative reaction path for the precursor with fluorine at key positions. Special interest is paid to the geometric properties along all reaction paths to determine the best parameterisation for developing a FF method.

To identify the functional which best describes such an auto-assembly reaction, three different functionals have been compared to CC calculations on a small system with and without the addition of empirical dispersion. This system was chosen to resemble the halogen assisted dehydrogenation of sp^2-hybridised carbon centres, which is believed to be a key mechanism in the auto-assembly of fullerene precursor molecules.

This thesis aims to identify which reaction paths of the precursor will most likely lead to an auto-assembly, how the molecule is deformed along this path and if it is possible to identify rules of thumb for the structure of possible precursor molecules and the choice and position of halogen atoms.

The remainder of this thesis is structured as follows: Chapter 2 will explain the theoretical background which is required to understand how the calculations were performed. It starts with the general properties of fullerenes followed by the methods by which they can be synthesised. The method how the intermediate geometries of the precursor molecule are calculated is explained, along with the graph-theoretical background that is needed to understand the procedure. The actual calculation of the three dimensional coordinates is described in section 2.5. After that, the quantum mechanical background for the understanding of the optimisation methods is given. Starting from the general solution of the N-body Schrödinger equation, then explaining the Hartree-Fock method, Coupled-Cluster and Density Functional Theory. This includes a short introduction to basis sets and their properties and how dispersion can be added to DFT. The general concept of Force Field (FF) methods and their limits of application is explained in section 2.7. This focuses on FF in the context of fullerene precursor molecules. Chapter 3 explains the methodology that was used to perform the calculations. It starts by describing how to obtain the required graph information from a fullerene unfolding, and continues by describing how the input geometries for the functional comparison and the two different precursor closing paths were obtained. Furthermore, it contains a description of the calculation of the analysed quantities. The results are presented and discussed in chapter 4. First the geometric and energetic description of the different functionals is assessed, and after that the results for the different closing paths are analysed and the possible implications presented. A summary of the results can be found in section 4.3.

1.1 Project Context

This thesis is part of a larger project exploring the mathematics, physics, and chemistry of polyhedral molecules, and fullerenes in particular, which is lead by James Avery at the University of Copenhagen (UofC) (web-page), and in collaboration with Peter Schwerdtfeger at Massey University, New Zealand (web-page). It is currently funded through Avery's VILLUM Experiment project 00023321 "Folding Carbon: A Calculus for Molecular Origami".

An overview of the "Folding Carbon"-project can be found here: . This thesis work investigates "Task U: Understanding and Modeling Fullerene Autoassembly", and provides preliminary work for "Task A: Simulating Autoassembly of Polyhedral Molecules".

For any number of carbon atoms all possible isomers consisting of only hexagonal and pentagonal faces can be generated by extremely fast and efficient algorithms within microseconds. One part of the big project is to develop fast parallel algorithms, which can optimise the obtained fullerene molecules via force field (FF) methods to find the best fitting three dimensional geometry. Another part of the project

is trying to build a prototype for a finite element solver on the non-Euclidean manifold, that is fullerene surfaces. Such a method could make it possible to solve wave equations directly on the manifold surface, without need for the three-dimensional geometry. Another possibility, which is currently explored is whether it is possible to describe the three-dimensional electron densities perpendicular to the fullerenes surface by a reduction to the two-dimensional electron density and an exponentially decreasing function. This would allow the density representation via quadratures and immensely simplify three-dimensional integration.

All these parts of the bigger project constitute a symphony of systematic assessment of all existing fullerene molecules. There are infinitely many of those and if it was possible to systematically predict their geometric shape from the bond graph, and molecular properties could be calculated efficiently from the surface manifold, it would be possible to scan the infinite space of fullerene molecules for a candidate with the required properties for a given application. All possible fullerene molecules could be asses and catalogued to construct a database with fullerenes of interest.

After the fullerenes of interest have been identified and their properties predicted, it would be interesting to determine, whether they could be synthesised. This is where this work fits in. The precursor molecules for the interesting fullerenes could be generated by making use of the general understanding of auto-assembling precursors. If it was possible to derive rules from the study of existing precursors, this understanding could be used to generate the best fitting precursor candidates, whose auto-assembly could then be simulated with an MD method.

1.2 Software

The software developed in this thesis can be found at https://github.com/HeuserB/Thesis_Folding_Carbon. The project contains:

1. Methods to generate fullerene unfoldings from the graph information
2. Methods to obtain the geometries of unfoldings along the closing path
3. Methods to produce GAUSSIAN [12] input files from the unfoldings
4. Methods to automatically extract information from ".log" files
5. Scripts to analyse the results
6. Scripts to visualise the results

Wherever appropriate, the methods will be incorporated into a central code base for the "Folding Carbon" project (web-page). As this project is a work in progress, the content is most likely to be changed in the future.

Theory

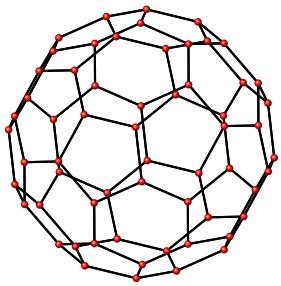
2

2.1 Fullerenes

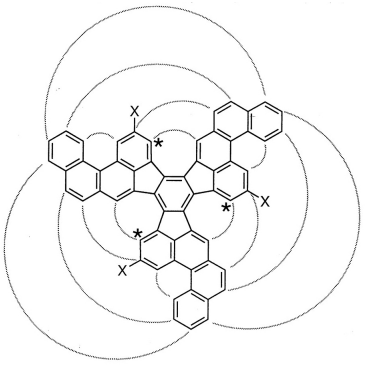
Carbon is one of the most interesting atoms, with several naturally occurring allotropes like graphene, graphite and diamond, which is the hardest natural material known to mankind. Another allotrope of carbon, the fullerenes were first discovered in nature as recent as 1985 [13] and have intrigued the scientific community ever since. Fullerenes are a class of molecules consisting only of carbon atoms, similar to graphene, but with a close surface structure. This way, they form convex polyhedral carbon cages with sp^2-hybridised carbon atoms. The most prominent fullerene is the C_{60} Buckyball, which has first been synthesised in 1984. It consists of 60 carbon atoms and has the shape of a truncated icosahedron (see figure 2.1a) and was synthesised via graphite vaporisation [14]. The number of possible fullerenes is infinite and the number of isomers grows as a function of the carbon atoms N in the order $\mathcal{O}(N^9)$ [15]. Fullerenes have very interesting chemical and electronic properties and the proposed applications are vast, ranging from already existing electron acceptors in solar cells to proposed efficient hydrogen storage and highly efficient drug delivery. However, the synthesis of fullerenes has been very difficult, with only a handful fullerenes available through a rational synthesis path, despite the infinite number of theoretically possible fullerenes. The following paragraphs give a quick introduction to the main characteristics of fullerenes and their geometrical properties.

2.1.1 Geometric Properties

Fullerenes are similar to graphene sheets, which have been wrapped around a three-dimensional object. Who, ever tried to wrap a spherical Christmas gift will have encountered the problem, that two dimensional undistorted flat surfaces are impossible to wrap around a three-dimensional object without, either leaving something visible or folding the paper. The mathematical term for this kind of flat wrapping paper is a developable surface, and it is defined as a smooth surface with zero Gaussian curvature. Gaussian curvature is an intrinsic property of all surfaces at all points. It describes, how "closed" a surface is and is formally defined as the product of the two principal curvatures. A surface, which is closing, has positive curvature (like the sphere) a surface which is opening has negative curvature (like a hyperboloid) and a surface, which does neither, has zero Gaussian curvature (like a flat surface or a cylinder). Any polyhedral surface, which is closed in three-dimensional space and homeomorphic to the sphere, has a total Gaussian curvature of 4π. The hexagonal structure of graphene does not introduce any Gaussian curvature, and therefore only allows shapes which have a curvature of zero at any point (like sheets or open nanotubes). In contrast, fullerenes do not only contain hexagonal faces but also pentagons and can form closed surfaces. As was shown in [6], the Gaussian curvature introduced by a single pentagon to a hexagonal plane, is $\frac{2\pi}{6}$, meaning that to close a fullerene exactly 12 pentagonal faces are required. The position of these pentagons uniquely determines the shape of a fullerene. The shape of the fullerene can be approximated by a twelve cornered polyhedron, although the faces are not perfectly planar.

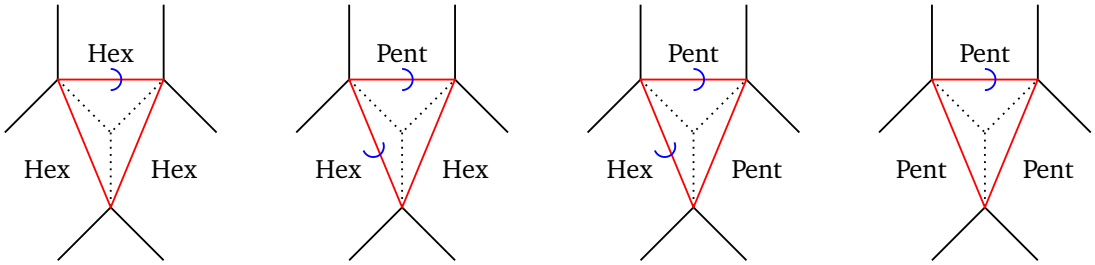


(a) C60 buckyball.



(b) Auto-assembling C60 precursor molecule, Reprinted from L. T. Scott *et al.*, „A rational chemical synthesis of c60“, *Science*, vol. 295, no. 5559, pp. 1500–1503, 2002, with permission.

Figure 2.1.: C60 Buckyball and precursor molecule.



(a) No pentagon. (b) One pentagon. (c) Two pentagons. (d) Three pentagons.

Figure 2.2.: Unique out of plane angles for the four different corner cases.

2.1.2 Chemical Bonds within Fullerenes

The faces of a fullerene, are either hexagon or pentagons, thus, the only bonds which occur are carbon-carbon bonds between a hexagonal and a hexagonal face, a hexagonal and a pentagonal face and two pentagonal faces. The isolated pentagon rule (IPR) [16] states, that the most stable fullerenes will be the ones, where no two pentagonal faces share an edge. The chemical bonds which exist in all fullerenes can be put in three categories: HH, HP, PP, with PP being by far the least prominent one. The angles which occur within any fullerene can be put into two categories, those which lie within a pentagon and those which lie in a hexagon. The out of plane angles (see section 2.7.1) can be put into four cases defined by the corner of three neighbouring faces: only hexagons, one pentagon, two pentagons, three pentagons. For the case of only hexagons and pentagons, there is only one unique angle while for the other two cases, there are two. The out of plane angles are defined as the angle between the faces of the red pyramid to the base of the pyramid in figure 2.2. The dotted lines form the tip of the pyramid and the red triangle forms the base of the pyramid. The angles are shown in blue and are defined as the angle between the base of the pyramid and the three top sited of the pyramid. In the first case, for example, all angles are the same, because this is a fully symmetric pyramid. In the second case, there is one angle for the pentagon to the bottom of the pyramid and one for the two hexagons. By knowing the optimal values for the bond lengths, the in-face angles and the out of plane angles, the geometry of any fullerene can be obtained for example by a force field method (see section 2.7). The out-of-plane angles which are described in section 2.7.1 are defined slightly different from those here, but the number of unique angles in a closed fullerene remain the same for all cases.

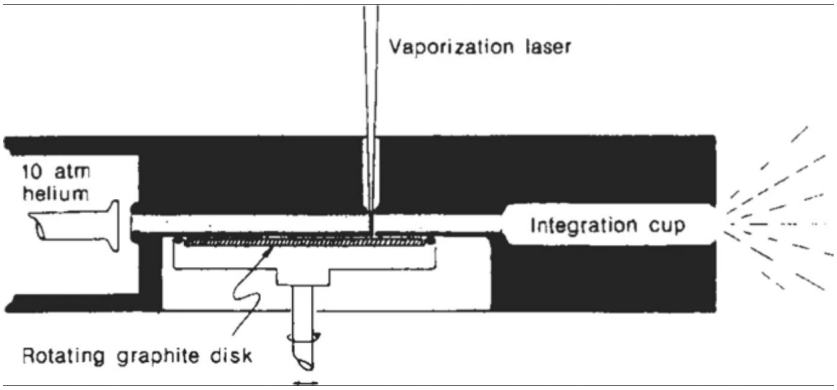


Figure 2.3.: Graphite Vaporisation, Reprinted from H. W. Kroto *et al.*, „C60: Buckminsterfullerene“, *Nature*, vol. 318, no. 6042, pp. 162–163, 1985, with permission.

2.2 Synthesis of Fullerenes

The most common methods for the synthesis of fullerenes are laser vaporisation or resistive carbon heating, but they do also occur naturally in outer space [17, 18]. Laser vaporisation of graphite as performed by Kroto et al. [1] uses an experimental setup, which is similar to the one depicted in fig. 2.3. In this method, a spinning disk of graphite is turned into a carbon plasma by a laser, and the plasma is then shot with a supersonic helium beam, which provides cooling for the plasma, allowing molecules to be formed. The same beam also "blows" the formed carbon molecules in the integration cup. This process is completely non-selective and therefore the integration cup will contain a multitude of different carbon molecules. Hence, mass spectroscopy is needed to sort the resulting carbon molecule by their mass. Kroto et al. found that many molecules, which were formed during this process had a very high symmetry, remarkable stability and consisted of 60 carbon atoms. They had discovered the previously predicted C_{60}-Ih Buckyball. The reason why this particular fullerene was formed, is that it is small and extremely stable, and the process is very non-selective. The fullerene formation only happens randomly when the carbon plasma is cooled with the helium beam. This way, the most stable molecules, are the most likely to occur.

The laser evaporation does not provide a method, to synthesise a specific fullerene, the results are simply a mixture of all fullerenes which can form spontaneously at high temperatures, with a strongly skewed distribution towards those with high stability. A selective method for the C_{60} buckyball synthesis has been developed by Krätschmer et al. in 1989 [19] and it made the production of this particular fullerene very easy. However, a rational synthesis for fullerene molecules was still far out of reach.

Resistive heating of carbon is another method to produce fullerenes from carbon materials with yields of up to 10 per cent for C_{60} and C70 molecules [20]. In this method, a graphite sheet is used as a conductor and exposed to an electric current. The graphite heats up and evaporates, and is placed in a helium atmosphere, in which the carbon condenses to fullerene molecules. This method has the same problem as the laser evaporation, of being non-selective and preferring the formation of small-sized fullerenes.

2.2.1 Rational Synthesis from Flat Precursor Molecules

The previous methods for fullerene synthesis all share some common problems. Neither are they selective to a single fullerene nor do they offer the opportunity to synthesise large, less stable fullerene

molecules. Of the millions of possible isomers of fullerenes, only a handful are formed by these methods, with C_{60}-Ih being the most likely. Producing a specific fullerene isomer which is suspected to have interesting properties is not an option. Therefore, a whole different approach is required when trying to synthesise the less stable elusive fullerenes. In 2002, Scott et al. developed a method for a rational synthesis of the icosahedral C_{60} fullerene[7]. The rational synthesis was achieved, by synthesising a polycyclic aromatic precursor molecule with flat geometry and chlorine atoms at key positions. This flat precursor molecule contained all 60 carbon atoms as well as hydrogen and chlorine atoms on the outline/periphery. The precursor molecule used by Scott et al. is shown in figure 2.1b, where the X indicates the position of the chlorine atom. By providing the activation energy with a laser pulse, it is possible to start a chain reaction which leads to the complete dehydrogenation of the molecule. In this chain reaction, the atom pairs on both sides of a cavity of the molecule (the fjords or gaps at the asterisks for example) form a bond with each other and leaving behind two aryl radicals on both sides of the cavity. Those radicals quickly form a bond together and close another carbon ring, which forces the molecule to introduce curvature. This curvature leads to a "zipping up" of the whole precursor molecule where the periphery atoms on two sides of cavities are brought close to each other from the previous cyclo-dehydrogenation and their removal from the precursor initiates the next cyclo-dehydrogenation.

Prior work of Scott et al. has proven that the introduction of chlorine atoms greatly improved the auto-assembly potential of small geodesic polyarenes. This understanding was used to add the chlorine atoms to the C_{60} precursor. This precursor molecule proved to auto-assemble selectively when activated with a laser pulse to the C_{60} fullerene. Thus, providing researchers with the first-ever rational synthesis of a fullerene molecule. But the yields of the reaction were very low with under 1% . In 2008, Otero et al. used a C_{60} precursor without halogen atoms ($C_{60}H_{30}$) on a platinum surface and were able to increase the yields of the auto-assembly up to nearly 100%autociteOtero2008. A yet unanswered question is, which mechanism makes the precursor molecule so selective, and how the choice an position of halogen affects the activation barrier of the precursor.

Making use of this concept of auto-assembling precursor molecules, it is conceivable to generalise the auto-assembly process to any kind of fullerene, especially the larger fullerenes, which are currently un-synthesisable. Those flat precursor molecules could then be produced by common rational synthesis paths and, exposed to laser pulses, would auto-assemble. The big problem is the extremely large number of possible precursor molecules. For the icosahedral C_{60} molecule, there are already 3.13×10^{18} non-identical unfoldings [21], and that is with restrictions of only having complete carbon rings. Finding the best precursor in this space is a high dimensional problem and finding rules for how a precursor molecule should look like is essential.

2.2.2 Flash Vacuum Pyrolysis

The auto-assembly of the precursor molecule is initiated by a laser pulse under vacuum. This method is called Flash Vacuum Pyrolysis (FVP) and is a common technique in the synthesis of organic molecules [22]. In this case, it provokes a chemical reaction with one educt (here the precursor molecule) and multiple products (the fullerene and periphery dimers). The precursor molecule is initially flat, meaning that the optimal geometry of the molecule in its current bond situation locates all the nuclei in a common plane. To dehydrogenate the carbon atoms, the activation barrier has to be overcome. Loosely speaking, one could say, the hydrogens and chlorines, which are repelling each other, have to be brought close

together to make a bond between those atoms energetically favourable over the bond to their respective carbon atom. This has to be done without severing the carbon-carbon bonds which are already present in the precursor. FVP activates the internal vibrational modes of the molecule and forces deformations on the geometry. Those internal modes will have strong deforming effects on the periphery atoms and are capable to deform the carbon skeleton to a certain degree by bending the angles. But in general, it has been shown, that the effect son the carbon rings itself are much smaller than for the attached molecules [23]. The main tunable parameters in a FVP induced reaction are the temperature of the laser and the exposure time, which restrict the amount of energy, which is transferred to the system.

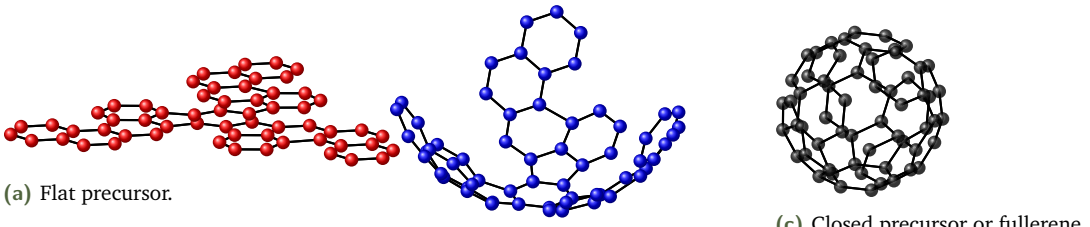
The perfect amount of energy would be just enough to introduce enough geometric deformation of the molecule to start a cyclo-dehydrogenation. This will introduce the curvature and start the chain reaction ("zipping up"). If the energy is too high, the carbon-carbon bonds will be severed and the precursor molecule would be broken or the deformations will be so large that the "wrong" carbon atoms form a bond, leaving the precursor in a very stable state. If, instead, the energy was too low the precursor molecule would just restore its initial geometry.

As any working precursor molecule is required to keep the original carbon-carbon bonds intact during the whole auto-assembly process, it is a good idea to assess a precursors potential by looking at the path, the carbon atoms have to take to bring the periphery atoms (H, Cl, etc.) close to each other. The further restriction of leaving the interatomic carbon bonds unaffected is imposed. A good way achieve both, is to treat the carbon skeleton as a rigid body (leaving the carbon-carbon bonds at their equilibrium lengths), which can only withstand a certain amount of deformation before braking. If the initial (flat) and final (closed fullerene) geometries of the precursor are known, such paths can be generated by interpolating between those in a fashion, which minimises the carbon-carbon bond deformation. The energies along this path are an important quantity for a precursor: The lower the relative energy barrier/s along this path is/are, the more likely a precursor molecule will auto-assemble to the closed fullerene. However, a too flat energy path might allow the precursor to reverse some of the steps along the way which would result in recrossing.

2.3 Describing Precursors as a Rigid Bodies with Hinges

To assess the quality of a precursor and its potential to auto-assemble to the closed fullerene molecule, it is important to look at the reaction path that a planar precursor has to take to form the required bonds. That is, the intermediate geometries which the planar precursor follows during the auto-assembly process.

There is no experimental data available that defines a clear reaction path for any auto-assembly as these chemical processes are neither spatially nor temporally resolvable with current methods. To obtain a possible path, other methods have to be employed. It is possible to perform quantum chemical calculations at distinct points along this path to understand how the energy develops, and how the precursor would react to the deformations which are introduced along the way. To perform these calculations an initial geometry for the precursor molecule is required at each point along the closing path, which can be used as an input. To determine a reasonable initial guess which minimises the energy while still being able to close the precursor the individual bond properties have to be considered.



(a) Flat precursor.

(b) Half closed precursor.

(c) Closed precursor or fullerene.

Figure 2.4.: Carbon cage of the C_{60}-Ih precursor molecule at three steps along the closing path.

The ring stabilities of hexagonal and pentagonal carbon rings are extremely high, forcing all carbon atoms, which are part of them to lie (roughly) within a plane. Therefore, the bonds between all carbon atoms are likely to close to their optimal values and so are the angles within the respective face. When trying to identify the closing path, which a fullerene precursor follows to initiate the cyclo-dehydrogenation and carbon-carbon bond formation, the existing faces are not to be destroyed. Therefore, a good way to enforce the precursor stability is to choose a path, which is close to the optimum for these faces, leaving them planar and the bond lengths and angles at the correct values. Hence, starting from the planar molecule, with the faces at the optimal geometry and face-face angles of $180°$, the face angles can subsequently be closed, until the closed fullerene shape is reached. This procedure will lead to all intermediate geometries of the unfolding, which leave the faces at their optimal geometry (compare figures 2.4a, 2.4b and 2.4c). These geometries can then be used as an input for the quantum chemical calculations.

It can be thought of the precursor molecule as carton model, where each face is represented as either a pentagonal or hexagonal face. This model is then glued together on the edges with duct tape so the faces can move. The precursor is now closed from a flat geometry to the final fullerene shape and this is closing is sampled. Each shot along the way can then be fed to a QC software. By describing the precursor as such rigid faces connected by hinges it is possible to enforce the restraints on the carbon cage. Section 2.5 will discuss in detail how the initial and intermediate geometries along the way can be obtained systematically.

The carbon atoms in fullerene and fullerene precursor molecules are all connected to exactly three other atoms. This regular behaviour allows the application of combinatorial methods to the geometry of fullerenes. Specifically, as was shown in [15], does this bond property allow the atoms and faces within fullerenes to be described as mathematical graphs. This concept has been generalised to all possible unfoldings, which are the cut-out polyhedrons put on the plane [6]. The following section 2.4 is going to present the main concepts and ideas from graph theory and introduce how they are related to rigid body model of fullerene precursors by prior work.

2.4 Graph-theoretical Background

Carbon atoms have four valence electrons and therefore always need to be part of four covalent chemical bonds to fulfil the octet rule. In fullerene molecules, this results in each carbon atom being bound to three other carbon atoms by two single and one double bond. This holds for closed surface fullerenes. Mathematically this can be represented as a graph. A graph $\mathcal{G}$ is a structure that consists of a set of nodes or vertices $\mathcal{V}$ which are pairwise connected to each other via some relation

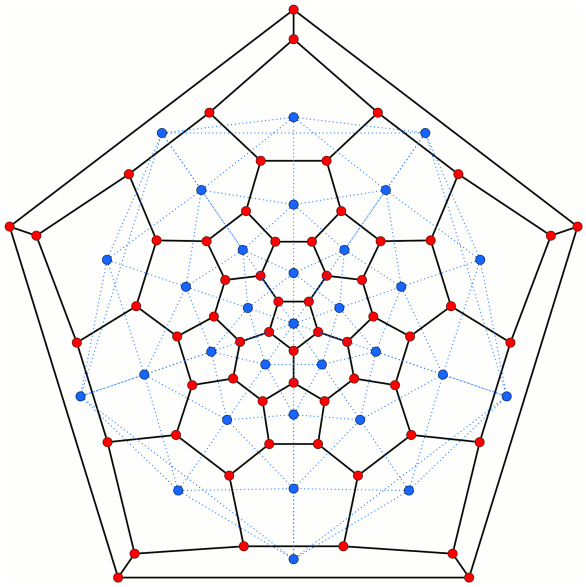


Figure 2.5.: Planar embedding of the C_{60}-Ih, atoms are shown as red dots and the faces as blue dots. Reprinted from P. Schwerdtfeger *et al.*, „The topology of fullerenes“, *WIREs Computational Molecular Science*, vol. 5, no. 1, pp. 96–145, 2015, with permission.

$\mathcal{E} \subseteq \{(u,v) \,|\, (u,v) \in \mathcal{V}^2 \wedge x \neq y\}$ called edge. If these pairs are unordered, meaning there is no difference between the edge $\mathcal{E} = (u,v)$ and $\mathcal{E}' = (u,v)$, $\mathcal{G}$ is called an undirected graph; if the edges are oriented $\mathcal{G}$ is called a directed graph or digraph. In the case of fullerenes the vertices represent the atoms while the edges of the graph substitute the chemical bonds. Representing the molecule in this fashion will lead to major simplifications when trying to obtain the geometries (see. 2.5).

A n-regular graph is a graph in which each vertex has the same number of neighbours n i.e. that each vertex is part of the same number of edges. The special case of a 3-regular graph is referred to as a cubic graph. Graphs can very intuitively be visualised by dots connected with lines, in the case of directed graphs the lines are replaced by arrows. A graph is n-vertex-connected if at least n vertices need to be removed to disconnect the graph and form two isolated subgraphs. If the edges of a graph can be drawn on the two-dimensional plane without intersections the graph is planar. Fullerene molecules can be represented as a cubic planar graph because each atom will be bound to exactly three other atoms.

The drawing of a planar graph on the plane without intersecting edges is called a plane graph or planar embedding of the graph. Every simple planar graph can also be drawn with straight lines according to Fáry's theorem [24]. The planar embedding of the C_{60}-Ih fullerene is displayed in figure 2.5, where it is represented as the black lines (bonds) connecting the red dots (carbon atoms).
The implications of the fact that fullerene bond graphs are three connected planar cubic graphs are profound because this means that, according to Alexandrov's uniqueness theorem the carbon atoms form a unique set of faces and there is a unique polyhedron associated with the graph [25, Chapter 3]. This means, that the faces of the three-dimensional polyhedron can be determined directly from the bond graph of a fullerene. Extensive research has been done on the generation of all possible fullerene bond graphs, as well as on the determination of the polyhedral structure from the graphs [3–5]. And with the software developed Schwerdtfeger et al. it is possible to generate fullerene graphs as fast as 10^5 isomers per second.

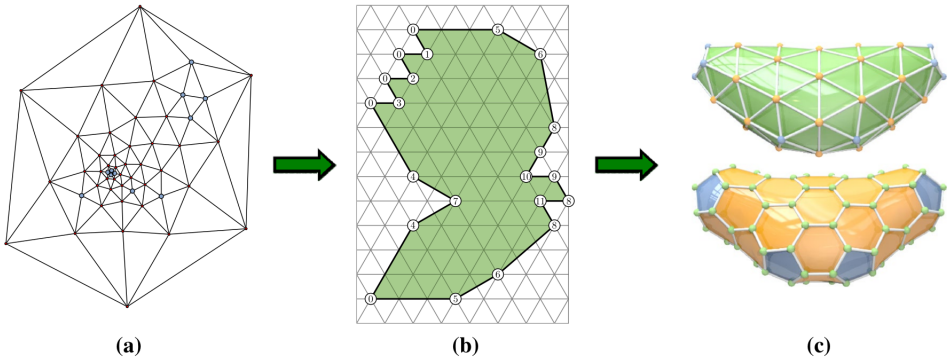


Figure 2.6.: C_{2v} C_{102} fullerene. Planar embedding of the dual graph (a) Unfolding (b) Dual and original Polyhedron (c). Reprinted from P. Schwerdtfeger *et al.*, „The topology of fullerenes“, *WIREs Computational Molecular Science*, vol. 5, no. 1, pp. 96–145, 2015, with permission.

2.4.1 Dual Graphs

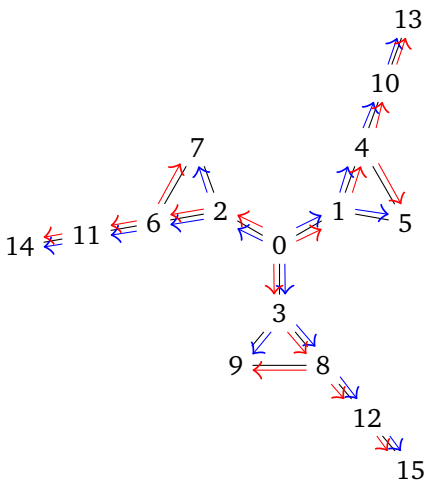
The dual of a graph is the graph, which for each face of the original graph has a vertex and an edge for every two neighbouring faces. In the case of a cubic graph, as a fullerene graph, the dual has a vertex for each face, and each of these vertexes has either five or six edges connected to it, depending on whether it is a pentagonal or hexagonal face. From a planar embedding, the dual of a graph can be easily obtained by placing a vertex in each face and connecting it to all neighbouring vertices, with which the face has at least two vertices in common. The dual graph of the C_{60}-Ih isomer is drawn in figure 2.5, where the blue dots represent faces and the blue lines the connecting edges.

2.4.2 Fullerene Unfolding and their Graphs

In [6] and [15], Avery discusses methods for the automated generation of fullerene unfoldings from their dual graph onto the plane of equilateral triangles (the "Eisenstein" plane). The distances on the fullerene surface are preserved on the inside of the fullerene unfolding. The procedure can be illustrating as taking the dual polyhedron and cutting it along the edges between the pentagon-nodes, which are the sites that induce curvature. This way there will be no curvature left and it can be unfolded onto a plane. An illustration of this unfolding for an isomer of the C_{102} fullerene is displayed in figure 2.6. The molecule can be seen in (c) bottom, with the dual polyhedron above. In (b) the unfolding of the dual polyhedron to the planes of equilateral triangles (Eisenstein plane) is shown.

In this dual graph, each vertex represents a face, while each triangle represents an atom. This way an unfolding of the dual fullerene graph has all the atoms represented on the plane, with the relative positions of the atoms along the surface preserved. But some bonds are cut, and hence some of the faces missing. The faces that are to be formed lie exclusively on the periphery and are the ones that will be formed by the cyclo-dehydrogenation.

It has to be the dual polyhedron that is unfolded by cutting, because if the polyhedron were to be cut out, this would duplicate the atoms at the periphery and place them in multiple places at once, which is not physically plausible. If the dual is unfolded, the faces will be duplicated instead of the atoms.



(a) Graph of the ceC60 unfolding and the different spanning trees.

Dual Graph	MST (BFS)	MST (DFS)
[1, 3, 2],	[1, 3, 2],	[1, 3, 2],
[4, 5, 0],	[4, 5],	[4],
[6, 7, 0],	[6, 7],	[6],
[8, 9, 0],	[8, 9],	[8],
[5, 1, 10],	[10],	[10, 5],
[4, 1],	[],	[],
[7, 2, 11],	[11],	[7, 11],
[2, 6],	[],	[],
[12, 9, 3],	[12],	[12, 9],
[3, 8],	[],	[],
[13, 4],	[13],	[13],
[6, 14],	[14],	[14],
[15, 8],	[15],	[15],
[10],	[],	[],
[11],	[],	[],
[12]	[]	[]

(b) Graph data as list of lists.

Figure 2.7.: Difference between Breadth- and Depth-First-Search at the example of a C_{60} unfolding.

By using this procedure the graphs of the dual unfolding can be obtained by cutting out the dual polyhedrons. This way, the graph and the dual graph of all possible unfoldings can be obtained from the fullerene graph directly with little computational effort. The space of possible unfoldings is extremely large, given that for every single fullerene, there are hundreds to thousands of different ways to cut bonds until the remaining faces are developable.

When the graph and the dual graph of the unfoldings are available, the planar, two-dimensional coordinates (or rather an initial guess for them) can be obtained by simple geometrical means and a few assumptions about bond lengths and angles (compare section 2.5). However, to close the face-face angles step by step, the dual graph has to be trimmed even further, to obtain a set of "hinges" (axes connecting two faces) which uniquely define the atoms three-dimensional positions. These are called hinges because when treating each face like a stiff object, and considering the face-face angles as free rotations, the mechanical analogon of a hinge comes to mind. Essentially, a set of hinges has to be found, which does not include any cycle, because a cycle (see figure 2.8 faces: $1 - 4 - 5$) will mean having more hinges then required to arrive at the closed fullerene and thereby lead to double rotations of certain vertices (for a detailed explanation see section 2.5.3).

2.4.3 Minimum Spanning Tree

To get this unique set of hinges, which provides the set of minimal rotations which have to be applied to each face. To arrive at the final fullerene geometry, the cycles have to be removed. A directed graph without cycles is required. Given a root face, each other face of the unfolding (or vertex of the dual) has to be reached by exactly one unique path along the edges. In graph theory, such a sub-graph is called a tree, because it branches out like a tree. The tree has to be directed outwards from the root face so that the last face at the end of an "arm" has no neighbours. Such an outwards directed tree can be used to define each faces position relatively to the parent face, therefore, any rotation applied to a face can be easily passed on to all its children and children's children, which results in a rigid body-like rotation regarding that face-face angle.

The use of an MST is very useful in this context, because it offers a way, to connect all faces of the unfolding by exactly one hinge to a parental face. Given the structure of each face (pentagon or hexagon) and the face-face angles, the coordinates of all vertices of the graph (that is the carbon atoms) are uniquely defined. Furthermore, should general rules for good precursors, like certain symmetries or arm lengths be found, those requirements could be encoded extremely fast into the generation of unfoldings, because the MST of the dual graph will contain all this information and a lot of algorithms have been derived for these.

The root face of an MST takes a special position in the tree, because it is the point of highest symmetry, and contains the only atoms, which will not be changed when applying the hinge rotations. All other faces and atoms' positions will be defined relative to the vertices of the root face. It is favourable, to pick as a root face the face in the unfolding with the highest symmetry. This can be implemented easily into the generation of fullerene unfoldings by picking a root face and for each removed edge, remove the edges which correspond to the wanted symmetry regarding the root face.

There are two different approaches to get a spanning tree from a graph, the first is via a Breadth-First Search (BFS) and the second via a Depth First Search (DFS). BFS starts at the root node and adds the directly connected nodes to the tree first, and then moves on to the nodes which are further away in the graph. The DFS starts at the root node, picks the first neighbour and explores all children of the children first, before moving to the next neighbour of the root node.

Figure 2.7a shows the dual of an icosahedral C_{60} precursor, with the black lines as the original graph. The three cycles at the faces $1-4-5$, $3-8-9$ and $2-6-7$ are removed by creating an MST. Figure 2.7b illustrates how the data is represented, as a list of lists. Each list element represents a face and the lists contain the information, which face is a neighbour of this element. The two different spanning trees created in a breadth-first and depth-first approach are also shown in those figures. In essence, the breadth-first approach explores nodes in order of their distance from the root node, while the depth-first approach explores neighbours at the bottom level of each path first.

With the graph of the fullerene, its dual and the spanning tree, mathematical representations are now available which enable quick and easy methods to obtain the planar geometries for any possible unfolding, as well as any intermediate geometry along the trivial closing path. How these geometries are obtained from the graph is described in detail in the following section.

2.5 Rigid Body Vertex Construction

After all graph-theoretical information was obtained from the fullerene bond graph, and an unfolding is created, the vertices' coordinates of the unfolding (not the dual) have to be calculated. This can be done first for the planar precursor. After the planar vertices are obtained, the hinge angles can be closed until the final positions of the vertices in the fullerene is reached. The intermediate geometries can be sampled and used as input for calculations. The implementation of this can be found here.

2.5.1 Planar Vertex Coordinates

As described in section 2.4, the unfolding of a fullerene can be represented as a graph which connects each face to all its neighbours. This representation of a fullerene unfolding as a graph, and especially as the MST of a graph, can be used to efficiently obtain the three-dimensional vertex coordinates. By using

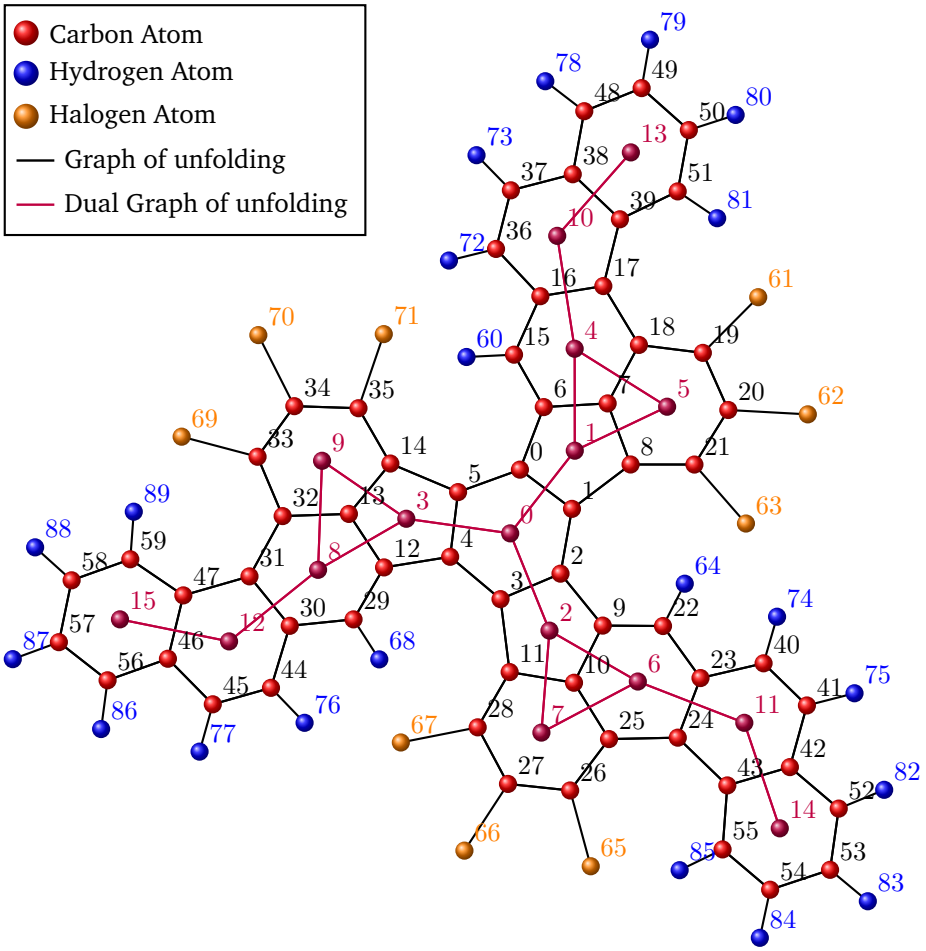


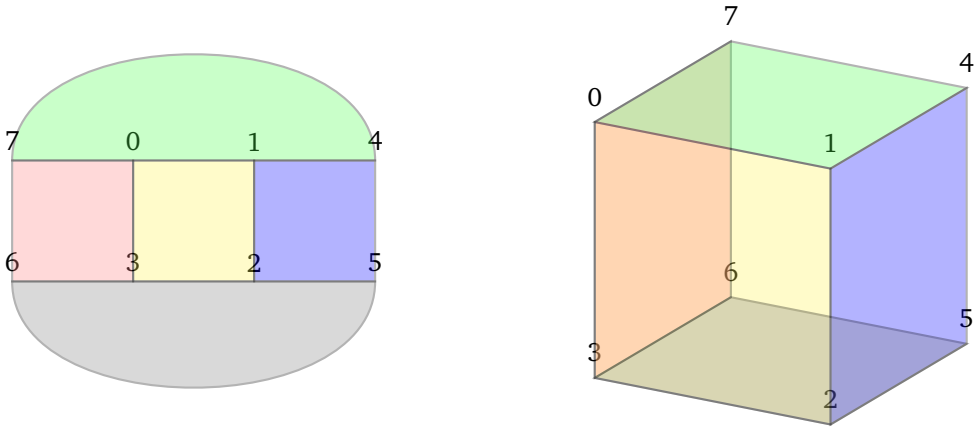
Figure 2.8.: C60 unfolding with Graph and Dual graph.

the dual of the unfolding, which connects the faces, it is possible to define internal hinges for a fullerene unfolding. Those hinges correspond to the edges of the dual graph. Each of these edges connects two different faces, and each of these faces has exactly two common carbon atoms. Hence, each hinge can be represented by a pair of two carbon atoms.

Figure 2.8 shows a C_{60} unfolding with the atom graph in black and the dual graph in purple, with the purple numbers representing the face ID. The dual graph here does not represent a MST because it contains three loops. A hinge can be found at each point where the dual graph intersects with the graph, so at each point where the blue and the purple lines cross. Not all of those hinges are present when the MST is used instead of the dual graph.

If all faces themselves are perfectly planar, that means all five points of a pentagonal face or all six for a hexagonal face lie within a plane defined by exactly one normal vector $\hat{\mathbf{n}}$. If this condition is strictly fulfilled for each face of the unfolding, the carbon coordinates of the unfolding are uniquely defined (in their local reference frame[1]) by the angles between the faces, the distances and the angles between the atoms. This becomes evident when a root face is picked, a MST created from that face, and the carbon coordinates constructed via the MST, a process which is described in the following paragraphs. By using the MST, each face of the unfolding can be reached by following the edges starting at the root face.

[1]Except for translation and rotations.



(a) Planar embedding of a cube.

(b) Three dimensional embedding of the graph.

Figure 2.9.: A cube.

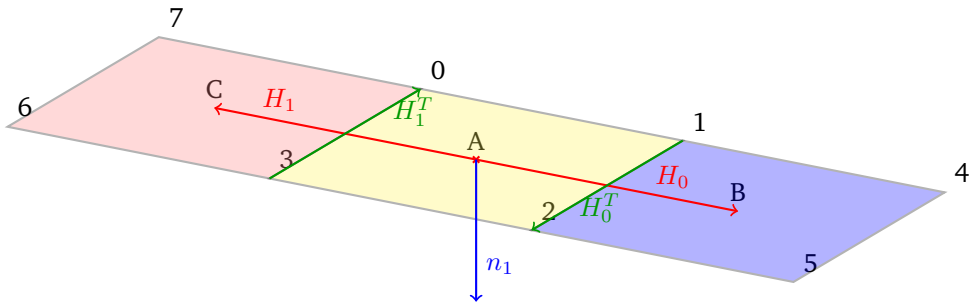


Figure 2.10.: Planar embedding of the cube unfolding. Note how open faces disappear in the unfolding.

To illustrate how this is done, the unfolding of a carbon cube is constructed. Figure 2.9 shows the planar and the 3D embedding of a simple cube, consisting of eight carbon atoms. The planar embedding shows the six faces in different colours, where the outside of the graph is an additional face shown in white colour. This face appears in the back of the cube in the 3D embedding.

To generate an unfolding of this carbon cube, two of the carbon bonds, which are represented as the black lines, have to be broken. By breaking the bonds between atoms 7 and 4, and 6 and 5, the cube is left with three faces (see fig 2.10). If now the MST is created, with the yellow face as a root face, the only directed edges that are in the tree are the ones that span from the root face A to faces B and C respectively (H_0 and H_1). If the bond length (cube length/width/height) is known and also the bond angles $(90°)$, the root faces coordinates can be determined by simple geometrical methods [2].

The coordinates of the root faces' vertices can be directly obtained from the face planarity, the bond lengths, and the bond angles. To construct the coordinates of all the vertices of the faces further down in the MST, it is useful to define the transversed hinges. As mentioned above, each directed edge of the MST can be interpreted as a hinge connecting two faces, and each of these two faces will always have exactly two vertices (atoms) in common. By defining the transversed hinge as the vector between those vertices in clockwise order of the root face, there is now a unique set of 3D vectors associated with the hinges (H_0^T and H_1^T). These transversed hinges are now vectors form atom to atom. For a planar

[2] In this case, it is just a regular square.

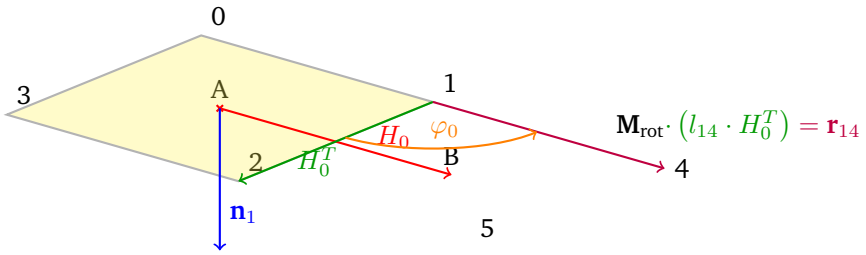


Figure 2.11.: Construction of planar carbon coordinates form the graph. Yellow face A is the root face, B the child face. H_0^T is the hinge vector, φ is the bond angle between atoms 2-1-4 and l_{14} the bond length between atoms 1-4.

face, there is a well defined normal vector, which is pointing into the paper plane for the yellow face, which can be obtained by using the cross product of two consecutive directed vertices within one plane. For example, the two vectors $\mathbf{r}_1 - \mathbf{r}_0 = \left(1,0,0\right)^{\text{T}}$ and $\mathbf{r}_2 - \mathbf{r}_1 \left(0,1,0\right)^{\text{T}}$ can be chosen, which leads to a normal vector $\hat{\mathbf{n}}_1 = \left(1,0,0\right)^{\text{T}} \times \left(0,1,0\right)^{\text{T}} = \left(0,0,-1\right)^{\text{T}}$.
Assuming that not only the root face but all other faces are also planar and lie within the same plane as the root face, the 3D coordinates of the vertices constituting faces further down in the MST can be created by the following procedure which is illustrated in figure 2.11:

1. Normalise and scale the hinge vector by the bond length $\left(l_{14} \cdot \left(H_0^{\text{T}}/|H_0^{\text{T}}|\right)\right)$
2. Create a rotation matrix that rotates around the normal vector of the parent face ($\hat{\mathbf{n}}_1$) and the bond angle φ_0, to determine the vertex of the next vertex in the child face ($\mathbf{r}_4$)

$$\mathbf{M}_{\text{rot}} = \begin{pmatrix} \cos\varphi_0 + {n_x}^2\left(1-\cos\varphi_0\right) & n_x n_y\left(1-\cos\varphi_0\right) - n_z\sin\varphi_0 & n_x n_z\left(1-\cos\varphi_0\right) + n_y\sin\varphi_0 \\ n_y n_x\left(1-\cos\varphi_0\right) + n_z\sin\varphi_0 & \cos\varphi_0 + {n_y}^2\left(1-\cos\varphi_0\right) & n_y n_z\left(1-\cos\varphi_0\right) - n_x\sin\varphi_0 \\ n_z n_x\left(1-\cos\varphi_0\right) - n_y\sin\varphi_0 & n_z n_y\left(1-\cos\varphi_0\right) + n_x\sin\varphi_0 & \cos\varphi_0 + {n_z}^2\left(1-\cos\varphi_0\right) \end{pmatrix}$$

.

3. Multiply the rotation matrix with the scaled hinge $r_{14} = \mathbf{M}_{\text{rot}} \cdot \left(l_{14} \cdot \left(H_0^{\text{T}}/|H_0^{\text{T}}|\right)\right)$.
4. Add the obtained vector to the preceding vertex $\mathbf{r}_1$ to obtain the new vertex coordinates $\mathbf{r}_4 = \mathbf{r}_1 + \mathbf{r}_{14}$.

If all faces are planar, and share the same normal vector $\hat{\mathbf{n}}$ steps 2-4 can be repeated to obtain all other vertices within a face, each time taking the resulting vector $\mathbf{r}_{14}, \mathbf{r}_{45}, \ldots$ as the vector which is rotated around $\mathbf{M}_{\text{rot}}$ and added to the results from the previous step $\mathbf{r}_4, \mathbf{r}_5, \ldots$.

2.5.2 Vertex Coordinates in Three Dimensional Space

In case each face lies within its own plane and neighbouring faces do not have identical normal vectors, more steps have to be performed after the vertices of each face have been calculated as described above. If the angle between the parent face and the child face is θ_0, another rotation matrix $\mathbf{M}_{\text{rot}}^{\text{face}}$ can be defined by the transversed hinge H_0^T and that angle in the same way as in step 2 above (see fig 2.12). Now the rotated face coordinates can be calculated via the following steps:

1. Calculate the connecting vector for the child faces nodes from one of the hinge nodes $\mathbf{r}_{14} = \mathbf{r}_4 - \mathbf{r}_1$ and $\mathbf{r}_{15} = \mathbf{r}_5 - \mathbf{r}_1$.

2. Construct the rotation matrix $\mathbf{M}_{\text{rot}}^{\text{face}}$ with the angle φ_0 and the transversed hinge vector H_0^T as axis.

3. Rotate the reference vectors by multiplying them with the rotation matrix $\mathbf{r}_{14'} = \mathbf{M}_{\text{rot}}^{\text{face}} \cdot \mathbf{r}_{14}$ and $\mathbf{r}_{15'} = \mathbf{M}_{\text{rot}}^{\text{face}} \cdot \mathbf{r}_{15}$.

4. Calculate the new vertices by adding the rotated vector to the reference vertex $\mathbf{r}_{4'} = \mathbf{r}_1 + \mathbf{r}_{14'}$ and $\mathbf{r}_{5'} = \mathbf{r}_1 + \mathbf{r}_{15'}$.

This can be applied to all vertices of the child face, including the two vertices which form the hinge (1 and 2 in the example) because the connecting vector is the null vector ($\mathbf{r}_{11} = \left(0,0,0\right)^{\text{T}}$) and the axis itself ($\mathbf{r}_{12} = H_0^T$) which will not be affected by the rotational matrix.

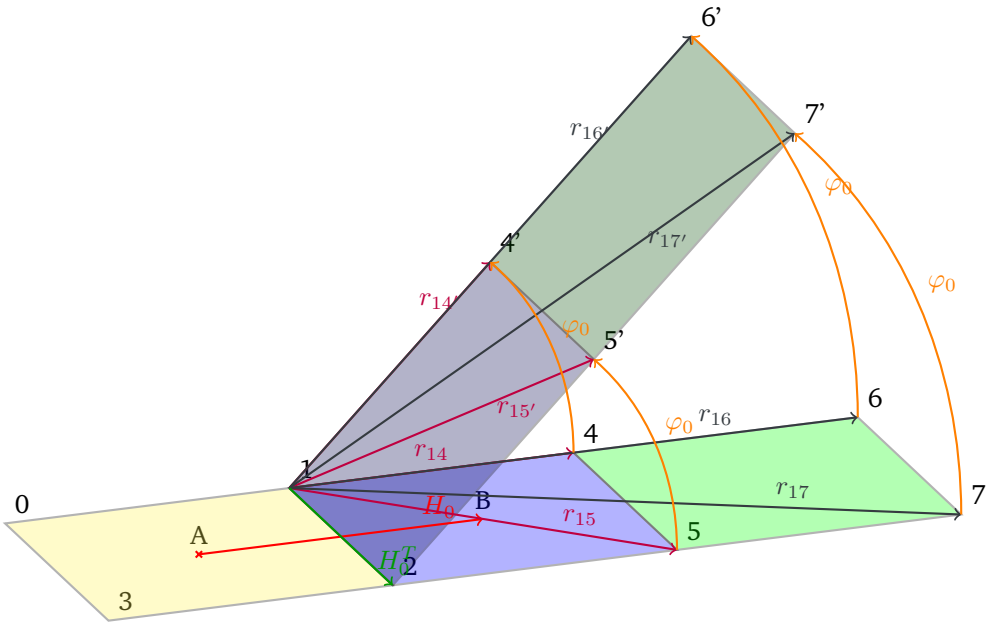


Figure 2.12.: Rotation of a child face for different-plane neighbouring faces and propagation to child face of child face.

By repeating the first process for all faces in the MST, all planar coordinates of the unfolding can be obtained. The face-face angles are all set to $180°$ by construction the vertices within a common plane. If now the final vertex positions for the closed fullerene, and thereby the final face-face angles are known, an up-folding path can be obtained by sampling each face-face angle between the initial value of $180°$ and the final angle θ_i and performing the rotation procedure for each step.

This process starts at the root face of the MST and rotates each child face by the required angle. Looking at figure 2.8, the process can be visualised as follows. First, the planar vertices are constructed, and after that, the rotation of the faces is applied. Starting at the root face 0 the vertex vectors of each face are calculated in the local reference frame which uses one of the hinge vertices as the reference origin.

In the example of the C_{60} unfolding in figure 2.8, that would mean, that for faces $1, 5, 4, 10, 13$ the reference origin would be 0, for faces $2, 6, 7, 11, 14$, the vertex 2 and for faces $3, 8, 9, 12, 15$ the vertex 4. All of the connecting vectors in the three local coordinate systems which are further down in the MST are rotated with the first three rotation matrices. Those matrices correspond to the three hinges:

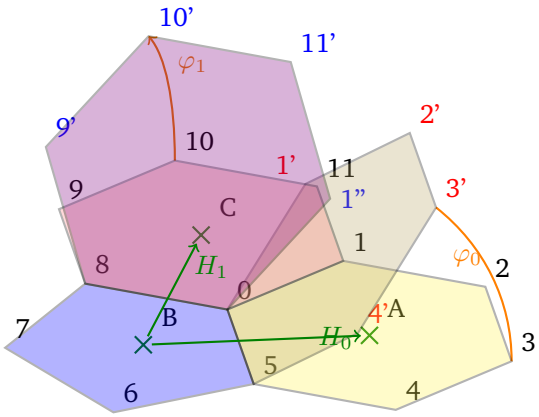


Figure 2.13.: Subgraph of the C_{60} unfolding with Hex-Hex-Pent corner.

$(0-1)$, $(0-2)$ and $(0-3)$. If other face-face angles are to be changed as well, the procedure has to be repeated after this, with the child faces.

2.5.3 Avoiding Double Rotations

There is one special case that has to be considered. When a vertex is part of two faces without being part of the hinge connecting those two faces (for example vertices $18, 25$ and 32 in the C_{60} in figure 2.8), applying the rotation procedures would result in a double rotation around two different axes. The reason is that at the corner between two hexagonal faces and one pentagonal face the face planarity assumption is only true if all three faces are in the same plane. To illustrate why that is so, such a subsystem is chosen and displayed in figure 2.13. The point 1 is part of both hexagonal child faces and will thus be rotated around two different rotation axes $\mathbf{r}_0 - \mathbf{r}_5$ and $\mathbf{r}_8 - \mathbf{r}_0$. Even if the rotational angle is identical, as in this case $\varphi_0 = \varphi_1 = 45.°$, the difference in the rotational axis will results in two different rotated points. The red points in fig 2.13 are the points for a rotation of hinge H_0 and the blue points for a rotation around H_1. The two different vertices for point 1 can be seen from the figure. The reason for this is, that the optimal bond lengths and bond angles within the hexagonal and the pentagonal faces, respectively can only be fulfilled for two different face angular settings. The first, and trivial case, is the planar surface corresponding to face-face angles of $180°$, the second and less trivial case is that of the truncated icosahedron, where the hexagon-hexagon angle is $\theta_{\mathrm{HH}} \approx 41.81°$ and the hexagon-pentagon angle $\theta_{\mathrm{HP}} \approx 37.38°$. If both rotations would be applied to the points 1 it would be first rotated around the first hinge and then the second, so it would not lie within either face. Therefore one rotation has to be chosen that is to be applied. Although the point will now only lie within one of the two faces, it is still a good guess for the geometries along the up-folding path. The bond length and angle of the face for which the rotation is applied will still be correct, while for the other face it will be different.

To identify the vertices for which this problem will occur, the MST needs to be considered. The problem will occur if two faces share a vertex that is not part of the hinge connecting the two faces. More precisely, if a vertex (that is one of the vertices forming a face of the dual graph) is part of two faces (which are vertices of the dual graph) with the same depth in the MST, and that vertex is not part of the parent hinges, only one of the hinge rotations can be applied to this vertex.

It has to be noted, that this problem is not only due to the procedure of constructing the three-dimensional coordinates by applying consecutive rotations, but it is inherent to the geometric properties

themselves. For a truncated icosahedron specifically, or any hexagon-hexagon-pentagon corner, the edge length and angle conditions can only be fulfilled simultaneously in the planar and one other case. Therefore, it is geometrically impossible to satisfy the bond length and bond angle conditions simultaneously if the three connected faces are not in the same plane. Also, the goal of this section was to obtain a guess for the geometries along the way, which can be used as inputs for the energy estimation, thus this method is suited for achieving this guess.

2.6 Quantum Mechanical Background

A molecule is an ensemble of protons, neutrons and electrons, which are all particles obeying the rules of Quantum Mechanics (QM), which dominates the behaviour of matter and photons in the microscopic scale. In the realm of QM, fundamental physical quantities do not take every value of a spectrum but only come in integer multiples of fundamental quanta. In 1926 Erwin Schrödinger proposed a differential equation, similar to Newtons equations of motion, which describes a quantum mechanical system via its wave function in the non-relativistic case. Solving this equation (or the Dirac equation in the relativistic case), leads to a wave function for a system which can be interpreted as the probability amplitude of physical properties of that system. The chemical properties, like geometry and energy of a molecule can be obtained by solving the Schrödinger equation of the system, containing the nuclei and electrons.

By guessing the most likely auto-assembly path, and performing a geometry optimisation at each step, it is possible to get an intuition about the behaviour of precursor molecules, and maybe identify the key mechanisms that make the auto-assembly of these molecules so selective.

Because a molecule is a QM system, its energy is defined by the Schrodinger equation, and to obtain the optimal energy, which corresponds to the optimal geometry, a solution for this equation has to be found. The following chapter introduces the fundamental approximations made and different approaches to solve the Schrödinger equation of a molecule, computationally.

2.6.1 From one to Many Particles

The Schrödinger equation can be written with the Hamiltonian operator $\mathsf{H}_{\text{total}} = \mathsf{T} + \mathsf{V}$ as:

$$\mathsf{H}\Psi = E\Psi, \tag{2.1}$$

where Ψ is the quantum mechanical wave function and solution to this eigenproblem with corresponding eigenvalue E, the energy of the system, with the kinetic energy operator T, and the potential energy operator V. The Hamiltonian operator is the sum of kinetic and potential energy, where the kinetic energy operator is $-\frac{1}{2m_i}\nabla_i^2$ and the potential energy is any external potential. The equation simply states, that at every point in time the sum of kinetic and potential energy has to be equal to a single constant scalar value E, essentially meaning energy conservation. The molecular wave function is a function of the nuclei's and electrons coordinates, and therefore a high dimension problem involving, even for small molecules, a large number of individual particles.

The solution Ψ_k to the Schrödinger equation can be found by turning it into a matrix equation. If $\{\Phi_\nu\}$ is an orthonormal basis of the solution space, such a matrix equation can be obtained by multiplying from the left and integrating with bra-ket notation:

$$\langle\Phi_\nu|\,\mathsf{T}+\mathsf{V}\,|\Psi_k\rangle = \langle\Phi_\nu|\,E_k\,|\Psi_k\rangle = E_k\,\langle\Phi_\nu|\Psi_k\rangle\,, \tag{2.2}$$

and because $|\Psi_k\rangle$ can be expanded in the basis set, and it is othonormal, this leads to:

$$\sum_\mu \langle\Phi_\nu|\,\mathsf{T}+\mathsf{V}\,|\Phi_\mu\rangle = E_k \sum_\mu \langle\Phi_\nu|\Phi_\mu\rangle = E_k c_\mu \delta_{\mu\nu}. \tag{2.3}$$

This has to be true for all basis functions and represents an eigenproblem, which can be written in a matrix:

$$\mathbf{Hc} = E_k\mathbf{c}. \tag{2.4}$$

If the single particle solution ψ_i is expanded in an M dimensional basis set $\phi_1, \cdots, \phi_M$:

$$\psi_i = \sum_{j=1}^{M} c_j^i \phi_j, \tag{2.5}$$

then a basis for the N-particle solution can be constructed by taking the tensor product of N such M-dimensional basis sets. If the particles are fermions and the combined wave function has to satisfy the antisymmetry under particle exchange, a way to construct the combined wave function Ψ is by taking a linear combination of all possible Slater determinants Φ_ν with $\nu = (i_1, i_2, ...i_N)$ of the single basis functions $\phi_1, \cdots, \phi_N$:

$$\Phi\nu(\mathbf{x}_1, ..., \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_{i_1}(\mathbf{x}_1) & \phi_{i_2}(\mathbf{x}_1) & \phi_{i_3}(\mathbf{x}_1) & \cdots & \phi_{i_N}(\mathbf{x}_1) \\ \phi_{i_1}(\mathbf{x}_2) & \phi_{i_2}(\mathbf{x}_2) & \phi_{i_3}(\mathbf{x}_2) & \cdots & \phi_{i_N}(\mathbf{x}_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{i_1}(\mathbf{x}_N) & \phi_{i_2}(\mathbf{x}_N) & \phi_{i_3}(\mathbf{x}_N) & \cdots & \phi_{i_N}(\mathbf{x}_N) \end{vmatrix}. \tag{2.6}$$

For a basis set of size M and N individual particles, there are exactly $\binom{M}{N}$ linearly independent Slater determinants, and the combined wave function would be a linear combination of those individual Slater determinants. The elements of the matrix from which the Slater determinant is constructed are so-called spin-orbitals. The dimension of the solution space of the combined wave function is, therefore, the binomial coefficient of M choose N. Trying to solve such a high dimensional problem with today's computational resources is a futile endeavour for all but the smallest systems, so to obtain a lower-dimensional problem, restrictions have to be made to the complexity.

To reduce the dimensionality of the problem, a common assumption made is to treat the nuclear and electronic Hamiltonian separate from each other. This method is known as the Born-Oppenheimer approximation and is based on the huge difference in nuclear and electronic masses. The mass of the nucleus exceeds the mass of the electron by several magnitudes and given the same kinetic energy, the movement of a single electron is much more rapid than of a nucleus. The Born-Oppenheimer approximation assumes the mass of the nucleus to be infinitely larger than the mass of the electrons. The electronic wave function would then only depend on the coordinates of the nuclei, and, given a perturbation in the nuclei's coordinates, immediately relax to the new ground state. By making this

approximation it is possible to treat the nuclei simply as an external potential to the electronic problem. Separating the Hamiltonian in that fashion results in a split-up of kinetic energy into two, and potential into three terms for electron-electron, electron-nuclei and nuclei-nuclei interactions:

$$\begin{aligned} \mathsf{H} &= \mathsf{T} + \mathsf{V} \\ \mathsf{H} &= \mathsf{T}_\mathrm{n} + \mathsf{T}_\mathrm{e} + \mathsf{V}_\mathrm{ee} + \mathsf{V}_\mathrm{ne} + \mathsf{V}_\mathrm{nn}. \end{aligned} \tag{2.7}$$

This leads to an electronic and a nuclear Hamiltonian:

$$\begin{aligned} \mathsf{H}_e &= \mathsf{T}_\mathrm{e} + \mathsf{V}_\mathrm{ee} + \mathsf{V}_\mathrm{ne} \\ \mathsf{H}_n &= \mathsf{T}_\mathrm{n} + \mathsf{V}_\mathrm{ne} + \mathsf{V}_\mathrm{nn}, \end{aligned} \tag{2.8}$$

because the of the potential energy between nuclei and electron a full separation of the problem is not possible and the problem contains a coupling term.

2.6.2 Hartree Fock Theory

Hartree Fock (HF) theory assumes, that the wave function can be expressed by a single Slater determinant Ψ. The separated Hamiltonian operator (eq. 2.8) can be sorted by the number of electrons that participate in each term. When inserting the kinetic and potential energy operators into the equation this leads to:

$$\mathsf{H}_\mathrm{e} = \underbrace{\sum_i \sum_{j>i} \frac{Z_i Z_j}{|R_i - R_j|}}_{\text{Zero electron part}} \underbrace{\sum_i \left(-\frac{1}{2}\nabla_i^2 - \sum_j \frac{Z_j}{|R_j - r_i|} \right)}_{\text{One electron part}} + \underbrace{\sum_i \sum_{j>i} \frac{1}{|r_i - r_j|}}_{\text{Two electron part}}. \tag{2.9}$$

The Hamiltonian can be written as:

$$\mathsf{H} = \sum_{i=1}^{N_e} \mathsf{h}_i + \sum_{i=1}^{N_e} \sum_{j>i}^{N_e} (\mathsf{J}_{ij} - \mathsf{K}_{ij}) + \mathsf{V}_\mathrm{nn} \tag{2.10}$$

with the core Hamiltonian:

$$\mathsf{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_j \frac{Z_j}{|R_j - r_i|}, \tag{2.11}$$

and the Coulomb- and exchange operators J and K. J is called the Coulomb operator, because it describes the energy associated with the repulsion of two electrons occupying their respective spin-orbital, while K is called the exchange operator. It describes the energy contribution by two electrons exchanging orbitals. Those two terms were inserted into equation 2.10 which simplifies the energy expression. The minus sign for the exchange energy term is due to the antisymmetry of the wave function. They are defined as:

$$\mathsf{J}_i \left|\phi_j(1)\right\rangle = \langle\phi_i(2)|\phi_i(2)\rangle \frac{1}{|r_{ij}|} \left|\phi_j(1)\right\rangle \tag{2.12}$$

$$\mathsf{K}_i \left|\phi_j(1)\right\rangle = \langle\phi_i(2)|\phi_j(2)\rangle \frac{1}{|r_{ij}|} \left|\phi_i(1)\right\rangle \tag{2.13}$$

The Coulomb operator J_i describes the mean repulsion energy from the electron occupying spin-orbital i on the electron occupying spin-orbital j. The exchange operator K_i described the mean repulsion energy where the electrons have been exchanged between the two orbitals. These operators treat the repulsion energies in the mean-field approximation. They are effective one-electron operators, which expresses the repulsion energy term for that particular electron as a sum of all the contributions from the other occupied spin-orbitals.

$$\mathsf{F}_i = \mathsf{h}_i + \sum_{j \neq i}^{N} (\mathsf{J}_j - \mathsf{K}_j) \tag{2.14}$$

This operator describes the kinetic energy of the electron, the energy in the potential of the nuclei, as well as the mean repulsion and exchange energy from all other orbitals. It is an effective one-electron operator. The energy of a single electronic orbital in the field of all other occupied spin-orbitals is described via the eigenvalue-equations:

$$\mathsf{F}_i \Phi_i = \sum_{j \neq i}^{N_{\text{el}}} \lambda_{ij} \Phi_j = \sum_{j}^{N_{\text{el}}} \lambda_{ij} \Phi_j \tag{2.15}$$

In this problem, each of the spin-orbitals itself depends on the solution of all other eigenvalue problems, what makes the problem non-linear. The restriction in the sum can be removed, because it can be shown, that the Coulomb and exchange energies cancel each other in the case of $j = i$. This results in a matrix of eigenvalues λ_{ij}, which can be diagonalised by a unitary transformation so that all off diagonal elements are zero :

$$\lambda_{ij} = \delta_{ij} \epsilon_i, \tag{2.16}$$

simplifying the system to:

$$\mathsf{F}_i \Phi'_i = \epsilon_i \Phi'_i \tag{2.17}$$

Because the solution of this eigenvalue depends on itself, the problem is a non-linear eigenvalue problem. The solution is found by iteratively solving the linear approximation until a stable solution is found. Therefore, the Hartree-Fock method is also referred to as a self-consistent field (SCF) method.

The set of orbitals in which the matrix becomes diagonal are called the canonical molecular orbitals (MOs), and the corresponding expectation values of the Fock operator ϵ_i are interpreted as the energies of those orbitals because the Fock operator describes the energy of a molecular orbital in the mean-field of all other occupied spin-orbitals.

To solve the Hartree-Fock equations, a reasonable initial guess for the molecular wave functions is needed. There is no analytic solution for the Schrödinger equation for any other system than the

hydrogen and the H_2^+ atom. The so-called Roothan-Hall equations[26, 27], are obtained by writing the MOs as a linear combination of the atomic orbitals:

$$\Phi_i = \sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha. \tag{2.18}$$

and inserting expansion into the Hartree-Fock equations 2.17:

$$\mathsf{F}_i \sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha = \epsilon_i \sum_{\alpha}^{M} c_{\alpha i}\chi_\alpha. \tag{2.19}$$

The linear expansion coefficients $\mathbf{c}_i$ are constants, and therefore not affected by the Fock operator. Multiplication from the left with the complex conjugate of all the atomic basis functions χ_α and integration of the equation, leads to a square matrix of equations with the dimensions of the number of basis functions:

$$\mathbf{F}_i\mathbf{c}_i = \epsilon_i\mathbf{S}\mathbf{c}_i, \tag{2.20}$$

where $\mathbf{F}_i$ are the Fock elements, $\mathbf{c}_i$ are the linear expansion coefficients, $\mathbf{S}$ the overlap integrals of the different basis functions χ_α and ϵ_i are the eigenvalues to the Fock operator.

$$\mathbf{F}_{\alpha\beta} = \langle\chi_\alpha|\,\mathsf{F}\,|\chi_\beta\rangle \tag{2.21}$$

$$\mathbf{S}_{\alpha\beta} = \langle\chi_\alpha|\chi_\beta\rangle \tag{2.22}$$

These equations provide a system, to relate the basis functions χ via the expansion coefficients $\mathbf{c}_i$ to the MO energies ϵ_i. By solving those equations a solution for the MO's Ψ_i can be expressed, and via the Slater determinant, the wave function of the combined system can be solved. It is important to appreciate, that solving the Roothaan-Hall equations leads to a total number of MOs, that equals the number of basis functions M used, while in the Fock matrix the number of contributing spin-orbitals is equal to the total number of electrons N of the system. To get the lowest energy wave function, the N lowest energy MOs are chosen.

The elements of the Fock matrix $\mathbf{F}_{\alpha\beta}$, contain the contribution from the core Hamiltonian h and the contribution from the two-electron integral part that describes the Coulomb and exchange terms.
In Restricted Hartree-Fock (RHF), the spatial component of two wave functions with opposite spin are forced to be the same, and the only difference is the spin component, while in Unrestricted Hartree-Fock (UHF), the spatial components of the wave functions with opposite spin are allowed to differ.

Figure 2.14 illustrates the Hartree-Fock procedure to solve the Schrödinger equation of a molecule. The obtained solution will be in form of the coefficients of the basis set expansion for each of the spin-orbitals in the Hartree product.
By using a set of basic orbital functions (see section 2.6.5), the Hartree-Fock system of eigenvalues can be solved iteratively until a self-consistent solution is found. This way, an approximation for the MO functions can be found and the energy of the system can be determined. By minimising this energy the optimal geometry of a system in the ground state can be found. In HF theory, the N-body wave function is described by a single Slater determinant, and the electron-electron repulsion is described in a mean-field approximation.

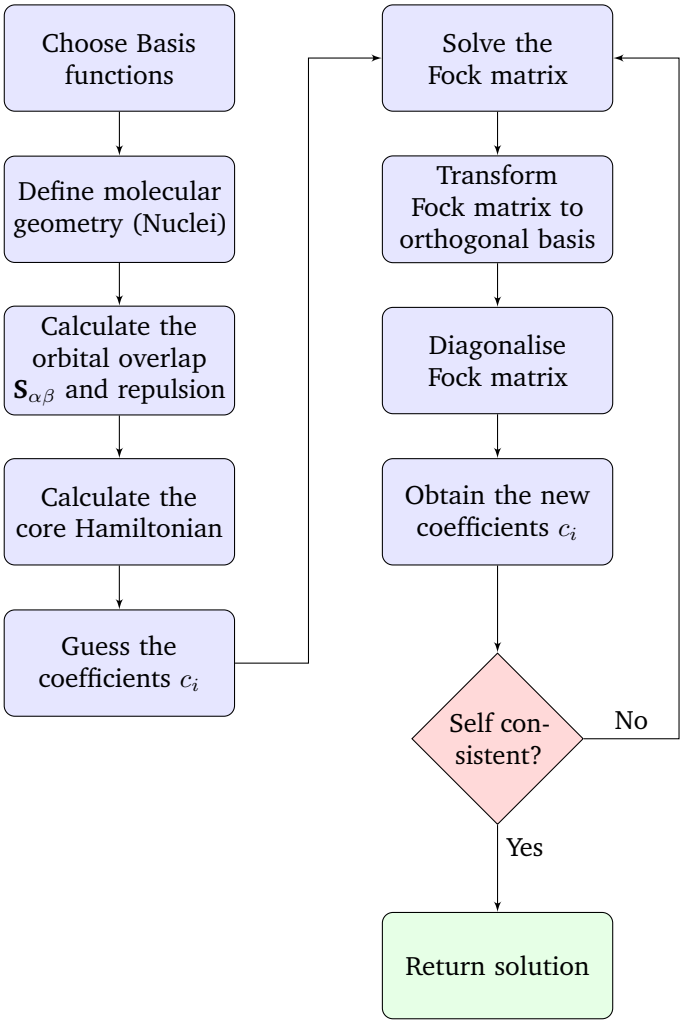


Figure 2.14.: Flow chart of the Self Consistent Field (SFC) method.

2.6.3 Coupled-Cluster Theory

The previously outlined Hartree-Fock method describes the electron-electron interactions in the mean-field approximation. This means, that one electrons' repulsion and exchange energy is described as the mean repulsion and exchange energies from all the other occupied spin-orbitals. Since the total wave function was defined by the single best-fitting Slater determinant, to improve on the HF solution, a good idea is to include other Slater determinants. The only MOs that were used to construct the combined wave function were the occupied ones, while the virtual orbitals were left unused. The electrons were therefore restricted to only occupy the non-virtual MOs.

The general approach to improve on HF results is to expand the solution for the combined wave function by excited Slater determinants. An excited Slater determinant is the determinant of a matrix of spin-orbitals, where some of the virtual orbitals are occupied by an electron, meaning that the matrix for a system of n electrons will include a ϕ_{n+1} orbital occupied by one of the electrons.

The linear combination of Slater determinants can be written as:

$$\Psi = a_0\Psi_{\text{HF}} + \sum_{i}^{\text{occ.}}\sum_{k}^{\text{vir.}} a_i^k \Psi_i^k + \sum_{i<j}^{\text{occ.}}\sum_{k<l}^{\text{vir.}} a_{ij}^{kl}\Psi_{ij}^{kl} \tag{2.23}$$

where ψ_{HF} is the Hartree-Fock solution, ψ_i^k is the Slater determinant, where electron i has been exited to occupy orbital k, ψ_{ij}^{kl} where electrons i and j have been excited to orbitals k ,l respectively, and so forth. This includes all possible excitations, from no electron occupying a virtual orbital to all of them occupying a virtual orbital, if the number of total MOs allows it. The methods using this kind of linear combination of Slater determinants are referred to as Configuration Interaction (CI) methods.

Coupled Cluster theory (CC) reduces the excitations to the consecutive application of a defined excitation operator **T**. The cluster operator **T** is usually defined as:

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + ...\mathbf{T}_\infty, \tag{2.24}$$

where:

$$\mathbf{T}_n = \frac{1}{(n!)^2}\sum_{i_1,i_2,\ldots,i_n}\sum_{a_1,a_2,\ldots,a_n} t_{a_1,a_2,\ldots,a_n}^{i_1,i_2,\ldots,i_n}\hat{a}^{a_1}\hat{a}^{a_2}\ldots\hat{a}^{a_n}\hat{a}_{i_n}\ldots\hat{a}_{i_2}\hat{a}_{i_1} \tag{2.25}$$

is the n-th excitation operator, which creates all possibile permutations of n excited electrons via the creation and annihilation operators $(\hat{a}_a^\dagger =)\hat{a}^a$ and $\hat{a}_i$. For example the $\mathbf{T}_1$ operator is the operator that creates all n single excited electronic Slater determinants and $\mathbf{T}_2$ is the operator that creates the linear combination of all possible two excited electronic wave functions Ψ_{ij}^{kl} with the corresponding coefficients t_{ij}^{kl}. These coefficients are commonly referred to as amplitudes in the CC method. Therefore, the full CI of all possible Slater determinants is "created" by the operator **T** plus the unexcited HF wave function:

$$\Psi_{\text{CI}} = (1 + \mathbf{T})\Psi_{\text{HF}} \tag{2.26}$$

If not the cluster operator, but instead its exponential is applied to the wave function, the results has some very interesting properties. The exponential of an operator is defined via the Taylor expansion of the exponential function:

$$\exp(\mathbf{T}) = \sum_{k=0}\frac{1}{k!}\mathbf{T}^k = 1 + \mathbf{T} + \frac{1}{2}\mathbf{T}^2 + \frac{1}{6}\mathbf{T}^3 + \ldots \tag{2.27}$$

And the exponential Ansatz yields:

$$\Psi_{\text{CC}} = \exp(\mathbf{T})\Psi_{\text{HF}}, \tag{2.28}$$

for the CC wave function. Because the cluster operator itself is a sum of all the different excitations (see eq. 2.24) this can be re-written as:

$$\exp(\mathbf{T}) = 1 + \underbrace{(\mathbf{T}_1 + \mathbf{T}_2 + \ldots)}_{\mathbf{T}} + \frac{1}{2}\underbrace{(\mathbf{T}_1\mathbf{T}_1 + \mathbf{T}_1\mathbf{T}_2 + \mathbf{T}_2\mathbf{T}_1 + \mathbf{T}_2\mathbf{T}_2 + \ldots)}_{\mathbf{T}^2} + \ldots \tag{2.29}$$

Using the full cluster operator as it is presented in equation 2.24 does not give any advantages over a full CI calculation. Therefore, in order to benefit from it, some restrictions have to be made, and a cut-off, or truncation, after a certain $\mathbf{T}_i$ has to be made. Usually all excitations in the order of 3 and higher are neglected, leaving $\mathbf{T}_{\text{trunc}} = \mathbf{T}_1 + \mathbf{T}_2$. This method is called Coupled-Cluster Single Double (CCSD), and the exponential of the reduced cluster operator becomes:

$$\begin{aligned}\exp(\mathbf{T}_{\text{trunc}}) =& 1 + (T_1 + T_2) + \frac{1}{2}\left(T_1^2 + T_1T_2 + T_2T_1 + T_2^2\right) \\ &+ \frac{1}{6}\left(T_1^3 + T_1T_2T_1 + T_2T_1^2 + T_2^2T_1 + T_1^2T_2 + T_1T_2^2 + T_2T_1T_2 + T_2^3\right) \\ &+ \cdots\end{aligned} \tag{2.30}$$

Although this expression formally only contains single and double excitation terms, the consecutive application of the single and double excitation operator can create wave functions with higher-order excitations. Sorting the terms from equation 2.30 by the excitation order, results in:

$$\exp(\mathbf{T}_{\text{trunc}}) = \underbrace{1}_{\text{No excitation}} + \underbrace{T_1}_{\text{Single excitation}} + \underbrace{\left(T_2 + \frac{1}{2}T_1^2\right)}_{\text{Double excitation}} + \underbrace{\left(\frac{1}{2}T_1T_2 + \frac{1}{2}T_2T_1 + \frac{1}{6}T_1^3\right)}_{\text{Triple excitation}} + \ldots \tag{2.31}$$

The big advantage in this representation is, that different to the linear expansion of Slater determinants, where each excitation has an independent coefficient a (see eq. 2.23), the amplitudes of higher-order excitations in the case of the exponential representation, are simply the product of lower-order excitations

To get the energy of the CC wave function, the Hamiltonian operator has to be applied on the modified wave function:

$$\begin{aligned}\mathcal{H}\exp(\mathbf{T})\left|\Psi_{\text{HF}}\right\rangle &= E_{\text{CC}}\exp(\mathbf{T})\left|\Psi_{\text{HF}}\right\rangle \\ \exp(-\mathbf{T})\mathcal{H}\exp(\mathbf{T})\left|\Psi_{\text{HF}}\right\rangle &= E_{\text{CC}}\left|\Psi_{\text{HF}}\right\rangle \\ \left\langle\Psi_{\text{HF}}\right|\exp(-\mathbf{T})\mathcal{H}\exp(\mathbf{T})\left|\Psi_{\text{HF}}\right\rangle &= E_{\text{CC}}\end{aligned} \tag{2.32}$$

It can be shown, that this holds not only for the full cluster operator, but also for the truncated one (see [28]). Therefore, the CC energy levels can be determined by solving the equation:

$$\left\langle\Psi_{\text{HF}}\right|\exp(\textbf{-}\mathbf{T})\mathcal{H}\exp(\mathbf{T})\left|\Psi_{\text{HF}}\right\rangle = E_{\text{CC}}, \tag{2.33}$$

which can be simplified to:

$$\langle \Psi_{\text{HF}} | \, \mathcal{H} \exp(\mathbf{T}) \, | \Psi_{\text{HF}} \rangle = E_{\text{CC}}. \tag{2.34}$$

To solve this equation, the amplitudes of the different excitations $t$are required. Those can be determined by projecting onto the excited Slater determinant instead of the ground state Hartree-Fock. The theorem of Brillouin, proposed in 1934, has proven that for a SCF solution, the matrix elements of the Hamiltonian between the ground state and a single excited state $\langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \Psi^k_{i_{\text{HF}}} \rangle$ are zero[29]. Furthermore, the Slater-Condon rules[30, 31], show that for a two-particle operator like the Hamiltonian the only matrix elements that are left are the two-electron integrals for spin-orbitals which differ by a maximum of two quantum numbers. Any higher-order permutation of spin-orbital coordinates will be zero. For the full exponential operator from equation 2.29, this leads to:

$$\begin{aligned}
E_{\text{CC}} &= \langle \Psi_{\text{HF}} | \, \mathcal{H} \left(1 + \mathbf{T}_1 + \mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2 \right) | \Psi_{\text{HF}} \rangle \\
&= \langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \Psi_{\text{HF}} \rangle + \langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \mathbf{T}_1 \Psi_{\text{HF}} \rangle + \langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \mathbf{T}_2 \Psi_{\text{HF}} \rangle + \frac{1}{2} \langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \mathbf{T}_1^2 \Psi_{\text{HF}} \rangle \\
&= E_{\text{HF}} + \cancel{\sum_{i}^{\text{occ}} \sum_{a}^{\text{virt}} t_i^a \langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \Psi_i^a \rangle} + \\
&\quad + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} \left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a \right) \langle \Psi_{\text{HF}} | \, \mathcal{H} \, | \Psi_{ij}^{ab} \rangle \\
&= E_{\text{HF}} + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{virt}} \left(t_{ij}^{ab} + t_i^a t_j^b + t_i^b t_j^a \right) \left(\langle \psi_i \psi_j | \psi_a \psi_b \rangle - \langle \psi_i \psi_j | \psi_b \psi_a \rangle \right)
\end{aligned} \tag{2.35}$$

Brillouin theorem

The last row in equation 2.35 applies the Slater-Condon rules for the two-electron Hamiltonian. It has to be noted, that the lack of higher-order excitation operators is not due to the truncation of the excitation operator, but is due to the Slater-Condon rules. The amplitudes, t_i^a and t_{ij}^{ab} are the amplitudes of the single and doubly excited wave functions.

Except for a full configuration interaction, which includes all excited Slater determinants contributing with individual amplitudes, CC is currently the best available ab initio method for computational chemistry. Except for the obvious contribution from the number of atoms, another limiting factor on the systems that modern supercomputers can solve, is the number of basis sets used, which will be discussed in section 2.6.5. With state-of-the-art basis sets and highly parallelised algorithms running on a supercomputer, it is possible to solve the Schrödinger equation via the CCSD method for systems consisting of up to a hundred atoms. It is possible to truncate the excitation to singly excited Slater determinants which would a leave Coupled-Cluster Single (CCS) method, or go up to triple and quadruple excitations, which is usually not feasible. Other methods, where the triple excitations are estimated with a less expensive method than CC exist and are referred to as CCSD(T), with a T in parenthesis to point out that those triple excitations are estimated using perturbation theory.

Although the CC method is an ab initio method and its results are yet to be matched by any other computational chemistry method, the system size of fullerenes (at least 60 atoms), restricts its application in this particular problem. It is not feasible to scan the vast precursor space with thousands of possible unfoldings with such an expensive method. Therefore, a cheaper method has to be used to make a fast

comparison of possible precursor molecules feasible. Density Functional Theory (DFT), which will be explained in the following section, provides such a tool. However, CC theory does provide a very robust method that leads to trustworthy results and if, in any way possible, the results obtained with it should be included as a reference solution for the problem at hand. In this work, the CC method was used to assess the accuracy of different functionals in the DFT method for the specific bond-formation and bond-breaking processes that occur in the auto-assembly of fullerene-precursor molecules.

2.6.4 Density Functional Theory

The two methods discussed above, Hartree-Fock and Coupled Cluster are both ab initio methods, using the wave function as a basis for the solution. The problem with those methods is the high dimensionality of the wave function. For each spin-orbital there are 4 coordinates, meaning that for a system of N electrons, there are at least $4N$ independent variables. It was described before, that the high dimensionality of the problem makes it infeasible to apply those methods to molecules that consist of more than a hundred atoms. A completely different approach to the determination of molecular energies developed, after the publication of the Hohenberg-Kohn theorem in 1964 [32, 33]. This theorem proofed, that for an inhomogenous electron gas in an external potential, there exists a unique functional, which relates the electron density to the energy of the ground state. In contrast to the rather obscure nature of a wave function, the electron density is a quantity, which can be observed. This functional is universal and independent from the external potential. By making the Born-Oppenheimer approximation, and treating the nuclear potential as the external potentials, it is proven that there exists a functional that relates the molecular electron density to the energy of the molecule. The big advantage in this method is, that the electrons are not treated as individual, interacting particles in their spin-orbitals, but the whole electronic state is treated as a three-dimensional function. This reduces the dimensionality drastically, making calculations of much larger molecules possible than were previously accessible through ab initio methods. There are two different approaches to DFT, one of which can do completely without orbitals, both are outlined in the following paragraphs.

Orbital-Free DFT

The methods, exploiting the Hohenheim-Kohn theorem, are called Density Functional Theory (DFT) methods. Although it has been proven, that a functional exists, which relates the three-dimensional electron density to the energy, the exact universal functional has yet to be found. The main findings of Hohenberg and Kohn lead to an energy expression in the form of:

$$E = \int \mathbf{V}(R)\rho(r)dr + \frac{1}{2}\int\int \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + \mathbf{G}\left[\rho(r)\right]. \tag{2.36}$$

Here $\mathbf{G}\left(\rho(r)\right)$ stands for the universal functional of the electron density, the first part represents the potential of the nuclei, and the second part the self repulsion Coulomb energy of the electron density. The external potential, can be obtained from the electron density itself via a functional V_{ne} that relates the cusp-like maxima in the electron density to the nucleis position and charge:

$$V_{ne}[\rho] = -\sum_{a}^{N_{\text{nuclei}}} \int \frac{Z_a(R_a)\rho(r)}{|R_a - r|}dr \tag{2.37}$$

with the nuclear coordinates R_a, and the nuclei's charge Z_a. there are no one electronic coordinates r_i anymore because there is now an electron density ρ that has to be integrated over the whole space. The

electronic density integrated over the whole space has to equal the number of electrons of the system. The electron-electron repulsion functional takes the form:

$$J[\rho] = \int\int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr'. \tag{2.38}$$

That is a trivial result from classical physics. Those two simple functional provide terms for the external potential and the electron-electron repulsion analogon. The Hamiltonian operator can be split into three individual parts $\mathcal{H} = \mathbf{V} + \mathbf{T} + \mathbf{U}$, with the potential **V**, the kinetic energy **T** and the electron-electron interaction energy **U**. The kinetic energy term is not trivial and to find an expression L. H. Thomas and E. Fermi [34, 35] which was later combined with results obtained by F. Bloch and P. Dirac [36, 37] to include an electron exchange energy, lead to the following functionals:

$$T_{\mathrm{TF}}[\rho] = \frac{3}{10}\left(3\pi^2\right)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(r) dr \tag{2.39}$$

$$K_{\mathrm{D}}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho^{\frac{5}{3}}(r) dr \tag{2.40}$$

The indices of the respective functionals are according to the person mainly associated with their discovery. Those results were derived from the assumption of a uniform electron gas, which is a bad approximation for a molecule. Another problem, that occurs if using the trivial Coulomb repulsion functional $J[\rho]$ is, that the density of every point will repel itself at any other point. The reason why this is a problem becomes evident when looking at a system with only one electron. A single electron should not have a repulsive energy contribution from itself. However, the description of an electron via the electronic density will have a repulsive contribution. The problem with the non-uniformity of the electron density can be attacked by adding correctional terms that depend on the derivative of the electron density, which has the problem that those correctional terms do not converge. A solution to the self repulsion problem is the introduction of so-called Fermi holes, which add an oppositely signed contribution to the exchange functional J.

In the DFT approach described above, it was possible to determine the energy of the system without making use of orbitals and reducing the many-dimensional HF method to a three-dimensional problem. Although this is a great achievement, the difficulties with the kinetic energy functional and the electron-electron repulsion are significant and the results obtained hereby do not have the required accuracy. Therefore a different approach, which re-introduces orbitals is described in the following.

Orbital DFT

Although the lack of orbitals has some obvious benefits like the reduction to three coordinates, the poor description of kinetic energy by orbital free DFT is a major disadvantage. W. Kohn and L. J. Sham took previous approaches from orbital free DFT further [33]. Their idea was, to describe the functional $\mathbf{G}[\rho(r)]$ as a sum of a fictional energy functional $\mathbf{T}[\rho(r)]$, which describes an explicitly non-interacting electron density and an exchange and correlation part $\mathbf{E}_{\mathrm{xc}}[\rho(r)]$, which contains all the correctional terms for the electron self repulsion and the kinetic energy interaction. The Hamiltonian operator for a non-interacting electron system can be written as a sum of the operators on every individual electron, without any two-electron operators. The eigenfunctions of the kinetic energy part of such a Hamiltonian operator are the Slater determinant of the N molecular orbitals of the electrons. This increases the number of independent variables to $3N$, which is still considerably lower than in e. g. CC theory. If the

electron density is now expressed via the basis set used for the construction of the Slater determinant, as their expectation value:

$$\rho(r) = \sum_{i=1}^{N} |\chi_i(r)|^2, \tag{2.41}$$

the kinetic energy functional can be simplified to:

$$\begin{aligned} E\left[\rho(r)\right] = & \underbrace{\sum_{i}^{N} \langle \chi_i | -\frac{1}{2}\nabla_i^2 | \chi_1 \rangle}_{\mathbf{T}_n i} - \underbrace{\sum_{i}^{N} \langle \chi_i | \sum_{j}^{\text{Nuclei}} \frac{Z_j}{r_i - r_k} | \chi_1 \rangle}_{\mathbf{V}_{ne}} \\ & + \underbrace{\sum_{i}^{N} \langle \chi_i | \int \frac{\rho(r')}{r_i - r'} dr' | \chi_1 \rangle}_{\mathbf{V}_{ee}} + \underbrace{\mathbf{E}_{XC}\left[\rho(r)\right]}_{\text{Exchange-Correlation Functional}} \end{aligned} \tag{2.42}$$

The first two terms simplify an expression whiteout the density, due to the construction of the electron density via the orthogonal orbitals, an N refers to the electron occupying the orbital χ_i. These terms are identical to the HF terms. It has to be noted, that this equation defines what the exchange and correlation energy is. Meaning, that in orbital-containing DFT, the energy of the system is described by the energy of the non-interacting electron density and an additional energy correction functional, which describes all other effects, like the exchange and correlation energy. Furthermore, as seen in equation 2.41, the density itself is dependent on the orbitals. This represents a set of equations, that have to be solved in the self-consistent approach similar to the HF method.

Because the definition of the exchange and correlation energy means, that this term contains all the complicated classical and non-classical interactions of the electron density, it is of paramount importance to find a functional, which describes these interactions well. Such a functional exists, according to the Hohenberg-Kohn theorem, but has not been found. The functionals, which Kohn and Sham introduced in 1965, is the so-called Local Density Approximation (LDA). In such a functional, the exchange and correlation potentials depend only on the local electron density and take the form of:

$$\mathbf{E}_{XC}\left[\rho(r)\right] = \int \rho(r) \left[\epsilon_X\left(\rho(r)\right) + \epsilon_C\left(\rho(r)\right)\right] \rho(r) dr. \tag{2.43}$$

This can be generalised to include the spin, and is then also referred to as Local Spin-Density Approximation (LSDA), by using two different spin density functionals for the two spin states α and β. Those result in two different spin densities, which are then normalised:

$$\zeta = \frac{\rho(r)_\alpha - \rho(r)_\beta}{\rho(r)_\alpha + \rho(r)_\beta}, \tag{2.44}$$

which is also called the spin polarisation. A quantity which is often seen in the formulation of functionals is the effective electron radius, which is the radius of a sphere in which the total charge of the electron density equals one electronic charge. So to say, it is the volume which contains one electron, as a function of space.

$$r_S(r) = \left(\frac{3}{4\pi\rho(r)}\right)^{\frac{1}{3}} \tag{2.45}$$

The exchange energies in LDA/LDSA are usually taken from the results for the homogeneous electron gas using Slater [38] or Dirac [37] exchange terms. The correlation energy has not been derived analytically, but determined via Quantum Monte Carlo integration calculations performed by Ceperly and Alder [39] in 1980. The same year, Vosko, Wilk and Nusair [40] used those calculations to construct local correlation functionals that fitted the results. In 1992, Perdew and Wang [41] used the same data set to obtain analytic functionals.

The LDA/LSDA use the homogeneous electron gas as the foundation for the electron density and to go beyond this approximation, the local non-homogeneity of the electron density has to be taken into account. This is done, by considering the derivatives of the electron density. Methods including the derivatives of the electron density are called Generalised Gradient Approximations (GGA). The usual way of constructing such a functional, is by starting with the LDA and include the dependency on the density gradient as a correctional term in the spirit of the Taylor expansion. Such a functional was proposed by A. D. Becke in 1988 [42], which included constraint on the electron density at the Fermi holes (same spin) and Coulomb hole (opposite spin). This functional describes the gradient dependent correction as:

$$\Delta\epsilon_X^{\mathrm{B88}} = -\beta\rho^{\frac{1}{3}}\frac{x^2}{1+6\beta x\sinh^{-1}(x)} \tag{2.46}$$

with the normalised un-directed gradient $x = \frac{|\nabla\rho|}{\rho^{\frac{4}{3}}}$. The only empirical parameter β was obtained by fitting to the same data set as the previous functionals. The functional has been shown to yield the correct asymptotic behaviour for large distances [43].

A functional, which does not add a correctional term to the LDA/LDSA, but aims at describing the correlation energy directly is the Lee-Yang-Parr (LYP) functional [44, 45]. This functional does not include the same spin correlation. Some functionals go beyond the first order derivative, including the Laplace operator and higher order terms. Instead of explicitly using the Laplace operator, a common approach is to use the kinetic energy density:

$$\tau = \sum_i^{\mathrm{occ}} \frac{1}{2}|\psi(r)|^2, \tag{2.47}$$

mainly for numerical reasons. Those functionals are called meta-GGA functionals.

Hybrid functionals are functionals, which describe the exchange and correlation energy of the electron density as a fractional contribution of different functionals. The initial Half-and-Half functional used by Becke in 1993 [46], calculated the exchange and correlation energy by by using a fraction of 50% Regular HF exchange (see equation 2.13), and 50% from the LSDA functional, resulting in a total exchange and correlation energy description of the form:

$$E_{\mathrm{X}}^{\mathrm{H+H}} = \frac{1}{2}E_{\mathrm{X}}^{\mathrm{HF}} + \frac{1}{2}\left(E_{\mathrm{X}}^{\mathrm{LSDA}} + E_{\mathrm{C}}^{\mathrm{LSDA}}\right). \tag{2.48}$$

This increases the complexity of the problem, because as was shown above, the calculation of the exchange operators matrix elements requires the integration over a product of two different occupied spin-orbitals. An even better approximation was found also by Becke in the same year as a combination of a HF exchange part, the LSDA exchange and correlation functional, a gradient corrections from the B88 and from the PW91 functional by Perdew and Wang [41, 47, 48]. This functional contains three

empirical parameter, which describe the relative importance of certain properties. The combined energy functional is called B3PW91 (Becke-3-parameter-Perdew-Wang-91) looks as follows:

$$E_{\mathrm{XC}}^{\mathrm{B3PW91}} = E_{\mathrm{XC}}^{\mathrm{LSDA}} + a\left(E_{\mathrm{X}}^{\mathrm{HF}} - E_{\mathrm{X}}^{\mathrm{LSDA}}\right) + b\delta E_{\mathrm{X}}^{\mathrm{B88}} + c\Delta E_{\mathrm{C}}^{\mathrm{PW91}}. \tag{2.49}$$

The parameter a reflects how much of the systems exchange energy can be taken as pure HF exchange, and the parameters b and c describe how much the system needs to be corrected for non-local effects. The three parameters were fitted for a data set and the values, which were obtained are $a = 0.20$, $b = 0.72$ and $c = 0.81$. Another approach used the 3 parameter LYP functional, which describes the complete correlation energy directly and not via a correctional term [49]. This functional is consequently called B3LYP and is one of the most used functional in computational chemistry. Interestingly, the functional made use of the same a, b and c parameter values as the B3PW91, using their respective physical interpretation and therefore scaling back the contribution of the LSDA correlation (compare eq. 2.50). This has been one of the most widely used functional in computational chemistry.

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-a)E_{\mathrm{X}}^{\mathrm{LSDA}} + aE_{\mathrm{X}}^{\mathrm{HF}} + b\delta E_{\mathrm{X}}^{\mathrm{B88}} + (1-c)E_{\mathrm{C}}^{\mathrm{LSDA}} + c\delta E_{\mathrm{C}}^{\mathrm{LYP}}. \tag{2.50}$$

Although the computational complexity of DFT methods, is comparable to HF methods when using a hybrid functional (which uses HF exchange), the results obtained through it, are much better, making DFT methods the best currently available method for large system quantum chemical calculations. Almost all methods, except for orbital-free DFT, presented above make use of a basis set to approximate the complicated form of MOs. The scaling of any method is therefore dependent on the size of this basis set. The following section will explain what kind of basis sets exist and what their properties are.

2.6.5 Basis Sets

As described above, all methods to solve the Schrödinger equation involve the approximation of the molecular orbitals by some basis set. The construction of the molecular wave function is then derived and expressed as a linear combination of those fundamental basis functions (LCAO method). Therefore, the results of any method are always restricted by the size of the basis set and it is important to pick a sufficiently large basis set for the chosen level of theory. This section describes the difference between basis sets, and which basis set is most appropriate for which problem.

The molecular orbitals, by which the complete electronic wave function is determined are usually expanded in a series of basis functions. These functions are required to be computationally easy to handle (allow quick integration) and should be able to yield good results for any type of atom. The spherical harmonics ($Y_{l,m}$) are a natural choice for a basis function because they are part of the analytic solution for the one-electron problem, Slater-Type orbitals (STO) uses the spherical harmonics with some modifications as a basis set. They are a good choice due to their good description of the orbitals but are computationally very expensive, so they are usually not feasible for methods calculating the two-electron integrals, like HF or Hybrid-DFT.

Another type of basis functions are so-called Gaussian-Type orbitals (GTO) which are three-dimensional Gaussian functions, centred at the nuclei. Integration over those GTOs is faster than for the STOs, but this computational advantage, comes at the cost, that the description of the atomic orbitals is not as good as for the STOs.

$$\chi^{\text{STO}}_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{n-1}\exp^{-\zeta r} \tag{2.51}$$

$$\chi^{\text{GTO}}_{\zeta,n,l,m}(r,\theta,\varphi) = NY_{l,m}(\theta,\varphi)r^{2n-2-l}\exp^{-\zeta r^2} \tag{2.52}$$

If the positions of the nuclei and the type of atoms are known, the choice of basis set and it defines the atomic orbitals. Each atomic orbital is then expressed as a linear combination of the basis set. Equations 2.51 and 2.52 show the form of the STO and GTO functions. For each orbital type ($1s$, $2s$, $2p$, ...), and each atom, there are several core basis functions which are used to construct the orbital. The parameter ζ is then fitted for each expansion and each atom. The term single ζ value, or single-ζ quality is used to describe a basis set, where for all atoms for each atomic orbital type there is exactly one ζ value which is optimised, for each of the core basis function.

For example, a basis set, which consists of three core GTO functions, and has single ζ quality, would have three different zeta values for the construction of a Hydrogen and Helium atom and each a linear expansion factor for each of the Gaussian functions. For double ζ quality, this number would rise to six. The Pople series [50] split up the orbitals, into an inner core part, and an outer valence part, which was furthermore subdivided into parts. For a second-row element, this would mean, that the $1s$ orbital is constituted, by 4 Gaussian functions, and the valence shell $2s$ has double ζ quality, with two Gaussian functions for the first ζ and one for the second. The valence shell $2p$ also has double ζ quality with two and one basis functions. Furthermore, the $2p$ orbitals are split up into three angular momentum components.

The polarisation functions are added to the basis set to describe the effect that bonds have on the electrons, however, the correlation effect also influences the probability of finding electrons in a certain region. Considering a hydrogen fluoride molecule, for example, the fluorine's electro-negativity (a measure for an atoms tendency to attract electrons) is much higher than for the hydrogen. The reason for this is the much larger positive charge residing in the fluorine atoms core. This miss-match in electro-negativity results in a much higher probability to find an electron close to the fluorine core than close to the hydrogen core. To account for this behaviour, it is necessary to allow the hydrogen's electron wave function to be further away from the core, or to diffuse into regions further away from the core. The functions, which are added to allow this diffusion are long-tail Gaussian functions (small ζ) and are indicated by a plus sign for heavy atoms and two plus signs for heavy atoms and hydrogen.

The basis set *3-21G**+*, for example, would mean a Pople like basis set describing the core with 3 GTOs, having double zeta quality (two numbers after the "-") for the valence part, with the first valence part represented by two GTOs and the second by one GTO. It will also include polarisation functions on all atoms, and diffusion functions on all but hydrogen. Basis sets can strongly influence the quality of quantum chemical calculations, and even the best level of theory, like full CI, will not yield good results if a poor basis set is chosen. However, the larger the basis set, the longer a calculation will take and it is, therefore, important to pick a basis set which suits the problem at hand.

2.6.6 Dispersion

Dispersion forces are weak, short-distance interactions like van der Waal forces, which occur between atoms. Most DFT methods fail to describe these forces accurately, which have an attractive long-range

part in the order of r^{-6}. The simplest way to add these forces is by adding a diatomic potential part with empirical parameters. To avoid short-range diversion of this potential it is custom to add a damping function, which scales back the attractive potential for small distances. Another problem is, that part of the short-range dispersion forces, might already be included in the functionals. Therefore, a scaling factor is included to reduce the contribution of this so-called empirical dispersion to avoid the double-counting of the dispersion contribution. These functional dispersion forms are also called Grimme dispersion, after their original proposer S. Grimme [51]. There are some functionals, which have been fitted to specifically include the effects of dispersion, like the Minnesota functionals *Mxy*. Although the addition of empirical dispersion for these functionals can still improve the results, they do include some double-counting, which has to be taken into consideration when interpreting the results [52].

The theoretical study of quantum chemical systems poses a big difficulty starting at the theoretical approximations which have to be made to make the problem solvable (Born-Oppenheimer Approximation, one-electronic Hamiltonian, limited CI,...), choosing a method to solve that problem (HF, CCSD, DFT,...) and picking a basis set to approximate the solution. At each step of this chain, approximations are made that take a step back from physical reality as a trade-off to make calculations possible. Although this might seem like the methods described above are very difficult to verify, there are methods which can increase faith in the results obtained. Comparing DFT results to CCSD results shows whether a calculation is consistent across different methods, and verifying a DFT calculation for different functionals shows that the result only depends slightly on the choice of functional. Also, choosing a large enough basis set appropriate for the problem at hand increases the reliability of the calculations. In summary, theoretical studies of molecular systems via computational chemistry is a tricky business, but there are methods to show that the results are trustworthy across a variety of different methods and these calculations offer a spatial and temporal resolution, which is impossible to obtain via experiments. To understand reactional mechanisms of a chemical process it is a good method which, taking some precaution, can lead to a better understanding of the reaction. This makes computational chemistry and especially DFT a good candidate to study the auto-assembly of fullerene precursor and can help to understand how this process works.

2.7 Force Field Methods

The QM methods presented earlier are computationally too expensive, to calculate the energies of hundreds of fullerene precursor molecules at multiple points along the up folding path. However, this is exactly the kind of calculation, which would allow an assessment of the precursor quality. Precursor, which have low energy barriers along their paths are much more likely to auto-assemble, then those with high activation barriers. Therefore, a fast method is needed, which can calculate a fine grid of molecular energies along the suspected closing path. Force Field (FF) methods provide such a tool and are outlined in the following.

In contrast to quantum mechanical ab-initio approximation theories, FF methods approximate the electronic energy as a parametric function of the nuclear coordinates mostly making use of the Born-Oppenheimer approximation. In FF methods, the electrons of each atom are not treated as individual particles but the energy is simply defined as a parametric function of the nuclei's coordinates. The parameters of these functions can be obtained by fitting them to computational results of higher-level theories like DFT or, in the case of small systems, even CC results. In the case of high-level ab initio

methods, the molecular bonds are the result of the solution of the Schrödinger equation. When using any FF method, the bonds have to be explicitly programmed into the calculations.

FF methods simplify any QM interaction and treat a system simply by using Newton mechanics. Atoms and bonds are represented via balls with masses and sizes, connected via springs with stiffness and equilibrium lengths. Depending on the FF implementation other forces like, bond angles, Coulomb forces or out of plane potentials can be included in the FF. There has been a lot of research in the field of FF methods or molecular mechanics (MM) since the 1930s, and many of the interatomic potential's parameters have widely accepted parametric values for bond distance and stiffness. It has to be noted, that all these parameters and individual forces are not based in the quantum mechanical theory, but should be seen as a simple heuristic which can be suited to categorise certain types of chemical bonds into clusters. Those clusters show similar geometrical properties, although, their quantum mechanical foundations do not make these similarities apparent.

Although the accuracy of MM results is not comparable to high-level ab initio, or even semi-empirical methods like DFT, they are very relevant due to their enormous advantages in computation time. The treatment of big molecular systems with high-level methods is, in general, not feasible, even with access to supercomputers and state of the art parallelisation techniques. The reason for this is the extremely high number of electron-electron interactions and the complicated integrals, resulting in scaling in the order of $\approx N^3$ for HF and DFT based theories.

The main idea behind a FF is that any function can be written as a Taylor series around the minimum, using its derivatives. If the molecular geometry is close to the optimum, the nuclear coordinates represent such a local minimum. The complicated quantum mechanical energy can be expanded in a Taylor series around this minimum using an inter-nuclear parameterisation. This approximation is only valid in the proximity of the minimum unless the complete Taylor series is used, which is practically impossible. Therefore to simulate the breaking and formation of bonds other approximations need to be made.

This section outlines, what the foundation of force fields methods is, which parameterisations can be used and what the nature of different interatomic energy functions should be to best represent the energies of fullerene precursor-like molecules.

2.7.1 Force Field Energy in Molecular Mechanics

The energy of a system in a Force Field method is composed of individual distortion terms, of which each is associated with a specific deformation of the geometry of the molecule. The optimal geometry of a molecule corresponds to the deep local minimum of the FF energy term. The force acting on an atom can be obtained by taking the gradient of the energy function regarding that atoms' nuclear coordinates. The potential energy of a molecule can be approximated by the contributions of individual chemical

entities like atoms, bonds and angles [53, 54]. The typical potential energy expression in a FF method looks like this:

$$V_{\text{FF}} = \sum_{\text{bonds}} V_r + \sum_{\text{bonds}} V_\theta + \sum_{\text{torsion}} V_\varphi + \sum_{\text{out-of-plane}} V_{\text{out-of-plane}} + \sum_{\text{not-bonds}} V_{\text{VdW}} + \sum_{\text{not-bonds}} V_{\text{C}} + \sum V_{\text{coupling}} \tag{2.53}$$

The coupling term in the potential refers to the energies associated with the coupling of two or more other energy contributions, such as two different bonds interacting with each other. The individual contributing potential in equation 2.53 can be expressed in a Taylor expansion around their respective minimum.

Bond Stretching Energy

The bond stretching energy term describes the energy associated with the stretching of a bond between two atoms A and B. Around an optimal bond length l_{AB} the Taylor expansion until the second-order derivative yields:

$$V_r(|\boldsymbol{r}_{AB}| - l_{AB}) = V(0) + \frac{dV}{dr}(|\boldsymbol{r}_{AB}| - l_{AB}) + \frac{1}{2}\frac{d^2V}{dr^2}(|\boldsymbol{r}_{AB}| - l_{AB})^2 + ... \tag{2.54}$$

The constant $V(0)$, which represents the energy of the two atoms at the equilibrium, is usually set to zero. Because the expansion was done at the point of the minimum, the first-order derivative is zero and leaves only the second-order term. The potential energy can, therefore, be simplified to:

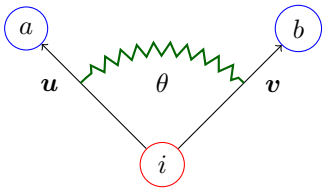
$$V_r(|\boldsymbol{r}_{AB}| - l_{AB}) = \frac{1}{2}k_{AB}(|\boldsymbol{r}_{AB}| - l_{AB})^2 \tag{2.55}$$

which takes the form of a harmonic potential with the force constant k_{AB} of the bond between atoms A and B. The harmonic potential has only one minimum and therefore it is only sufficient to describe a bond between atoms that are not broken or formed during the calculations. The Taylor expansion of the potential is therefore only valid in the close vicinity around the actual local minimum. If new bonds are to be formed or old bonds to be broken with the application of a force field, the harmonic approximation is insufficient and higher-order terms have to be included. As this is computationally more expensive and does not necessarily reproduce the correct behaviour in the dissociation limit, other custom made interatomic potentials have been invented. A widely used potential is the Morse potential [55]:

$$V(r) = D_{AB}(1 - e^{-\alpha(|\boldsymbol{r}_{AB}| - l_{AB})})^2 \tag{2.56}$$

with D_{AB} as the dissociation energy and the constant $\alpha = \sqrt{k_{AB}/2D_{AB}}$. The main assumption, for

(a) Two atoms connected by a linear spring.



(b) Three atoms connected by an angular spring.

Figure 2.15.: The two different bond springs.

the auto-assembly of fullerene precursor, is, that the carbon bonds, which are already present in the

precursor molecule "survive" the auto-assembly, meaning that these bonds will not be broken in the closing process. This is a reasonable assumption for two different reasons. Firstly, the stability of benzene and cyclopentane is very high [56], making braking of those rings unlikely and secondly, the selectivity of the C_{60} precursor strongly suggests that the bonds, which are present in the precursor do not break in the closing process, because otherwise, other fullerene molecules would be able to form during the FVP. Therefore, the assumption of a harmonic potential is justified for the interatomic carbon potentials, for which a bond is already present in the precursor. However, there are interatomic carbon-carbon potentials which should and indeed can not be represented via this approximation. Those are the ones, between which bonds are to be formed during the auto-assembly process. Those carbon atoms are, in the precursor bond to either hydrogen or halogen atom and the automated quality assessment simulation should be able to break these carbon-halogen/carbon-hydrogen bonds and instead form a carbon-carbon bond. Moreover, the respective hydrogen or halogen atom should itself be able to form a bond with another hydrogen or halogen atom to be separated from the precursor body.

Hence, for the interatomic bonds which are to be broken during the closing, it is necessary to use a potential, which can reproduce this kind of behaviour, like the Morse potential. So the approximation for the bond stretching energies, which should be used, are harmonic potentials, for the carbon-carbon bonds, which are already present in the precursor molecules, and Morse potentials for those, where a bond needs to be formed.

All other interatomic potentials, where bonds are neither formed nor broken it is possible to use either a Morse potential, to allow a minimum at the correct bond length or simple Coulomb repulsion. Under the assumption, that as soon as a carbon-carbon bond is formed, this bond will not be broken and should thus be described as a harmonic potential, the interatomic potentials for the carbon atoms, which are close to the minimum of the Morse potential could be replaced with the harmonic potential. The tolerance, around the optimal distance for which this potential form exchange is done, is an important parameter, which could strongly affect the results of a FF method.

The optimal distances for the bonds area another very important value. There are a lot of literature values available for average bond lengths, which could be used. However, the fullerene precursor molecules present a very specific type of molecule, with reoccurring properties for every single molecule. For example, every precursor will have periphery C-C bonds, those which are on the outline of the precursor, and internal C-C bonds, which connect either a pentagonal with a hexagonal, a hexagonal with a hexagonal or a pentagonal with a pentagonal face. Hence the bond lengths for those different types of C-C bonds should not be taken from an average literature value, but rather determined from ab initio or other accurate calculations for fullerene precursor-like molecules. Furthermore, the optimal bond lengths are not necessarily constant during the closing process. In the closed fullerene, there will be no periphery bonds present, so those bond lengths are likely to change as a function of the precursor shape (how much curvature is present, and how closed it is). Thus, an interesting idea is to look at the results for optimal bond lengths along the suspecting closing path to include them to a FF method.

Angular Bending Energy

In addition to the distance between two atoms in a bond, the optimal geometries for molecules also show typical behaviour for the angle between three atoms. The water molecule, for example, has a minimum at a $\angle(HOH) \approx 105°$. Therefore, the bond angle between three atoms is a natural choice

for a force field parametrisation. The Taylor expansion is usually performed in the same way as for the bond stretching potential and terminated after the second order, leading to:

$$V_\theta = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \tag{2.57}$$

The problem with a potential in this form is the singularity, which arises in the gradient at an angle of $180°$ and a common workaround is to use the Taylor expansion around the $\cos\theta$:

$$V_\theta = \frac{1}{2} k_\theta (\cos\theta - \cos\theta_0)^2 \tag{2.58}$$

The derivative of the $\cos$-function is zero at the minimum and the maximums on both sides, therefore, the gradient of this potential is zero at the maximal deviation, which results in the absence of a restoring force. In practice, this means that, once an angular bond has reached a value of $180°$, the forces regarding that angular bond are going to be zero. This is because any derivative of the potential will, via the chain rule, involve a multiplication by the derivative of the cosine, which is zero.

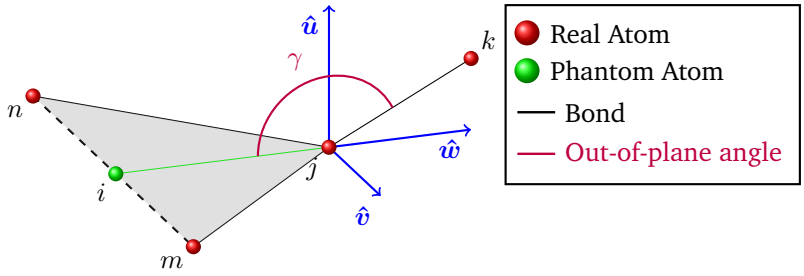


Figure 2.16.: Out of plane parametrisation
The vectors $\hat{v}$ and $\hat{w}$ lie within the gray plane.

Out of Plane Energy

The out of plane energy term is used to describe the planarity of a system involving more than three atoms. Sp^2 hybridised carbon centres, which are surrounded by three other atoms form a plane in the optimal geometry unless otherwise disturbed. To reproduce this planarity an expression is needed, that increases the energy of the molecule as one of those atoms is being pushed out of the plane. As was shown in [57], the out of plane energy term can be interpreted as a special case of bond bending. Therefore, a local coordinate system has to be established for each vertex. The first vector of the local coordinate system is obtained by:

$$\hat{\boldsymbol{u}} = \frac{\mathbf{r}_{mj} \times \mathbf{r}_{jn}}{|\boldsymbol{r}_{mj} \times \mathbf{r}_{jn}|}. \tag{2.59}$$

This vector defines the $\boldsymbol{n}-\boldsymbol{j}-\boldsymbol{m}$-plane, and the next coordinate axis is obtained, by using the component of $\mathbf{r}_{jk}$ which lies within this plane. This is done by taking $\mathbf{r}_{jk}$ and subtracting the projection on the $\hat{\boldsymbol{u}}$-axis:

$$\begin{aligned} \mathbf{r}_{jk(w)} &= \mathbf{r}_{jk} - (\mathbf{r}_{jk} \cdot \hat{\boldsymbol{u}})\hat{\boldsymbol{u}} \\ \hat{\boldsymbol{w}} &= \frac{\mathbf{r}_{jk(w)}}{|\mathbf{r}_{jk(w)}|}. \end{aligned} \tag{2.60}$$

The third, and final coordinate vector is obtained by the cross product of the other two:

$$\hat{\boldsymbol{v}} = \hat{\boldsymbol{w}} \times \hat{\boldsymbol{u}}. \tag{2.61}$$

Now a phantom atom $\boldsymbol{i}$ is placed on the $\hat{\boldsymbol{w}}$-axis, between the atoms $\boldsymbol{n}$ and $\boldsymbol{m}$, essentially, it is placed on the intersection of the $\hat{\boldsymbol{w}}$-axis and the line connecting $\boldsymbol{m}$ and $\boldsymbol{n}$ at the fraction f along $\mathbf{r}_{mn}$. This way the out of plane bend can be interpreted as a case of the bond-bending energy. The coordinates of the point $\boldsymbol{i}$ can be obtained by the two equations:

$$\mathbf{r}_i = (1-f)\boldsymbol{r}_m + f\mathbf{r}_n \tag{2.62}$$

$$f = \frac{\mathbf{r}_{mj}\hat{\boldsymbol{v}}}{\mathbf{r}_{mj}\hat{\boldsymbol{v}} + \mathbf{r}_{jn}\hat{\boldsymbol{v}}} \tag{2.63}$$

Face Planarity

Carbon rings tend to stay within one plane. Therefore, each face can be described by a mean plane and it would make sense for an MD simulation for fullerene molecules to include an additional energy term for this face planarity. This would result in a restoring force acting on each atom to be pushed in the same plane. To find the plane which described the face best, the normal vector of the mean plane has to be found under the restriction that it minimises the orthogonal distance from the plane to each point within the face. The equation which has to be satisfied if all vertices $\mathbf{r}_i$ of a face lie within a plane with normal vector $\boldsymbol{n}$:

$$\boldsymbol{n} \cdot \mathbf{r}_i = d \tag{2.64}$$

and if the vertices are transformed to the centre of mass $\mathbf{r}_{CM}$, the scalar d becomes zero:

$$\boldsymbol{n} \cdot (\mathbf{r}_i - \mathbf{r}_{CM}) = 0. \tag{2.65}$$

For each point i in the face, the squared orthogonal distance form the plane to the point is given by:

$$\mathbf{r}_\perp = \frac{|\boldsymbol{n} \cdot (\mathbf{r}_i - \mathbf{r}_{CM})|^2}{\boldsymbol{n} \cdot \boldsymbol{n}} \tag{2.66}$$

If the matrix M is defined as all the coordinates of all the faces vertices, centred to the center of mass $r_i - r_{CM}$ the system of equations can be written as matrices:

$$\mathbf{r}_\perp = \frac{\boldsymbol{n}^T(\boldsymbol{M}^T\boldsymbol{M})\boldsymbol{n}}{\boldsymbol{n}^T\boldsymbol{n}}. \tag{2.67}$$

Defining $A = M^T M$ leads to:

$$\mathbf{r}_\perp = \frac{\boldsymbol{n}^T\boldsymbol{A}\boldsymbol{n}}{\boldsymbol{n}^T\boldsymbol{n}}, \tag{2.68}$$

which is an eigenvalue problem and the solution is given by the eigenvector corresponding to the smallest eigenvalue. The parameter d can be afterwards obtained by projecting the centre of mass onto the normal vector, leading to $d = r_{CM} \cdot n_0$. The smallest corresponding eigenvalue is a good measure for how planar a face is. In case of a perfectly planar face, where all vertices lie within a single plane, the eigenvalue will be zero.

2.7.2 Gradient of the Energy

To describe the forces acting on the system, the definition of the force $\mathbf{F} = -\nabla\mathbf{V}$ as the negative gradient of the potential is used. If the potential energy of the system is described as in equation 2.53 as a sum of individual contributions, the derivative of the potential energy is simplified also to the form of a summation over the different gradients, and the force acting on each atom, is simply the sum over all contributions, from bonds, angles, out of plane etc. Therefore the individual gradients of those potential energy contributions are derived in the following.

When using any harmonic potential, the gradient can be split up into the functional derivative and the derivative of the inner function, essentially the chain rule. The functional derivative of a potential of the form $\mathbf{V} = \frac{1}{2}(x_0 - x)^2$ is $(x_0 - x)$ while the inner derivative is the derivative of the quantity x itself. Therefore, for harmonic potentials, only the inner derivatives will be derived.

Bond Stretching Gradient

The bond stretching between points A and B is a function of the distance between two points. If this distance is derived by one of the two points coordinates, this results in:

$$\nabla_{\mathbf{r}_A}\mathbf{r}_{AB} = \nabla_{\mathbf{r}_A}(\mathbf{r}_B - \mathbf{r}_A) = -\mathbb{1}. \tag{2.69}$$

In case of the Morse potential (see eq. 2.56, the functional derivative is given as:

$$\nabla_{\mathbf{r}_A} V_{\text{Morse}} = 2\alpha D_{AB}(e^{-\alpha(|\boldsymbol{r}_{AB}|-l_{AB})} - e^{-2\alpha(\bar{r}_{AB}-\tilde{r}_{AB})}), \tag{2.70}$$

which has to be multiplied with $\nabla_{\boldsymbol{r}_A}\boldsymbol{r}_{AB}$ to obtain the complete gradient.

Angular Bending Gradient

The cosine of the angle for an angular spring system s as shown in figure 2.15b is $\cos(\theta) = \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}}$. With the two unit vectors $\hat{\boldsymbol{v}} = \frac{\vec{\boldsymbol{v}}}{|\vec{\boldsymbol{v}}|}$ and $\hat{\boldsymbol{u}} = \frac{\vec{\boldsymbol{u}}}{|\vec{\boldsymbol{u}}|}$. The gradient of a unit vector can be derived as:

$$\begin{aligned}\nabla_{\mathbf{r}_A}\frac{\mathbf{r}_{AB}}{|\mathbf{r}_{AB}|} =& \frac{1}{|\mathbf{r}_{AB}|}\nabla_{\mathbf{r}_A}\mathbf{r}_{AB} + \mathbf{r}_{AB}\nabla_{\mathbf{r}_A}\left(|\mathbf{r}_{AB}|^{-1}\right) \\ & \frac{1}{|\mathbf{r}_{AB}|}\left(\mathbf{r}_{\hat{AB}} \otimes \mathbf{r}_{\hat{AB}} - \mathbb{1}\right) = -\nabla_{\mathbf{r}_B}\mathbf{r}_{\hat{AB}},\end{aligned} \tag{2.71}$$

with using the identity matrix and the tensor product. The angle θ between the three points a, i and b can be defined via the cosine similarity $\cos(\theta) = \hat{\boldsymbol{u}} \cdot \hat{\boldsymbol{v}}$. The derivatives of the cosine regarding the three points involved are given by:

$$\begin{aligned}\nabla_{\mathbf{r}_a}\cos\theta &= \hat{\boldsymbol{v}} \cdot \nabla_{\mathbf{r}_a}\hat{\boldsymbol{u}} + \cancelto{0}{\hat{\boldsymbol{u}} \cdot \nabla_{\mathbf{r}_a}\hat{\boldsymbol{v}}} \\ &= \frac{1}{|\boldsymbol{u}|}\left(\hat{\boldsymbol{v}} - \hat{\boldsymbol{u}} \cdot \cos\theta\right)\end{aligned} \tag{2.72}$$

$$\begin{aligned}\nabla_{\mathbf{r}_b}\cos\theta &= \cancelto{0}{\hat{\boldsymbol{v}} \cdot \nabla_{\mathbf{r}_b}\hat{\boldsymbol{u}}} + \hat{\boldsymbol{u}} \cdot \nabla_{\mathbf{r}_b}\hat{\boldsymbol{v}} + \\ &= \frac{1}{|\boldsymbol{v}|}\left(\hat{\boldsymbol{u}} - \hat{\boldsymbol{v}} \cdot \cos\theta\right)\end{aligned} \tag{2.73}$$

$$\begin{aligned}\nabla_{\mathbf{r}_i} \cos\theta &= \hat{\boldsymbol{v}} \cdot \nabla_{\mathbf{r}_i} \hat{\boldsymbol{u}} + \hat{\boldsymbol{u}} \cdot \nabla_{\mathbf{r}_i} \hat{\boldsymbol{v}} + \\ &= -\left(\nabla_{\mathbf{r}_a} \cos\theta + \nabla_{\mathbf{r}_b} \cos\theta\right)\end{aligned} \tag{2.74}$$

Out of Plane Gradient

The out of plane gradient is a special case of an angular potential gradient with a virtual atom. The gradient regarding the points $\boldsymbol{j}$ and $\boldsymbol{k}$ are identical to the one in the case of an angular spring. The gradient in regard to the virtual atom $\boldsymbol{i}$ (compare fig. 2.16) is itself a function of the points $\boldsymbol{m}$ and $\boldsymbol{n}$. The partial derivative of the out of plane potential for the points $\boldsymbol{m}$ and $\boldsymbol{n}$ are given by:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{m}} &= \frac{\partial}{\partial \boldsymbol{i}} \frac{\partial \boldsymbol{i}}{\partial \boldsymbol{m}} \\ \frac{\partial}{\partial \boldsymbol{n}} &= \frac{\partial}{\partial \boldsymbol{i}} \frac{\partial \boldsymbol{i}}{\partial \boldsymbol{n}}.\end{aligned} \tag{2.75}$$

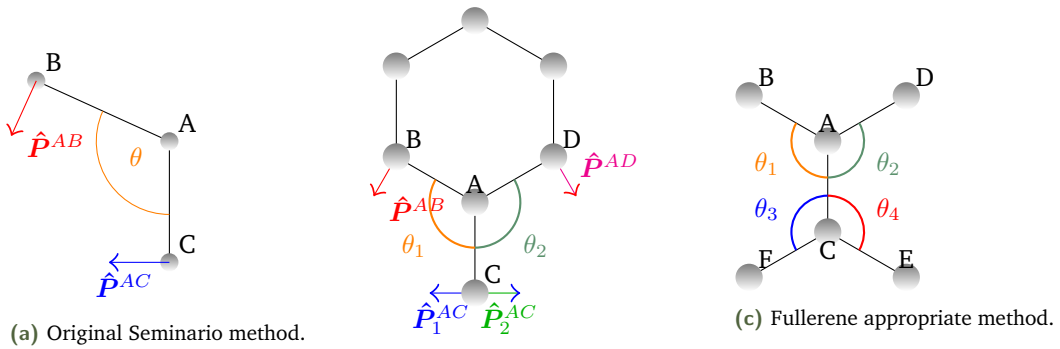
With $\frac{\partial \boldsymbol{i}}{\partial \boldsymbol{m}} = 1 - f$ and $\frac{\partial \boldsymbol{i}}{\partial \boldsymbol{n}} = f$.

2.7.3 Obtaining Force Constants

If a fast FF method is to be applied for the assessment of precursor quality, the geometries and energies associated with deformation of the precursor should be as accurate as possible. To obtain the force constants for the individual potential contributions, the Hessian matrix of the molecule has to be calculated with an adequate method. The Hessian matrix is the matrix of all second-order derivative combinations of a function (see eq. 2.76). In the case of a molecule, this could be the energy of the system as a function of the nuclear coordinates. The coordinates x_1 through x_n are the $3N$ Cartesian coordinates of the N atoms in the molecule. The contribution $\frac{\partial^2 f}{\partial x_i \partial x_j}$ can be interpreted as how the energy changes as a function of x_j when the system is changed a small step in the x_i direction. Therefore, it describes the change of the energy as a function of x_j under perturbation of the x_i coordinate.

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{2.76}$$

The change of the energy as a function of x_j, is equivalent to the force acting on the system in the direction of x_j, via the definition of the force via the gradient of the energy. Hence, if the gradient of the energy for a perturbation in x_i is considered, this describes how the system would react to minimise the energy after a change in x_i. The Seminario-Method [58] offers a way to relate the Sub-Hessian of a system to force constants, which are typical for a FF parameterisation (bonds, angles, out-of-plane, ...). A refined Seminario-Method was introduced in 2018, which offers a more accurate method to obtain some of the constants [59].



(a) Original Seminario method.

(b) Augmented Seminario method.

(c) Fullerene appropriate method.

Figure 2.17.: Illustration of the different methods for angular force constant calculation.

The original Seminario method uses the sub-Hessian matrix of two atoms A, B in a chemical bond (see equation 2.77). The eigenvalues and eigenvectors of the submatrix are the magnitude and direction of the restoring forces acting on the atom A, when moving the respective partner atom B. The Hessian matrix is computed in the local minimum and therefore the first-order term of the Taylor expansion of the potential energy will disappear. Therefore, the potential energy regarding the bond length between atoms A and B takes the form of a harmonic potential in the vicinity of the local minimum. How atoms A reacts to a displacement of atom B is described by the derivative of the system regarding the two points (equation 2.78). However, the sub-Hessian matrix is obtained from the numerical derivative of the quantum mechanical energy function. Therefore, it will contain all force contributions acting on atom A when atom B is displaced. This can not simply be interpreted as describing the bond stretching force constants, because displacement of atom B will also change all angles in which the atoms A and B are involved, as well as the out-of-plane angles and so on.

$$H_{AB} = \begin{bmatrix} \frac{\partial^2}{\partial x_A \partial x_B} & \frac{\partial^2}{\partial x_A \partial y_B} & \frac{\partial^2}{\partial x_A \partial z_B} \\ \frac{\partial^2}{\partial y_A \partial x_B} & \frac{\partial^2}{\partial y_A \partial y_B} & \frac{\partial^2}{\partial y_A \partial z_B} \\ \frac{\partial^2}{\partial z_A \partial x_B} & \frac{\partial^2}{\partial z_A \partial y_B} & \frac{\partial^2}{\partial z_A \partial z_B} \end{bmatrix} V \tag{2.77}$$

$$\begin{aligned} V_{AB} &= \frac{1}{2} H_{AB} \left(\boldsymbol{r}_{AB} - \tilde{\boldsymbol{r}}_{AB} \right)^2 \\ k_{AB}^{\text{tot}} &= -\frac{\partial^2}{\partial \boldsymbol{r}_A \partial \boldsymbol{r}_B} V_{AB} = -H_{AB} \end{aligned} \tag{2.78}$$

The contribution of the force in the direction of the A-B bond has to be found. This can be done by projecting the Hessian matrix onto the A-B unit vector. This can be achieved by scaling the projection of all eigenvectors onto the bond direction with the corresponding eigenvalue. Assuming that the three eigenvalues and eigenvectors of the sub-Hessian matrix are $\lambda_i^{AB}, \hat{\boldsymbol{\nu}}_i^{AB}$, the total force constant for a bond in the Seminario method is given by:

$$k_{AB} = \sum_i^3 \lambda_i^{AB} |\hat{\boldsymbol{u}}^{AB} \cdot \hat{\boldsymbol{\nu}}_i^{AB}|. \tag{2.79}$$

Loosely speaking, this is the sum of the magnitudes of all normal modes in the direction of the A-B bond. A similar procedure leads to the calculations of the angular force constants. Here the eigenvectors are

not projected onto the bond direction, but on the direction perpendicular to the two bond directions, but within the plane defined by the three atoms, which form the angle (see figure 2.17a). The reason for this is that the angular force will act in this direction. The perpendicular vectors $\hat{\boldsymbol{P}}^{AB}$ and $\hat{\boldsymbol{P}}^{AC}$ are projected onto the two pairwise sub-Hessian matrices H_{AB} and H_{CB}, and the combined force constant in the Seminario method is calculated by treating it as a system of two springs connected in series. This leads to an angular force constant in the form of:

$$\frac{1}{k_\theta} = \frac{1}{R_{AB}^2 \sum_i^3 \lambda_i^{AB} |\hat{\boldsymbol{P}}^{AB} \cdot \hat{\boldsymbol{\nu}}_i^{AB}|} + \frac{1}{R_{AC}^2 \sum_i^3 \lambda_i^{AC} |\hat{\boldsymbol{P}}^{AC} \cdot \hat{\boldsymbol{\nu}}_i^{CB}|} \tag{2.80}$$

The method proposed in [59] uses a slightly different approach for the calculation of the angular force constants, which also takes into account the coupling of multiple angles applied to the same atoms. This method is required in a case as one depicted in figure 2.17b, where there are two individual bond angles. In such a case, the displacement of atom C will not only affect one single angle but will result in a change in both angles and the system has to be treated as the deformation of two coupled angular springs. Although this has been proven to improve vibrational frequency results in the cases, where a single atom is bound to a carbon ring (being part to two angular springs), the method does not take into consideration the cases in which an atom is fully surrounded and therefore part of four angular springs as it is in fullerene-like molecules (see figure 2.17c). In the case of the periphery angular springs, this will not lead to any miscalculations because this is the special case of a single atom attached to a carbon ring.

Such a method would provide a systematic way to calculate the force constants directly from the Cartesian Hessian matrix and the bond information. The harmonic force constants can be obtained at any level of theory, which offers the calculation of a Hessian matrix, essentially any Quantum Chemistry method (CC, DFT, HF), making the results comparable across a wide variety of theories. The equilibrium values of the potentials, namely the bond lengths, bond angles and out of plane distances, can be obtained from the optimal geometry itself.

2.7.4 Validity of the Potential Approximations

As explained above, the FF parameterisation is essential a Taylor expansion of the energy in an internal coordinate system. These coordinates are the bond lengths, angles and out-of-plane angles etc. The Taylor expansion is performed at the point of optimal geometry, which represents a local energy minimum. Therefore, the first-order terms disappear and the second-order terms yield an accurate description of the local minimum. However, as the system moves further and further away from this local minimum, the expansion around the original minimum will not be sufficient to describe the energy surface anymore. If all the higher-order derivatives of the energy function where know, the terms could be included to ensure a good approximation further and further away from the x_0. However, the big advantage of the description of the system as a harmonic potential around the minimum is the simplicity of the potential and that only the second-order derivatives are required. Therefore, the harmonic approximation should be used whenever adequate to describe an internal potential.

As mentioned above the carbon-carbon bonds of any fullerene precursor molecule should not be broken during the auto-assembly process and the bond lengths for these bonds can be expected to not leave their local minimum. Although these bonds are required to stay in the potential well, special attention should still be paid to the theoretical calculations of the precursor molecule. If at any point along the closing path the bond lengths between the carbon atoms should indicate, that the corresponding optimal

values changes, this could be added to the MD simulation in the form of a potential functional, in which the optimal value for the carbon-carbon bond itself is described as a function of how closed the fullerene is. The same reasoning inspires a thorough investigation of the bond angles along the closing path.

For the out-of-plane values, the situation is different. By using planar precursor molecules the initial state of the molecule has inherently out-of-plane angles of $180°$, this does not hold for the final product. Due to the geometric properties of fullerene molecules, the out-of-plane angles in the closed case is not planar. Rather will the out-of-plane angles move from one local minimum to the next and the description of the energy contribution as a harmonic potential will be wrong. An out-of-plane study for the optimal intermediate geometries will be necessary to understand how these angles behave along the way. In the end, after all, other potential forms and parameter values have been performed, it is also possible to compare the obtained FF energy of the geometries along the path to the one from the DFT calculations and fit a potential for the out-of-plane angle contribution.

In summary, FF methods offer a very fast and simple way to determine molecular energies and simulate molecular dynamics. They provide an intuitive understanding of molecular forces in internal coordinates and can be used to perform many calculations for numerous different molecules. By making use of calculations performed with ab initio methods it is possible to obtain parameters, like bond lengths and angles, which can be transferred between systems. All these qualities make FF methods and especially molecular dynamics a natural choice for the automated assessment of fullerene precursor molecules.

Method

3

To generate the initial geometries for the fullerene closing path, the Minimal Spanning Tree is needed. The method by which the MST was obtained from the graph and its dual is described below. To evaluate the quality of each functional and basis set, a CCSD calculation was performed on a smaller system, with similar properties to the precursor molecules. The setup for these calculations will be described in this section. After the functionals and basis sets, which lead to the best results on the small system have been identified, the C_{60} precursor molecule was chosen and its geometry optimised along a suspected up folding path, to find the geometries, along the way, which correspond to the minimal energy.

3.1 Obtaining Relevant Graph Information

The spanning tree for the fullerene unfolding was calculated in a breadth-first approach. The reason, why this was preferred over the depth-first approach is, that the hinges were required to be as close to the root face as possible. The data of the unfolding was represented as a list of lists. For the C_{60} unfolding, this results in a list of 60 lists, where each of those lists has three or fewer entries stating which atoms is the neighbour or the atom. When looking only at the carbon atoms in figure 3.3 this becomes clear. Atom C0 has the neighbours 1, 5 and 6, so the first list (neighbours of atom C0) will contain those numbers indicating the neighbours given in clockwise order. Atom C28 instead has only two neighbours, so the 28th list of the graph will look like $[11,27]$. The graph of the molecule will include the periphery atoms (halogens and hydrogens) and will for each carbon atom have exactly three neighbours so their octet is fulfilled. The dual graph is represented in the same way. The dual of the carbon unfolding and the precursor molecule will be the same because the periphery atoms do not constitute a face and will therefore not introduce any new vertices or edges to the dual graph.
The algorithm for creating the spanning-tree looks like this:

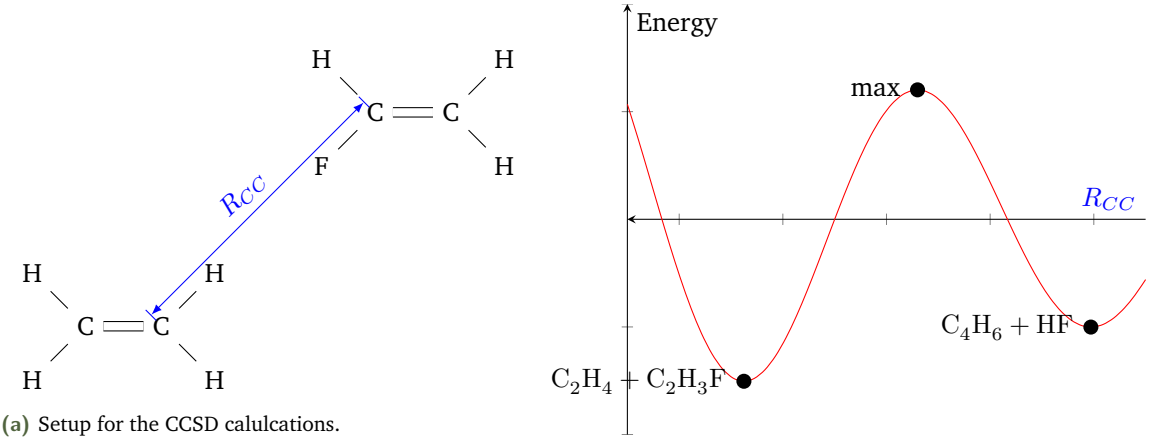
```
def spanning_tree(dual_graph):
    queue = []
    visited = [False] * len(dual_graph)
    queue.append(root)
    tree = [ [] for i in range(len(dual_graph)) ]
    visited[root] = True
    while queue:
            # Take a vertex from the queue
            s = queue.pop(0)
            # Get all adjacent vertices of the dequeued vertex s. If a adjacent
            # vertex has not been visited, then mark it visited and queue it
            for i in dual_graph[s]:
                if visited[i] == False:
                    tree[s].append(i)
                    queue.append(i)
                    visited[i] = True
    return tree
```

This will return a directed spanning tree of the dual graph, which for each face contains all the immediate child faces. Having obtained the tree, the hinges can now be defined. Each hinge can be represented

as two tuples, the first tuple contains the two faces which are connected, so the vertices of the dual graph, in the order: parent face - child face. This information is until now purely graph theoretical and to define the rotational axis from section 2.5 they have to be augmented with vertex information from the original graph. Therefore, for each hinge a second tuple is defined by using the two vertices (carbon atoms) which the connected faces share. These axes should be given in a consistent order, which means they should all point in a direction, such that a rotation around them for a given angle will result in all of them, to either open or close the fullerene. The vertices of the faces are given in clockwise order seen from above and the best way to obtain a consistent orientation for the axes is to let them point in the clockwise direction of the parent face. Each hinge is therefore represented as a cross-like structure with the dual edge crossing over the edge. For a given tuple of parent and child face i,j and the corresponding face lists in clockwise orientation an algorithm could look like this:

```
def hinge_axis(i,j,faces):
    hinge = []
    flip = False
    found_first_vertex = False
    # Iterate through all vertices of the parent face i
    for u in faces[i]:
        found_vertex = False
        # Iterate through all vertices of the child face j
        for v in faces[j]:
            # If parent and child face share the vertex, add it to the hinge
            # axis and set found_first_vertex and found_vertex to True
            if u == v:
                hinge.append(u)
                found_first_vertex = True
                found_vertex = True
                break
        # Is the axis complete ?
        if len(hinge) == 2:
                break
       # If the parents face vertex is not found in the child face
        if found_vertex == False:
            # But the first vertex has been found
            if found_first_vertex == True:
                # Vertices in between have been skipped and the axis is not in
    clockwise order -> flip it!
                flip = True
    if flip == True:
        return hinge[::-1]
    else:
        return hinge
```

By assigning an axis tuple to each hinge, the hinges are now complete and contain all information needed to rotate the child-faces. Each hinge represents the change of the angle between two faces, while all other parts of the molecule are treated as a rigid body. To efficiently apply these rotations to the entire molecule, it is useful to define all vertices of the graph, which will be affected by a hinge. This way, the rotation matrix can be calculated once for the required angle, all affected vertices can be expressed in the local frame of reference (relative to the hinge axis) and these connecting vectors rotated. For a single hinge, this can be done in a very simple way, by starting at the child face of a hinge and exhaustively (following all paths) go through the tree. For each face in the tree, all corresponding vertices of the graph are appended to the list.



(a) Setup for the CCSD calulcations.

(b) Schematic illustration of a possible reaction path.

Figure 3.1.: Reaction for the C_2H_4 and C_2H_3F molecules.

As explained in section 2.5.3 after this list was created, the vertices which appear twice in the tree for hinges, which have the same distance from the root face have to removed from the affected vertices of one of them. This can be easily achieved by looking at all hinge levels, which are defined as the number of steps they are away from the root face and remove the duplicates within them unless they constitute the axis of that hinge.

3.2 Comparison of Different Functionals

As described in section 2.6.4, the choice of functional for a given system is not trivial. There is no one best functional, which is superior under all circumstances. Different functionals excel under different conditions. All functionals are approximations and the different description of exchange and correlation energies make some functionals better suited for a problem than others. If computational resources were unlimited, the best way to find the best fitting functional for a given system is to compare all of them with a CC calculation and pick the one, which leads to results closest to the CC calculation. However, resources are limited, and a CC calculation of a system like the C_{60} precursor molecule, which is the smallest fullerene precursor found so far, is not even feasible in a CC calculation. Therefore, a smaller system had to be chosen to compare the results of different functionals. The small system, however, should, to a large extent, reproduce the kind of chemical properties, that are present in the full precursor molecule.

The main feature of the auto-assembling C_{60} precursor was the halogen assisted cyclo-dehydrogenation (HAD), which is the name of the process, in which the hydrogen atoms, which are attached to the carbon-cycles, are removed from the precursor molecule and form a bond with a halogen atom. There are also bonds formed between two hydrogen atoms, but the main focus of this work was set on the halogen-hydrogen interaction because, as was shown by the works of L. H. Scott, the addition of halogen atoms greatly improves the yields of the auto-assembly reaction. So the main process of interest is the formation of a halogen-hydrogen bond, more specifically the formation of a bond between a halogen and hydrogen, which are both in a bond with an sp^2 hybridised carbon atom. Such carbon bonds are often interpreted as a mixture between the lower energy s and two p orbitals, which lie higher in energy, resulting in the three sp^2 hybridised orbitals. Those orbitals will form σ bonds (electron density highest between the atoms), while the last remaining "unmixed" p orbital will form a π bond (electron density

highest left and right of the bond axis). So, to reproduce this kind of carbon-hydrogen/halogen bond breaking, followed by a halogen-hydrogen bond formation, two sp^2 hybridised carbon centres, with an attached halogen and hydrogen are required. One of the simplest and smallest molecules containing a sp^2 hybridised carbon atom is ethene C_2H_4. To simulate the bond formation between a hydrogen and a halogen atom, another ethene has to be added to the setup, where one of the hydrogen atoms has been replaced by a halogen atom. For the CCSD calculations, fluorine has been chosen, due to the lower number of electrons, which makes the calculations faster. The goal of these simulations was to find the optimal geometries for such a system and find the point at which the formation of the halogen halide happens, first with a CCSD calculation to have a reference value and then with different functionals to compare the results.

The functionals which have been chosen to compare to the CC results are the well known and widely used B3LYP functional. This is a functional which is comparatively cheap to compute and it is possible to do multiple calculations in a short time. The second functional which was chosen was the CAM-B3LYP functional proposed by Handy et al., which adds some long-range corrections to the B3LYP functional [60]. It is more expensive than the pure B3LYP functional but it has shown to improve the results especially for charge-transfer excitations. The third functional which was chosen for comparison was the M062X functional developed by Truhlar's research group at the University of Minnesota. The M062X functional has a strong focus on non-local exchange and its parameters are optimised only for the non-metal elements of the periodic table. Although the M062X functional, like the other two functionals, does not include an explicit description of dispersion terms, it was optimised for results including dispersion, which means that its parameters can be assumed to have a small contribution towards the dispersion corrections. All three functionals were studied with and without the addition of explicit Grimme D3 dispersion.

The calculations were performed with the GAUSSIAN G16 software [12], by performing a geometry optimisation. This geometry optimisation finds the optimal geometry for a molecule, given an input geometry. This optimisation optimises the geometry to the closest local minimum in the potential energy surface because the global minimum is the same for any system of atoms. The idea was, to move the ethene and the modified ethene (with hydrogen replaced by fluorine) closer and closer together and determine the point at which the CCSD calculation results in the formation of the halogen-halide. At this point, it is energetically favourable for the fluorine and hydrogen to form a molecule themselves and the carbon and carbon to form a bond. The initial geometric setup for the calculations was obtained by first optimising the ethene molecule and the C_2H_3F with the CCSD method, and then placing them at a certain carbon-carbon distance as shown in figure 3.1a. If an optimisation without any restrictions would have been performed for such an initial configuration, the repulsion of the bonded H and bonded F atoms would place the two individual molecules further apart, at the distance corresponding to the minimal energy.

To hold the CC distance at a constant value, and look at how the rest of the molecule behaves under those restrictions, the calculations were performed with some coordinates frozen. This means, that the software does not optimise those nuclei's coordinates, but all the atoms which are not frozen can be moved in the optimisation process. This is a very helpful tool for this kind of calculations. The C_{60} auto-assembly is activated by FVP, which means, that the laser pulse provides the activation energy needed to overcome the energy barrier at the chemical transition state. If a normal optimisation was

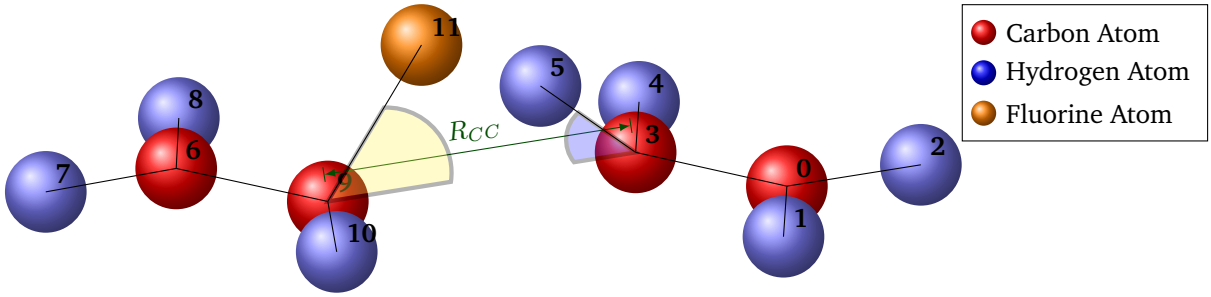


Figure 3.2.: Setup for the CCSD calulculations.

to be performed, before the top of the hill was reached the system would just be optimised towards the closest local minimum, following the downhill direction, and all calculation would end up having the same geometry. If it is assumed, that the reaction path for the halogen capture, followed by the carbon-carbon bond formation, looks somewhat like the graph in figure 3.1b, where the x-axis is the carbon-carbon distance R_{CC}, then every point between the origin and the maximum would be optimised to the local minimum.

To determine, how accurately the different functionals represent the energy of this particular system, all the coordinates of the carbon atoms were frozen. This is a reasonable thing to do because the C-C distance for the two existing double bonds within the C_2H_4 and C_2H_3F are very close to the distances in the resulting butane molecule, and the bond angle is also very close to the H-C-C angle. Therefore, for the correct carbon-carbon distance, all the carbon atoms within the butane molecule would be in the right position. This means, that by fixing the carbon atoms along the path, the hydrogen and fluorine atom are now placed outside of the otherwise planar molecule due to the orbital overlap. This angle will increase, the closer the two carbon atoms are, until, at a certain distance, there will be an orbital overlap between the carbon atoms as well as for the H and F atoms and they will form a bond respectively. By scanning along the carbon-carbon distance, those distance and the intermediate angles can be calculated, as well as the activation barrier, and the results can be compared to DFT calculation, to see which functional yields the results, closest to the CCSD calculations. This line search offers a way to assess the functional accuracy, in the case of a halogen-hydrogen capture attached to sp^2 hybridised carbon atoms.

To speed up the calculations and avoid small interatomic distances, the H and F atoms have been rotated outside of the plane, which is formed by the four carbon atoms at an angle as a function of the C-C distance, obtained from previous work of the author. Figure 3.2 shows how those two atoms haven been rotated out of the plane as a function of the distance R_{CC}. The calculations were performed for carbon-carbon distances between 1.2 and 2.49Å in steps of 0.01Å to offer a high spatial resolution for the functional comparison.

3.2.1 Comparison of Geometric Properties

As mentioned before the optimal values and force constants for any FF method, should, if not experimentally accessible, be calculated from an ab initio method or a method like DFT. The best method available for systems of fullerene size is DFT and the parameters should be determined by such a method. To pick the functional, which best reproduces the geometrical properties of molecules, which are similar to fullerene precursor, the optimal geometry results for the different functionals have been compared, to

those of the CCSD calculations.

As the C9 and C3 (figure 3.2) atom come closer and closer together, the H and F atoms which are bound to them will start repelling each other. The bond lengths can be expected to change during this process. If the two carbon atoms are close enough, the repulsion between the H - F and H - H, will be strong enough to force the atoms out of the plane, causing out-of-plane angles which are different from the usual flat $180°$ for sp^2 hybridised carbon centres. Any FF method used for the quality assessment of fullerene should be able to reproduce this kind of behaviour withing reasonable uncertainties. Thus, the functional which is used to calculate the precursor optimal geometries along the path should yield geometric properties similar to those of the CCSD. To compare the geometric results of different functionals, some key parameters (distance, angles and out-of-plane angles) have been chosen and calculated from the final geometries along the path.

3.2.2 Comparison of Reaction Enthalpy

Comparing the results for energy calculations of different functionals is impossible. Each functional uses a different approach to represent the exchange and correlation energy and this will lead to different absolute energy values. The same holds for a CC calculation. Although, all of them do describe the relative energy of the system in a similar way, which lead to geometric minima at the same internal coordinates, the actual energy values for these minima can not be compared. Nevertheless, the energies of a system are an extremely important quantity and if the energy barriers of fullerene precursors along a given path are to be studied by a DFT method, the functional with the highest accuracy should be chosen. As the absolute energies do not have a meaning in this comparative context, it is necessary, to come up with a quantity, which does not depend on the exact description of the system's energy. Such a quantity is the standard enthalpy of a chemical reaction. The enthalpy is defined as the energy of the product of a reaction minus the energy of the reactants. It was chosen as a comparative quantity over the activation barrier because the calculation of the activation barrier requires the calculation of the energy of the transition state. The transition state is generally not well described by a single Slater determinant but it a superposition of multiple determinants. Therefore, a method describing the system as a single determinant will result in a very poor description of the energy and the more stable quantity of the reaction enthalpy was chosen for functional performance. The two different states representing the reactants and the products of the reaction are, firstly, the two molecules placed at the maximal studied distance, and secondly, at the first distance, where the HF molecule has been formed.

3.3 DFT Study of Optimal C_{60}-Precursor Geometry

The planar C_{60} precursor molecule which was used by L. T. Scott et al is displayed in figure 3.3, where the red lines show the bounds, which are formed during the auto-assembly. Each of those red lines is associated with the formation of a HCl or H_2 molecule. As one of the main goals of this work is to identify, what the auto-assembly mechanism exactly looks like, the C_{60} precursor molecule was chosen to perform the DFT calculations. This molecule has been proven to selectively auto-assembly to the C_{60} icosahedron under FVP (see section 2.2.2), and thus provides a natural choice for this study.

The "zipping up" process is started by the FVP, meaning that the laser pulse provides the energy needed to overcome the activation barrier and start the auto-assemble process. The laser provides energy, which activates the oscillatory modes of the precursor molecule which distribute in the molecule, forcing it out of the energy minimum, and introducing intra-molecular deformations and vibrations. These

deformations then lead to a halogen assisted hydrogen capture, followed by a carbon-carbon bond formation. This newly introduced carbon-carbon bond then starts a chain reaction, which results in a total dehydrogenation of the precursor molecule and the closing of the carbon cage. Although this is a suspected reaction path, the selectivity of the auto-assembly strongly suggests that this chain reaction is taking place during the "stitching up" [7].

The deformations, associated with the smallest energies, are in general the most likely to occur. Therefore, deformations which push the molecular geometries far from the equilibrium are much less likely than those which occur close to the minimal energy geometry. Furthermore, the selectivity of the precursor suggests that no "wrong" carbon-carbon bonds are formed in the process. That means the only carbon-carbon bonds, which are formed, are the red lines depicted in figure 3.3. This indicates, that the hydrogen captures, which occur first and start the chain reaction are the ones where the "correct" H and Cl molecule form a bond. The most likely candidates for this are the chlorine and hydrogen atoms at the cavities because they are already closest to each other (shortest arcs in fig. 3.3). The inmost cavities are the ones, which are most likely to be closed first. That is, because all the forces which are acting on the outward arms, will exercise a much larger torque on the central part of the precursor due to the lever effect and the stiffness of the arms due to the face stability.

To close the first cavity, the most inwards hinges, adjacent to the root face (orange hexagon) were closed step by step with the procedure described in section 2.5. The vertices have been rotated until the required C-C distance was reached. To place the two carbon atoms close together the three hinges around the root node have to be rotated beyond the optimal value of $142.62°$ of the truncated icosahedron. This could have been achieved, by first rotating the maximal amount around the first hinge, and then rotating about the hinges further down in the MST, but it was decided against it because those rotations would introduce local curvature at the carbon atoms $7, 10$ and 13. For the fluorine-containing precursor, the first three hinges were instead simply bent beyond the optimal value, forcing the carbon atoms closer together.

When the hinges are closed, the hydrogen and halogen atom at the inmost cavity will get close to each other. To avoid large repulsion energies along the closing path, in the optimal geometry they will be places them apart from each other. There are three different possibilities for how this can be done. The two atoms can be placed outside of the bowl, inside of the bowl, or they can be on opposite sites. Because the GAUSSIAN optimisation process optimises towards the closest local minimum, the choice of the initial placement of the two atoms is very important. If both would be placed inside of the geodesic cage, to be flipped outside, both atoms would have to be rotated gradually to the outside of the bowl, because the optimisation will follow the low paths along the energy surface. As the two atoms will both be moved simultaneously, the repulsion will gradually increase. Therefore, it is extremely unlikely that the optimisation will change the inside/outside state of the two atoms. There are (most likely) two local minima in the geometric configuration, one with them being placed inside the bowl and one with them being placed outside. The case where they are placed on opposite sides was not considered in this study because it was suspected, that this will be the least likely to achieve the halogen-hydrogen capture. To calculate the energies and geometries of both local energy minima, two initial geometries were provided, both having the exact same carbon coordinates, but where the two atoms were placed either inside or outside of the geodesic cage.

The geometries were sampled as a function of the carbon carbon-distance between the atoms C29-C25, C35-C15, C21-C22, and the hydrogen and halogen atoms were rotated out of the plane using the same method as in the functional comparison. Figure 3.4 illustrates one of the intermediate initial geometries. The three red lines are the carbon-carbon distance, for which the geometries were sampled in steps of 0.05Å from 1.8Å to 3.08Å. An additional interval from 1.25Å to 1.8Å was calculated for the fluorine case. Using the carbon distance as a quantity to sample is preferable over the angle because the angle will not be easily transferable to larger precursor molecules, while the carbon-carbon distance between two atoms which are expected to form a bond is an intrinsic quantity of any fullerene precursor molecule.

The initial geometry for the flat precursor molecule was obtained from a geometry optimisation performed with the GAUSSIAN software, the CAM-B3LYP functional and the Pople basis set 6-311G++(d,p). The calculations along the suspected auto-assembly path were performed using fluorine and chlorine as a halogen in a wide range of the C-C distance, to compare the two results. Although the initial geometry was calculated with Cl in the precursor, the same initial geometry was used for the precursor containing F, with the only difference being, that the C−Cl vectors have been scaled with the vacuum C−F bond length. This was done, because the initial geometries of the calculations only provide a guess for the optimisation, and the geometry of the precursor molecule can be expected to be mainly dominated by the carbon cage, with the choice of halogen only making a slight difference to the optimal geometry. The functional, which was used, was the B3LYP functional with Grimme dispersion (D3), because the computation times are much lower than for CAM-B3LYP, and made a large number of calculations possible.

The geometry corresponding to the minimum energy of the precursor molecule is the flat structure. This is already known and it does not make sense, to perform a full geometry optimisation at each step, because that can be expected to result in the same flat geometry for every initial setting. The focus of these calculations should provide some insight into how the molecule behaves if the curvature is forcefully introduced. Therefore, some of the nuclear coordinates had to exempt from the optimisation process to ensure the molecule will not simply relax to the planar geometry. However, these restrictions should be kept as small as possible to give the system as much freedom as possible. Therefore, the only carbon coordinates which were frozen during the calculations, are those for which the carbon distance was sampled: (C21, C22, C28, C29, C35, C15).
The study was performed for both different halogens, fluorine and chlorine at their respective cavity position, to study how the choice of halogen influences the bond behaviour.

3.4 Closing the C_{60} Precursor with Atoms Placed Inside of the Bowl

Because no halogen halide formation, followed by a carbon-carbon bond formation was be achieved by closing the innermost hinges the calculations have been repeated, with fluorine as a halogen and a slightly different initial setting. The coordinates of the carbon atoms were left exactly as before, but the fluorine and hydrogen atoms at the inmost cavity were not rotated outside of the closing cage but instead placed on the inside of the bowl. In this case, it was not possible, to close the hinge as much as before because the hydrogen and fluorine atoms get much closer to each other at smaller distances, which will cause the GAUSSIAN software to crash. The geometries were sampled for carbon-carbon distances from 1.6Å to 2.95Å in steps of 0.05Å. Figure 3.5 shows how the initial setting with the halogen

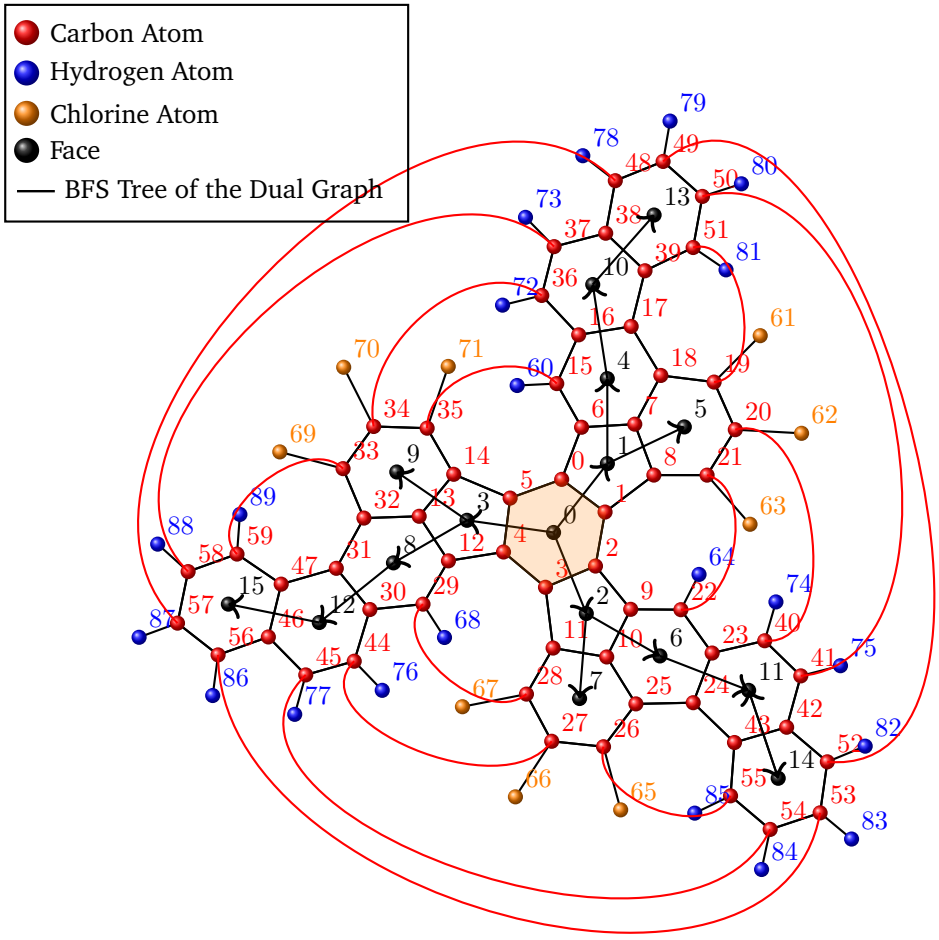


Figure 3.3.: Auto-assembling C60 precursor molecule used by L. T. Scott et al [7].

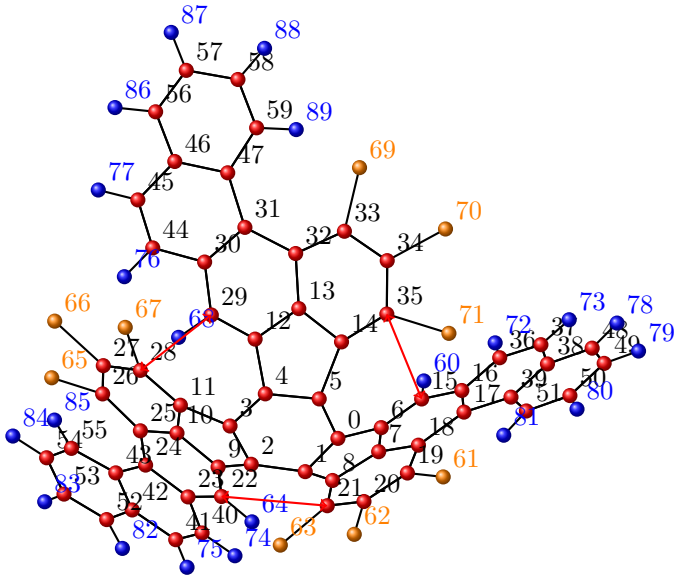


Figure 3.4.: C60 precursor closed at the three hinges, adjacent to the root face.

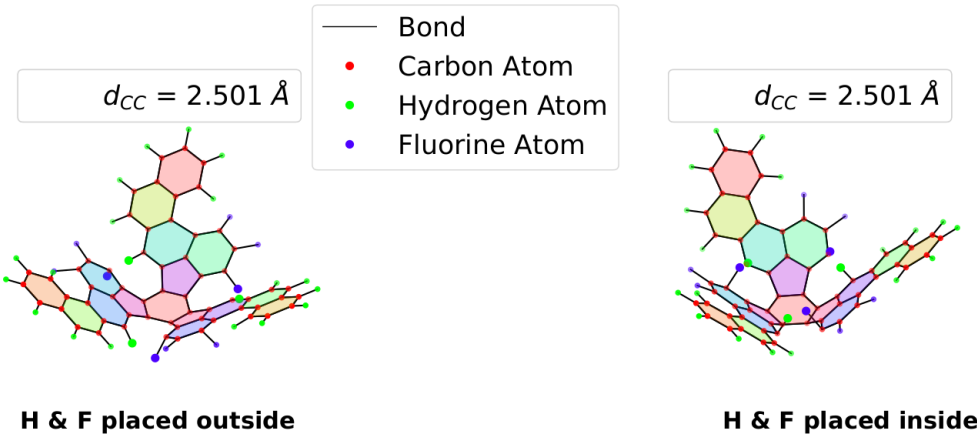


Figure 3.5.: Initial geometry a carbon carbon distance of 2.501Å with the hydrogen and fluorine atoms placed inside and outside of the bowl.

and hydrogen atoms placed on the inside differs from the setting where they have been placed on the outside. The carbon positions for both settings are exactly the same while the fluorine and hydrogen atoms at the inner cavity have been places on opposite sides.

3.5 Finding the Optimal Parameter Values along a Path

To find accurate potential descriptions for MD simulations, it is important to first understand which shape the potentials should have. The different cases for the potentials are the ones explained in section 2.7.1. It could be, that some of the bond lengths change their optimal values along the way, and some of the bonds are entirely broken. To find out how the individual bonds behave along a closing path the optimised geometries along the path have been studied in detail. Although some of the carbon atoms have been frozen and their coordinated are thereby not optimised along the way, that was never done for two carbons which are bound together. Therefore, all neighbouring atoms of a carbon atom were always able to relax into the position corresponding to minimal energy. If along the closing path for the precursor molecule, the optimal bond lengths or angles for certain types of carbon bonds (hexagon-hexagon, hexagon-pentagon,...) should change, this should be visible when systematically analysing the optimal geometry path. Any MD simulation should be considering this change in optimal bond length. Special interest has to be paid to the periphery because this kind of bond lengths can be expected to largely differ from previously performed diffraction studies of the closed C_{60} icosahedron. The pentagon-hexagon and hexagon-hexagon bond lengths are already present in the closed molecule and can be studied with great accuracy, while the periphery bonds can not. It was analysed, whether the periphery bonds can also be put into clear categories as the interior lengths.

The face stability was an assumption made on the general knowledge, that carbon rings have a strong restoring force pushing all atoms in a plane. To verify this assumption along the whole way of the auto assembly, a least-squares plane was fitted to each face of the C_{60} unfolding and the squared error of all the points from that plane analysed to see if at any point along the way there is a large discrepancy between the optimal result and the planar face assumption.

Results and Discussion

4

4.1 Functional Comparison for a Fluorine Assisted Dehydrogenation of an sp^2 Hybridised Carbon Atom

The optimised geometries for the ethene and C_2H_3F molecule were sampled along the distance between the carbon atoms C3 and C9. Figure 4.1 shows the optimal geometry for a distance of $R_{CC} = 2.49$Å. As can be seen, the fluorine atom and the hydrogen atoms arrange themselves in two parallel planes which each contain the carbon spine of their respective molecule. The planes are arranged to each other in such a way, that it ensures the maximal distance between the H5 and F11 and the H4 and H10 atoms. This is exactly what was to be expected when freezing all the carbon atoms. The hydrogen atoms and the fluorine atom are arranged relative to the carbon atoms of their belonging molecules in a way so that the inter-molecular repulsion is minimised.

Figure 4.1 also illustrates the carbon-carbon distance R_{CC} along which the geometries were sampled, as well as the main geometric quantities of interest. The bond stretching is investigated at all bonds the carbon atoms C3 and C9 are part of. Those are the atoms, which are going to form a bond at some point along the way. The bond distance between the carbons is fixed, therefore it has not been included in the analysis. Another quantity which is of interest for the comparison of the geometries obtained by different density functionals are the out-of-plane angles. These calculated as described in equation 2.7.1. Special interest has to be paid to the out-of-plane angle of the fluorine and the hydrogen atom, which will form a molecule with the fluorine atom. The fluorine out-of-plane angle is calculated by the angle between the connecting vector from the carbon atom C9 to the fluorine atom onto the plane formed by the carbon atoms C9, C6 and the hydrogen atom H10. The plane is displayed in figure 4.1 in blue colour and the corresponding out-of-plane angle is drawn as a blue line. The out-of-plane angle and the corresponding plane for the hydrogen atom are shown in dark red colour.

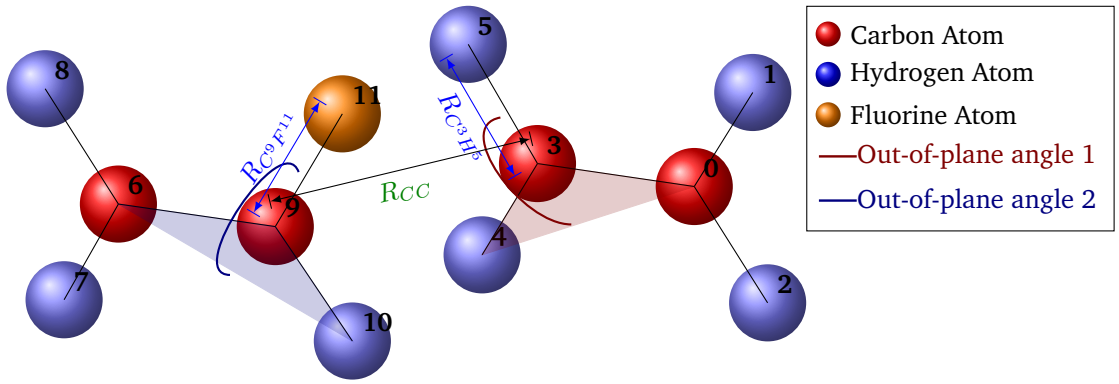


Figure 4.1.: Optimised geometry for the CC calculation at a CC distance of 2.49Å.

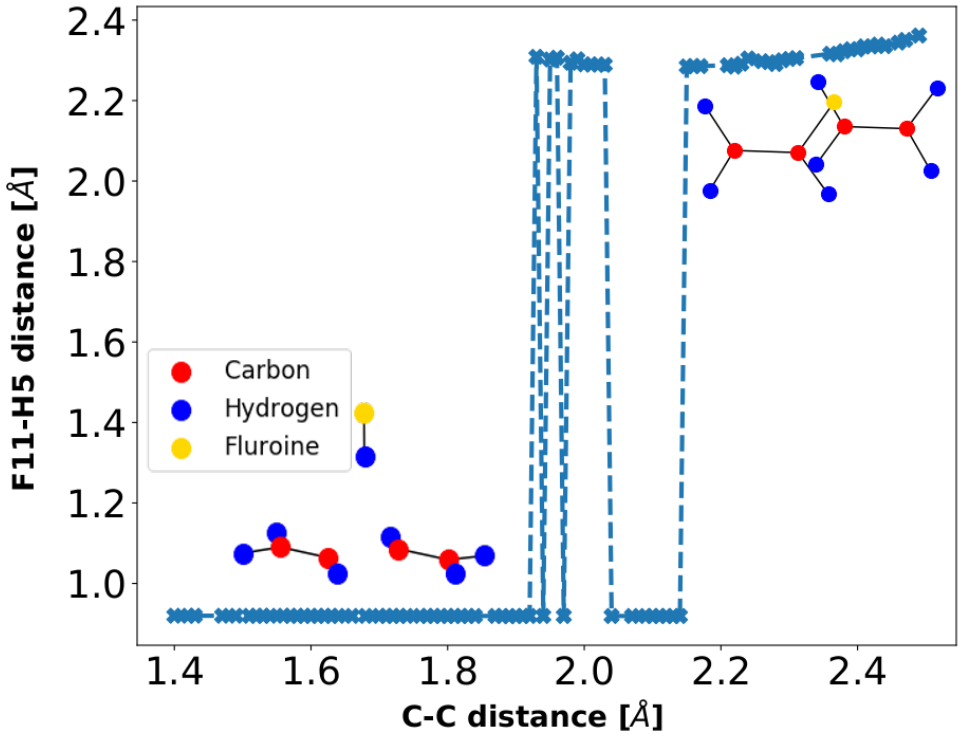


Figure 4.2.: HF distance of the optimised geometry as a function of the CC distance.

4.1.1 Comparison of the Optimal Geometries for CC and DFT

To analyse the results properly it is important to find out at which points at the line search the formation of the HF molecule is energetically favourable. At these points, the hydrogen and fluorine will be removed and placed apart from the rest of the two molecules. The easiest way to inspect at which points this has happened is, to look at the distance between the F11 and the H5 atom. If this distance is close to the bond distance of the HF molecule, the molecule can be expected to be formed. The molecules might be rotated around one of the carbon axes and therefore the distance between the F11 and the H5 molecule is not necessarily the correct distance to compare with, but the F11 might instead form a molecule with the H4 atoms. Therefore, the minimal F11 - H5 / F11 - H4 distance was chosen as an indicator of the HF formation.

Figure 4.2 shows the distance between the F11 and the H5 / H4 atom depending on which one is smaller, for the calculations which terminated normally. The optimal geometries for tow intact initial molecules (top right) and a case where the HF has been removed from the rest (bottom left) are displayed. The graph shows the very clear behaviour, that for large carbon-carbon distances (to the right) both initial molecules are simply rotated to the point of minimal repulsion. The distance between H and F is almost constant until a carbon-carbon distance of 2.15Å. In the region above 2.15Å, the H-F distance remains almost constant. If the carbon atoms move closer to each other, but the H-F distance is constant, there must be out-of-plane bending and bond stretching involved.
This trend of constant H-F distance is broken between carbon distances of 2.14Å and 2.04Å. The small distance between the H and F atoms are because of the formation of the HF molecule. Below this,

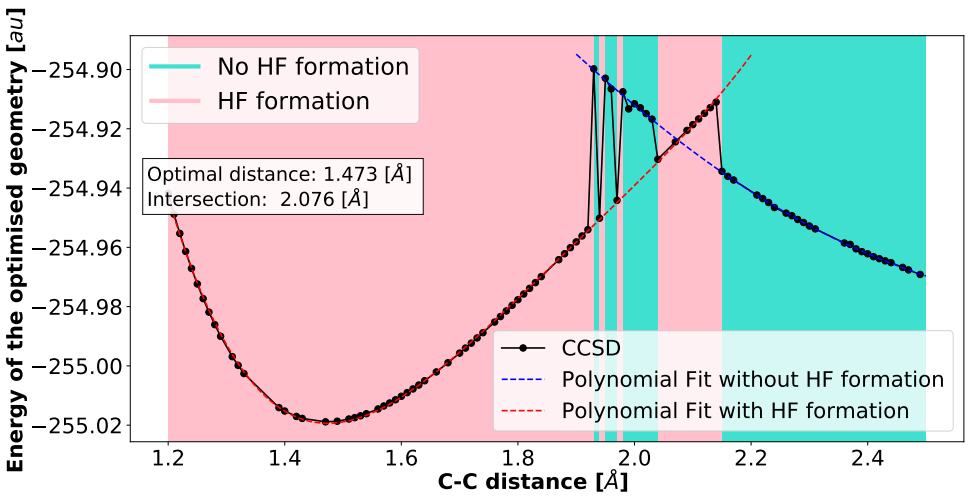


Figure 4.3.: Energies of the optimal geometries of the CC calculations.

there are some points, where the HF is formed and some where it is not. The two cases for the HF distance can be put into two very clear categories. One, where the HF is formed and one where it is not. Therefore, the state of the system can be easily assessed by looking at the HF distance. The threshold for the HF distance below which the HF molecule is expected to have formed was set to $r_{\text{thresh}} = 1.5$Å. This value was more ore less arbitrary because the two categories are so easily distinguishable that any value between the two distance clusters could have been chosen.

To understand why most of the calculations in the range of medium carbon-carbon distances between 1.92Å and 2.04Å do not form the hydrogen fluorine but it is formed at the larger distances between 2.04Å and 2.14Å, it is helpful to look at the energies of the CC calculations. Figure 4.3 shows the energies along the line search path, where the background colour indicates, whether the hydrogen fluorine has been formed or not.
It is clearly visible, that the energy curve can be described as a linear combination of the two cases. One potential for the energies of the molecule where no HF has been formed and one, for the cases where the molecule has not. Both potentials look polynomial, and to find a description of the two potential wells, polynomials have been fitted to the two. A fourth-order polynomial was fitted to part in which the HF molecule is formed, and a quadratic polynomial was used for the energies without HF formation. It is can be expected that at the point where the two lines intersect, the actual turning point is reached. Left of this point, the formation of the hydrogen fluorine is energetically favourable, while on the right the energy of the state with the ethene and C_2H_3F molecule intact has lower energy. The reason why the calculations in some cases terminate in a state of higher energy is, that the optimisation procedure does not find the global minimum of a set of atoms, it simply optimises to the closest local minimum or the local minimum it reaches first during the optimisation. This might be due to bad initial positioning of the atoms. Therefore, a calculation for a single carbon-carbon distance of a system should never be trusted completely but always be regarded in context. However, the collective results of the line search paint a very clear picture of the optimal geometry along the path.

The minimum of the energy well where the HF molecule has been formed (the bottom of the red curve) is at 1.473Å and therefore very close to the reported CC bond length of 1.476 [61]. The two polynomials

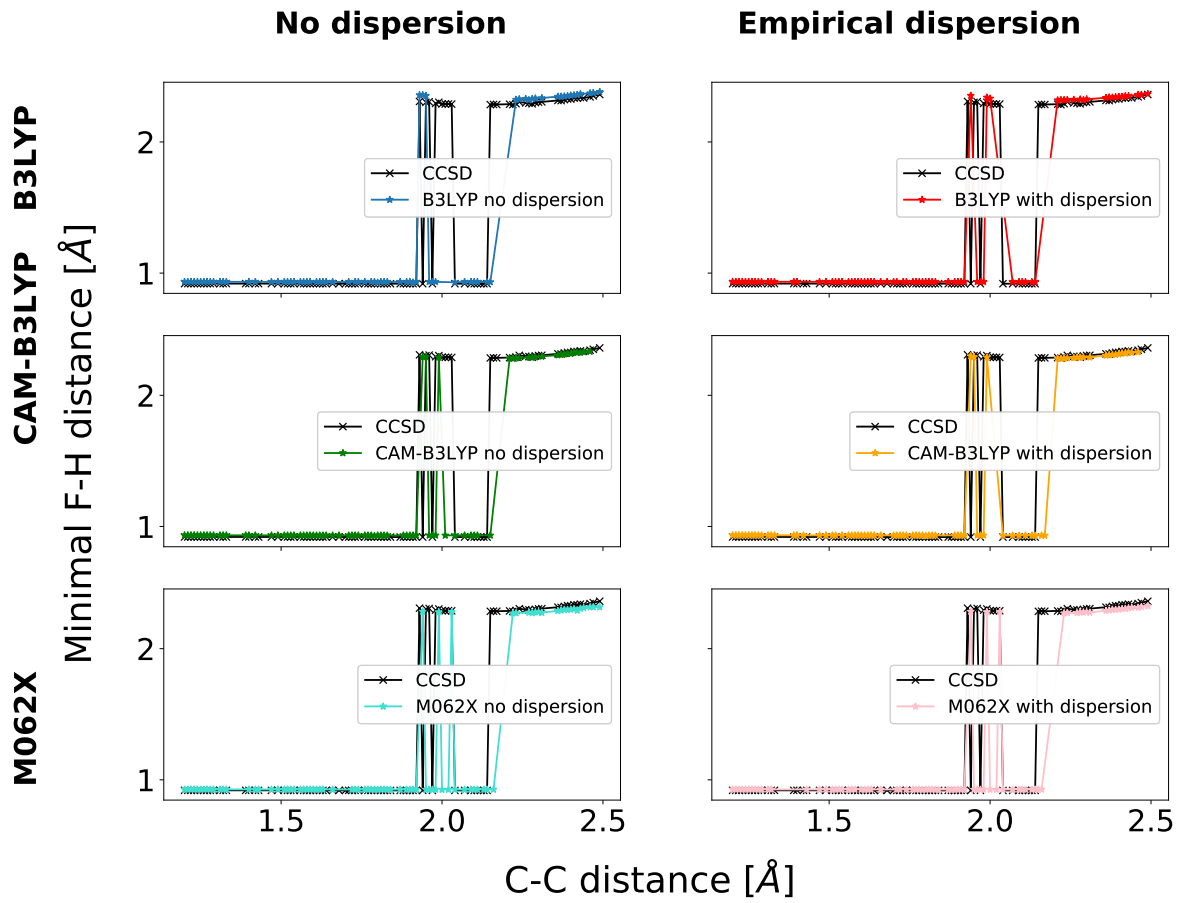


Figure 4.4.: Minimal F-H distance for the different functional with and without dispersion in comparison to CCSD method.

intersect at a carbon-carbon distance of 2.076Å. For the geometric comparison of DFT to CC results, only calculation that end up in the same final state should be compared.

After the CCSD results have been discussed and the restrictions for a reasonable comparison have been identified, the obtained geometries can be compared to the results that were obtained with the different density functionals. To compare the functionals the formation of the HF as a function of the carbon-carbon distance was investigated first. This was done to determine at which points the calculations of CC and DFT end up in the same minimum and the geometries can be compared.

Figure 4.4 shows the distance results, where the GAUSSIAN program terminated normally. The rows indicate the three different functionals, and the two columns stand for the calculations with and without dispersion. All the distances are plotted on top of the CC results to make a comparison possible. It is easy to see, that the H-F distances follow the exact same trend as the CC results for all functionals with and without dispersion. This is a reassuring result, showing that all methods lead to results in the same ballpark. However, there are clear differences for when the fluorine capture is happening. This capture is visualised by the steep drop which occurs when following the coloured curve from right to left. The distance between H and F follow the CC results for all functionals with and without dispersion in very good agreement. This is not only true for the state in which they are simply repelled and do not form a molecule, but it is also true for the case in which they do form a molecule. However, this is the less

interesting of both cases, because for the study of fullerene precursor molecule an accurate description of the HF molecule is no big feature.

As explained before the reason for the formation of the HF molecule might be the process of the optimisation rather than the energetical favourability of that state. Therefore, the intersection of the two potential curves for the two different final states will lead to a more accurate distance below which the HF formation is preferable.

Such an analysis of the energies has been performed for each functional and dispersion iteration and is displayed in figure 4.5. The figure shows all results for the three functionals with and without dispersion. The background colour indicates if the corresponding HF distance of the final geometry is above or below 1.5Å. For each line search, a fourth-order polynomials have been fitted to the respective data set with HF formation and a second order polynomial for no HF formation. The exact shape of the energy function is not that important, rather the point at which the two potentials meet, as well as the position and energy of the minimum for the HF forming potential.

For each functional and dispersion, the intersection point of the two plots was calculated and printed in the plot. This is the distance, below which the formation of a HF molecule should consequently lead to lower energy geometries. The C3-C9 distance corresponding to the minimal energy was calculated, by taking the minimum value of the red fit. This distance should be equivalent to the optimal bond length of the single carbon-carbon bond in 1,3 Butadiene. The calculated distances are displayed in figure 4.6a in comparison to the CCSD result plotted as a line.

For the 1,3 Butadiene molecule, the B3LYP and CAM-B3LYP functionals lead to similar values of 1.467Å only differing in the fourth digit after the decimal. The M062X functional yields optimal carbon-carbon distances of 1.470Å, with the dispersion corrected functional leading to almost the same result 1.470Å. In general, all of these values are very close to the CCSD result with the worst functional, pure CAM-B3LYP yielding a distance of 1.4675Å which is only 0.006Å away from the CCSD result. Although there is almost no significant difference between B3LYP and CAM-B3LYP, the M062X functional is certainly closer to the CCSD result. All these values were determined from a fourth-order polynomial fit and it can not be said with confidence that any functional would be preferable to another from this data and the addition of explicit dispersion only slightly improves the results for CAM-B3LYP and M062X, while the effect on the normal B3LYP functional is negligible.

Of more interest is the distance below which the HF will be removed from the rest because this halogen assisted dehydrogenation of the sp^2 hybridised carbon atom is the key process in the precursor auto-assembly of fullerene molecules. Therefore, any functional which will be used to perform calculations of the precursor molecules should accurately describe this "capture" distance between atoms C3 and C9. This distance is shown in figure 4.6, again, in relation to the CCSD results. The B3LYP functional yields the results closest to CCSD with the addition of empirical dispersion slightly improving over the pure functional. The other two, more complex functionals, lead to intersection distances further away from the CCSD calculation with pure M062X being furthest away and underestimating the intersection point by 0.028Å. When looking at the fitted polynomials in figure 4.5, it is clear that especially the fit without HF formation has very few data points at distances smaller than the intersection point. This can lead to rather bad fits and the differences are so small, that a real distinction between the quality of the functionals by this data would not be justified. They all lead to similar results for both the capture and the 1,3 Butadiene distance with dispersion showing a slight improvement over the pure functionals.

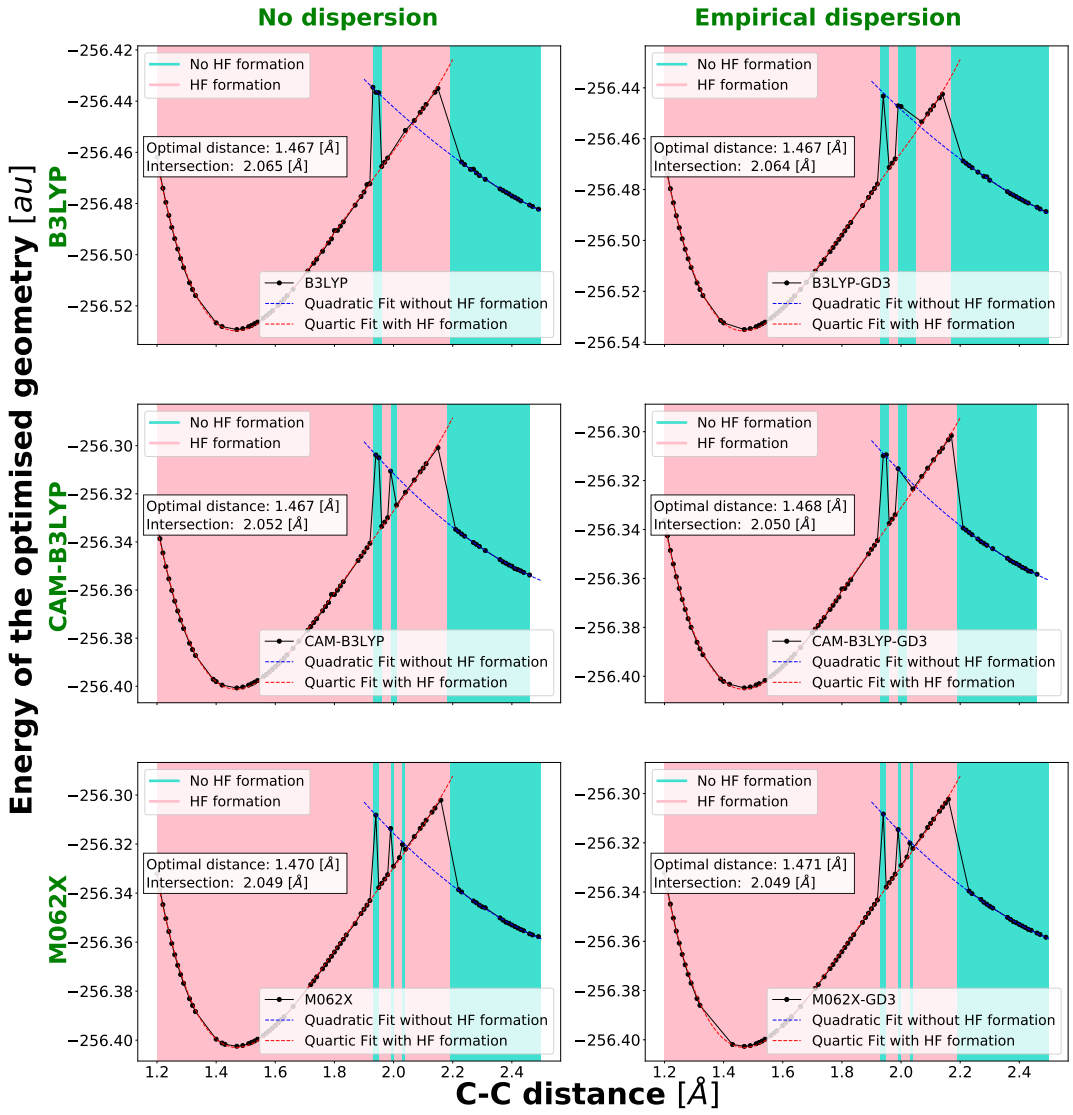
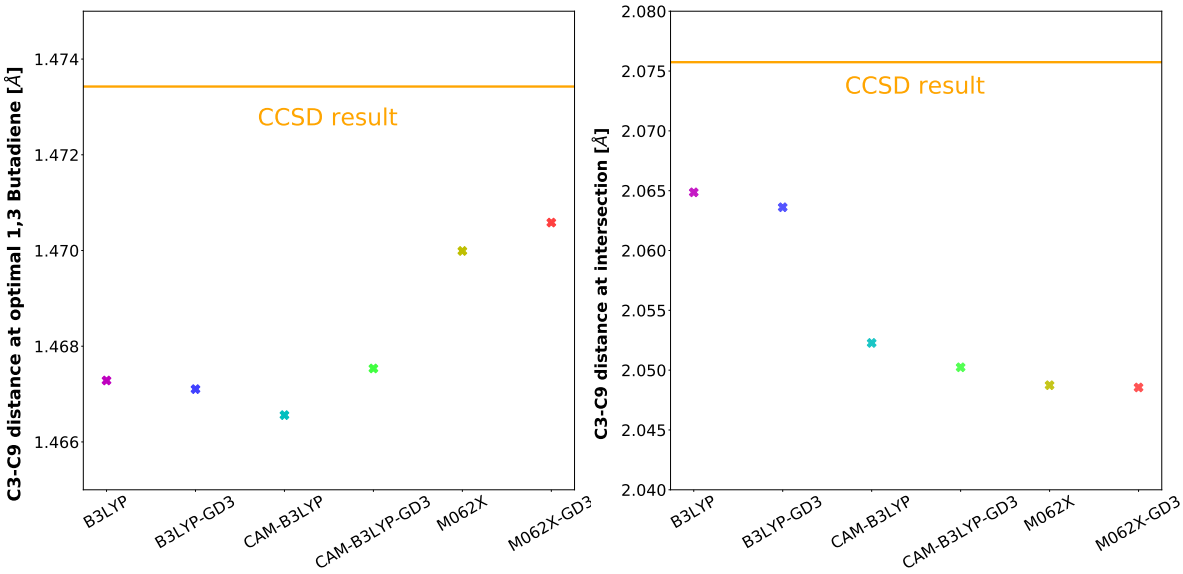


Figure 4.5.: Energies of the optimal geometries of the DFT calculations.



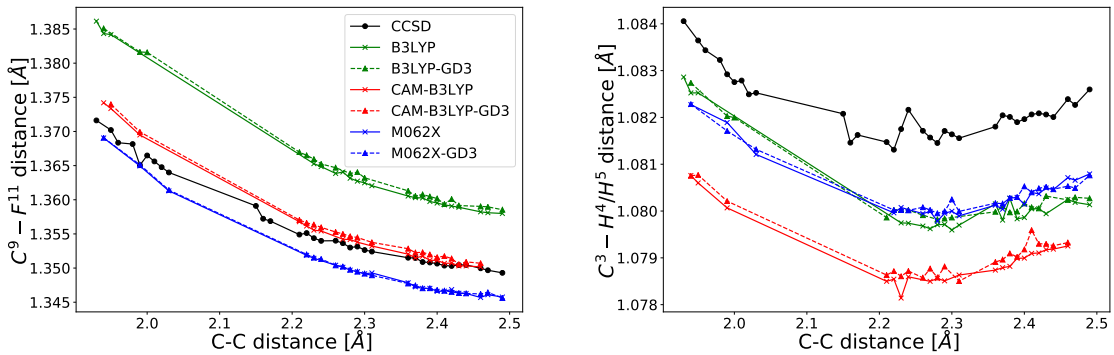
(a) Optimal distance in the 1,3 Butadiene molecule. **(b)** Capture distance.

Figure 4.6.: Results of the functional calculations in relation to the CCSD result.

Because dispersion effects do depend on the system size, the addition of empirical dispersion can be expected to have a bigger effect on the complete fullerene precursor.

Bond Stretching

It is also of interest to inspect the C-F and C-H bond lengths and identify those functionals, which best describe the stretching of them when the two molecules approach each other and the atoms start repelling. The distance between the H and F molecule stays almost constant along the whole path (except of course the HF molecule was formed) which indicated that some significant bond length and angular stretching must be involved because the carbon atoms do move closer together. The only way for the H and F to remain at a constant distance would be, by being pushed away from each other as the carbon atoms approach. The two distances between C9-F11 and C3-H4/H5 can be analysed at all



(a) Distance between the carbon and fluorine atom. **(b)** Distance between the carbon and hydrogen atom.

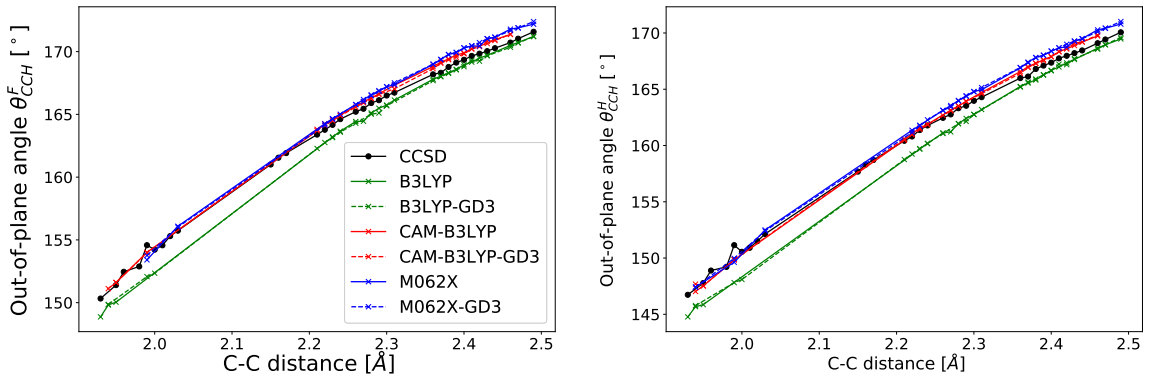
Figure 4.7.: Distances along the line search path for the DFT and CCSD calculations.

points, where the HF molecule has not been formed. It might seem counter-intuitive to also include the results where the formation of a HF molecule is energetically favourable, but it is not. This only means, the optimisation procedure for the CCSD calculation did get stuck in the 'wrong' local minimum and if the DFT calculations did get stuck in the same minimum the results can also be included in the analysis. The only requirement is, that the calculation for which the distances were analysed did end in a state without the formation of the halogen-fluorine.

Figure 4.7a shows the distance between the C9 and the F11 atom for the optimised geometries along the path. The figure 4.7b shows the distance between the C3 atom and either the H4 or the H5 atom, depending on which one is closest to the F11 atom. This way it is ensured, that even if the molecules were rotated, the distance between the atom which is repelled mainly by the F11 atom is displayed. These two distances describe how the carbon-fluorine and carbon-hydrogen bonds are stretched when the two molecules approach each other.

The fluorine bond starts at a value $r_{FC} = 1.349$Å for the CCSD calculation which is the same as the literature value reported in [62]. The initial values for the functionals CAM-B3LYP and M062X are very close to this value with and without explicit dispersion while the pure B3LYP functional is overestimating it by roughly 0.01Å. As the molecules get closer and closer together, the carbon-fluorine bond is stretch up to a value of $r_{FC} = 1.371$Å at a C9-C3 distance of $r_{CC} = 1.93$Å. This stretching is not linear but seems to follow the curve of a second-order polynomial. All different functionals reproduce the same kind of behaviour with and without dispersion, but the absolute value for the B3LYP functional keeps to be off by a constant 0.01Å. The absolute value for the CAM-B3LYP and the M062X functionals however, do lie close to the CCSD results. The CAM-B3LYP is in better agreement with the CCSD calculations for larger carbon-carbon distances while the M062X does slightly improve at smaller distances. While the B3LYP functional overestimates the C-F distance, the CAM-B3LYP's overestimation is greatly reduced, and the M062X functional slightly underestimated the distance. When comparing the B3LYP and CAM-B3LPY results, it does make sense that the CAM-B3LYP description is better than the simple B3LYP, because it includes more long-range interaction.

However, when considering the bond stretching of the hydrogen carbon bond (figure 4.7b) this trend is not continued. First, the bond is not consistently stretched for the CC result, it starts at a distance of $r_{CH} = 1.083$Å decreases first and then increases again as a function of the carbon-carbon distance. Also, the CAM-B3LYP performs worst for this bond, while the B3LYP and M062X functionals are equally good. Although it seems like the error between the CCSD and the DFT calculations is larger, special attention should be paid to the scaling of the axis. The variability in the H-C bond stretching is much smaller than for the C-F bond. The difference, between the minimal and maximal C-H distance for the CCSD calculations is $\Delta r_{CH}^{CCSD} = 0.0027$Å, compared to $\Delta r_{CF}^{CCSD} = 0.022$Å for the C-F bond. So the variability is one order of magnitude larger in the fluorine bond. This is because the hydrogen atom is extremely small and will be bound very tightly to the carbon atom, while the fluorine atom is much larger and will allow much more room for bond stretching, without braking. For these reasons, the focus should be put on the C-F bond and the M062X and CAM-B3LYP functionals can be expected to describe the bond stretching more accurately, while empirical dispersion has virtually no effect on the results.



(a) Angle in the C_2H_3F molecule.

(b) Angle in the ethene molecule.

Figure 4.8.: Results of the two out-of-plane angles for the bond forming H and F molecule.

Out-of-plane Bending

The bond stretching is not the only deformation which allows the molecule to avoid large repulsion energies from the hydrogen and fluorine atoms. Another deformation which is introduced into the system is the out-of-plane bending of those atoms. In a perfectly symmetric sp^2 hybridised carbon atom the three bonds are arranged in a plane with out-of-plane angles of $180°$[1]. As a notation to indicate the out-of-plane angle between the vector C9-F11 and the C9-C6-H10 plane (compare figure 4.1) θ^F_{CCH} is used, while for the out-of-plane angle for the other molecule the notation θ^H_{CCH} is used. Here again, the H atom was chosen which was closest to the F atom and only final geometries without HF formation were chosen.

The results for both angles are shown in figures 4.8a and 4.8b. Both out-of-plane angles start for all functionals and the CCSD calculations at $170°$. This is reasonable because the molecules are already repelled by each other and although the distance between the two carbon atoms is $r_{CC} = 2.49$Å this repulsion can be expected to push the atoms out of the plane. The out-of-plane angle decreases almost linearly as a function of the carbon-carbon distance. The two more complicated functionals CAM-B3LYP and M062X are in close agreement to the angles obtained with the CCSD calculation in the lower end of the spectrum, where the B3LYP functional underestimated the out-of-plane angles, which means it overestimated the out of plane bend of the two atoms. In the upper end of the carbon-carbon distance, the B3LYP functional does a much better job, even better than the other two functionals. Although the absolute results slightly differ, the overall trend of the out-of-plane angle is very consistent for all methods. The addition of empirical dispersion seems to have almost no effect on all geometric quantities, this might be because of the small system size dispersion effects are very small.

4.1.2 Comparison of the Reaction Enthapy

Finally, one of the most important qualities that the density functional should be able to describe accurately is the energy of the system. The precursor quality depends heavily on the energy of the geometries that the precursor molecule takes along the path. As described before, the restrictions set by the stability of the carbon bonds put restraints on these geometries and a way to obtain geometries satisfying these constraints was presented above.

[1] For the definition of the out-of-plane angles see 2.7.1

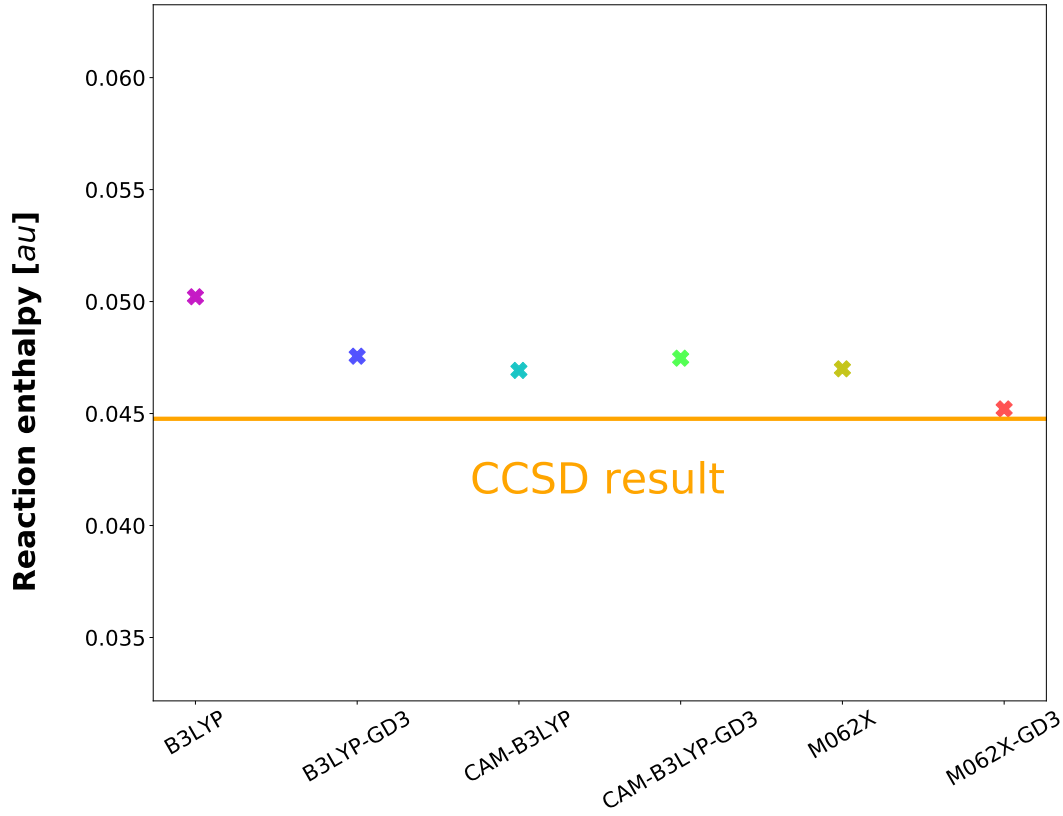


Figure 4.9.: Reaction enthalpy comparison for different density functionals and CCSD.

The reaction enthalpy offers a stable, well comparable quantity. This enthalpy was calculated by taking the energy difference from the system at the largest calculated carbo-carbon distance of $d_{CC} = 2.49$Å and the energy at the potential minimum determined by the red fitting curve. The differences between the energies are displayed in figure 4.9. The energy difference obtained by the CCSD calculation is 0.0448 Hartrees. The B3LYP functional does the worst job at approximating the enthalpy with values of 0.0502 and 0.0476 Hartrees without and with dispersion, with the addition of explicit dispersion considerably improving upon the result. The CAM-B3LYP and M062X functionals lead to almost the same reaction enthalpy, where the addition of dispersion improves the M062X result significantly while it does the opposite for CAM-B3LYP. This is somehow surprising because the addition of dispersion should lead to a clear improvement for all functionals, least of which for M062X. After all, the Minnesota functionals parameters have been fitted for dispersion corrected results. Although there is no clear best functional, it is evident that the long-range Coulomb corrections introduced by the CAM-B3LYP functional certainly improve upon B3LYP. The difference in energies is much larger and more evident than for the capture distance or the 1,3 Butadiene bond length. There is a clear improvement over B3LYP by using either CAM-B3LYP or M062X, probably due to the better description of long-range interactions. As mentioned before the size of dispersion effects will be larger for the whole fullerene precursor molecules and the effect of Grimme dispersion should be evaluated in more detail on a larger system.

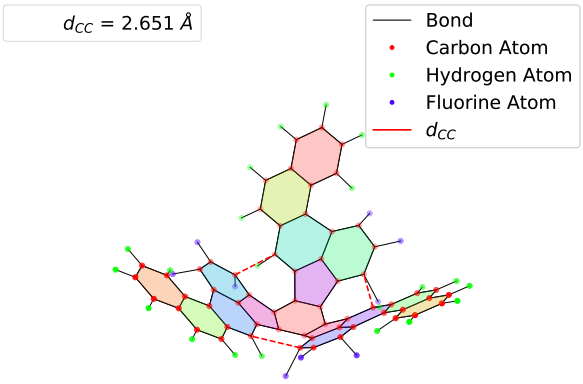


Figure 4.10.: Input geometry for the closing if the three innermost hinges with fluorine and hydrogen rotation outside of the bowl.

4.2 Geometries Along an Auto Assembly Path for the C_{60} Precursor Molecule

After it was verified, that the choice of functional does not affect the qualitative results of the bond lengths and angles, the results from the B3LYP calculations along the folding path for the C_{60} precursor were analysed. Although the carbon atoms at the innermost cavity were required to stay in their position to ensure that the curvature will be introduced to the molecule, no neighbouring (or even second neighbouring) atoms were frozen so the bond lengths, angles and out-of-plane angles of the final geometry can be treated as independent from the freezing. The analysis was only performed on GAUSSIAN calculations which terminated normally and the results have been analysed for their geometric properties, especially the behaviour of the suggested FF parameters (bond lengths, angles, face planarity) along the closing path as well as the energies of the optimised system.

4.2.1 Closing Precursor Molecule at the First Hinge

To first understand how the closing of the innermost hinges affects the overall geometry of the precursor molecule a visual comparison of the optimised geometries is most helpful. Figure 4.10 shows the initial geometry for a carbon-carbon distance of $d_{CC} = 2.651$Å. The geometries that were obtained by the B3LYP DFT optimisation with the hydrogen and fluorine atom placed outside of the bowl in the initial setting are displayed in figure 4.11. As described in section 3.3, the closing of the inmost hinges was not sampled along the hinge angle itself but instead as a function of the three carbon-carbon distances, which will first form a bond (red lines in figure 3.4). Those distances were sampled in steps of 0.05Å and ten optimisations along the path were chosen to illustrate how the molecule deforms when being folded up.

It can be seen, how for the first calculation (top left) the precursor starts at a geometry very close to the initial completely flat one, but it can already be seen clearly, how the repulsion between the hydrogen and fluorine atom force them outside of the bowl. Here it is important to remember again,

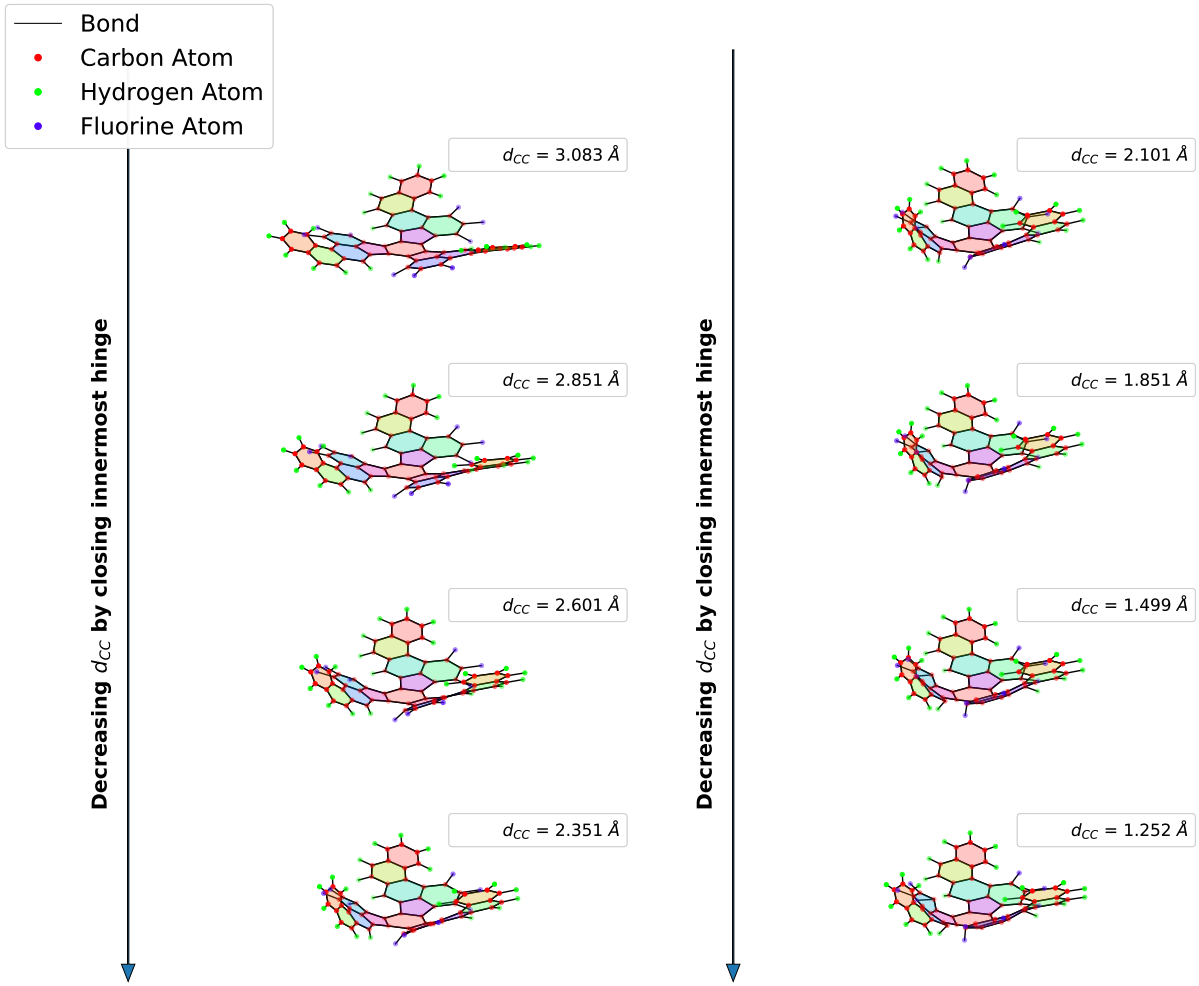


Figure 4.11.: Optimised geometries for the C_{60} precursor molecule with fluorine in key positions, sampled as a function of the carbon carbon distance.

that the two carbon atoms which the H and F are attached to, were not allowed to be optimised freely, because otherwise, the repulsion would have just flattened out the molecule. When considering the end of the "arms", namely the four outermost hexagonal faces (seen from the root face) it becomes clear how the small deformation close to the root face already has a strong impact on the outermost arm deformation. The faces are visibly twisted around an axis. A closer look is taken at one of the arms for the $d_{CC} = 3.083$Å in figure 4.12 where the rest of the molecule is indicated by a grey face. The axis around which the rotation is occurring is visualised as a black arrow.

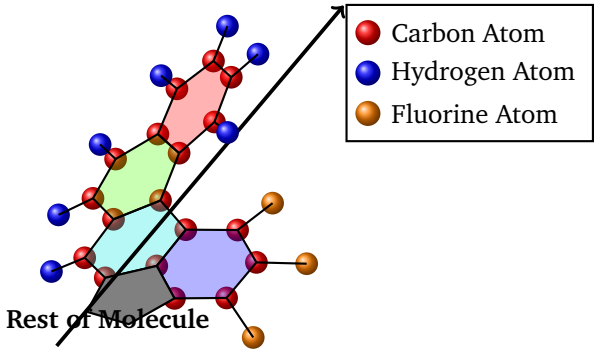


Figure 4.12.: Illustration of the face-twisting around an axis introduced by the closing of the innermost hinge.

As the carbon atoms come closer to each other this twisting deformation increases steadily and so does the bending of the fluorine and hydrogen atom. This is exactly what was expected, because the closer the two carbon atoms are, the stronger the repulsion between H and F will be. The last calculation was performed at a distance of $d_{CC} = 1.252$Å which is already below the optimal bond length of the two carbon atoms, which is reported as 1.39Å in the free benzene ring [63]. Although a carbon-carbon bond could have been expected to have formed at this distance, the molecule seems to be stable enough to prevent the breaking of the carbon-fluorine and carbon-hydrogen bonds. The energy due to repulsion and orbital overlap does not seem to be enough to break the bond and create the HF molecule(s).

The results for the calculations with chlorine as the halogen in the key positions are shown in figure 4.13. As mentioned before, the hinge was not overstretched in this case, but only closed until the optimal value of $142.62°$ was reached. The smallest carbon-carbon distance, which could be obtained for this angle is 1.801Å. It is clear from the figure, that for the chlorine case, the twisting and bending of the arms and the atoms in the cavity is larger than when fluorine is used as a halogen. This is easily understood when the bond lengths for the respective halogen carbon bond is considered. The bond length for a carbon-fluorine bond is around $r_{CF} = 1.35$Å while the chlorine carbon bond has an optimum value of $r_{Cl} = 1.77$Å. This means, that for the same carbon-carbon distance in the cavity, the chlorine atom will be considerably closer to the hydrogen atom and the repulsion therefore much larger. As the repulsion between the halogen and the hydrogen atom is probably the main source introducing the arm twisting, it is only natural that this will be larger for a longer carbon halogen bond.

For both studied halogen atoms, the introduction of curvature through the closing of the innermost hinges results in significant deformation of the previously planar molecule which can be described as a twisting of the precursor arms around an axis. This deformation reminds of the helix structure of DNA. For all distances and both halogens, the initial placement of the halogen and hydrogen atom outside of the bowl is not changed. The repulsive energy which would be introduced, should these atoms be

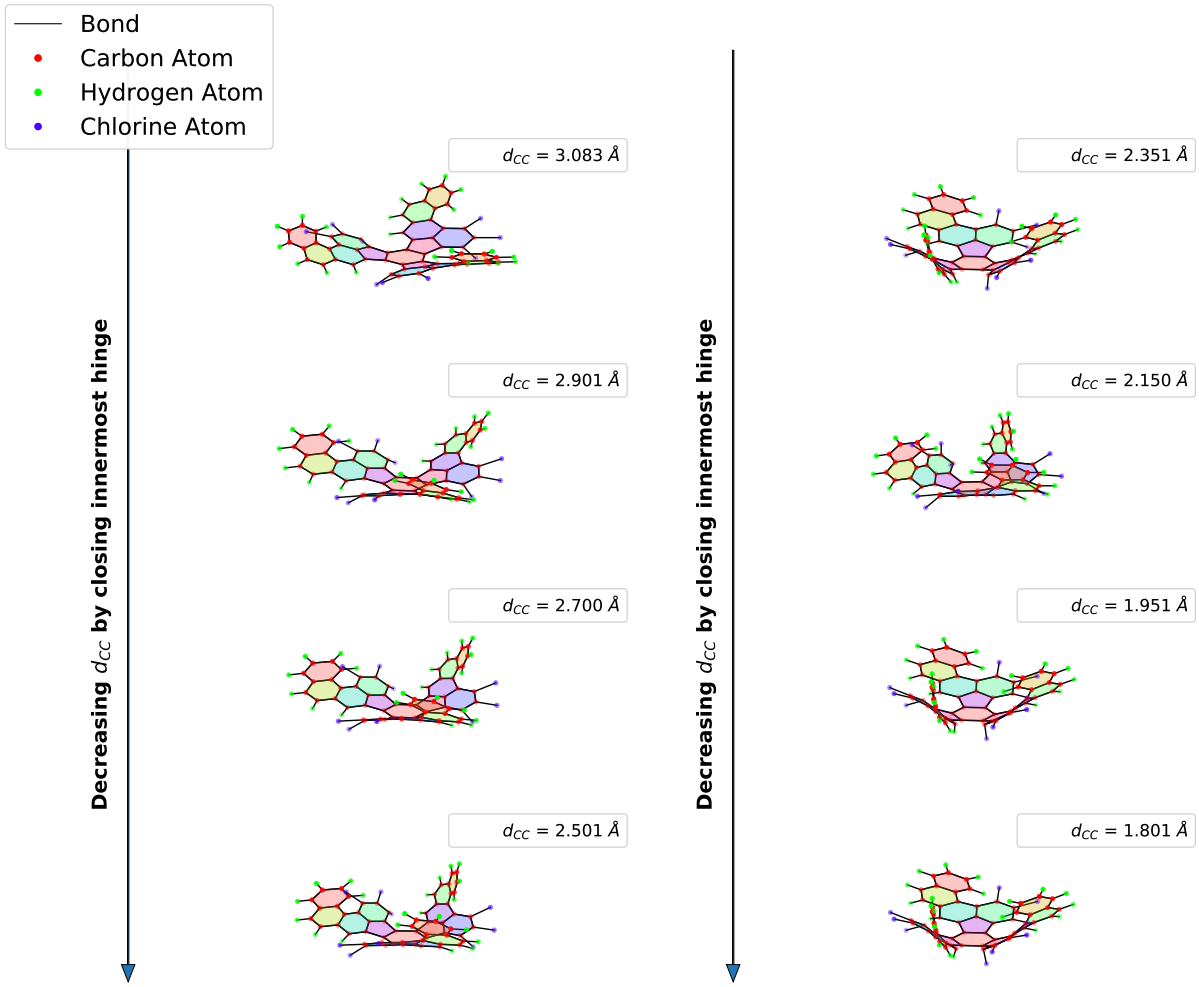


Figure 4.13.: Optimised geometries for the C_{60} precursor molecule with chlorine in key positions, sampled as a function of the carbon carbon distance.

moved to the inside of the bowl prevents the atoms to be "flipped" to the inside.

Energy along the Path

To assess the reliability of the geometry optimisation, the energy of each final geometry along the path is a good quality to compare. By looking at the overall energy development along this smooth closing path it is possible to find the points at which the optimisation might have got stuck in a local minimum. The energies along the closing path are displayed in figure 4.14. The offset of the y-axis should be pointed out for both different plots, which indicate, that the energies differ in the order of 0.5 Hartrees. In the calculations where chlorine was placed in the precursor, the plot ends at a distance of 1.801Å. Some of the optimised geometries corresponding to the points along the curves can be found in the figures 4.11 and 4.13.

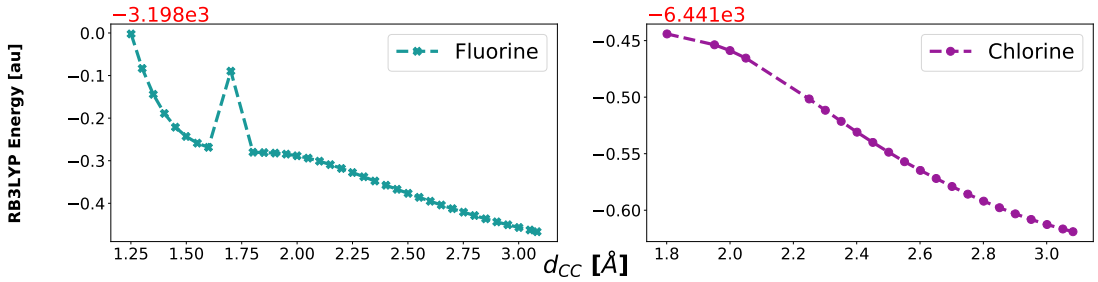


Figure 4.14.: Energies of the optimised geometries plotted against the carbon carbon distance for the fluorine and chlorine containing precursor molecules.

The energies which correspond to the final geometries for the fluorine-containing precursor molecule have a very smooth shape, starting with an almost sine-like increase in energy as the hinge is closed, reaching an energy plateau in the region between $d_{CC} = 1.9$Å, to $d_{CC} = 1.6$Å before rising quadratically below carbon-carbon distances of $d_{CC} = 1.6$Å. This very smooth behaviour is interrupted only by a single outlier at a distance of 1.7Å. Here the energy is significantly higher than would be expected considering the rest of the curve. To understand whether this is a point where the optimisation is stuck in a local minimum the final geometry was inspected. It was found that in this final geometry, the atoms C15, C22, and C29 which sit in the cavities of interest (compare figure 3.3), have formed a bond with both, the hydrogen and a fluorine atom, leaving them in a state where they have more electrons than needed to fulfil the octet rule. This is not a reasonable final state and this data point was excluded from further analysis.

The chlorine-containing precursor energies follow the same sine-like behaviour as the ones with fluorine. It seems like they would reach their respective plateau at a value right below the last calculated distance of 1.8Å.

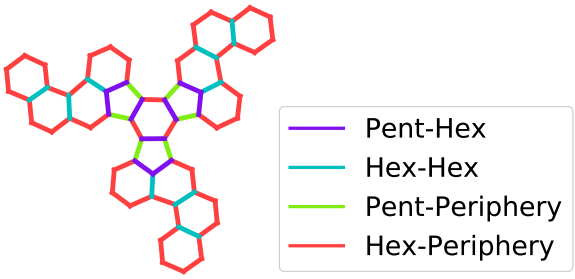


Figure 4.15.: The four different bond types which occur in fullerene precursor molecules.

FF Parameters along the Path

The carbon-carbon bonds look like they remain almost constant along the whole path and so do the faces, but a visual assessment of the optimal geometries is not sufficient to justify that the assumed FF parameterisation will indeed be valid along the path. Some parameters might be significantly distorted along the way, which would mean the FF energy term regarding those internal parameters can not be expressed by making the harmonic approximation. After the outlier was removed from the data, the remaining geometries were analysed for the behaviour of the FF parameters.

The bond lengths were put into four categories. Hexagon bonds on the periphery, pentagon bonds on the periphery, bonds for neighbouring pentagon and hexagon and bonds for two neighbouring hexagons. Those bond types are illustrated in figure 4.15. Because the arms are symmetric, it is enough to look at the behaviour of one arm. This symmetry has been confirmed along the way and the bond lengths are equivalent up to the numerical error in the order of 10^{-9}Å. Considering the symmetry of the precursor this leaves only two independent periphery-pentagon, two pentagon-hexagon, three hexagon-hexagon and seventeen hexagon-periphery bond distances. Therefore, some of the results will not have much statistical significance, but it is still interesting to look at the overall trend of the lengths.

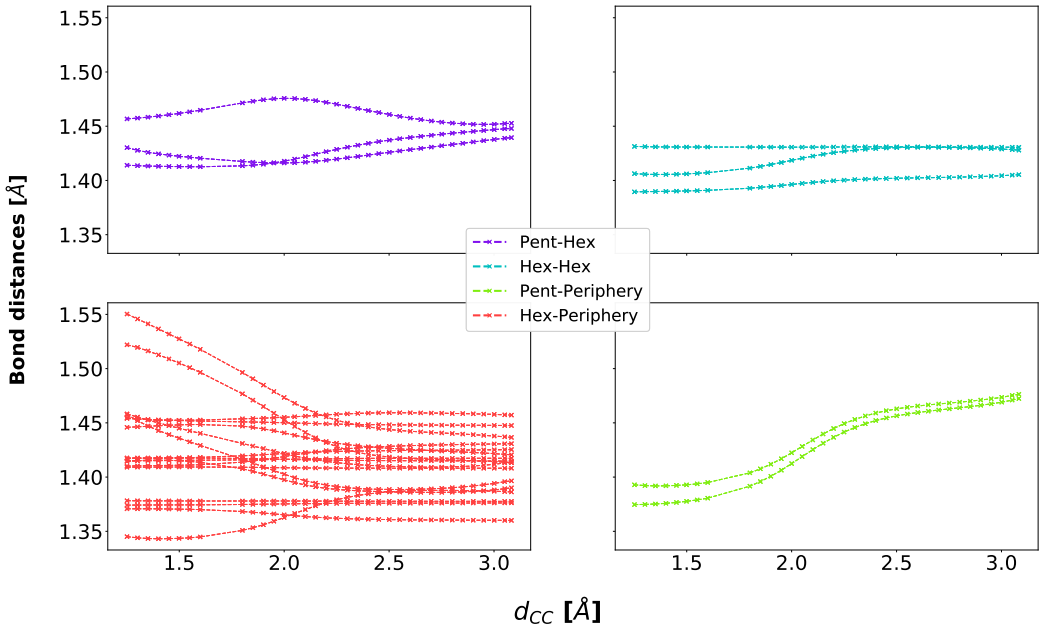
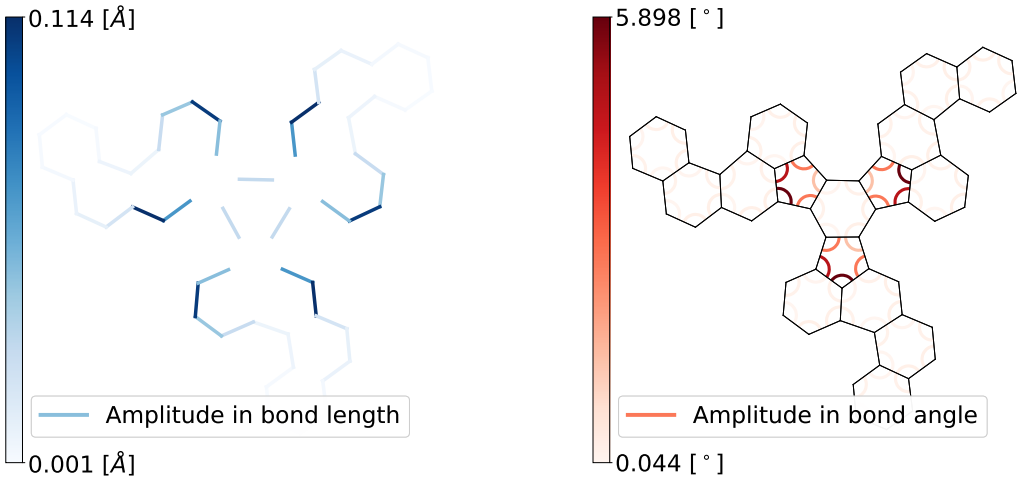


Figure 4.16.: Development of the four different types of bond lengths along the path.

The development of these four different bond types is displayed in figure 4.16 as a function of the carbon-carbon distance. It can be seen, that the pentagon-hexagon (top left) bonds are almost constant along the whole path, and so is the hexagon-hexagon bond (top right). The pentagon-periphery bonds start at a value of $d_{P-Per} = 1.47$Å and decrease to a value of $d_{P-Per} = 1.39$Å which is a change of roughly 0.1Å. Most of the hexagon-periphery bond lengths do remain constant over the whole range but it is visible that some of them are stretched and some of them are shortened along the way. In general, the spread for this kind of bond is rather large with initial values (at the largest carbon-carbon distance of $d_{CC} = 3.08$Å) ranging from $d_{H-Per} = 1.35$Å to $d_{H-Per} = 1.45$Å. To understand while most of the hexagon periphery bonds remain constant, but some of them are significantly deformed, the individual bonds have been analysed for the maximal deformation which they experience along the way. Figure 4.17a shows a plot of each bond coloured by how large the amplitude is in which they vary along the path. This makes it very obvious, that the bonds, which experience the largest deformations are the ones connected to the carbon atoms, which are bound to the fluorine and hydrogen atom in the innermost cavities. The other bonds are hardly affected at all. The reason why these bonds are stretched and contracted, is again, due to the high energies associated with the repulsion of the halogen and hydrogen atoms. These are also the carbon atoms whose coordinates were frozen during the optimisation process

to enforce the molecule to be closed. If they had not been frozen, the final geometry would be flat and there would be no change in the bond lengths.



(a) Maximal amplitude of the bond lengths along the path. **(b)** Maximal amplitude of the bond angles along the path.

Figure 4.17.: Maximal absolute difference of bond lengths and angles along the outside closing path for fluorine.

The same analysis was performed for the bond angles. Here there are only two types of angles, those within a pentagon and those within a hexagon. The development of those two different types of angles, as the molecule is closed, is displayed in figure 4.18. The angles within the hexagon remain almost entirely constant along the path, while the angles within the pentagon do vary. Some of them decrease and some increase. How big the difference in the maximal absolute angular difference along the path is, becomes even more evident when considering figure 4.17b. Here, each angle within the molecule has been coloured by its respective maximal difference along the closing path. While the angles within the pentagons changes by up to $\Delta\theta_{Pent} = 6°$, the variability within the hexagonal faces is negligible. This could mean, that the angular force field constant for the hexagonal ring is higher than for the pentagonal ring because a weaker force allows larger deformations. It could also mean that the deforming forces acting on the pentagons are stronger. To determine what is true, a similar study could be performed on a larger precursor with pentagons that are not involved in the dehydrogenation.

The last geometric quantity that was analysed along the path, was the planarity of the faces. For each final geometry along the carbon-carbon distance, the least squared problem described in section 2.7.1 was solved to obtain a normal vector and an intersection parameter d. The eigenvalue corresponding to that normal vector is the best quantity to assess how "flat" a face is. However, to provide a quantity which is more intuitively understood, the mean distance from all points to the least-squares plane were also calculated in Å. Figures 4.19 and 4.20 show how the planarity of the faces evolve as a function of the distance, and which faces are experiencing the largest absolute difference in planarity along the path. The three faces which contain the three fluorine atoms (darkest shade of green in fig. 4.20) are the ones which experience the largest change along the way. However, no face is deformed by values larger 0.08Å in terms of a mean distance from the mean plane.

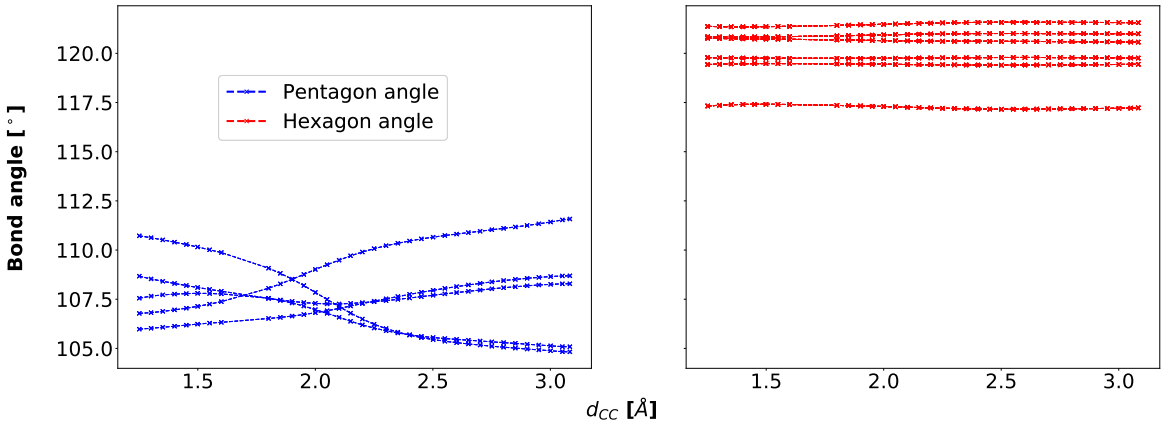


Figure 4.18.: Development of angles within pentagon and hexagon faces along the path.

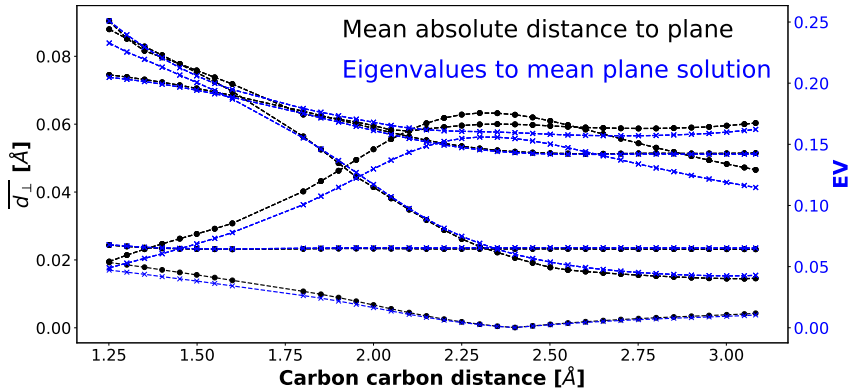


Figure 4.19.: Development of the face planarity in terms of mean distance from the plane and described by the smallest eigenvalue of the least squares problem.

In summary, the three analysed geometric quantities which were suggested as the main parameters in a FF description of fullerene precursors, namely bond length, bond angles and face planarity show a smooth behaviour along the whole spectrum of the precursor closing path. The faces do remain almost entirely constant for all sampled carbon-carbon distances, suggesting a strong restoring force regarding any planarity affecting deformations. The hexagon angles show the same behaviour, while the pentagon angles do vary more along the way. Any force constants which are obtained from a QM Hessian matrix calculation should be able to reproduce this behaviour with strong face planarity constants and possibly angular spring constants, which are stronger within a hexagon than a pentagon.

Although the bond lengths do change along the path in the order of 0.1Å, the bond lengths after the halogen halide formation will most likely be much closer to the distances obtained from C_{60} diffraction studies. This means, that despite the bond stretching along the path, the description of the bonds by a harmonic potential is justified, because there is no reason to believe, that another local energy minimum for the bond lengths is passed as the fullerene is closed.

The bond lengths have been manually put into four categories, but when trying to verify this assumption by using a k-mean clustering algorithm with four centres, it was not possible to clearly distinguish these categories all along the way. This might be due to the very small statistical significance (only two data points for pentagon-periphery bond up to symmetry) and for any larger fullerene precursor this

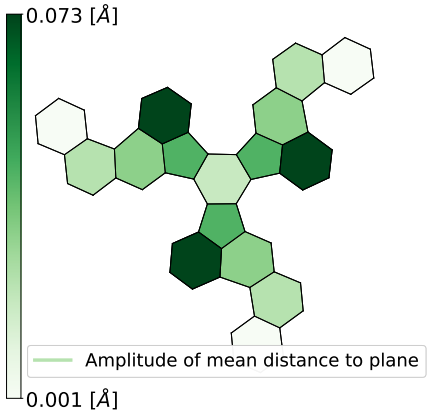


Figure 4.20.: Range of the mean absolute distance of all vertices of a face from the mean plane along the outside closing path with fluorine.

clustering should be repeated when the final geometries are obtained.

All deformations which occur can be attributed to the deformations which have to be introduced to overcome the barrier and initiate the cyclo-dehydrogenation of the carbon atoms. The further away the bonds, angles and faces are from the cavities, the better the description by a harmonic potential with a single optimal value is. This analysis along the path has also been performed for the choice of chlorine as a halogen and yielded similar results. The obtained plots can be found in appendix A.1 . This raises the hope that a MD simulation of the precursor auto-assembly will be possible by using the simple harmonic potential.

4.2.2 Dehydrogenation Along the Assembly Path

As the energies of the precursor molecules show very smooth behaviour along the closing path, there is no indication that a halogen halide is formed at any point along the way. This has indeed been verified, by studying the optimal geometries. There is no clear measure for whether two atoms form a bond or not, but as mentioned before, the distance between two atoms is a good indicator for this. The hydrogen and halogen atoms do not come reasonably close to each other, nor do they move significantly far away from their respective carbon atom. As the optimal geometries depicted in the figures 4.11 and 4.13 clearly show, the arms of the precursor molecule are twisted and the hydrogen and halogen atoms bend out of the plane to avoid small interatomic distances associated with large repulsive energies. The goal of this study was to find the point along the closing path at which the formation of the halogen halide would be energetically preferable to the precursor molecule. From the optimisation results, it does not look like there is such a point along the way. Although the optimisation is not a simulation of a chemical reaction, the results are an indicator of how the molecule would arrange itself, would a deformation be introduced by thermal energy. If the innermost hinges were to be closed by the thermodynamic movement it is likely, that the hydrogen and halogen atoms would be flipped to the outside of the carbon bowl and the arms of the molecule would be twisted. If the hinges were to be closed further, the bending and deformation of the arms and halogen and hydrogen atoms would simply increase. At no point along this way, does it look like the formation of a halogen halide gives an obvious energetic advantage over the full precursor molecules. It could have been tried to stretch the halogen-carbon and

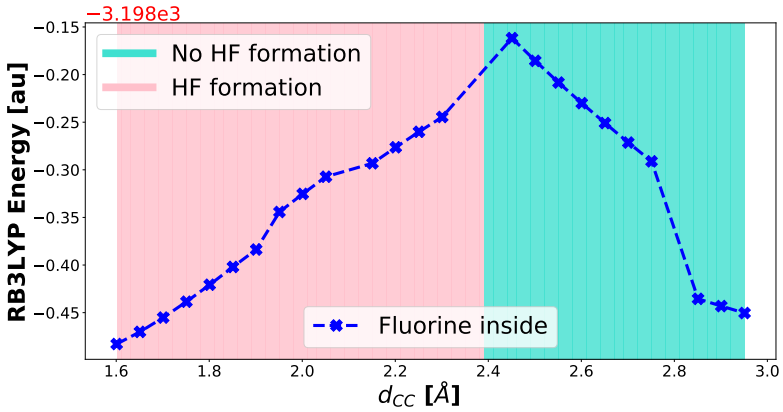


Figure 4.21.: Energies of the system after geometry optimisation with the fluorine and hydrogen atoms placed inside of the bowl.

hydrogen-carbon bonds in the initial setting to force the bonds to break, but because the results from the functional comparison showed a very small variation of these bond lengths, it was decided against it.

4.2.3 Placing the Halogen and Hydrogen Atom on the Inside

Because no halogen halide was formed, even for very small carbon-carbon distances it was decided to study the system where the fluorine and hydrogen atoms are not placed on the outside, but on the inside of the forming bowl. This has the advantage of allowing the carbon atoms to stay as faces with the correct bond angles and lengths as well as keeping the fluorine carbon and hydrogen carbon lengths unchanged. The only quantity which will change by flipping the atoms on the other side is the out of plane angles. As these have shown to be less stiff than the bond lengths this was preferred over a bond stretching because it poses a more probable path for the dehydrogenation.

The development of the final geometries for decreasing carbon-carbon distance is displayed in figure 4.21. Starting from the largest carbon-carbon distance, the first three points show a rise in energy, which is the same as for the situation where the fluorine and hydrogen atoms were placed outside of the bowl. The final geometries of those first three data points are indeed identical to the optimisation with fluorine and hydrogen placed outside. It is believed that for carbon distances above $d_{CC} = 2.851$Å, the two fluorine and hydrogen atoms with their respective bond lengths of $d_{CF} \approx 1.36$Å and $d_{CH} \approx 1.1$Å still have enough space to be flipped on the other side of the bowl during the optimisation process.

This is not the case for the following seven data points in the range from $d_{CC} = 2.451$Å to $d_{CC} = 2.751$Å. Here the energy curve seems to follow a much steeper increase. The final geometries show that in these cases, the fluorine and hydrogen atoms were placed on opposite sides of the bowl. The hydrogen on the outside and the fluorine on the inside. This placement of the hydrogen atom on the inside increases the energy significantly over the placement on the outside, because there is simply less space on the inside, leading to higher repulsive energies.

For distances below $d_{CC} = 2.301$Å, there is a sudden significant drop in energy, which continues down to the last sampled carbon distance of $d_{CC} = 1.601$Å. This energy drop is associated with the formation of the three HF molecules in the innermost cavities. The three formed molecules are removed from the

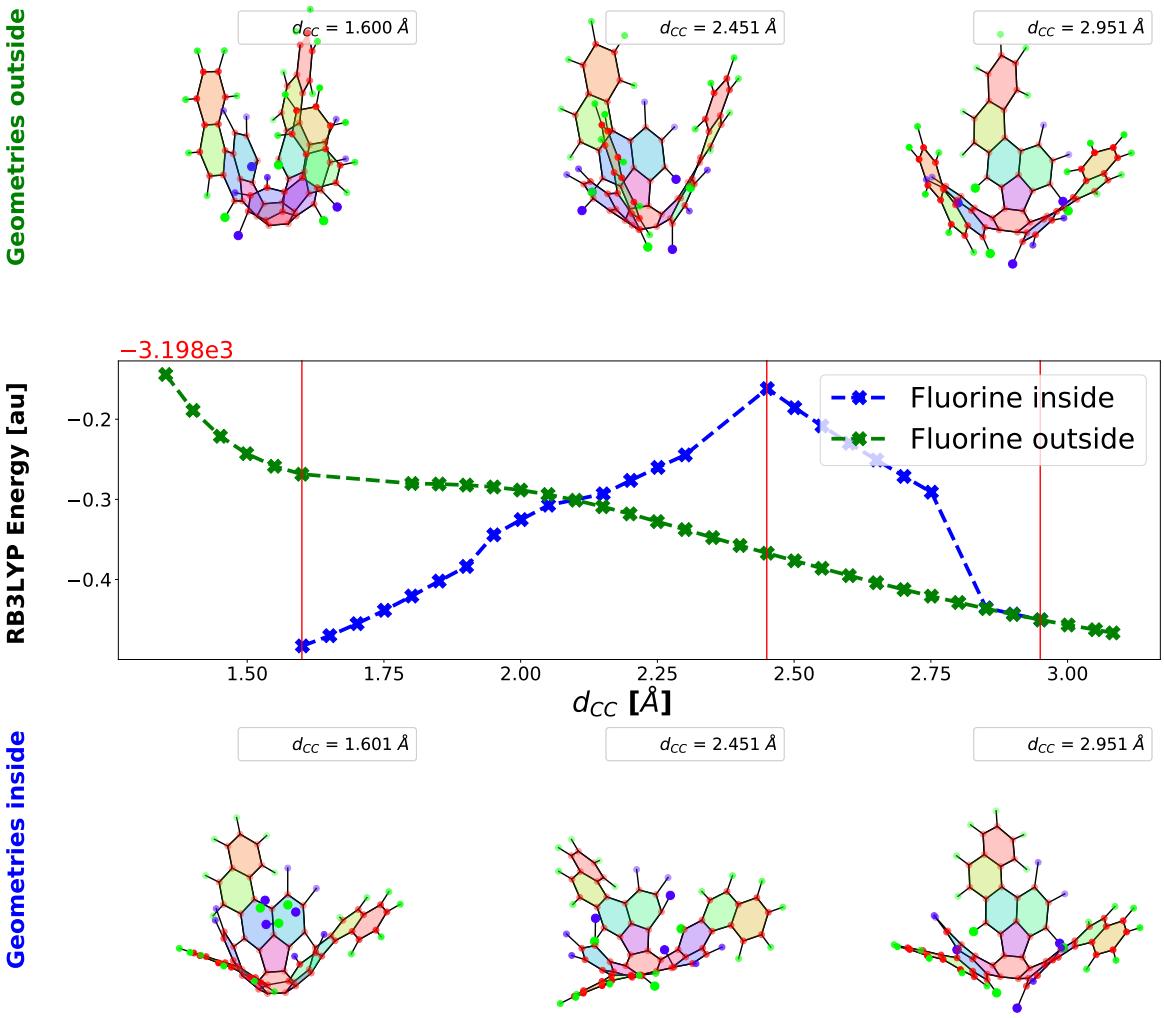


Figure 4.22.: Comparison of the optimised geometries and the corresponding energies of the two different settings with fluorine and hydrogen inside and outside.

rest of the precursor molecule. The formation of the HF molecule is energetically preferable over the state where the two atoms are rotated for distances below $d_{CC} = 2.301$Å.

The continued decrease in energy for further closing of the innermost hinge strongly suggests, that as soon as the HF molecules have been removed and the three new carbon-carbon bonds have been formed, the precursor will continue to close the hinges. This trend can be expected to continue for further closing of the hinge until the optimal benzene ring bond length of $d_{CC}^{Hex} = 1.4$Å is reached. The formation of the bond between the carbon atoms can be expected to pull the two ends of the cavity closer together and introduce the curvature.

Figure 4.22 shows a comparison of the energy development of the inside system (below) and the system where the atoms have been placed on the outside of the bowl (above). It is clear, both from the energy and the corresponding geometries, that in the upper end of the spectrum the final results are identical. Then the outside settings experience a much higher increase in energy, due to the placement of the hydrogen atoms on the inside until it forms the HF molecules and the energy significantly drops.

Although these results were obtained by performing a geometry optimisation and are not to be confused with a simulation, they can still help to understand the underlying process of the auto assembly. Due to the validity of the Born-Oppenheimer approximation, the electronic state of the system can be expected to react almost instantaneously to any deformations introduced by thermal movement. Therefore, if a deformation along the three innermost hinges was introduced to the molecule, the system can be expected to arrange itself in the geometry which is energetically most preferable regarding this deformation. This will result in a somehow smooth reaction path. One can also think of slow nuclei that are very inert and slow the overall movement of the carbon cage.

If the hinges were to be closed without rotating the inner hydrogen and fluorine atoms, they will most likely be pushed out of the forming bowl and to break the corresponding C-F and C-H bonds, the carbon atoms will have to be brought together very closely. Even when bringing the two carbon atoms as close as $d_{CC} = 1.252$Å the HF formation could be achieved during the optimisation process. This deformation was associated with a relative rise in the energy of the system of 0.464 Hartrees. To form the bond, the two carbon atoms will have to be brought even closer together to overcome the barrier along the green curve.

If instead, the two atoms were to be forced to stay on the inside of the bowl while closing the hinges, the blue path is taken and the formation happens between a carbon-carbon distance of $d_{CC} = 2.301$Å and $d_{CC} = 2.451$Å (more than one Å more!). The relative rise in energy drops to a value of 0.289 Hartrees. This is a huge difference and suggests that the reaction path, in which the halogen and hydrogen atoms are forced to stay within the bowl, although steeper at first, has a much smaller total activation barrier than if the atoms were left free. It is assumed, that any mechanism that can "trap" the fluorine and hydrogen atoms inside of the bowl would lead to lower activation barriers.

Interestingly, the geometry which Otero et al [11] have calculated for the $C_{57}N_3H_{30}$ precursor on a platinum (111) surface somehow reminds of such an "inside"-reaction path. When the precursor molecule was deposited on the platinum surface, the carbon atoms experienced an attractive force from the surface, while the periphery atoms were repelled. This results in an optimal geometry on the surface similar to the one displayed in figure 4.23, where all periphery atoms would point away from the surface

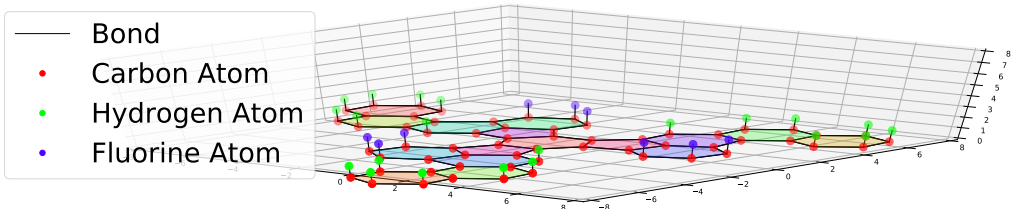


Figure 4.23.: Schematic illustration of a surfaced attached C_{60} precursor with maximal repulsion between periphery atoms and the surface.

and the carbon faces are flattened out on the surface (x-y plane).

If such a precursor was now closed on the innermost hinges, these hinges could only close by removing the arms upwards from the surface. The periphery atoms at the cavities would necessarily have to be trapped on the inside of the precursor. This would mean, that all precursor molecules would have to take the reaction path with the atoms placed on the inside of the bowl. This is a possible explanation, why the surface assisted dehydrogenation portraits extremely high yields with nearly 100 $\%$ and why a non-surface assisted auto assembly would most likely take the "outside" reaction path with a significantly higher activation barrier and lower yields.

The geometric properties of the inside-reaction path have been analysed as before and yielded very similar results to the outside-paths for both halogens, with a maximal hexagon periphery bond range of 0.074Å, a negligible hexagon-angle range, a maximal pentagonal-angle range of $2.18°$ and least-squares face distance range of 0.099Å. The results can be found in appendix A.2.

The inside reaction path has shown to dehydrogenate the precursor molecule at carbon-carbon distances where the outside path did not. To understand how the choice of halogen would affect such an inside reaction path the differences between fluorine and chlorine should be considered. The bond length between carbon and chlorine is larger than the bond length between carbon and fluorine. If inserted at the cavity, the chlorine atom would thereby be closer to the hydrogen atom on the other side of the cavity. Because it is closer the closing of the hinge would lead to stronger bending of the chlorine atom inside of the bowl than for the fluorine atom. Although this could be seen as a better starting position for the reaction to happen, because a larger out of plane bending of the chlorine atom would make the formation of HCl energetically favourable at a larger distance, there are two effects which could make fluorine the better choice as a halogen.

The first is the bond stability. The fluorine carbon bond is much stronger than the chlorine carbon bond. Although this is not preferable for the finished precursor, it also has to be kept in mind that the precursor molecule has to be synthesised. When thinking about large fullerenes, in the range of several hundred carbon atoms, the bond stability is an important advantage of fluorine over chlorine. The higher stability would mean, that the precursor is more stable during the synthesis.

The other advantage of the fluorine atom is the shorter bond length. A larger bond length between carbon and halogen atom might be associated with a lower activation barrier, but it could negatively influence the selectivity of the carbon-carbon bond formation. If the halogen atoms are less stiffly bound and have more opportunity to move around, the chance that they will accidentally form a molecule with the "wrong" hydrogen atom increases. This would be followed by the wrong carbon-carbon bond

formation and render the precursor useless.

As for the general shape of precursor molecules, the obtained results have shown, that as soon as the three HF atoms are formed, the new carbon bond will result in lower energies for a more closed precursor. As soon as the carbon-carbon bonds are formed the molecule can be believed to introduce curvature and thus make further dehydrogenation more likely.

To generalise this to all precursor molecules, the understanding of curvature from above has to be revisited. In fullerene molecules, the curvature is introduced by the pentagons, and more precisely, by pentagons which are surrounded by at least three adjacent hexagons. In the C_{60} precursor the pentagons have three neighbouring hexagons, but only two of them adjacent. As soon as the carbon-carbon bonds are formed, there are now four neighbouring hexagons around each pentagon. For the carbon atoms of the newly formed hexagons to be at the benzene ring bond length, the curvature of the molecule has to be introduced.

In general, the structure of any precursor should be in a way, such that the first bonds which are to be formed introduce as much curvature as possible. So it is preferable for the cavities, to sit around a pentagon or for the bonds which are formed, to form a pentagon. Furthermore, hinges on which the largest torque acts should be the one responsible for bringing the carbon atoms close to each other. Those hinges will be the ones with the longest arms connected (most torque). Interestingly, all precursors which were described in chapter 1 fulfil these criteria.

As the C_{60} precursor is a very well working precursor, the distance between carbon atoms which should form the first bond in any precursor, could be required to be at roughly the same length of $d_{CC} = 3.45$Å (flat molecule). If the distance were to be much larger, it is unlikely to trap a periphery atom inside of the precursor. Larger distance would also make the formation of the wrong carbon bonds more likely.

4.3 Summary

The comparison of the different results from the functionals to the CC calculations has shown, that all three chosen functionals B3LYP, CAM-B3LYP and M062X can qualitatively reproduce the geometric structure of the two molecules that was obtained by the CCSD optimisation. This was found with and without the addition of explicit dispersion terms. All six calculations also yielded carbon-carbon distances in the 1.3 Butadiene molecule that were both, close to the literature value and the CCSD results, with the M062X functional being significantly better than the other two. For the capture distance at which the halogen assisted dehydrogenation is happening, no clear favourite can be determined due to the low statistic significance of the quadratic fit. In general, all of these values are very close to each other.

Regarding the geometric properties, it can be said, that although all functionals can determine the overall development of the bond lengths and out of plane bending correctly, the M062X functional with and without dispersion is consistently close to the CCSD results, while CAM-B3LYP does estimate the out of plane bends well but is less accurate for the description of the carbon-hydrogen bond stretching. Because this bond experiences very small relative changes, the CAM-B3LYP functional can still be recommended for determining geometric quantities of fullerene precursors.

$\Delta(\Delta H)$[Hartrees]			
	B3LYP	CAM-B3LYP	M062X
No dispersion	0.0054	0.0022	0.0022
GD3	0.0028	0.0027	0.0004

Table 4.1.: Difference in reaction enthalpy (ΔH) for the three different functionals and the CCSD calculations.

The energy comparison for the reaction enthalpy of the different functionals lead to much larger differences. The B3LYP functional leads to results that are furthest away from the CCSD result. Both other functionals yield better results. The addition of dispersion brings the M062X enthalpy closets to the CCSD calculations, but it negatively affects the CAM-B3LYP result. All the values can be found in table 4.1. Although the addition of empirical dispersion did not consequently lead to better results, it improved the accuracy in two of three cases and can be expected to yield better results for larger systems with more dispersive contributions. The B3LYP functional does not seem to be an appropriate choice for the energetic description of the dehydrogenation reaction and further calculations regarding the energy should be performed with the more expensive functionals CAM-B3LYP or M062X.

At no point along the closing path of the precursor molecule did the calculations lead to a cyclo-dehydrogenation. For both studied halogens the precursor geometry reacted with large out-of-plane bending of the periphery atoms at the cavities and an overall twisting of the arms. Although the deformations were associated with large rises in energy, the halogen-halide formation was not preferred over the precursor deformation. Even for carbon-carbon distances of 1.801Å (chlorine) and 1.251Å (fluorine) did the carbon-halogen and carbon-hydrogen bonds remain intact. This is remarkable because the lowest analysed carbon-carbon distance 1.251Å is significantly smaller than the optimal hexagon bond length.

The placement of the key periphery atoms on the inside of the closing bowl did have a very large effect on the reaction path. The inside placement was associated with much larger energies than the outside placement, but the halogen halide formation and followed cyclo-dehydrogenation was energetically favourable at a carbon-carbon distance between 2.3Å and 2.45Å and further closing of the hinges lead to a significant decrease in the energy of the system. This suggests that the inside placement of the periphery atoms could initiate the cyclo-dehydrogenation at much smaller deformations of the precursor molecule. Any method that could force the precursor to take such a reaction path could be expected to result in higher yields for the auto-assembly rate.

For all studied precursor molecules and closing paths, the bond lengths between the carbon atoms remained very close to their initial values. The maximal amplitude occurred in the periphery carbon bonds of the hexagon rings with 0.114Å. The angles within the hexagonal and pentagonal rings remained equally stable with a much higher amplitude for the pentagon rings. The maximal amplitude in angular deformation was $5.9°$. The faces remained remarkable planar along the whole path, with a maximal average distance from the least square plane of 0.1Å. The small deformations for these quantities suggest, that a Taylor expansion around those three minimum values is valid even for large deformations of the hinges of the precursor molecules. This can justify the assumption that a closing path would most likely have large deformations acting on the hinges while leaving bonds, angles and faces close to their optimal values. Any force field parameterisation implemented in the future should include harmonic

energy contributions from these three quantities.

Conclusion and Outlook

5

In this thesis, a model for the description of fullerene precursor molecules as rigid faces is developed. The method provides the possibility to obtain flat precursor geometries directly from the bond graph, as well as being able to obtain all intermediate geometries if the closed fullerene shape is known. A software package is provided that generates all intermediate precursor geometries with all possible periphery atoms. The software can produce GAUSSIAN input files systematically from these geometries. GAUSSIAN ".log" files can be automatically processed and analysed, which greatly simplifies the analysis. The repository also contains a prototype for MD implementation for fullerene precursor auto-assembly, along with a qt based visualisation tool. This software should be considered in the larger context of the Folding Carbon Project.

A summary of all results can be found in section 4.3.

The investigation of the performance of different functionals has been compared to the very accurate reference results from CCSD calculations. The overall agreement of DFT and CCSD is a strong indicator of the reliability of the results. By performing an extensive and spatially highly resolved comparison of the geometric and energetic descriptions of the functionals, this work should be considered in the choice of functional for future DFT calculations on fullerene precursor molecules.

By analysing two different reaction paths, it was found that the position of the periphery atoms of precursor molecules during the auto-assembly can strongly influence the activation barrier of the reaction. In general, any method that forces the periphery atoms inside the closing precursor can be expected to be preferable over an outside placement. Surface catalysation is believed to be such a method. For all possible precursors, a placement of the halogen atoms that introduces curvature after the first cyclo-dehydrogenation is suspected to lead to higher auto-assembly rates. This introduction of curvature is suspected to be the starting point for the zipper-like chain reaction.

This work found fullerene precursor molecules to have remarkably stable carbon bonds and in-face angles, along with very high in-plane stability. Those properties are expected to make the simulation of auto-assembly of fullerene precursors accessible through FF methods. The harmonic approximation can be expected to hold for the three geometric quantities analysed. Section 2.7.3 should be seen as an inspiration to how the angular force constants could be calculated from QM frequency calculations.

To determine the effects of dispersion on the results, a DFT studies with and without dispersion could be compared on a large system. The magnitude of the dispersion effect can be estimated from that. To find the best fitting parameter values for the three suggested FF contributions, a larger fullerene unfolding could be optimised and the bond lengths and angles assessed by a clustering algorithm. Once an adequate projection of the Hessian matrix has been found a frequency calculation should be performed on a large precursor molecule with the functional that best described the energy. The force constants obtained by the projection can be inserted into the MD prototype and the optimal geometries

calculated from it compared to the geometries obtained in this work.

As this work did only explore the first three cyclo-dehydrogenations, it would be of interest to take this study further along the whole auto-assembly path and the complete dehydrogenation of the precursor molecule. This would also put the energy barriers in context of other dehydrogenations. Because the hinges with chlorine as halogen atom were not closed further then the values they have in the closed fullerene, it would be interesting to determine what would happen if they were to be closed further. Also could a few calculations of the chlorine containing precursor molecule be performed with the inside placement. This might help to determine how exactly the choice of halogen affects the precursor.

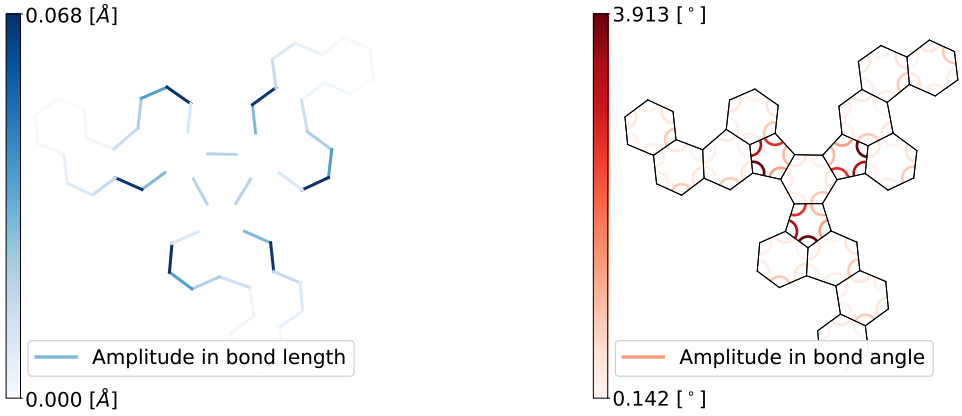
The fast automated assessment of precursor quality can be used to scan the vast precursor space for the most probable candidates. Once they are identified, a synthesis path for the precursor molecules can be developed to systematically produce molecules with a high possibility to auto-assemble. This could mean that it would be possible to produce specific fullerene isomers in macroscopic quantities via rational synthesis. Combined with the automated prediction of molecular properties of fullerenes from the graph (see Folding Carbon Project) this could enable researchers and engineers to automatically find fullerenes that match their demand and systematically produce them from scratch. Given the predicted applications of fullerene molecules, this could mean an enormous step in nanotechnology.

Appendices

Appendix A

A

A.1 Geometric Quantities Along the Closing Path with Chlorine in the Precursor Molecule's key positions



(a) Maximal amplitude of the bond lengths along the path. **(b)** Maximal amplitude of the bond angles along the path.

Figure A.1.: Maximal absolute difference of bond lengths and angles along the outside closing path for chlorine.

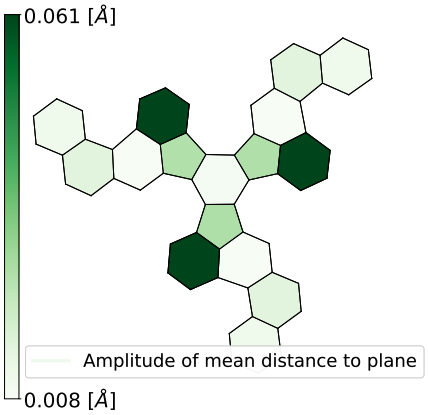
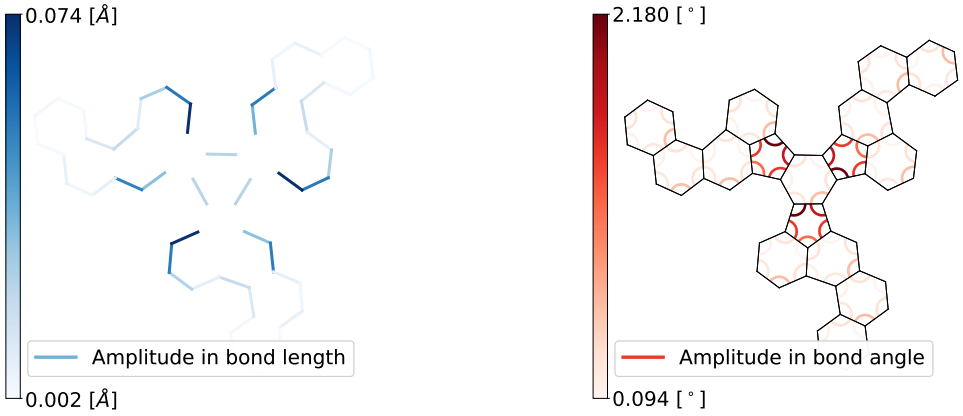


Figure A.2.: Range of the mean absolute distance of all vertices of a face from the mean plane along the outside closing path with chlorine.

A.2 Geometric Quantities Along the Closing Path for the Inside Fluorine Path



(a) Maximal amplitude of the bond lengths along the path. **(b)** Maximal amplitude of the bond angles along the path.

Figure A.3.: Maximal absolute difference of bond lengths and angles along the inside closing path for fluorine.

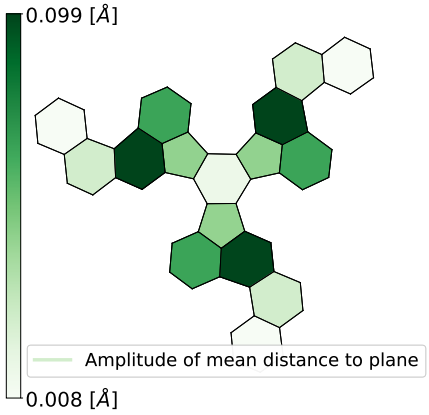


Figure A.4.: Range of the mean absolute distance of all vertices of a face from the mean plane along the inside closing path with fluorine.

Bibliography

A

[1] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, „C60: Buckminsterfullerene", *Nature*, vol. 318, no. 6042, pp. 162–163, 1985.

[2] P. W. Fowler and D. E. Manolopoulos, *An Atlas of Fullerenes*, 2nd ed. Mineola, New York: Dover Publications Inc., 2006.

[3] M. Hasheminezhad, H. Fleischner, and B. D. Mckay, „A universal set of growth operations for fullerenes", eng, *Chemical Physics Letters*, vol. 464, no. 1-3, pp. 118–121, 2008.

[4] G. Brinkmann, J. Goedgebeur, and B. D. Mckay, „The generation of fullerenes.", eng, *Journal of chemical information and modeling*, vol. 52, no. 11, pp. 2910–2918, 2012.

[5] P. Schwerdtfeger, L. Wirz, and J. Avery, „Program fullerene: A software package for constructing and analyzing structures of regular fullerenes", *Journal of Computational Chemistry*, vol. 34, no. 17, pp. 1508–1526, 2013.

[6] J. E. Avery, „Wave equations without coordinates i: Fullerenes", *Rendiconti Lincei. Scienze Fisiche e Naturali*, vol. 29, no. 3, pp. 609–621, Sep. 2018.

[7] L. T. Scott, M. M. Boorum, B. J. McMahon, S. Hagen, J. Mack, J. Blank, H. Wegner, and A. de Meijere, „A rational chemical synthesis of c60", *Science*, vol. 295, no. 5559, pp. 1500–1503, 2002.

[8] K. Y. Amsharov and M. Jansen, „A c78 fullerene precursor: toward the direct synthesis of higher fullerenes", *The Journal of Organic Chemistry*, vol. 73, no. 7, pp. 2931–2934, Apr. 2008.

[9] K. Amsharov and M. Jansen, „Synthesis of a higher fullerene precursor—an "unrolled" c84fullerene", *Chem. Commun.*, pp. 2691–2693, 19 2009.

[10] M. A. Kabdulov, K. Y. Amsharov, and M. Jansen, „A step toward direct fullerene synthesis: C60 fullerene precursors with fluorine in key positions", *Tetrahedron*, vol. 66, no. 45, pp. 8587–8593, 2010.

[11] G. Otero, G. Biddau, C. Sánchez-Sánchez, *et al.*, „Fullerenes from aromatic precursors by surface-catalysed cyclodehydrogenation", *Nature*, vol. 454, no. 7206, pp. 865–868, Jul. 2008.

[12] M. J. Frisch, G. W. Trucks, H. B. Schlegel, *et al.*, *Gaussian~16 Revision C.01*, Gaussian Inc. Wallingford CT, 2016.

[13] P. R. Buseck, S. J. Tsipursky, and R. Hettich, „Fullerenes from the geological environment", *Science*, vol. 257, no. 5067, pp. 215–217, 1992.

[14] E. A. Rohlfing, D. M. Cox, and A. Kaldor, „Production and characterization of supersonic carbon cluster beams", *The Journal of Chemical Physics*, vol. 81, no. 7, pp. 3322–3330, 1984. eprint: https://doi.org/10.1063/1.447994.

[15] P. Schwerdtfeger, L. N. Wirz, and J. Avery, „The topology of fullerenes", *WIREs Computational Molecular Science*, vol. 5, no. 1, pp. 96–145, 2015.

[16] A. H. Francis, „An atlas of fullerenes by p. w. fowler (university of exeter), d. e. manolopoulos (university of nottingham). oxford: new york. 1995. viii + 392 pp. $98.00. isbn 0-19-855787-6.“, *Journal of the American Chemical Society*, vol. 118, no. 21, pp. 5161–5161, Jan. 1996.

[17] J. B. Howard, A. L. Lafleur, Y. Makarovsky, J. T. McKinnon, and M. E. Johnson, „Fullerenes c60 and c70 in flames.“, English, *Nature*, vol. 352, pp. 139–141, 1991, SubjectsTermNotLitGenreText - ANTE; Spectrophotometry; Fullerenes; Flames.

[18] M. A. Cordiner, H. Linnartz, N. L. J. Cox, *et al.*, „Confirming interstellar c60+ using the hubble space telescope“, *The Astrophysical Journal*, vol. 875, no. 2, p. L28, Apr. 2019.

[19] W. Krätschmer, „The story of making fullerenes“, *Nanoscale*, vol. 3, pp. 2485–2489, 6 2011.

[20] R. Otsuki, S. Nasu, R. Fujimori, K. Anada, K. Ohhashi, R. Yamamoto, K. Fujii, and K. Ohkubo, „Preparation of fullerenes by resistive heating vaporization method“, *Journal of the Japan Society of Powder and Powder Metallurgy*, vol. 51, no. 8, pp. 622–625, 2004.

[21] T. Horiyama and W. Shoji, „The number of different unfoldings of polyhedra“, in *Algorithms and Computation*, L. Cai, S.-W. Cheng, and T.-W. Lam, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 623–633.

[22] C. Wentrup, „Flash vacuum pyrolysis: Techniques and reactions“, *Angewandte Chemie International Edition*, vol. 56, no. 47, pp. 14 808–14 835, 2017. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201705118.

[23] D. Kvaskoff, H. Lüerssen, P. Bednarek, and C. Wentrup, „Phenylnitrene, phenylcarbene, and pyridylcarbenes. rearrangements to cyanocyclopentadiene and fulvenallene“, *Journal of the American Chemical Society*, vol. 136, no. 43, pp. 15 203–15 214, Oct. 2014.

[24] I. Fáry, „On straight-line representation of planar graphs“, *Acta Sci. Math. (Szeged)*, vol. 11, pp. 229–233, 1948.

[25] A. Alexandrov, *Convex Polyhedra*, eng, 1st ed. 2005., ser. Springer Monographs in Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg.

[26] C. C. J. Roothaan, „New developments in molecular orbital theory“, *Rev. Mod. Phys.*, vol. 23, pp. 69–89, 2 Apr. 1951.

[27] G. G. Hall, „The molecular orbital theory of chemical valency. viii. a method of calculating ionization potentials“, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 205, no. 1083, pp. 541–552, 1951.

[28] T. Helgaker, P. Jørgensen, and J. Olsen, „Coupled-cluster theory“, in *Molecular Electronic-Structure Theory*. John Wiley & Sons, Ltd, 2014, ch. 13, pp. 648–723. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119019572.ch13.

[29] L. Brillouin, „Les bases de la théorie électronique des métaux et la méthode des champs self-consistents“, ger, *Helvetica Physica Acta*, vol. 7, no. [2], 1934.

[30] J. C. Slater, „The theory of complex spectra“, *Phys. Rev.*, vol. 34, pp. 1293–1322, 10 Nov. 1929.

[31] E. U. Condon, „The theory of complex spectra“, *Phys. Rev.*, vol. 36, pp. 1121–1133, 7 Oct. 1930.

[32] P. Hohenberg and W. Kohn, „Inhomogeneous electron gas“, *Phys. Rev.*, vol. 136, B864–B871, 3B Nov. 1964.

[33] W. Kohn and L. J. Sham, „Self-consistent equations including exchange and correlation effects“, *Phys. Rev.*, vol. 140, A1133–A1138, 4A Nov. 1965.

[34] L. H. Thomas, „The calculation of atomic fields“, *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 23, no. 5, pp. 542–548, 1927.

[35] E. Fermi, „Zur quantelung des idealen einatomigen gases“, *Zeitschrift für Physik*, vol. 36, no. 11, pp. 902–912, Nov. 1926.

[36] F. Bloch, „Bemerkung zur elektronentheorie des ferromagnetismus und der elektrischen leitfähigkeit“, *Zeitschrift für Physik*, vol. 57, no. 7, pp. 545–555, Jul. 1929.

[37] P. A. M. Dirac, „Note on exchange phenomena in the thomas atom“, *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 26, no. 3, pp. 376–385, 1930.

[38] J. C. Slater, „A simplification of the hartree-fock method“, *Phys. Rev.*, vol. 81, pp. 385–390, 3 Feb. 1951.

[39] D. M. Ceperley and B. J. Alder, „Ground state of the electron gas by a stochastic method“, *Phys. Rev. Lett.*, vol. 45, pp. 566–569, 7 Aug. 1980.

[40] S. H. Vosko, L. Wilk, and M. Nusair, „Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis“, *Canadian Journal of Physics*, vol. 58, no. 8, pp. 1200–1211, 1980. eprint: https://doi.org/10.1139/p80-159.

[41] J. P. Perdew and Y. Wang, „Accurate and simple analytic representation of the electron-gas correlation energy“, *Phys. Rev. B*, vol. 45, pp. 13 244–13 249, 23 Jun. 1992.

[42] A. D. Becke, „Correlation energy of an inhomogeneous electron gas: A coordinate-space model“, *The Journal of Chemical Physics*, vol. 88, no. 2, pp. 1053–1062, 1988. eprint: https://doi.org/10.1063/1.454274.

[43] G. Ortiz and P. Ballone, „Pseudopotentials for non-local-density functionals“, *Phys. Rev. B*, vol. 43, pp. 6376–6387, 8 Mar. 1991.

[44] C. Lee, W. Yang, and R. G. Parr, „Development of the colle-salvetti correlation-energy formula into a functional of the electron density“, *Phys. Rev. B*, vol. 37, pp. 785–789, 2 Jan. 1988.

[45] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, „Results obtained with the correlation energy density functionals of becke and lee, yang and parr“, *Chemical Physics Letters*, vol. 157, no. 3, pp. 200–206, 1989.

[46] A. D. Becke, „A new mixing of hartree–fock and local density-functional theories“, *The Journal of Chemical Physics*, vol. 98, no. 2, pp. 1372–1377, 1993. eprint: https://doi.org/10.1063/1.464304.

[47] A. D. Becke, „Density-functional exchange-energy approximation with correct asymptotic behavior“, *Phys. Rev. A*, vol. 38, pp. 3098–3100, 6 Sep. 1988.

[48] A. D. Becke, „Density-functional thermochemistry. iii. the role of exact exchange“, *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993. eprint: https://doi.org/10.1063/1.464913.

[49] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, „Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields“, *The Journal of Physical Chemistry*, vol. 98, no. 45, pp. 11 623–11 627, Nov. 1994.

[50] R. Ditchfield, W. J. Hehre, and J. A. Pople, „Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules“, *The Journal of Chemical Physics*, vol. 54, no. 2, pp. 724–728, 1971. eprint: https://doi.org/10.1063/1.1674902.

[51] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, „A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu“, *The Journal of Chemical Physics*, vol. 132, no. 15, p. 154 104, 2010. eprint: https://doi.org/10.1063/1.3382344.

[52] L. Goerigk, „Treating london-dispersion effects with the latest minnesota density functionals: Problems and possible solutions“, *The Journal of Physical Chemistry Letters*, vol. 6, no. 19, pp. 3891–3896, Oct. 2015.

[53] U. Dinur and A. T. Hagler, „New approaches to empirical force fields“, in *Reviews in Computational Chemistry*. John Wiley & Sons, Ltd, 2007, pp. 99–164. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470125793.ch4.

[54] J. Hill, C. M. Freeman, and L. Subramanian, „Use of force fields in materials modeling“, in *Reviews in Computational Chemistry*. John Wiley & Sons, Ltd, 2007, pp. 141–216. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470125939.ch3.

[55] P. M. Morse, „Diatomic molecules according to the wave mechanics. ii. vibrational levels“, *Phys. Rev.*, vol. 34, pp. 57–64, 1 Jul. 1929.

[56] B. M. Rode, „On the relative stability of cycloalkanes“, *Monatshefte für Chemie / Chemical Monthly*, vol. 112, no. 8, pp. 911–916, Aug. 1981.

[57] K. J. Miller, R. J. Hinde, and J. Anderson, „First and second derivative matrix elements for the stretching, bending, and torsional energy“, *Journal of Computational Chemistry*, vol. 10, no. 1, pp. 63–76, 1989. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540100107.

[58] J. M. Seminario, „Calculation of intramolecular force fields from second-derivative tensors“, *International Journal of Quantum Chemistry*, vol. 60, no. 7, pp. 1271–1277, 1996. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/\%28SICI\%291097-461X\%281996\%2960\%3A7\%3C1271\%3A\%3AAID-QUA8\%3E3.0.CO\%3B2-W.

[59] A. E. A. Allen, M. C. Payne, and D. J. Cole, „Harmonic force constants for molecular mechanics force fields via hessian matrix projection“, *Journal of Chemical Theory and Computation*, vol. 14, no. 1, pp. 274–281, Jan. 2018.

[60] T. Yanai, D. P. Tew, and N. C. Handy, „A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp)“, *Chemical Physics Letters*, vol. 393, no. 1, pp. 51–57, 2004.

[61] G. Herzberg, *Electronic spectra and electronic structure of polyatomic molecules*, English. New York: Van Nostrand, 1966.

[62] D. O'Hagan, „Understanding organofluorine chemistry. an introduction to the c–f bond“, *Chem. Soc. Rev.*, vol. 37, pp. 308–319, 2 2008.

[63] G. E. Bacon, N. A. Curry, and S. A. Wilson, „A crystallographic study of solid benzene by neutron diffraction“, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 279, no. 1376, pp. 98–110, 1964.

Acknowledgements

I want to thank James Avery for his patience and the time he dedicated to my many questions. He has shown incredible endurance in his will to make me understand Quantum Mechanics. I further want to thank him for arousing my interest in the field of fullerene research.
I would like to thank Kurt Mikkelsen for his crucial insight into the subject of Chemistry, without whom this work would not have been possible.
I also want to thank my office/garden colleagues for their joyful company, and the good memories.
My gratitude and affection belongs to all the wonderful friends, old and new, that have made my time in Copenhagen one of the most enjoyable experiences of my life.
Finally, I want to thank my family, not only for their unconditional support and love, but also for making me the person I am.