UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS
DEPARTMENT OF MATHEMATICS

Nino Bašić

ALGEBRAIC APPROACH TO SEVERAL FAMILIES OF CHEMICAL GRAPHS

Doctoral thesis

Adviser: Prof. Dr. Tomaž Pisanski
Coadviser: Prof. Dr. Patrick W. Fowler, FRS

Ljubljana, 2016

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA MATEMATIKO

Nino Bašić

ALGEBRAJSKI PRISTOP K OBRAVNAVI NEKATERIH DRUŽIN KEMIJSKIH GRAFOV

Doktorska disertacija

Mentor: prof. dr. Tomaž Pisanski
Somentor: prof. dr. Patrick W. Fowler, FRS

Ljubljana, 2016

Izjava o avtorstvu

Podpisani Nino Bašić izjavljam:

- da sem doktorsko disertacijo z naslovom *Algebrajski pristop k obravnavi nekaterih družin kemijskih grafov* oz. v angleškem jeziku *Algebraic approach to several families of chemical graphs* izdelal samostojno pod mentorstvom prof. dr. Tomaža Pisanskega in somentorstvom prof. dr. Patricka W. Fowlerja ter
- da Fakulteti za matematiko in fiziko Univerze v Ljubljani dovoljujem objavo elektronske oblike svojega dela na spletnih straneh.

Ljubljana, 30. 4. 2016 Podpis:

Abstract

This work is an attempt to establish a stronger link between mathematics and chemistry and also to introduce discrete structures, e.g. *maps*, into the field of mathematical chemistry.

We present the *Hückel Molecular-Orbital* theory and focus our attention to the notion of *free valence*. It is assumed in the literature that the maximum π *bond number* (i.e., the total π bond order around a sp^2 carbon atom) that can be theoretically obtained (on any centre in any sp^2 π system) is no larger than $\sqrt{3}$. This statement does not appear to have been formally proved. We obtained some partial results. We also provide empirical evidence on the behaviour of maximum π bond number as a function of vertex count, n, of chemical graphs and describe the family of graphs that realises local maxima for small n.

In 2013, a group of scientists led by Roman Jerala successfully designed a *self-assembled tetrahedral polypeptide*. We describe a suitable mathematical model for self-assembly of polypeptide structures. We also provide a *dynamic programming* algorithm for enumeration of *strong traces*, i.e., double traces of a graph that have additional properties.

In 2012 the interesting family of *convex benzenoids* was introduced by Cruz et al. We present several equivalent definitions of convex benzenoids and explain some of their properties. In OEIS the sequence A116513 by A. C. Wechsler represents their enumeration. S. Reynolds enumerated and listed them all up to 250 hexagons. Our study independently verifies their enumeration. Furthermore, we stratify their generation into what we call the *fundamental families of convex benzenoids*. We provide an algorithm which extends the table up to 10^6 hexagons.

In this work we also revisit *coronoids*, in particular *multiple coronoids*. We consider a mathematical formalisation of the theory of *coronoid hydrocarbons* that is solely based on incidence between hexagons of the infinite hexagonal grid in the plane.

We also consider *perforated patches*, which generalise coronoids: in addition to the hexagons of any benzenoid, other polygons may also be present. Just as coronoids may be considered as benzenoids with holes, perforated patches are patches with holes. Both cases, coronoids and perforated patches, admit a generalisation of the *altan* operation that can be performed at several holes simultaneously. A formula for the number of Kekulé structures of a generalised altan can be derived easily if the number of Kekulé structures is known for the original graph. *Pauling Bond Orders* for generalised altans are also easy to derive from those of the original graph.

Math. Subj. Class. (2010): 92E10, 05C10, 05C90.

Keywords: altan, generalised altan, iterated altan, benzenoid, coronoid, patch, perforated patch, Kekulé structure, Pauling Bond Order, Pentagonal Incidence Partition, map trace.

Povzetek

To delo je poskus vzpostavitve krepkejše povezave med matematiko in kemijo. Polega tega poskušamo diskretne strukture, kot so npr. zemljevidi, uveljaviti v matematični kemiji.

Najprej predstavimo *Hücklovo teorijo molekulskih orbital*, posebno pozornost pa namenimo konceptu *proste valence*. V literaturi je pogosto predpostavljeno, da je največje π *vezno število* (tj. skupna vsota redov π vezi, ki izhajajo iz nekega sp^2 ogljikovega atoma), ki ga je mogoče teoretično doseči (pri poljubnem atomu v poljubnem sp^2 π sistemu), največ $\sqrt{3}$. Vendar vse kaže, da ni bila ta domneva nikoli formalno dokazana. V tem delu smo uspeli dobiti nekatere delne rezultate. Poleg tega postrežemo z izračuni obnašanja π veznega števila kot funkcije števila atomov n v družini kemijskih grafov in opišemo družino grafov, ki dosežejo lokalne maksimume za manjše vrednosti parametra n.

Leta 2013 je skupina znanstvenikov pod vodstvom Romana Jerale uspešno izdelala *samosestavljiv polipeptid*, ki se je zložil v *tetraeder*. Najprej podamo matematični model, ki je primeren za opis samosestavljanja. Nato predstavimo algoritem, ki s pomočjo *dinamičnega programiranja* našteje *krepke obhode*, tj. dvojne obhode, ki imajo še neke dodatne lastnosti.

Leta 2012 so Cruz in sodelavci uvedli zanimivo družino *konveksnih benzenoidov*. V tem delu predstavimo več ekvivalentnih definicij konveksnih benzenoidov in nekatere njihove lastnosti. V enciklopediji OEIS zaporedje z oznako A116513, ki ga je definiral A. C. Wechsler, predstavlja njihovo enumeracijo. S. Reynolds je preštel in poiskal vse primerke, ki imajo največ 250 šestkotnikov. Naša študija neodvisno potrdi pravilnost njihove enumeracije. Konveksne benzenoide razdelimo v tako imenovane *fundamentalne družine*, generiranje pa opravimo v vsaki družini posebej. S takšnim pristopom z lahkoto preštejemo vse konveksne benzenoide, ki imajo do 10^6 šestkotnikov.

V tem delu se posvetimo tudi *koronoidom*, še posebej *večkratnim koronoidom*. Predstavimo matematično formalizacijo teorije *koronoidnih ogljikovodikov*, ki temelji zgolj na sosednosti med šestkotniki neskončne šestkotniške mreže v ravnini.

Nekaj pozornosti namenimo še *naluknjanim obližem*, ki posplošijo koronoide. Poleg šestkotniških smejo imeti tudi lica drugačnih dolžin. Tako kot lahko koronoide obravnavamo kot benzenoide z luknjami, lahko tudi naluknjane obliže obravnavamo kot obliže z luknjami. Na enih in na drugih lahko naredimo posplošeno operacijo *altan*, ki poteka na več luknjah hkrati. Izpeljemo formulo, ki prešteje Kekulejeve strukture posplošenega altana, če je število Kekulejevih struktur originalnega grafa že znano. Tudi *Paulingov red vezi* lahko enostavno izračunamo za altan, če že od prej poznamo njihove vrednosti v osnovnem grafu.

Math. Subj. Class. (2010): 92E10, 05C10, 05C90.

Ključne besede: altan, posplošeni altan, iterirani altan, benzenoid, koronoid, obliž, naluknjan obliž, Kekulejeva struktura, Paulingov red vezi, petkotniška incidenčna particija, obhod zemljevida.

Contents

Chapter 1

Introduction

Recent work in *mathematical chemistry* has developed various tools for analysis of various physical properties and effects of importance in physical chemistry and molecular physics. Recent developments have drawn attention from mathematicians working in applications of *linear algebra* in graph theory [143, 144]. On the other hand, several tools from algebraic and topological graph theory have been shown to be suitable for construction, description and analysis of carbon-based structures, such as *hexagonal systems* [102, 52] and *fullerenes* [70, 186]. In this work, we treat several problems in mathematical chemistry from a mathematical point of view. We also present several chemical concepts that establish a stronger link between mathematics and chemistry. Among various families of *molecular graphs*, *hexagonal systems* will be given special attention.

Chapter 2 is a preliminary chapter in which we define several mathematical objects that are needed to follow the thesis. In addition to *graphs*, we also present *polygonal surfaces* and *maps* which are used in Section 3.5 and in Chapters 5 and 6. Section 2.4 is devoted to *point groups*. We describe the *Schönflies notation* that is normally used by chemists. Section 2.6 is an overview of results from spectral graph theory which play an important role in the *Hückel Molecular-Orbital theory* presented in Section 3.2. We define important concepts such as *atomic charge*, *bond order* and *free valence*.

Chapter 3 introduces various chemical concepts via graph theory. It is aimed at mathematicians to give them better understanding of the chemistry that lies behind the mathematical problems which we consider in this work. Chapter 3.1 describes the basics of *chamical bonding*. In Chapter 3.2 we start off with the *Schrödinger wave equation* and make a series of approximations that lead us to the *Hückel theory*. This is the simplest quantum-chemical theory of unsaturated systems that still gives meaningful results. We show that Hückel theory is strongly linked to linear algebra. Section 3.3 introduces *fullerene graphs* which are an important class of chemical graphs. They are used to model *fullerenes*, i.e., allotropes of carbon that are composed of closed-cage molecules that contain only pentagonal and hexagonal rings. Here, we introduce the *Pentagonal Incidence Partition* which is a refinement of the concept of *IPR fullerenes*. We also briefly consider symmetries of fullerenes. In Chapter 3.4 we treat *Kekulé structures*, i.e., valence-bond structures on conjugated hydrocarbons in which every C atom is involved in a double bond. A Kekulé structure is a synonym for *perfect matchings*. The theory of matchings in graphs has been

widely studied in the past. We describe algorithms that can determine in polynomial time whether a graph has a perfect matching. *Determinants* are essential tools in linear algebra. They also turn out to be important for counting the perfect matchings. We also describe a polynomial-time algorithm for counting the perfect matchings in *planar graphs*, e.g. *fullerenes* and *coronoids*. This algorithm is based on the *Kasteleyn's Theorem*. In Section 3.4.3 we briefly consider the *conjugated-circuit* approach to π*-electron currents*.

Section 3.5 is devoted to *polyhedral self-assembly*. In 2013, a group of scientists led by Roman Jerala successfully designed a self-assembled tetrahedral polypeptide. We describe a suitable mathematical model for self-assembly of polypeptide structures. Fijavž, Pisanski and Rus defined an important notion of *strong traces*. Those object are of great importance to biochemists, because they tell them how to arrange the peptides on the polypeptide chain to obtain the desired self-assembled structure. A *dynamic programming* algorithm for enumeration of *map traces* (which are closely related to strong traces) is introduced.

In Chapter 4 we investigate the *Coulson conjecture on maximal bond order*. This long-standing conjecture effectively claims that the maximum π *bond number* in any subcubic graph is no larger than $\sqrt{3}$. We provide some partial results and some empirical evidence on the behaviour of maximum π bond number. Moreover, we describe the family of chemical graphs that realises local maxima for graphs on small number of vertices. We also provide results on computer searches within fullerenes and benzenoids.

Chapter 5 treats *hexagonal systems* and their generelisations. We describe a new approach to hexagonal systems that is solely based on incidences between hexagons of the *infinite hexagonal grid*. We describe an algorithm which finds an embedding of a coronoid in the infinite hexagonal grid (or decides that there is no such embedding). We also define a description of benzenoids via their *boundary-edges code* that was popularised by Hansen et al. Recently, a special family of *convex benzenoids* was studied by Gutman et al. [47, 102]. We provide several equivalent definitions of a convex benzenoid and give rigorous proofs of their equivalence. We also investigate the enumeration problem of convex benzenoids. We introduce the families of *infinite benzenoids* and *infinite convex benzenoids*. We show that the former has *uncountably* many members, whilst, the latter has only *countably* many members. Kekulé structure of several families of benzenoids are also considered. We describe tools for determining the existence of a Kekulé structure in a given benzenoid.

In Chapter 6, we shift attention from benzenoids to the more general subcubic planar graphs that we called *patches*, which generalise the *fullerene patches* of Graver et al. A mathematical formalisation which is based on the treatment of benzenoids and coronoids in Chapter 5.4 is given here. We generalise coronoids to *perforated patches*, i.e., to patches with several disjoint holes.

Chapter 7 is devoted to a class of graphs called *altans*. We study *iterated altans* and show their connections with *nanotubes* and *nanocaps*. The term 'altan' was recently coined [148] to describe a particular type of conjugated π-system, defined by a notional expansion of the annulene-like perimeter of a parent hydrocarbon. Mathematical formalisation [99, 100] gives an operation that can be applied to *any* planar graph to produce the altan of the parent graph, and to predict consequent changes in various properties of mathematical/chemical interest. We also apply successive altan operations, not to a single *perimeter* (or *peripheral root*), but to a collection of disjoint perimeters. In particular, a composite operation of this

type applies well to general coronoids, which, unlike benzenoids, may possess more than one perimeter. In addition to the outer perimeter they have a perimeter for each of the holes. We call this operation a *generalised altan*. Owing to its generality it applies to *single* coronoids, i.e., to coronoids possessing exactly one corona hole, and *multiple* coronoids, i.e., coronoids possessing more than one corona hole. It seems that in the past investigations of coronoids mostly *single coronoids* were considered. We also consider Kekulé structures of generalised altans. It turns out to be easy to determine the number of Kekulé structures of a generalised altan, if the number of Kekulé structures are known for the original graph.

The central theme of the dissertation is the development of various tools described above and their applications in practice. An important part of the investigation is focused on automatic computation of geometric, topological and algebraic parameters from pure combinatorial descriptions of molecular graphs. In this respect, concise descriptions of graph families are crucial ingredients. Consultation with chemists help to ensure relevance to applications and dissemination of the results to the user community.

Chapter 2

Preliminaries

In this short chapter, in order to keep this work self-contained, we will define the basic mathematical objects that will be needed to follow the thesis. A self-confident reader may skip this chapter and use it as a reference should the need arise.

2.1 Graphs and pregraphs

Traditionally, a *graph* is defined as a pair $G = (V, E)$, where V is the set of *vertices* and E is a subset of $\binom{V}{2}$, i.e., E is a collection of (unordered) pairs of vertices. Elements of E are called *edges*. The cardinality of $V(G)$ is called the *order* of graph G and the cardinality of $E(G)$ is called the *size* of graph G. Occasionally, we will need more general objects called *pregraphs*.

Definition 2.1. *A* pregraph G *is a quadruple* $G = (V, S, i, r)$ *where* V *is the set of vertices,* S *is the set of* semiedges *(also known as* darts*),* $i\colon S \to V$ *is a mapping that assigns the* initial vertex *to each semiedge, and* $r\colon S \to S$ *is an involution which we call the* reversal involution*.*

Let s be a semiedge. Then $e = \{s, r(s)\}$ is an edge. It is called a *proper edge* if $|e| = 2$ and a *half edge* if $|e| = 1$. If the involution r has no fixed points, then G has no half edges and is called a graph. Elements of $\{i(s), i(r(s))\}$ are the *end vertices* of edge e. Note that graphs defined in this way also allow *parallel edges* (edges with the same set of end vertices) and *loops* (edges with only one end vertex). A graph without loops or parallel edges is a *simple* graph.

Degree of a vertex v in a pregraph G is defined as

$$\deg_G(v) \coloneqq |i^{-1}(v)| = |\{s \in S \mid i(s) = v\}|. \tag{2.1}$$

When the pregraph under consideration is clear from the context, we simply write $\deg(v)$. Note that

$$\sum_{v \in V(G)} \deg(v) = |S(G)|. \tag{2.2}$$

When G is a graph, we can write the above equation as $\sum_{v \in V(G)} \deg(v) = 2|E(G)|$. This equality is known as the *handshaking lemma*. A vertex of degree 0 is called an *isolated*

vertex. Let $\delta(G)$, $d(G)$ and $\Delta(G)$ denote the minimum, average and the maximum degree in G, respectively. A graph is called *regular* of degree k, or simply *k-regular*, when every vertex has degree k. A *cubic* graph is a synonym for a 3-regular graph. If $\Delta(G) \leq 3$ then G is called a *subcubic* graph.

Example 2.1. Let $V = \{v_1, v_2, v_3\}$ and $S = \{s_1, s_2, \ldots, s_9\}$ for the pregraph P. Mappings i and r are given in the Table 2.1. The drawing of pregraph P is shown in Figure 2.1. The

s	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9
$i(s)$	v_1	v_3	v_2	v_2	v_2	v_3	v_1	v_1	v_3
$r(s)$	s_8	s_5	s_6	s_7	s_2	s_3	s_4	s_1	s_9

Table 2.1: Initial vertex map and the reversal involution for pregraph P.

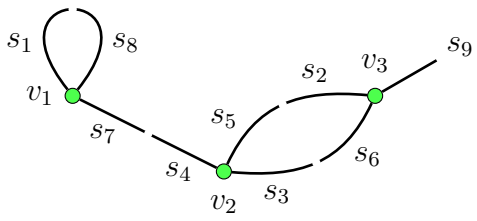


Figure 2.1: Pregraph P

edge $\{s_9\}$ is a half edge, whilst all other edges are proper; $\{s_2, s_5\}$ and $\{s_3, s_6\}$ are parallel edges and $\{s_1, s_8\}$ is a loop.

If vertices u and v of a graph G are connected by an edge $e = uv$, we say that u and v are *adjacent* and their adjacency relationship is denoted by $u \sim v$. The set of all vertices that are adjacent to a given vertex u is called the *neighbourhood* of u, denoted $G(u)$, i.e., $G(u) = \{v \in V(G) \mid v \sim u\}$. A *walk* of length k in a graph G is a sequence

$$W = v_0\ e_0\ v_1\ e_1\ \ldots\ v_{k-1}\ e_{k-1}\ v_k, \tag{2.3}$$

where $v_i \in V(G)$ and $e_i \in E(G)$ such that $e_i = v_i v_{i+1}$. If graph G is simple then the walk is uniquely determined by a sequence of vertices $v_0\ v_1\ \ldots\ v_{k-1}\ v_k$ (if $v_i \sim v_{i+1}$ then there exists exactly one edge e_i such that $e_i = v_i v_{i+1}$). If vertices of a walk W are pairwise distinct then W is called a *(v_0, v_k)-path* in G (or just *path*). If all vertices of W are pairwise distinct except for $v_0 = v_k$ then W is called a *cycle* (of length k) in G.

Let G and H be two graphs. Graph H is a *subgraph* of G, denoted $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. By $G - u$ we denote the subgraph obtained from G by removing vertex u together with its incident edges. If $S \subseteq V(G)$ then $G - S$ denotes the subgraph that is obtained from G by removing all vertices in S. If $e \in E(G)$ then $G - e$ denotes the subgraph obtained from G by removing the edge e. If $F \subseteq E(G)$ then $G - F$ denotes the subgraph that is obtained from G by removing all edges in F. A subgraph H of a graph G is an *induced subgraph* if $E(H) = E(G) \cap \binom{V(H)}{2}$, i.e., if it contains all those edges of G

that have both end vertices in $V(H)$. A subgraph H of a graph G is a *spanning subgraph* if $V(H) = V(G)$.

If $e = uv$ is an edge of the graph G then G/e denotes the graph which is obtained from G by removing the edge e and identifying vertices u and v. This operation is called an *edge contraction*. Let $F \subseteq E(G)$. Then G/F denotes the graph that is obtained from G by consecutively contracting edges in F.

The *distance* between vertices u and v of a graph G, denoted $d_G(u, v)$, is the length of the shortest (u, v)-path. A graph G is *connected* if there exists a path between every pair of vertices; otherwise it is called *disconnected*. A subset $S \subseteq V(G)$ is a *separating set* if $G - S$ is disconnected. The *connectivity* of G, denoted $\kappa(G)$, is the minimum size of some $S \subseteq V(G)$ such that $G - S$ is disconnected or a single vertex. A graph G is *k-connected* if $k \leq \kappa(G)$. The *diameter* of a connected graph G, denoted $\operatorname{diam}(G)$, is the maximum distance between two vertices of G, i.e.,

$$\operatorname{diam}(G) = \max_{u \in V(G)} \max_{v \in V(G)} d_G(u, v). \tag{2.4}$$

The *girth* of a graph G, denoted $\operatorname{girth}(G)$, is the length of the shortest cycle in the graph G. If the graph G is *acyclic* (i.e., it has no cycles) then its girth is defined to be infinity.

A *complete graph*, K_n, is a graph on n vertices such that $E(K_n) = \binom{V(K_n)}{2}$, i.e., every two vertices are connected by an edge. A *path graph*, P_n, is a graph with $V(P_n) = \{1, 2, \ldots, n\}$ and $E(P_n) = \{\{i, i+1\} \mid i = 1, \ldots, n-1\}$. A *cycle graph*, C_n, is a graph with $V(C_n) = \{0, 1, \ldots, n-1\}$ and $E(P_n) = \{\{i, i+1\} \mid i = 0, \ldots, n-1\}$ where the arithmetic is done modulo n.

An *automorphism* of a pregraph is a pair of bijections $(\varphi\colon V \to V, \psi\colon S \to S)$ such that $i(\psi(s)) = \varphi(i(s))$ and $r(\psi(s)) = \psi(r(s))$ for each $s \in S$. Note that if (ψ, φ) and (ψ', φ') are two automorphisms then $(\psi \circ \psi', \varphi \circ \varphi')$ is also an automorphism. The set of all automorphisms together with above composition is a group. It is called the *automorphism group* (also called the *full automorphism group*) of pregraph G and is denoted by $\operatorname{Aut}(G)$. In the case of simple graphs, mapping ψ is redundant and the automorphism is determined by the permutation φ on V. Let $G = (V, E)$ and $G' = (V', E')$ be two graphs. If there exists a bijection $\varphi\colon V(G) \to V(G')$, such that $uv \in E(G)$ if and only if $\varphi(u)\varphi(v) \in E(G')$, then graphs G and G' are said to be *isomorphic*, denoted $G \cong G'$, and φ is a *graph isomorphism*. When $G = G'$, each isomorphism is an automorphism of graph G. Most of the time, the labeling of vertices is not important and therefore we do not make distinction between isomorphic graphs.

A *directed graph* or a *digraph* is a graph where the edges have a direction associated with them. *Directed edges* are more commonly called *arcs*. They are ordered pairs of vertices. This means that $(u, v) \neq (v, u)$ if u and v are two distinct vertices of G.

Ocassionally, we need the notion of an *empty graph*, i.e., the graph whose set of vertices is the empty set.

2.1.1 Trees and bipartite graphs

Definition 2.2. *A graph G is* bipartite *if its vertex set can be partitioned $V(G) = V_1 \sqcup V_2$ such that for every edge $e \in E(G)$ one of its end vertices belongs to V_1 and the other belongs*

to V_2.

In the above expression $\sqcup$ stands for disjoint union and $V_1 \sqcup V_2$ is called the *bipartition* of the vertex set $V(G)$. In Section 2.6, we will see that bipartite graphs have certain nice properties. Coronoid graphs, which will be the centre of our attention in Chapter 5, are also bipartite. The following characterisation of bipartite graphs was proved by König:

Theorem 2.1 (Theorem 1.2.3 in [145])**.** *A graph is bipartite if and only if it contains no cycle of odd length.* □

Definition 2.3. *A* forest *is an* acyclic *graph (i.e., a graph without any cycles). A connected forest is called a* tree.

Trees are among the most important classes of graphs in discrete mathematics, chemistry and computer science. Note that every tree is bipartite (by Theorem 2.1). The following characterisation of trees can be found in almost any reference on graph theory:

Proposition 2.2 (Proposition 1.2.1 in [145])**.** *Let G be a graph on n vertices. Then the following statements are equivalent:*

(i) G is a tree.
(ii) G is connected and has $n-1$ edges.
(iii) G contains no cycles and has $n-1$ edges.
(iv) G is connected, but $G-e$ is disconnected for every $e \in E(G)$.
(v) G has no cycles but addition of any edge results in a graph with a cycle.
(vi) Any two vertices of G are connected by exactly one path.

□

A *leaf* of a tree is a vertex of degree 1. Every tree with $n \geq 2$ vertices has at least two leaves. The path graph P_n, $n \geq 2$, is a tree with exactly 2 leaves. If a tree with $n \geq 2$ vertices has exactly 2 leaves then it is isomorphic to some path P_n. A *caterpillar* is a tree for which identifying and removing all the leaves (together with incident edges) produces a path graph. In other words, it is a tree in which there exists a path that contains every vertex of degree two or more. An example is in Figure 2.2.

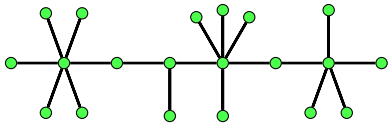


Figure 2.2: A caterpillar

The *complete bipartite* graph $K_{n,m}$ is the graph with the maximum number of edges among all bipartite graphs G such that $|V_1| = n$ and $|V_2| = m$ where $V(G) = V_1 \sqcup V_2$ is the bipartition. Graph $K_{1,n}$ is called the *star graph*.

Complete bipartite graphs can be generalised naturally to complete multipartite graphs. *Complete multipartite* graph $K_{n_1,n_2,\ldots,n_s}$, $s \geq 2$, $n_i > 0$ for $1 \leq i \leq s$, is a graph on $n_1 + n_2 + \cdots + n_s$ vertices, where $V(G) = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_s$ such that $|V_i| = n_i$; vertices $u \in V_i$ and $v \in V_j$ are adjacent if and only if $i \neq j$. If this notion is extended to $s = 1$ then

K_n denotes a *totally disconnected* graph on n vertices. In our definition, we require $s \geq 2$ and notation K_n is reserved for a complete graph on n vertices. (Also, note that $K_{1,1} = K_2$, $K_{1,1,1} = K_3$, etc.)

A *rooted tree* is a tree in which one vertex has a "special status". This special vertex is called the *root* of the tree.

2.1.2 Generating small graphs

Libraries **nauty** and **Traces** [136, 158, 137] are the best known tools for determining the automorphism group of a graph (or a digraph) and for testing graphs for isomorphism. They are written in C and were developed respectively by Brendan D. McKay (nauty) and Adolfo Piperno (Traces). Both libraries can be obtained from `http://pallini.di.uniroma1.it/` free of charge.

The **nauty** library comes with a set of utilities called **gtools**. One of them is **geng** which can generate non-isomorphic graphs very quickly. In principle, it generates all graphs for a given number of vertices n. By using appropriate command-line options, it is possible to set bounds for the number of edges, set bounds for the minimum and the maximum degree, generate only connected graphs, etc. This means that **geng** can generate trees, connected subcubic graphs and various other classes of graphs of interest to us.

2.1.3 Planar graphs

We usually draw figures of graphs in such way that vertices are represented by points and each edge is represented by a line which connects the corresponding pair of points. A figure of a graph is considered to be "nice" if no two lines (or just a few of them) intersect each other. In this section we will give formal mathematical definition of a "drawing" of a graph and present some basic results.

By a *curve*, we mean a continuous image of the unit interval $[0, 1]$ into a topological space X. A *closed curve* is a continuous image of circle S^1 into the space X. A (closed) curve is *simple* if it does not intersect or touch itself.

Theorem 2.3 (Jordan Curve Theorem, Theorem 10.1 in [19])**.** *Any simple closed curve C in the plane $\mathbb{R}^2$ partitions the plane into two disjoint path-connected open sets.* □

A detailed treatment of the above theorem can be found in the paper by Thomassen [194].

Embedding of a graph G into the space X is given by an injective mapping $\varphi\colon V(G) \to X$ (that assigns a point in X to every vertex of the graph) and a family of continuous mappings $\{\varphi_e\colon [0,1] \to X \mid e \in E(G)\}$ such that:

(i) for every $e = uv \in E(G)$ we have $\{\varphi(u), \varphi(v)\} = \{\varphi_e(0), \varphi_e(1)\}$;
(ii) $\varphi_e|_{(0,1)}$ is injective for every $e \in E(G)$;
(iii) $\varphi_e((0,1))$ and $\varphi_{e'}((0,1))$ are disjoint if $e \neq e'$ $(e, e' \in E(G))$;
(iv) $\varphi_e((0,1))$ and the set $\varphi(V(G))$ are disjoint for every $e \in E(G)$.

Images of vertices and edges (subsets of X) are called *points* and *lines*, respectively. While every graph can be embedded in the 3-dimensional space $\mathbb{R}^3$, this is not true for the plane

$\mathbb{R}^2$. A graph that admits embedding in $\mathbb{R}^2$ is called a *planar graph*. Plainly speaking, a graph is planar if its figure can be drawn on the paper in such a way that no two lines cross each other. It is well-known that graphs K_5 and $K_{3,3}$ are *nonplanar*. A *plane* graph is a planar graph together with a fixed embedding. Coronoid graphs, which we will consider in Chapter 5, and fullerenes, which we will consider in Section 3.3, are plane graphs.

Let graph G be embedded in the plane $\mathbb{R}^2$. When the lines and points are removed, the plane is divided into one or more connected components called the *faces* of the embedding. Note that every face f is an open set and its boundary is a disjoint union of some points and lines. The subgraph of G that corresponds to those points and lines is called the *boundary* of face f. There is always precisely one unbounded face and the rest are bounded. The unbounded face is called the *outer face*; the other faces are called *inner faces*. The set of all faces will be denoted $F(G)$. A vertex $v \in V(G)$ is incident to a face $f \in F(G)$ if v is in the boundary of f. Similarly, an edge $e \in E(G)$ is incident to a face f if e is in the boundary of f. Two faces that share an edge are adjacent. If edge e lies on some cycle of G, then e is incident with two faces; otherwise it is incident with exactly one face.

Proposition 2.4 (Proposition 4.2.6 in [61])**.** *In a 2-connected plane graph G, the boundary of every face is a cycle in G.* □

The degree of a face f, denoted $\deg(f)$, is the number of edges in its boundary. If an edge e is incident with only one face, then it contributes 2 to its degree. The following equality is a "dual" analogue of the handshaking lemma:

$$\sum_{f \in F(G)} \deg(f) = 2|E(G)|. \tag{2.5}$$

Theorem 2.5 (Theorem 10.4 in [19])**.** *A graph G is embeddable on the plane if and only if it is embeddable on the sphere.* □

Theorem 2.5 can easily be proved using the *stereographic projection*. Sometimes, it is more convenient to consider the embedding on the sphere. In the plane, the unbounded face has a "special status". On the sphere, every face is bounded, so none of them is distinguished from the others. Using the stereographic projection and by rotating the sphere on which the embedding resides one can easily obtain:

Corollary 2.6 (Proposition 10.5 in [19])**.** *Let G be a planar graph and let f be a face in some planar embedding of G. Then there exists an embedding where f is the outer face.* □

A graph may also be embedded in other surfaces. Let Σ be a surface and G a graph. An embedding of G in Σ is *cellular* if every face is homeomorphic to an open disc in $\mathbb{R}^2$. The following theorem relates the number of vertices, edges and faces in a cellularly embedded graph:

Theorem 2.7 (Euler's Formula, [145])**.** *Let G be a graph that is cellularly embedded in a surface Σ. Then*

$$|V(G)| - |E(G)| + |F(G)| = \chi(\Sigma), \tag{2.6}$$

where $\chi(\Sigma)$ is the Euler characteristic of the surface Σ. □

Euler characteristic χ is a *topological invariant* of the surface in which the graph is cellularly embedded, i.e., homeomorphic surfaces have the same Euler characteristic. It is well known that $\chi(S^2) = 2$ and the Euler formula for planar graph follows:

$$|V(G)| - |E(G)| + |F(G)| = 2. \tag{2.7}$$

Other topological types of surfaces will be considered in Section 2.2.

For convenience, let us give the embedding of a graph G into space X as a mapping $\Phi\colon V(G) \cup (E(G) \times [0,1]) \to X$, such that $\Phi(v) = \varphi(v)$ for $v \in V(G)$ and $\Phi(e,t) = \varphi_e(t)$ for $e \in E(G)$ and $t \in [0,1]$. A plane graph can be treated as a pair (G, Φ), provided that $X = \mathbb{R}$ or $X = S^2$. Let Φ and Ψ be two embeddings of a planar graph G into S^2. Embeddings Φ and Ψ are *topologically equivalent* if there exists a homeomorphism $\mu\colon S^2 \to S^2$, such that $\Phi = \mu \circ \Psi$.

Example 2.2. Figure 2.3 shows four embeddings of a graph in S^2. Embeddings (c) and

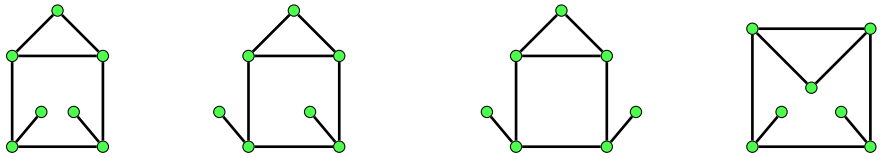


Figure 2.3: Four embeddings of a graph.

(d) are topologically equivalent. Embedding (a) is not topologically equivalent to any other embedding in the figure. Embedding (b) is also not topologically equivalent to any other in the figure. □

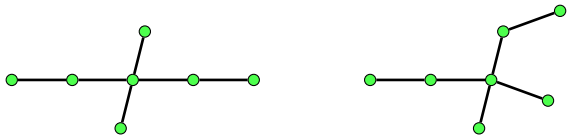
Let $F_\Phi(G)$ and $F_\Psi(G)$ denote the faces of plane graphs (G, Φ) and (G, Ψ), respectively. Let $\nu\colon F_\Phi(G) \to F_\Psi(G)$ be a bijection. Mapping ν is said to preserves vertex-face incidence when $v \in V(G)$ is incident with $f \in F_\Phi(G)$ if and only if v is incident with $\nu(f)$. Similarly, ν preserves the edge-face incidence when $e \in E(G)$ is incident with $f \in F_\Phi(G)$ if and only if e is incident with $\nu(f)$. If there exists a bijection $\nu\colon F_\Phi(G) \to F_\Psi(G)$ that preserves incidence of vertices and edges with faces then Φ and Ψ are called *combinatorially equivalent*.

Note that a homeomorphism of the sphere preserves incidence of vertices, edges and faces. This means that two topologically equivalent embeddings are always combinatorially equivalent.

Example 2.3. Figure 2.4 shows two embeddings of a graph that are combinatorially equivalent, but not topologically equivalent. □

Theorem 2.8 (Theorem 4.3.1 in [61])**.** *Let G be a 2-connected planar graph. Every two embeddings of G that are combinatorially equivalent are also topologically equivalent.* □

Theorem 2.8 tells us that the two notions of equivalence of planar graph coincide in the class of 2-connected planar graphs.



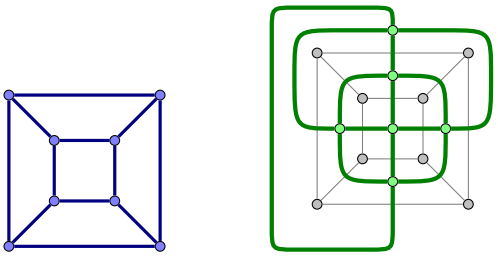
(a) (b)

Figure 2.4: Two combinatorially equivalent embeddings of a tree that are not topologically equivalent.

Theorem 2.9 (Whitney's Theorem, Theorem 4.3.2 in [61])**.** *Any two planar embeddings of a 3-connected graph are equivalent.* □

Definition 2.4. *Let G be a plane graph. The* plane dual*, denoted G^*, of G is a plane graph that can be obtained in the following way: for every face f of G, place a new vertex inside face f. Those new vertices are the vertices of the plane dual. For every edge e of G, create an edge e^* that connects the vertices of G^* which correspond to the two faces that are adjacent to e in G. If e is adjacent to only one face, create a loop at the corresponding vertex. Edge e^* is drawn in such a way that it crosses edge e exactly once.*

Example 2.4. The graph Q_3 in Figure 2.5 is skeleton of the 3-dimensional cube. Its plane dual Q_3^* is skeleton of the octahedron. □



(a) Q_3 (b) Q_3^*

Figure 2.5: The cube Q_3 and its plane dual Q_3^*. In the drawing of Q_3^* one can also see the silhouette of Q_3 (drawn with light gray lines).

2.2 Polygonal surfaces and maps

Let $\{P_i\}_{i\in I}$ be a collection of *polygons*. The polygons do not need to be embedded in a topological space. We may keep this at an abstract level and define polygonal complexes in a purely combinatorial manner. The *boundary* of a polygon is a finite cyclic sequence of edges. Two consecutive edges meet in a common vertex. The boundary of length n may be viewed as a cycle graph C_n. Let us orient the (boundary) edges of polygons. We may partition the set of all oriented edges of those polygons and identify those edges that belong to the same block of the partition. In this way we obtain an (*abstract*) *polygonal complex* Π. If all blocks of the partition are of size 2 and the complex is connected then this polygonal complex is called a *polygonal surface*. If there are also blocks of size 1 in addition of those

of size 2, this complex is called a *polygonal surface with boundary*. Polygons of a polygonal complex are often called 2-*cells* or *faces*. The graph that is obtained from taking vertices and edges of the polygonal complex is called its 1-*skeleton*. For more on this subject the reader is referred to [162].

Example 2.5. Let there be a collection of 3 polygons with boundaries of lengths 5, 4 and 6 (see Figure 2.6(a)). Their edges are labeled with letters $a, b, c, \ldots, o$. We oriented the edges as indicated by arrows in Figure 2.6(a). The orientations which are not indicated in the figure may be chosen arbitrarily. Let the partition of the oriented edges be such that $\{c, f\}$

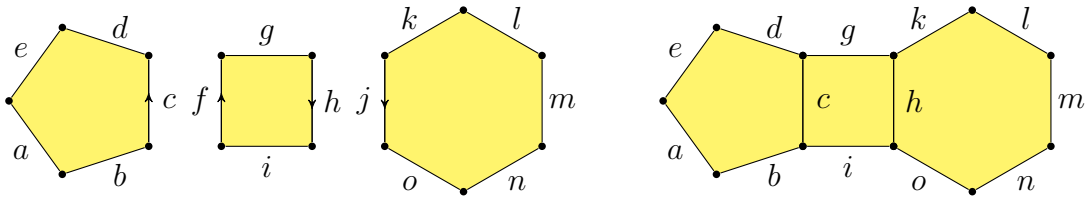


Figure 2.6: A polygonal complex which is also a surface with boundary.

and $\{h, j\}$ are two blocks of size 2 and all the other blocks are of size 1. When the edges are identified, polygonal complex shown in Figure 2.6(b) is obtained. It has 11 vertices, 13 edges and 3 polygons (faces). □

The concept of polygonal surfaces is general enough to represent polyhedra, types of topological surfaces, infinite plane tilings and also cellular embeddings of graphs into surfaces.

Example 2.6. Figure 2.7 shows four different surfaces that can be obtained from a single 4-sided polygon by choosing appropriate edge labels and orientations. The *sphere* $\mathcal{S}$ is

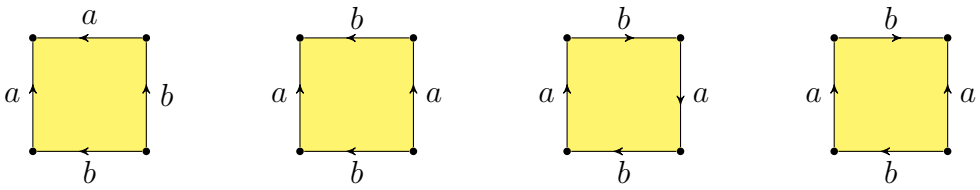


Figure 2.7: Representation of surfaces with polygonal complexes.

shown in Figure 2.7(a). Surface $\mathcal{T}$ in Figure 2.7(b) is the *torus*. Surface $\mathcal{P}$ in Figure 2.7(c) is called the *projective plane* and surface $\mathcal{K}$ in Figure 2.7(d) is called the *Klein bottle*. □

Suppose that we are given two polygonal surfaces Σ_1 and Σ_2, containing the polygons P_1 and P_2, respectively, such that P_1 and P_2 have the same number of boundary edges, say ℓ, and no two edges of P_1 or P_2 are matched. We can create a new polygonal surface, called the *connected sum* of Σ_1 and Σ_2 with respect to P_1 and P_2, denoted $\Sigma_1 \#_{P_1,P_2} \Sigma_2$, as follows. Let us remove P_1 from Σ_1 (obtaining a polygonal surface with boundary) and traverse the

boundary. While traversing the boundary, orient the edges in the direction of traversal and label them by $e_1, e_2, \ldots, e_\ell$. Remove P_2 from Σ_2 and label its boundary in the same way as in the case of Σ_1. The two surfaces with boundary can be glued together along the boundary with respect to the labels and orientations on the boundary edges. The surface obtained in this way is $\Sigma_1 \#_{P_1,P_2} \Sigma_2$. If we choose some other polygons P_1' and P_2' then we may obtain a polygonal complex $\Sigma_1 \#_{P_1',P_2'} \Sigma_2$ which is different from $\Sigma_1 \#_{P_1,P_2} \Sigma_2$, but the two surfaces are always homeomorphic. Most of the time we do not distinguish between homeomorphic surfaces and therefore we write simply $\Sigma_1 \# \Sigma_2$.

Let us define $\Sigma^n \coloneqq \underbrace{\Sigma \# \Sigma \# \cdots \# \Sigma}_{n}$ and $\Sigma^0 \coloneqq \mathcal{S}$. Note that $\mathcal{S} \# \Sigma = \Sigma$ for every surface Σ and that $\mathcal{K} = \mathcal{P} \# \mathcal{P}$ and $\mathcal{P} \# \mathcal{P} \# \mathcal{P} = \mathcal{T} \# \mathcal{P}$. It is also easy to see that

$$\chi(\Sigma_1 \# \Sigma_2) = \chi(\Sigma_1) + \chi(\Sigma_2) - 2. \tag{2.8}$$

It has been known for a long time that all topological types of closed surfaces can be described as connected sums of $\mathcal{S}$, $\mathcal{T}$ and $\mathcal{P}$. Here is the famous classification theorem:

Theorem 2.10 (Classification of Surfaces, Theorem 4.6 in [162])**.** *Every closed surface has the topological type of one of the following:*

(i) the sphere $\mathcal{S}$;
(ii) a connected sum of n, $n \geq 1$, tori $\mathcal{T}^n$;
(iii) a connected sum of n, $n \geq 1$, projective planes $\mathcal{P}^n$.

Moreover, $\chi(\mathcal{S}) = 2$, $\chi(\mathcal{T}^n) = 2 - 2n$ and $\chi(\mathcal{P}^n) = 2 - n$. □

The surfaces $\mathcal{S}$ and $\mathcal{T}^n$ are *orientable* surfaces (two sided), whilst $\mathcal{P}^n$ is *nonorientable* (one sided). Two surfaces are of the same type if and only if they have the same orientability and the same Euler characteristic. For more on this subject consult the reference [145] or reference [162].

There is another combinatorial description of surfaces with which it is sometimes more natural to work. Imagine that in the centre of each polygon in Figure 2.6(b) we draw a point and then draw lines from this point to vertices of the boundary and to midpoints of boundary edges (see Figure 2.9(a)). If we cut the polygons (i.e., perform the *barycentric subdivision*) along those lines we obtain *flags*. We can think of a flag as a right triangle connecting a vertex, the midpoint of an edge and the centre of a face with its right angle at the midpont of edge. A polygon of length n gives rise to $2n$ flags. A map can be formally defined in the following way:

Definition 2.5. *A* map *M is a quadruple $M = (\Phi, \tau_0, \tau_1, \tau_2)$, where Φ is the set of flags, and τ_0, τ_1 and τ_2 are three fixed-point-free involutions on Φ with the following two properties:*

(i) $\tau_0\tau_2 = \tau_2\tau_0$ is fixed-point-free, and

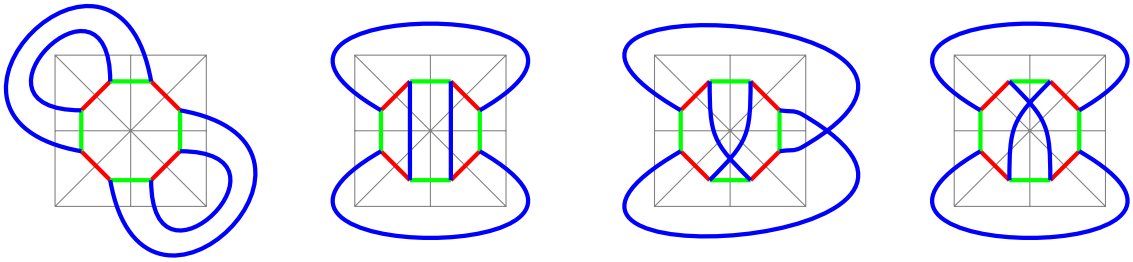
(ii) the group $\langle \tau_0, \tau_1, \tau_2 \rangle$ acts transitively on Φ.

The orbits $V(M)$ of $\langle \tau_1, \tau_2 \rangle$ are vertices, the orbits $E(M)$ of $\langle \tau_0, \tau_2 \rangle$ are edges and the orbits $F(M)$ of $\langle \tau_0, \tau_1 \rangle$ are faces of the map M. We will denote the vertex, edge and face that corresponds to flag $\phi \in \Phi$ by v_ϕ, e_ϕ and f_ϕ, respectively. We can therefore write

$V(M) = \{v_\phi \mid \phi \in \Phi\}$, $E(M) = \{e_\phi \mid \phi \in \Phi\}$ and $F(M) = \{f_\phi \mid \phi \in \Phi\}$. The pair $(V(M), E(M))$ determines a graph which is called the 1-*skeleton* of map M and is denoted by Skel(M). The map M represents the surface in which Skel(M) is cellularly embedded. If there are no loops in Skel(M) or its dual, the map M is called *flag-simple* and each flag can be uniquely identified by a triple (v_ϕ, e_ϕ, f_ϕ).

The *flag graph* of map M is an edge-coloured cubic graph whose set of vertices is the set of flags Φ. The edge set is determined by the involutions τ_0, τ_1 and τ_2. The edges induced by τ_0, τ_1 and τ_2 will be coloured green, red and blue, respectively. For more on the subject of maps the reader is referred to references [162] and [164].

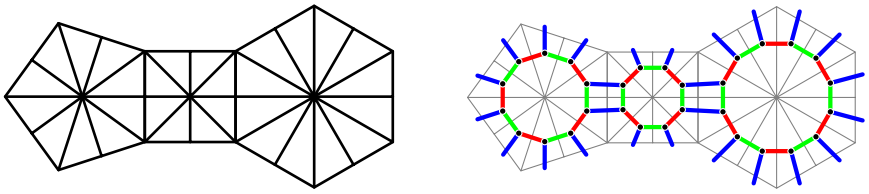
Example 2.7. Figure 2.8 shows flag graphs obtained from the polygonal complexes in Example 2.6. □



(a) $\mathcal{S}$ (b) $\mathcal{T}$ (c) $\mathcal{P}$ (d) $\mathcal{K}$

Figure 2.8: Flag graphs of surfaces from Example 2.6.

Example 2.8. Take the polygonal complex from Example 2.5. It represents a surface with boundary which can be subdivided as shown in Figure 2.9(a). To describe surfaces with



(a) (b)

Figure 2.9: Barycentric subdivision and flag pregraph of polygonal complex from Example 2.5.

boundary, one has to use *flag pregraphs*. Flags on the periphery are adjacant to just two other flags; thus there are pendant blue half edges. The above definition of a map has to be relaxed to allow fixed-points in involution τ_2. □

Flag graphs were used in mathematical chemistry for the first time by Dress and Brinkmann [63]. Later, they were also used by Balaban and Pisanski for describing benzenoids [159].

2.3 Chemical graphs

The notion of a *chemical graph* is somewhat vague. In principle, this can be any graph that has its use in chemistry. *Molecular graphs* are graph-based representations of molecules. Graphs may not only represent molecules, but also chemical processes. In this work, we will focus on certain families of molecular graphs.

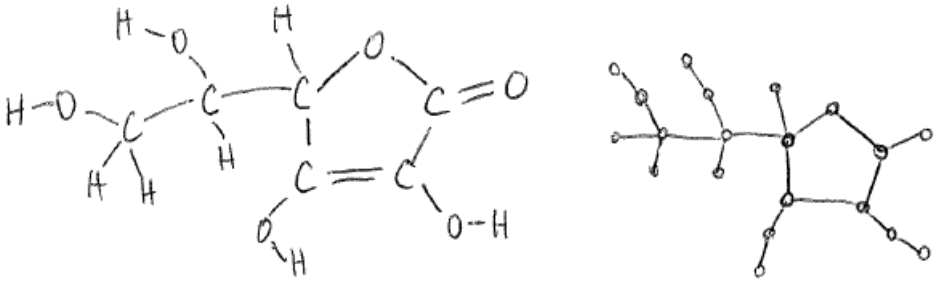
Suppose a molecule contains n atoms labeled with integers $1, 2, \ldots, n$. In the simplest possible model, there may be

(a) either a chemical bond between atoms i and j, or

(b) no chemical bond between atoms i and j.

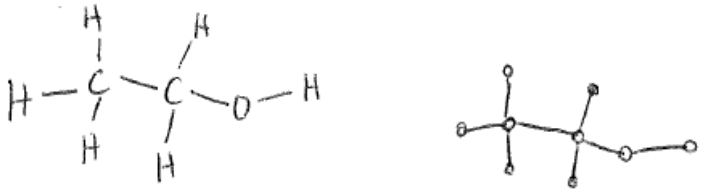
When we model a molecule with a graph, so that each vertex represents an atom and edges represent chemical bonds, we call such a graph the *complete molecular graph*.

Example 2.9. Figure 2.10 displays the structural formula of vitamin C (ascorbic acid) and its complete molecular graph. Note that this graph contains a cycle of length 5. Another example is ethanol (drinking alcohol) in Figure 2.11. Its complete molecular graph is a tree. □



(a) Structural formula (b) Complete molecular graph

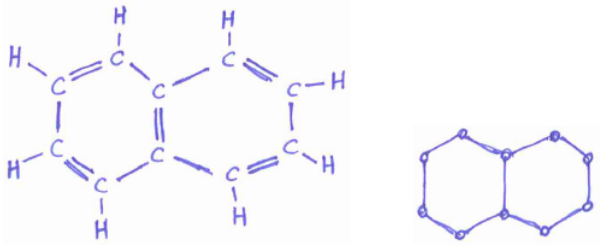
Figure 2.10: Vitamin C



(a) Structural formula (b) Complete molecular graph

Figure 2.11: Ethanol

Sometimes, only those atoms that form the framework of the molecule are considered, i.e., certain atoms (usually H atoms) are neglected. Such molecular graphs are called *skeleton graphs* or *depleted graphs*. Skeleton graphs of fully conjugated systems are sometimes called *Hückel graphs*. We will use them in Section 3.2 when discussing the Hückel theory.



(a) Structural formula (b) Hydrogen-depleted graph

Figure 2.12: Naphthalene

Example 2.10. Figure 2.12 displays the structural formula of *naphthalene* and its Hückel graph, i.e., its hydrogen-depleted graph. However, this structural formula does not accurately describe bonding in naphthalene molecule. □

In the simplest model that is described above, there is no distinction between different elements (from the periodic table). Moreover, we assumed that there exists only one type of chemical bond (two atoms are either bonded or they are not). This simple model may be upgraded if necessary. We could assign additional information to vertices or edges (or both). An *edge-weighted graph* is a graph G together with a function $w\colon E(G) \to X$, where X is some set. (Edge-weighted digraphs are called *networks*.) The function w assigns to every edge a value from X. Similarly, a *vertex-weighted graph* is a graph G together with a function $w\colon V(G) \to X$. We could also use both vertex- and edge-weights to obtain graphs with additional data on both vertices and edges. The edge weights can be numerical values (e.g. $X = \mathbb{R}$) which could represent bond strength or bond length. The vertex weights could be natural numbers, i.e., $X = \mathbb{N}$. In this case the mapping w could be used to assign atomic numbers to vertices.

Some authors define chemical graphs as molecular graphs of fully conjugated π systems, i.e., simple connected subcubic graphs. In this context, a *chemical tree* would be a subcubic tree.

Numerous other types of graphs have been introduced in mathematical chemistry. Balaban defined the *characteristic* (*dualist*) *graph* of a benzenoid system [9]. This graph is the plane dual of the benzenoid graph (without the vertex that corresponds to the unbounded face) in which the geometric information (i.e., angles between edges) is retained. Gutman defined *Gutman trees* for non-branched catacondensed benzenoids [96]. A Gutman tree can be used to obtain the sextet polynomial of its corresponding benzenoid. Clar graphs [98] were also introduced by Gutman and they play an important role in the sextet theory of Clar.

2.4 Molecular symmetry and point groups

One possible approach to the study of symmetries in molecules would be via symmetries of their corresponding molecular graphs. Symmetries of a graph G are elements of its automorphism group $\mathrm{Aut}(G)$. Automorphism groups of graphs are *permutation groups*.

This is a widely studied class of groups [34]. However, a molecular graph only tells us which atoms are bonded and which are not. Geometric information is completely disregarded.

We can take a different approach that takes the geometry into account. Let us represent an atom by a pair $(E_i, \mathbf{x}_i)$, where E_i is a chemical element (from the *periodic table*) and $\mathbf{x}_i$ is its position in $\mathbb{R}^3$. Then a molecule consisting of k atoms is represented by a set

$$\mathcal{M} = \{(E_1, \mathbf{x}_1), (E_2, \mathbf{x}_2), \ldots, (E_k, \mathbf{x}_k)\}. \tag{2.9}$$

This set contains one element for each atom in the molecule. We also require that

$$|\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}| = k, \tag{2.10}$$

i.e., two different atoms can not reside in exactly the same position in space. With other words, $\mathbf{x}_i = \mathbf{x}_j$ implies $i = j$.

Example 2.11. Let us represent the molecule of *methane*, whose molecular formula is CH_4 (see Figure 2.13):

$$\mathcal{M}_{CH_4} = \{(C, (0,0,0)), (H, (0, \tfrac{2\sqrt{2}}{3}, -\tfrac{1}{3}), (H, (\tfrac{\sqrt{6}}{3}, -\tfrac{\sqrt{2}}{3}, -\tfrac{1}{3})),$$
$$(H, (-\tfrac{\sqrt{6}}{3}, -\tfrac{\sqrt{2}}{3}, -\tfrac{1}{3})), (H, (0,0,1))\}.$$

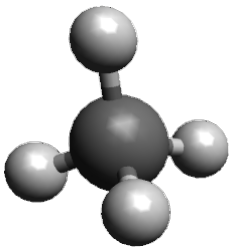


Figure 2.13: Methane

□

Let $\varphi\colon \mathbb{R}^3 \to \mathbb{R}^3$ be an isometry of the 3-dimensional Euclidean space. Define

$$\varphi(\mathcal{M}) \coloneqq \{(E_1, \varphi(\mathbf{x}_1)), (E_2, \varphi(\mathbf{x}_2)), \ldots, (E_k, \varphi(\mathbf{x}_k))\}. \tag{2.11}$$

If $\varphi(\mathcal{M}) = \mathcal{M}$, we will say that φ is a *symmetry operation* on $\mathcal{M}$. There may exist a non-trivial element $\varphi\colon \mathbb{R}^3 \to \mathbb{R}^3$ that fixes $\mathcal{M}$ element-wise. In other words, there may exist two distinct isometries $\varphi_1, \varphi_2\colon \mathbb{R}^3 \to \mathbb{R}^3$ such that $\varphi_1(\mathbf{x}_i) = \varphi_2(\mathbf{x}_i)$ for all $i = 1, \ldots, k$. Even if that happens, we still treat elements φ_1 and φ_2 as two distinct symmetry operations on $\mathcal{M}$. Instead of points, we could also use spheres (or balls) to represent atoms. In such circumstances, because the identity is the only isometry that fixes a sphere pointwise, there would be no doubt that the above φ_1 and φ_2 are indeed two distinct symmetry operations. However, the use of spheres would be an unnecessary complication.

The (unweighted) *centroid* of $\mathcal{M}$ is a point in $\mathbb{R}^3$ defined as

$$c(\mathcal{M}) \coloneqq \tfrac{1}{k}\left(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k\right). \tag{2.12}$$

If $\varphi(\mathcal{M}) = \mathcal{M}$ then φ must fix the centroid of the molecule, i.e., $c(\mathcal{M}) = c(\varphi(\mathcal{M}))$. Therefore, we can assume that $c(\mathcal{M}) = (0,0,0)$ (the coordinate system can be repositioned) and consider only those isometries of $\mathbb{R}^3$ that fix the origin of the coordinate system. It is well known that those isometries comprise a group which is isomorphic to the *orthogonal group* $O(3)$. The group $O(3)$ has been widely studied in the past. This group is isomorphic to the group of real orthogonal 3×3 matrices. The determinant of an orthogonal matrix is either 1 or -1. The matrices of determinant 1 form an important subgroup of $O(3)$ called the *special orthogonal group* $SO(3)$. Elements of $SO(3)$ are sometimes called *pure rotations*.

Example 2.12. Let $\mathcal{M}_{\mathrm{CH_4}}$ be the representation of methane molecule as in Example 2.11. Let

$$\varphi_1 = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \varphi_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

The reader can verify that $\varphi_1(\mathcal{M}_{\mathrm{CH_4}}) = \mathcal{M}_{\mathrm{CH_4}}$, whilst, $\varphi_2(\mathcal{M}_{\mathrm{CH_4}}) \neq \mathcal{M}_{\mathrm{CH_4}}$. This means that φ_1 is a symmetry operation on $\mathcal{M}_{\mathrm{CH_4}}$, whilst φ_2 is not.

Note that $c(\mathcal{M}_{\mathrm{CH_4}}) = (0,0,0)$. □

We are therefore interested in the elements $\varphi \in O(3)$ with the property that $\varphi(\mathcal{M}) = \mathcal{M}$. It is clear that $\{\varphi \in O(3) \mid \varphi(\mathcal{M}) = \mathcal{M}\}$ is a subgroup of $O(3)$. It is the *symmetry group* of $\mathcal{M}$. Those groups are often called *point groups*, because there is a point in the space (in our case the origin of $\mathbb{R}^3$) that is fixed by every symmetry operation. We will describe the *Schönflies notation* [184] which is a system of notation that is normally used by chemists [77]. We have already formally defined a symmetry operation on $\mathcal{M}$. Plainly speaking, it is a movement of the molecule such that after it has been carried out, the molecule looks exactly as it did in its original position. (If we looked away during this movement, we would have not been able to tell whether it was actually performed or not.) Here, we will only enumerate the possible types of *symmetry groups* of molecules and give some of their basic properties. We will follow [39]. For an in-depth threatment of this subject the reader may consult any of numerous references, such as [39, 45].

It turns out that only the following five types of symmetry operations need to be considered:

a) E – *identity*;
b) σ – *reflection* in a plane;
c) i – *inversion* through the centre (also called *central inversion*);
d) C_n – *proper rotation* (i.e., rotation about an axis);
e) S_n – *improper rotation* (i.e., rotation about an axis followed by reflection in the plane perpendicular to the axis of rotation).

The *symmetry element* is a line, plane or point with respect to which one or more symmetry operations may be carried out. We will briefly describe each of the operations. For each of those operations we will also describe the associated symmetry element. All symmetry elements will have at least one point in common – the centroid of the molecule.

- **Identity.** This is the trivial operation that fixes the whole space $\mathbb{R}^3$, i.e., nothing is moved. It is denoted E and is given by the $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ matrix.

- **Reflection.** The symbol for this operation is σ. The symmetry element associated with a reflection is the plane that is fixed by σ. For example, if we take as the symmetry element the plane that is spanned by the x-axis and the y-axis then the reflection associated with this symmetry element is given by the $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ matrix. Note that any planar molecule has at least one plane of symmetry, i.e., its *molecular plane* which is the plane that contains all of its atoms. Note that $\sigma^2 = E$.

- **Inversion.** The symbol for this operation is i. The symmetry element associated with the inversion is the point that is fixed by the operation (in our case the origin of the coordinate system). It is given by the $\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ matrix. Note that $i^2 = E$.

- **Proper rotation.** Let $n \in \mathbb{N}$ and $n \geq 2$. If there exists a line such that rotation by $\frac{2\pi}{n}$ about this line is a symmetry operation then this line is the symmetry element called a *proper axis* (of rotation) or a *n-fold axis*. This operation is called a proper rotation and is denoted C_n. It generates n symmetry operations:

$$C_n, C_n^2, C_n^3, \ldots, C_n^{n-1}, C_n^n = E.$$

  For example, the rotation by $\frac{2\pi}{n}$ about the z-axis is given by the $\begin{bmatrix} \cos\frac{2\pi}{n} & -\sin\frac{2\pi}{n} & 0 \\ \sin\frac{2\pi}{n} & \cos\frac{2\pi}{n} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ matrix.

- **Improper rotation.** Let $n \in \mathbb{N}$ and $n \geq 2$. If there exists a line such that rotation by $\frac{2\pi}{n}$ about this line followed by a reflection in the plane perpendicular to this line is a symmetry operation then this line is a symmetry element called an *improper axis* (of rotation). This symmetry operation is called improper rotation and is denoted S_n. In principle, there are infinitely many planes that are perpendicular to a line. Because the operation has to fix the origin, the plane is uniquely determined (the one that contains the origin). For example, the matrix $\begin{bmatrix} \cos\frac{2\pi}{n} & -\sin\frac{2\pi}{n} & 0 \\ \sin\frac{2\pi}{n} & \cos\frac{2\pi}{n} & 0 \\ 0 & 0 & -1 \end{bmatrix}$ gives the improper rotation by $\frac{2\pi}{n}$ about the z-axis. Note that S_n can be expressed as $S_n = C_n\sigma_h = \sigma_h C_n$ where C_n is a proper rotation about the same axis and σ_h is the plane that is perpendicular to this axis of rotation. It is easy to verify that C_n and σ_h commute. Also, note that $S_n^n = (C_n\sigma_h)^n = C_n^n\sigma_h^n = \begin{cases} E & \text{if } n \text{ is even;} \\ \sigma_h & \text{if } n \text{ is odd.} \end{cases}$

The symmetry element will be denoted by the same symbol as used for its corresponding symmetry operation. For example, the plane that corresponds to the operation σ will also be denoted by σ.

The group $O(3)$ has infinitely many subgroups. However, up to conjugacy in $O(3)$, the symmetry groups of molecules can be classified into several infinite families and some sporadic cases. First, we will consider *linear molecules*, i.e., molecules whose atoms are collinear.

If there is only one atom in the molecule then its symmetry group is the whole $O(3)$. If there are two or more atoms that are collinear then this line is an axis of rotation. Suppose that the molecule is positioned in such a way that its axis of rotation coincides with the z-axis. Then $\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a symmetry operation for every α. Also, any plane containing this axis of rotation is also a symmetry element (associated with the reflection in that plane). There are two options to consider:

i) *The molecule consists of two equivalent halves* (e.g., carbon dioxide with formula O=C=O). In this case there are infinitely many C_2 axes (every line through the origin that is perpendicular to the axis of rotation) and a horizontal plane of symmetry. This group is designated $\mathbf{D}_{\infty h}$.

ii) *The molecule does* not *consist of two equivalent halves* (e.g., hydrogen cyanide with formula H−C≡N). Then there are no other operations in addition to those already described. The group is called $\mathbf{C}_{\infty v}$.

We use bold fonts for groups in order to avoid confusion with symmetry operations. There exist other kinds of infinite point groups, but they can not appear as symmetry groups of molecules (which have finitely many atoms).

Proposition 2.11. *If the molecule has at least 3 non-collinear atoms then its symmetry group must be finite.*

Proof. Let $\mathcal{M}$ be a k-atom molecule with its atoms positioned at $\mathbf{x}_1, \ldots, \mathbf{x}_k$ (see equation (2.11)). We will show that there are finitely many symmetry operations on $\mathcal{M}$. Let $\varphi \in O(3)$ be a symmetry operation on $\mathcal{M}$, i.e., $\varphi(\mathcal{M}) = \mathcal{M}$. There exist two indices $1 \leq i, j \leq k$, $i \neq j$, such that vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ are linearly independent. Without loss of generality, we can assume that the two linearly independent vectors are labeled $\mathbf{x}_1$ and $\mathbf{x}_2$. By the definition of a symmetry operation on $\mathcal{M}$, there exist indices i_1 and i_2, $1 \leq i_1, i_2 \leq k$, such that $\varphi(\mathbf{x}_1) = \mathbf{x}_{i_1}$ and $\varphi(\mathbf{x}_2) = \mathbf{x}_{i_2}$. There are only finitely many ways to choose i_1 and i_2 (some of them may not be admissible). Because φ is a linear mapping, it is uniquely determined on the plane that is spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$. There are exactly two points in $\mathbb{R}^3$ that are simultaneously at unit distance from the origin and at unit distance from the plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$. Let us denote the two points by $\mathbf{y}_1$ and $\mathbf{y}_2$. Note that $\mathbf{y}_1$ is linearly independent of $\mathbf{x}_1$ and $\mathbf{x}_2$. The mapping φ will be uniquely determined on the whole space $\mathbb{R}^3$ by the image of $\mathbf{y}_1$. Because φ is an isometry, there are only two options: either $\varphi(\mathbf{y}_1) = \mathbf{y}_1$ or $\varphi(\mathbf{y}_1) = \mathbf{y}_2$. □

The following groups can be generated by a single symmetry element:

- The simplest finite group is denoted $\mathbf{C}_1$. This is the trivial group (of order 1). This means that the molecule is completely *asymmetric*.

- $\mathbf{C}_s$ is the group of order 2 that is generated by only one reflection σ. $\mathbf{C}_i$ is another group of order 2 that is generated by the inversion i. (Algebraically, i.e., if we consider them as abstract groups, both $\mathbf{C}_s$ and $\mathbf{C}_i$ are isomorphic. However, neither is the conjugate of the other in $O(3)$.) Researchers who are involved with abstract groups would denoted them by Z_2 (*cyclic group* of order 2).

- $\mathbf{C}_n$, $n \in \mathbb{N}$, $n \geq 2$, is a group generated by a single proper axis of rotation C_n. This group of order n would be denoted Z_n in the abstract world.

- $\mathbf{S}_n$, $n \in \mathbb{N}$, is a group that is generated by a single improper axis of rotation. The definiton of $\mathbf{S}_n$ implies that n must be even. (If n is odd then $S_n^n = \sigma_h$ which gives rise to another symmetry element.) Note that S_2 is equivalent to i and in this case the group is called $\mathbf{C}_i$ rather than $\mathbf{S}_2$. In the abstract world, this group of order n is isomorphic to Z_n.

Next, we consider groups where there is only one n-fold axis (proper or improper) where $n \geq 3$ and also other symmetry elements are present. The n-fold axis will be called the *principal axis*:

- If in addition to a proper n-fold axis, C_n, there is also a 2-fold axis, C_2, perpendicular to C_n then there are n 2-fold axes. Operations C_n and C_2 generate a group of order $2n$ denoted $\mathbf{D}_n$, where $n \in \mathbb{N}$, $n \geq 2$. In abstract group theory, this group is called the *dihedral group* and is denoted D_n (or Dih_n).

In what follows, the z-axis will coincide with the principal axis. The plane perpendicular to this axis will be denoted σ_h and called the horizontal plane whilst vertical planes (the ones that include the princpal axis) will be denoted σ_v or σ_d.

- $\mathbf{C}_{nh}$, $n \in \mathbb{N}$, $n \geq 2$, is a group of order $2n$ that is generated by C_n and σ_h. Note that this group is abelian (C_n and σ_h commute). In abstract group theory, this is $Z_n \times Z_2$ (direct product of two cyclic groups).

- $\mathbf{C}_{nv}$, $n \in \mathbb{N}$, $n \geq 2$, is a group of order $2n$ that is generated by C_n and σ_v. If there is one plane of symmetry that contains C_n, there must be n in total. If n is odd then all of them belong to the same conjugacy class and these planes are denoted σ_v. If n is even then there are two conjugacy classes of $\frac{n}{2}$ vertical planes. Planes of one of the classes are denoted σ_v and planes of the other class are denoted σ_d. In the abstract sense, this group is isomorphic to D_n.

- When we add a σ_h to the group $\mathbf{D}_n$ we obtain the group $\mathbf{D}_{nh}$, $n \in \mathbb{N}$, $n \geq 2$, which is of order $4n$. Another way of obtaining $\mathbf{D}_{nh}$ is by adding σ_v to the group $\mathbf{C}_{nh}$ (or adding σ_h to the group $\mathbf{C}_{nv}$). In the abstract sense, this group is isomorphic to $D_n \times Z_2$.

- If we add a dihedral plane σ_d to the group $\mathbf{D}_n$, we obtain a group of order $4n$ denoted $\mathbf{D}_{nd}$, $n \in \mathbb{N}$, $n \geq 2$. This dihedral plane contains the principal axis and dissects the angle between two consecutive C_2 axes. In the abstract group theory, this group is isomorphic to D_{2n}.

The Schönflies notation is rather unfortunate in the sense that n in $\mathbf{D}_{nd}$ or $\mathbf{C}_{nv}$ is a natural number whereas d and v are just symbols. As this notation is well established and widely used, we do not consider ourselves commissioned to change it.

If there is more than one high-order axis (high-order here means of order 3 or more). There are only 7 such groups:

- $\mathbf{T}_d$ is the full symmetry group of the *tetrahedron*. It is of order 24. $\mathbf{T}_d$ has 5 conjugacy classes:
  - Element E, as always, is in its own class.
  - There are three S_4 axes of improper rotation passing through centres of non-incident edges. They generate operations S_4, $S_4^2 = C_2$ and S_4^3. Operations S_4 and S_4^3 are in the same conjugacy class which has 6 elements in total. The three operations C_2 comprise another conjugacy class.
  - There are four C_3 axes passing through apex and the center of the opposite face, generating C_3 and C_3^2 operations that are both in the same conjugacy class which has 8 elements in total.
  - Finally, there are 6 planes of symmetry that include an edge and bisect the centre of the opposite edge. Thus, there are 6 σ_d operations in the last conjugacy class.

  The entire set of operations can be listed by conjugacy classes in the following way:

$$E, 8C_3, 3C_2, 6S_4, 6\sigma_d.$$

- $\mathbf{T}$ is the pure rotational subgroup of $\mathbf{T}_d$, i.e., $\mathbf{T} = \mathbf{T}_d \cap SO(3)$. It is of order 12 and consists of the following 4 conjugacy classes: $E, 4C_3, 4C_3^2, 3C_2$. Note that C_3 and C_3^2 are in the same conjugacy class in $\mathbf{T}_d$, but not in $\mathbf{T}$.

- $\mathbf{T}_h$ is another group of order 24 that can be obtained from $\mathbf{T}$ by adding a set of planes of symmetry σ_h which contain pairs of C_2 axes. It has the following 8 conjugacy classes: $E, 4C_3, 4C_3^2, 3C_2, i, 4S_6, 4S_6^5, 3\sigma_h$.

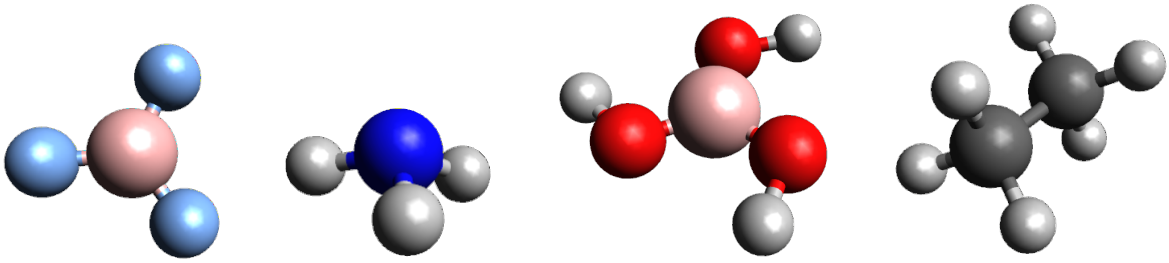
- $\mathbf{O}_h$ is the full symmetry group of the *octahedron* (or *cube*). It is of order 48 and has 10 conjugacy class:
  - As before, there is the identity element E in its own class.
  - There are three S_4 axes passing through opposite apices of the octahedron, generating S_4, $S_4^2 = C_2$ and S_4^3. Operations S_4 and S_4^3 belong to the same conjugacy class that has 6 elements in total. Operations C_2 comprise a class with 3 elements.

- There are three C_4 axes that are collinear with the above S_4 axes. They generate operations C_4, $C_4^2 = C_2$ and C_4^3. Operations C_4 and C_4^3 are in the same conjugacy class that has 6 elements in total. The C_2 operations were already generated above by S_4.
- There are 6 C_2' axes bisecting opposite edges. Those 6 elements comprise another conjugacy class. (They were denoted C_2' to distinguish them from the above C_2 operations.)
- There are four S_6 axes passing through centres of opposite faces generating operations S_6, $S_6^2 = C_3$, $S_6^3 = i$, $S_6^4 = C_3^2$ and S_6^5. Operations S_6 and S_6^5 are in the same conjugacy class that has 8 elements in total. Similarly, C_3 and C_3^2 are also in the same class that has 8 elements. Operation i is in its own conjugacy class.
- Another class consists of three planes of symmetry σ_h passing through 4 coplanar apices.
- Finally, there is a class with 6 planes of symmetry σ_d containing two opposite apices and bisecting two opposite edges.

The entire set of operations by conjugacy classes is:

$$E, 6S_4, 3C_2, 6C_4, 6C_2', 8S_6, 8C_3, i, 3\sigma_h, 6\sigma_d.$$

- $\mathbf{O}$ is the pure rotational subgroup of $\mathbf{O}_h$, i.e., $\mathbf{O} = \mathbf{O}_h \cap SO(3)$. It is of order 24 and consists of the following 5 conjugacy classes: $E, 3C_2, 6C_4, 6C_2', 8C_3$.
- $\mathbf{I}_h$ is the full symmetry group of the *dodecahedron* (or *icosahedron*). It is of order 120 and its elements reside in 10 conjugacy classes:
  - The identity E is in its own class.
  - There are six S_{10} axes passing through centres of opposite faces of the dodecahedron generating operations S_{10}, $S_{10}^2 = C_5$, S_{10}^3, $S_{10}^4 = C_5^2$, $S_{10}^5 = i$, $S_{10}^6 = C_5^3$, S_{10}^7, $S_{10}^8 = C_5^4$ and S_{10}^9. Operations S_{10} and S_{10}^9 belong to the same class that has 12 elements. Similarly, S_{10}^3 and S_{10}^7 belong to the same class that has 12 elements in total. Operations C_5^2 and C_5^3 belong to the same conjugacy class that also has 12 elements in total. Another such pair is C_5 and C_5^4 comprising a conjugacy class with 12 elements. And the remaining operation i is in its own class.
  - There are ten S_6 axes passing through opposite pairs of vertices generating operations S_6, $S_6^2 = C_3$, $S_6^3 = i$, $S_6^4 = C_3^2$ and S_6^5. Operation i was already generated. Operations S_6 and S_6^5 are in the same class which has 20 elements in total. Similarly, operations C_3 and C_3^2 are in the same class which has 20 elements in total.
  - There are 15 C_2 axes bisecting opposite edges and they constitute another conjugacy class.
  - Finally, there are 15 planes of symmetry σ bisecting opposite edges that constitute the last conjugacy class.



(a) BF_3 (b) NH_3 (c) H_3BO_3 (d) C_2H_6

Figure 2.14: A few examples of molecules posessing symmetry.

The entire set of operations by conjugacy classes is:

$$E, 12S_{10}, 12S_{10}^3, 12C_5, 12C_5^2, i, 20S_6, 20C_3, 15C_2, 15\sigma.$$

- $\mathbf{I}$ is the pure rotational subgroup of $\mathbf{I}_h$, i.e., $\mathbf{I} = \mathbf{I}_h \cap SO(3)$. It is of order 60 and consists of the following 5 conjugacy classes: $E, 12C_5, 12C_5^2, 20C_3, 15C_2$.

Example 2.13. The symmetry group of the methane molecule in Figure 2.13 is $\mathbf{T}_d$. We have already seen that *carbon dioxide* (with structural formula O=C=O) has the symmetry group $\mathbf{D}_{\infty h}$. We have also seen that *hydrogen cyanide* (with structural formula H−C≡N) has the symmetry group $\mathbf{C}_{\infty v}$.

In Figure 2.14 there are some more examples. *Boron trifluoride* (with molecular formula BF_3) has the symmetry group $\mathbf{D}_{3h}$. *Ammonia* (with molecular formula NH_3) has the symmetry group $\mathbf{C}_{3v}$. *Boric acid* (with molecular formula H_3BO_3) has the symmetry group $\mathbf{C}_{3h}$ and *ethane* (with molecular formula C_2H_6) has the symmetry group $\mathbf{D}_{3d}$.

Pictures of molecules in Figure 2.14 were created in the **Avogadro** program which can be obtained from `http://avogadro.cc/` free of charge. □

In geometry, there is an important notion of *chirality*. Plainly speaking, a geometric object is *chiral* (it has chirality) if it is not identical to its mirror image. This notion can be formally defined in the language of symmetry groups. An object is *achiral* if its symmetry group has at least one *orientation-reversing isometry*, otherwise it is *chiral*. In other words, if G is the full symmetry group of an object then that object is chiral if and only if $G \cap SO(3) = G$. For example, you can not make your right hand look like your left hand by only rotating your right hand, i.e., your hands are chiral. However, if you look at your right hand in a mirror, it looks exactly like your left hand.

This has also considerable importance in chemistry. Back in the 50's, a German pharmaceutical company developed *thalidomide* (with molecular formula $C_{13}H_{10}N_2O_4$) which was proclaimed the "wonder drug" for morning sickness and was as such issued to thousands of women worldwide. It turned out that thalidomide prevented the proper growth of the fetus, resulting in horrific birth defects such as malformed limbs [197]. Now, it is known that thalidomide is chiral and has two *enantiomers*[1], i.e., the (R)-thalidomide and

[1] According to the IUPAC "Gold Book" [138], an *enantiomer* is one of a pair of molecular entities which are mirror images of each other and are not superimposable (i.e., one of a pair of chiral molecules).

the (S)-thalidomide. The latter caused the severe side-effects. It interconverts in the human body where both enantiomers are formed. This made the disaster inevitable [78].

2.5 Rotagraphs and fasciagraphs

In chemistry, a *polymer* is a large molecule composed of many small molecules called *monomers*. This concept was applied to graph theory in order to define *polygraphs* which are graphs composed of smaller building blocks referred to as *monographs*. It is no surprise that polygraphs were used to describe polymers graph-theoretically [167].

Let $G_0, G_1, \ldots, G_{n-1}$ be arbitrary graphs and let $X_i \subseteq V(G_i) \times V(G_{i+1})$ for $i = 0, \ldots, n-1$ (the indices are considered modulo n). Relations X_i represent edges connecting vertices of G_i to vertices of G_{i+1}. A *polygraph* $\Omega_n = \Omega_n(G_0, \ldots, G_{n-1}; X_0, \ldots, X_{n-1})$ over *monographs* G_0, G_1, ..., G_{n-1} is defined as follows:

$$V(\Omega_n) = V(G_0) \sqcup V(G_1) \sqcup \cdots \sqcup V(G_{n-1}), \tag{2.13}$$

$$E(\Omega_n) = E(G_0) \cup X_0 \cup E(G_1) \cup X_1 \cup \cdots \cup E(G_{n-1}) \cup X_{n-1}. \tag{2.14}$$

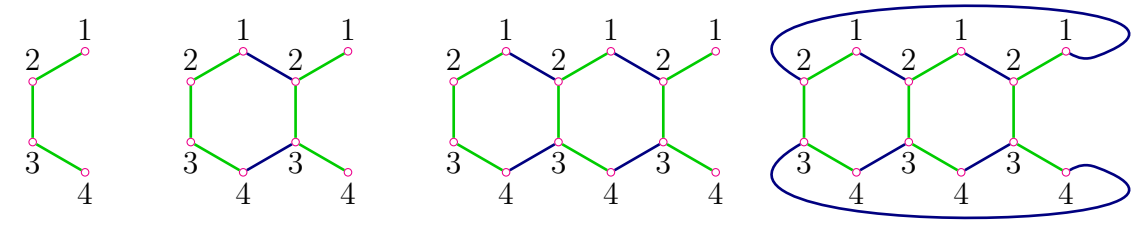
Operator $\sqcup$ in the above definition denotes the disjoint union. This means that even if we take $G_0 = G_1$, two disjoint copies of this graph will be created when the polygraph is formed. Of course, $E(G_0)$ and $E(G_1)$ will connect vertices from their corresponding copies of the graph. For the above polygraph Ω_n and for $i = 0, 1, \ldots, n-1$ we define:

$$D_i = \{u \in V(G_i) \mid \exists v \in V(G_{i+1}) : uv \in X_i\}, \tag{2.15}$$

$$R_i = \{u \in V(G_{i+1}) \mid \exists v \in V(G_i) : vu \in X_i\}. \tag{2.16}$$

Now assume that $G_0 = G_1 = \ldots = G_{n-1}$ and $X_0 = X_1 = \ldots = X_{n-1}$. Let us denote $G \coloneqq G_0 = \ldots = G_{n-1}$ and $X \coloneqq X_0 = \ldots = X_{n-1}$. Then the polygraph $\Omega_n(G_0, \ldots, G_{n-1}; X_0, \ldots, X_{n-1})$ is called a *rotagraph* and is denoted $\omega_n(G; X)$. The polygraph $\Omega_n(G_0, \ldots, G_{n-1}; X_0, \ldots, X_{n-2}, \emptyset)$ is called a *fasciagraph* and is denoted $\gamma_n(G; X)$. A fasciagraph is very similar to a rotagraph, but is without edges between the first and the last copy of the monograph. We may also write $\gamma_n(G; X) \subseteq \omega_n(G; X)$.

Example 2.14. Let $G = P_4$, i.e., a path on 4 vertices with $V(P_4) = \{1, 2, 3, 4\}$, and let $X = \{(1, 2), (4, 3)\}$. Graphs $\gamma_1 \cong P_4$, γ_2, γ_3 and ω_3 are shown in Figure 2.15. □



(a) $\gamma_1 \cong P_4$ (b) γ_2 (c) γ_3 (d) ω_3

Figure 2.15: Examples of rota- and fasciagraphs

Many more examples can be found in [165].

Results on various graph theoretical invariants of rota- and fasciagraphs were obtained in the recent past. Klavžar and Žerovnik [122] proposed a general method based on path algebras for solving various algorithmic questions about rota- and fasciagraphs. Babić, Graovac, Mohar and Pisanski [7, 8] have developed a general method for determining the matching polynomial of a polygraph that can be applied to calculation of other *graph invariants* (*topological indices*) [54, 116, 121, 177].

2.6 Spectral graph theory

Spectral graph theory is a branch of mathematics that studies relations between properties of graphs and spectra of their associated (adjacency) matrices. We will give some basic terminology and results. Our main reference is the monograph [32] by Brouwer and Haemers. Another good reference which is more application-oriented is the book [49] by Cvetković, Doob and Sachs.

Let G be a simple graph with vertices labeled $1, 2, \ldots, n$. The *adjacency matrix*, denoted $A(G)$, is an $n \times n$ matrix indexed by the vertices of G where

$$A_{uv} = \begin{cases} 1, & \text{if } u \sim v; \\ 0, & \text{otherwise.} \end{cases} \tag{2.17}$$

The (ordinary) *spectrum of a graph* of a graph G, denoted $\sigma(G)$, is the spectrum of the adjacency matrix $A(G)$, i.e., its set of eigenvalues together with their multiplicities. If an eigenvalue has its multiplicity greater than 1 then it is called a *degenerate* eigenvalue; otherwise it is called *non-degenerate*. The eigenvalues will be denoted by λ_i, $1 \leq i \leq n$, and ordered in non-increasing order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The i-th eigenvalue of graph G is sometimes denoted $\lambda_i(G)$. Graphs G and G' are called *cospectral* if $\sigma(G) = \sigma(G')$.

Note that the spectrum does not change if we relabel the vertices of graph G. The *characteristic polynomial* of G, denoted $p_G(\lambda)$, is the characteristic polynomial of $A(G)$, i.e.,

$$p_G(\lambda) = \det(\lambda I - A(G)). \tag{2.18}$$

Example 2.15. Let G be a graph on 6 vertices as shown in Figure 2.16. Let us determine

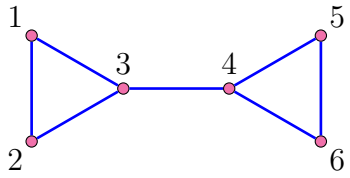


Figure 2.16: A connected graph on 6 vertices.

its spectrum. Its adjacency matrix is

$$A(G) = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{2.19}$$

Its characteristic polynomial is then

$$\begin{aligned} p_G(\lambda) &= \det(\lambda I_{6\times 6} - A(G)) = \lambda^6 - 7\lambda^4 - 4\lambda^3 + 11\lambda^2 + 12\lambda + 3 \\ &= (\lambda+1)^2(\lambda-\sqrt{3})(\lambda+\sqrt{3})(\lambda-1+\sqrt{2})(\lambda-1-\sqrt{2}). \end{aligned} \tag{2.20}$$

The spectrum of G is $\sigma(G) = \{-\sqrt{3}, -1, -1, 1-\sqrt{2}, \sqrt{3}, 1+\sqrt{2}\}$. Note that each component

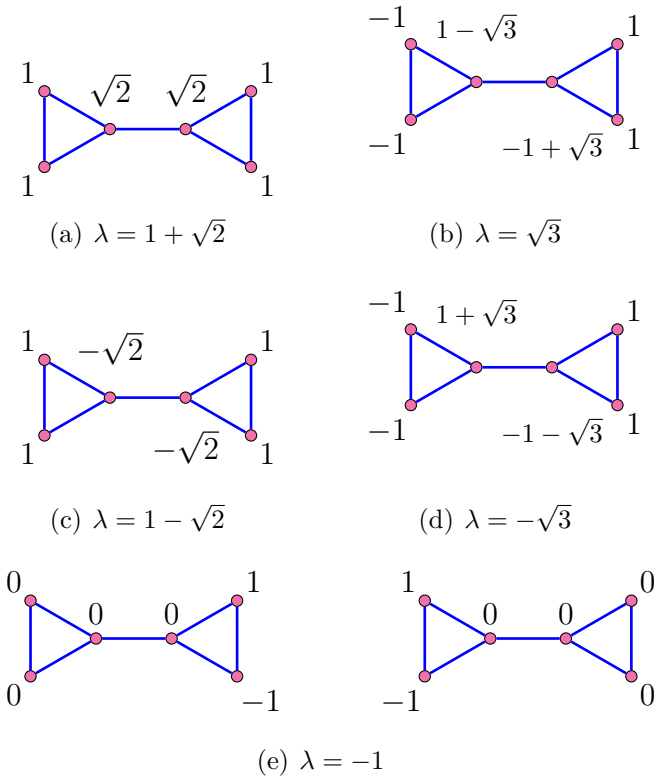


Figure 2.17: Eigenvalues and their corresponding eigenvectors.

of an eigenvector corresponds to a vertex of the graph. The eigenvectors that correspond to eigenvalues determined above are shown in Figure 2.17. Note that the eigenvalue -1 has multiplicity 2 and is therefore a *degenerate* eigenvalue. □

Let $\mathbf{c}$ be the eigenvector associated with the eigenvalue λ of a graph G. From $A(G)\mathbf{c} = \lambda\mathbf{c}$ it follows that

$$\sum_{v \in G(u)} \mathbf{c}_v = \lambda \mathbf{c}_u, \tag{2.21}$$

for every $u \in V(G)$, i.e., if we consider a vertex u of the graph then its component in the eigenvector, denoted $\mathbf{c}_u$, multiplied by the corresponding eigenvalue equals the sum of components of neighbours of u. The reader can verify that this holds for the eigenvectors in Figure 2.17. Sometimes, the easiest way to find the eigenvector is by finding the numbers that satisfy the (2.21) rule.

As G is an undirected simple graph, $A(G)$ is real and symmetric. Therefore, all its eigenvalues are real. Since $A(G)$ has zero diagonal,

$$\operatorname{tr}(A(G)) = \sum_{i=1}^{n} \lambda_i(G) = 0. \tag{2.22}$$

In equation (2.22), $\operatorname{tr}(A(G))$ denotes the *trace* of the adjacency matrix of G.

The adjacency matrix of a k-regular graph has row sums equal to k, i.e., $A(G)\mathbf{1} = k\mathbf{1}$, where $\mathbf{1}$ is the all-ones vector. We see that k must be an eigenvalue of $\mathbf{1}$. The next proposition tell us that a graph with a large diameter will have many distinct eigenvalues:

Proposition 2.12 (Proposition 1.3.3 in [32])**.** *Let G be a connected graph with diameter d. Then G has at least $d+1$ distinct eigenvalues.* □

It is also possible to define spectra of directed graphs. The definition is similar to the case of undirected graphs. As the adjacency matrix of a directed graph is not necessarily symmetric, its eigenvalues are in general complex numbers.

The vertices of the bipartite graph can be labeled in such a way that $V_1 = \{1, 2, \ldots, m\}$ and $V_2 = \{m+1, m+2, \ldots, n\}$, where V_1 and V_2 are blocks of the bipartition. Then adjacency matrix $A(G)$ has the following form:

$$A(G) = \begin{bmatrix} 0_{m\times m} & B(G) \\ B(G)^\intercal & 0_{(n-m)\times(n-m)} \end{bmatrix}, \tag{2.23}$$

where $B(G)$ is an $m \times (n-m)$ matrix called the *biadjacency matrix*. It is easy to see that the spectrum of a bipartite graph is symmetric with respect to 0. If $\begin{bmatrix} u_1 & \ldots & u_m & v_1 & \ldots & v_{n-m} \end{bmatrix}^\intercal$ is an eigenvector with eigenvalue λ then $\begin{bmatrix} u_1 & \ldots & u_m & -v_1 & \ldots & -v_{n-m} \end{bmatrix}^\intercal$ is an eigenvector associated with the eigenvalue $-\lambda$. The converse also holds:

Proposition 2.13 (Proposition 3.4.1 in [32])**.** *Let G be an undirected graph.*

(i) Graph G is bipartite if and only if for each eigenvalue $\lambda \in \sigma(G)$, also $-\lambda \in \sigma(G)$ with the same multiplicity, i.e., if $\sigma(G)$ is symmetric with respect to 0.

(ii) If G is connected with the largest eigenvalue λ_1 then G is bipartite if and only if $-\lambda_1$ is an eigenvalue of G.

□

2.6.1 Spectra of certain families of graphs

Here are spectra for the following families of graphs:

- **The complete graph K_n.** The spectrum is $\{(n-1)^1, (-1)^{n-1}\}$. Here, the notation λ^k means that λ appears with multiplicity k. As K_n is $(n-1)$-regular, $n-1$ is one of its eigenvalues with eigenvector $\mathbf{1}$. It is not hard to see that $\begin{bmatrix}-1 & 1 & 0 & 0 & 0 & \dots & 0\end{bmatrix}^\intercal$, $\begin{bmatrix}-1 & 0 & 1 & 0 & 0 & \dots & 0\end{bmatrix}^\intercal$, $\begin{bmatrix}-1 & 0 & 0 & 1 & 0 & \dots & 0\end{bmatrix}^\intercal$, $\dots$, $\begin{bmatrix}-1 & 0 & 0 & \dots & 0 & 0 & 1\end{bmatrix}^\intercal$ are eigenvector for the eigenvalue -1.

- **The complete bipartite graph $K_{m,n}$.** The spectrum is $\{\sqrt{mn}, 0^{m+n-2}, -\sqrt{mn}\}$. Suppose that degree-$n$ vertices are labeled $1, \dots, m$ and degree-m vertices are labeled $m+1, \dots, m+n$. Let $\mathbf{c}$ and $\mathbf{c}'$ be vectors defined as

$$\mathbf{c}_i = \begin{cases}\sqrt{n}, & \text{if } 1 \le i \le m;\\ \sqrt{m}, & \text{otherwise;}\end{cases} \qquad \text{and} \qquad \mathbf{c}'_i = \begin{cases}\sqrt{n}, & \text{if } 1 \le i \le m;\\ -\sqrt{m}, & \text{otherwise.}\end{cases}$$

  It is not hard to see that $\mathbf{c}$ and $\mathbf{c}'$ are eigenvectors associated with eigenvalues $\sqrt{mn}$ and $-\sqrt{mn}$, respectively. There are $m+n-2$ remaning vectors associated with the eigenvalue 0. These are $\mathbf{d}^{(k)}$ for $2 \le k \le m$ and $\mathbf{e}^{(k)}$ for $m+2 \le k \le m+n$, where

$$\mathbf{d}^{(k)}_i = \begin{cases}1, & \text{if } i = 1;\\ -1, & \text{if } i = k;\\ 0, & \text{otherwise;}\end{cases} \qquad \text{and} \qquad \mathbf{e}^{(k)}_i = \begin{cases}1, & \text{if } i = m+1;\\ -1, & \text{if } i = k;\\ 0, & \text{otherwise.}\end{cases}$$

- **The cycle C_n.** Let us first determine the eigenvalues of the *directed cycle* $\vec{C}_n$. Let $V(\vec{C}_n) = \{0, 1, \dots, n-1\}$. The arcs of $\vec{C}_n$ are $(i, i+1)$ for $i = 0, \dots, n-1$ (all numbers are considered modulo n). It is easy to see that

$$\begin{bmatrix}1 & \zeta & \zeta^2 & \zeta^3 & \dots & \zeta^{n-1}\end{bmatrix} \tag{2.24}$$

  is an eigenvector for eigenvalue $\zeta \in \mathbb{C}$ if $\zeta^n = 1$. In other words, eigenvalues of $\vec{C}_n$ are precisely the complex n-th roots of the unity. Note that $A(C_n) = A(\vec{C}_n) + A(\vec{C}_n)^\intercal$. It is not hard to see that (2.24) is also an eigenvector of $A(C_n)$ associated with the eigenvalue $\zeta + \zeta^{-1}$. By using de Moivre's formula, we obtain the numbers $2\cos\left(\frac{2\pi j}{n}\right)$ for $j = 0, \dots, n-1$, which comprise the spectrum of C_n. The spectrum of a cycle can be visualised as shown in Figure 2.18. The dots on the circle are complex numbers $2e^{k\pi i/n}$ for $k = 0, \dots, n-1$. Their real parts (projections on the x-axis) are the eigenvalues of C_n. A slightly different version of this diagram is known in chemistry as the *Frost-Musulin circle* [79].

- **The path P_n.** Interestingly, the eigenvalues of P_n can be obtained from the eigenvalues of C_{2n+2}. Let $\mathbf{u}(\zeta) = \begin{bmatrix}1 & \zeta & \zeta^2 & \dots & \zeta^{2n+1}\end{bmatrix}^\intercal$. Let $\zeta = e^{j\pi i/(2n+2)}$ where $0 \le j \le 2n+1$. Note that $\zeta^{2n+2} = 1$ and that $\mathbf{u}(\zeta)$ and $\mathbf{u}(\zeta^{-1})$ have the same

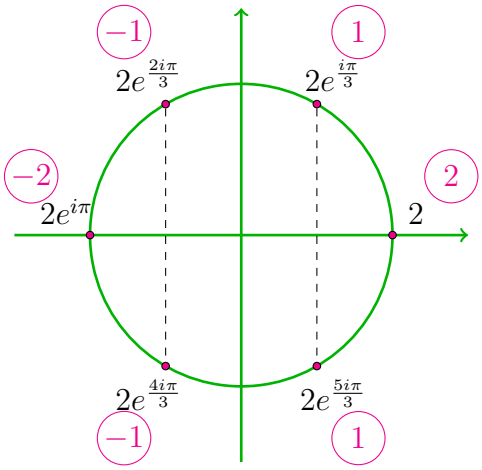


Figure 2.18: Visualisation of the spectrum of C_6.

eigenvalue $2\cos\left(\frac{\pi j}{n+1}\right)$. Therefore, $\mathbf{u}(\zeta) - \mathbf{u}(\zeta^{-1})$ is also an eigenvector for this same eigenvalue. For $\zeta \neq \pm 1$, this latter vector has two zero components at distance $n+1$ in C_{2n+2}. It induces the eigenvector on P_n. This can be obtained using the (2.21) rule by removing the two vertices whose eigenvector entries are 0. Thus, the spec-

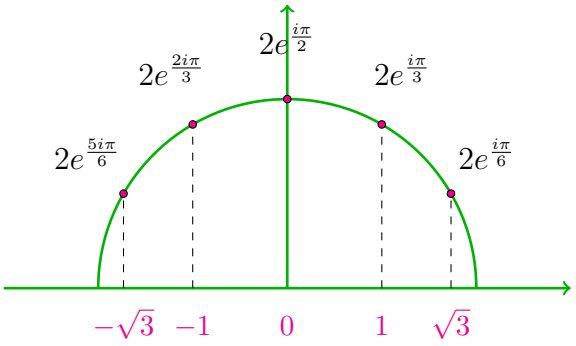


Figure 2.19: Visualisation of the spectrum of P_5.

trum of P_n contains values $2\cos\left(\frac{\pi j}{n+1}\right)$ for $j = 1, \ldots, n$ and is, on the example of P_5, visualised in Figure 2.19. The dots on the circle are complex numbers $2e^{k\pi i/(n+1)}$ for $k = 1, \ldots, n$. Their real parts are the eigenvalues of P_n.

2.6.2 Interlacing eigenvalues

Let A be a matrix. A *symmetric minor* of A is a submatrix B obtained from A by deleting some rows and their corresponding columns.

Theorem 2.14 (Interlacing eigenvalues, Theorem 6.2.1 in [130])**.** *Let A be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and let B be a $(n-k) \times (n-k)$ symmetric*

minor of A with eigenvalues $\mu_1 \geq \cdots \geq \mu_{n-k}$. Then

$$\lambda_i \geq \mu_i \geq \lambda_{i+k} \tag{2.25}$$

for $i = 1, 2, \ldots, n-k$. □

In the language of graph theory this can be stated as follows:

Corollary 2.15 (Proposition 3.2.1 in [32])**.** *Let G be a graph on n vertices and H an induced subgraph of G on $n-k$ vertices. Then the eigenvalues of H interlace those of G, i.e.,*

$$\lambda_i(G) \geq \lambda_i(H) \geq \lambda_{i+k}(G) \tag{2.26}$$

for $i = 1, 2, \ldots, n-k$. □

The *Perron-Frobenius theorem* is an important theorem in linear algebra. It tells us that in the spectrum of an $n \times n$ matrix M with non-negative entries there exists a non-negative real eigenvalue λ_1, such that $|\lambda| \leq |\lambda_1|$ for all $\lambda \in \sigma(M)$, i.e., λ_1 obtains maximum absolute value among all eigenvalues. Moreover, the eigenvalue λ_1 has a non-negative real eigenvector associated with it, called the *Perron-Frobenius eigenvector*. Larry Page and Sergey Brin (the founders of Google Inc.) computed the Perron-Frobenius eigenvector of the web graph [21] and became billionaires. Let us state this important theorem in the language of graph theory:

Theorem 2.16 (Perron-Frobenius Theorem, Proposition 3.1.1 in [32])**.** *Every graph G has a real eigenvalue $\lambda_1(G)$ with non-negative real corresponding eigenvector, such that for each $\lambda \in \sigma(G)$ we have $|\lambda| \leq \lambda_1(G)$. If graph G is connected then $\lambda_1(G)$ is non-degenerate. Moreover, the value $\lambda_1(G)$ does not increase when vertices or edges are removed from G.* □

By combining Corollary 2.15 and Theorem 2.16, an important result on the spectrum of a graph is obtained:

Corollary 2.17 (Proposition 3.1.2 in [32])**.** *Let G be a connected graph with the largest eigenvalue λ_1. If G is regular of degree k then $\lambda_1 = k$. Otherwise,*

$$\delta(G) < d(G) < \lambda_1 < \Delta(G), \tag{2.27}$$

where $\delta(G)$, $d(G)$ and $\Delta(G)$ are the minimum, average and the maximum degree of G, respectively. □

2.6.3 Nullity of Graphs

Let $\mathcal{E}(\lambda_i)$ denote the eigenspace corresponding to eigenvalue λ_i. A graph G is said to be *singular* if $A(G)$ is singular, i.e., if at least one of its eigenvalues is zero. The eigenspace $\mathcal{E}(0)$ is called the *nullspace* of $A(G)$. Vectors in the nullspace of $A(G)$ are called *kernel eigenvectors*.

Definition 2.6 ([187])**.** *The* nullity *of a singular graph G, denoted $\eta(G)$, is the dimension of the nullspace of $A(G)$, i.e., the multiplicity of the zero eigenvalue of G.*

A vertex r is *core* for the eigenspace $\mathcal{E}(\lambda_i)$ corresponding to eigenvalue λ_i if there is a non-zero entry $\mathbf{c}_r$ for *some* vector $\mathbf{c} \in \mathcal{E}(\lambda_i)$; otherwise the vertex is *core-forbidden*.

Example 2.16. In Section 2.6.1 we have determined the eigenvalues (and their corresponding eigenvectors) of the complete bipartite graph $K_{n,m}$. We have seen that $\eta(K_{n,m}) = n + m - 2$. If $n + m > 2$ then the graph $K_{n,m}$ is singular. Moreover, every vertex is a core vertex for every eigenspace.

Consider the graph on 6 vertices from Example 2.15. From Figure 2.17 it is clear that vertices 1, 2, 5 and 6 are core vertices for eigenspace $\mathcal{E}(-1)$, whilst, vertices 3 and 4 are core-forbidden. □

Definition 2.7 ([188])**.** *A graph G is a* nut graph *if it is singular with nullity one and none of the components of a kernel eigenvector is zero.*

Example 2.17. In Figure 2.20, there are two examples of nut graphs. The reader can

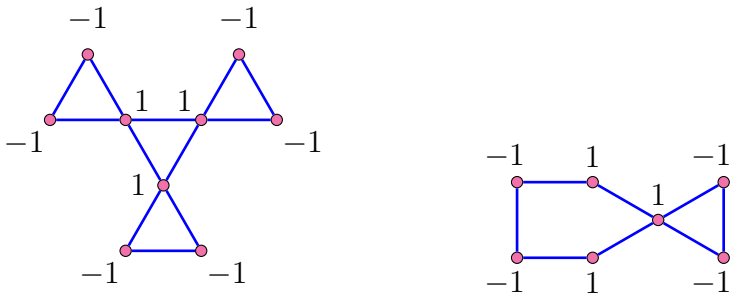


Figure 2.20: Two small nut graphs.

determine their eigenvalues as an exercise and verify that they are indeed singular with nullity one. The numbers next to vertices are the corresponding components of the (non-normalised) eigenvector that belongs to the eigenvalue 0. □

The *inertia* of a matrix A is the ordered triple (n_+, n_0, n_-), where n_+ is the number of positive eigenvalues of A, n_- is the number of negative eigenvalues of A, and n_0 is the number of zero eigenvalues of A. Inertia of a graph G is the inertia of its adjacency matrix $A(G)$. Note that $n_+ + n_- + n_0 = n$, where n is the order of graph G and $\eta(G) = n_0$.

Chapter 3

Chemical Concepts via Graph Theory

Chemistry is all about *bonding*. Without knowing the basics of bond formation, introduction of chemical concepts via graph theory would not make much sense to the reader. We will first introduce, at a very superficial level, the valence bond method. Then, we will discuss the molecular orbital (MO) theory, which offers a more sophisticated model of bonding.

The two main categories of bonds are the *covalent* bond, in which two electrons are *shared* between two bonding atoms, and the *ionic* bond, in which electrons are not shared (one atom gives its electron to the other atom). Most carbon-based compounds are *covalently* bonded, therefore we will only consider covalent bonds here.

3.1 Valence Bond Theory

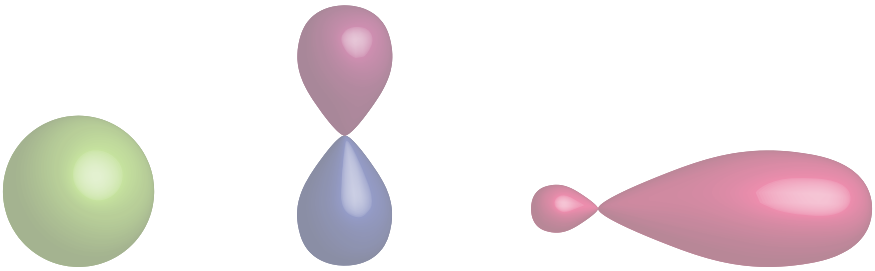
The concepts described in this section can be found in any general chemistry textbook such as [155]. We recommend the book by P. W. Atkins [6].

The essence of an atom is the number of *protons* in its *nucleus*, which is called the *atomic number*[1]. For example, the C atom has 6 protons in its nucleus. Protons are positively charged. In order for an atom to be neutral, it has to have the same number of (negatively charged) *electrons*. When the number of protons does not balance the number of electrons, the atom itself is charged and is called an *ion* (*anion* if negatively charged and *cation* if positively charged).

Electrons reside in *shells* that surround the nucleus. The first shell is of the lowest energy, is the closest to the nucleus and can host up to 2 electrons. The second shell is higher in energy and can host up to 8 electrons. The third shell is even higher in energy and can host up to 18 electrons, etc. Electron shells are further divided into (*atomic*) *orbitals*. An orbital is a region of space in which an electron is found[2]. Each orbital can host up to 2 electrons. A shell can be thought of as a collection of orbitals. The shape of an orbital is important in chemical bonding. The two main types of atomic orbitals (in organic chemistry) are s and p orbitals (see Figure 3.1). The s orbital is spherical, whilst, the p orbital is "dumbbell" shaped. For each atom one can write down its *electron configuration*,

[1] We ignore the presence of *neutrons*.

[2] More precisely, it is a one-electron wavefunction, solving an effective Schrödinger equation for an electron in the presence of the others, and defining the distribution of the electron in space.



(a) s orbital (b) p orbital (c) sp^3 orbital

Figure 3.1: The shape of s, p and sp^3 orbitals.

i.e., the distribution of electrons of that atom in atomic orbitals. For example, the electron configuration of sulfur is $1s^2\,2s^2\,2p^6\,3s^2\,3p^4$. It is a sequence of terms of the form $3s^2$, where the letter denotes the shape of the orbital, the number in front of the letter tells us the shell that includes the orbital and the superscript is the number of electrons occupying that orbital. The $2p$ level consists of three individual p orbitals, the p_x, p_y and p_z orbitals, which are of equal energy in free space and are called *degenerate orbitals*. The set of degenerate orbitals is referred to as the *subshell*. The degenerate orbitals of the p subshell are arranged in such a way that they point in three mutually perpendicular directions in space. The electron configuration of sulfur can be abbreviated as $[\mathrm{Ne}]\,3s^2\,3p^4$, where [Ne] stands for the electronic configuration of neon, the nearest noble gas atom.

The outermost shell of an atom is called the *valence shell*. The electrons in the valence shell are called the *valence electrons*. They are the most important electrons for bonding. The ones in the inner shells are called *core* electrons. The valence bond theory treats chemical bond as an overlap of orbitals of valence shell. In order to form a chemical bond, two half-filled orbitals may superimpose to share electrons. They interact in order to minimize the overall energy of the system.

There is a systematic approach to describe electrons in an atom. It turns out [92] that each electron can be described completely using four *quantum numbers* (n, l, m_l, m_s). Each electron in an atom has a different quadruple of quantum numbers. The *principal quantum number*, n, corresponds to the electron shell. Possible values of number n are positive integers, i.e., $n \in \{1, 2, 3, \ldots\}$. The largest value of n for an atom corresponds to the valence shell. The *angular quantum number* (sometimes called *orbital quantum number*), l, describes the subshell. The value of l ranges from 0 to $n-1$. In chemistry, $l = 0$ is called an s orbital, $l = 1$ is called a p orbital[3], $l = 2$ is called a d orbital, etc. The *magnetic quantum number*, m_l, describes the specific orbital within the subshell. The value of m_l ranges from $-l$ to l. We can immediately see that the s subshell contains only one orbital ($m_l \in \{0\}$), the p subshell contains three orbitals ($m_l \in \{-1, 0, 1\}$), the d subshell contains five orbitals ($m_l \in \{-2, -1, 0, 1, 2\}$), etc. The *spin projection quantum number*, m_s, describes the spin of the electron within an orbital. The value of m_s is in the range $-s, -s+1, \ldots, s-1, s$, where s is the *spin quantum number* of the particle. An electron has spin number $s = \frac{1}{2}$,

[3]It is a subshell that comprises orbitals p_x, p_y and p_z. It would be correct to call it a p subshell.

therefore $m_s \in \{-\frac{1}{2}, \frac{1}{2}\}$. Table 3.1 summarises the above discussion.

As we will limit our consideration to hydrocarbons, we will not go beyond the p subshell. Six dominant elements that are common to all living organisms are carbon, hydrogen, nitrogen, oxygen, phosphorus and sulfur. CHNOPS [5] is an acronym for those basic elements in organic chemistry. The first n rows of Table 3.2 are the quantum numbers of the n electrons in a (neutral) atom with atomic number n. The table consists of 18 rows which is sufficient for all CHNOPS elements.

Name	Symbol	Meaning	Range
Principal quantum number	n	shell	$n \geq 1$
Angular quantum number	l	subshell	$l \in \{0, 1, \ldots, n-1\}$
Magnetic quantum number	m_l	orbital	$m_l \in \{-l, -l+1, \ldots, l\}$
Spin projection quantum number	m_s	spin	$m_s \in \{-\frac{1}{2}, \frac{1}{2}\}$

Table 3.1: The four quantum numbers for electrons in an atom.

n	l	m_l	m_s	n	l	m_l	m_s
1	0	0	1/2	2	1	1	−1/2
1	0	0	−1/2	3	0	0	1/2
2	0	0	1/2	3	0	0	−1/2
2	0	0	−1/2	3	1	-1	1/2
2	1	-1	1/2	3	1	0	1/2
2	1	0	1/2	3	1	1	1/2
2	1	1	1/2	3	1	-1	−1/2
2	1	-1	−1/2	3	1	0	−1/2
2	1	0	−1/2	3	1	1	−1/2

Table 3.2: The first n rows in the table are quantum numbers of the n electrons in a neutral atom with atomic number n.

Example 3.1. Consider the *hydrogen sulfide* molecule with molecular formula H_2S. The electronic configuration of hydrogen is $1s^1$. Its only orbital is half-filled. We have seen above that the electron configuration of sulfur is $[\text{Ne}]\, 3s^2\, 3p^4$. It has 6 valence electrons. The $3s$ orbital is completely filled. One of the three degenerate $3p$ orbitals, say p_z, is completely filled and the other two host one electron each. The $1s$ orbital of an H atom and one of the half-filled $3p$ orbitals of S, say p_x, overlap to form the bond. This can be represented as an *orbital diagram* (see Figure 3.2). Note that each of the orbitals of each atom contains 2 electrons. The p_z orbital of S does not participate in bonding. This pair of valence electrons that are not shared with another atom is called a *lone pair*. □

The electron configuration of carbon is $[\text{He}]\, 2s^2\, 2p^2$. The $2s$ orbital is completely filled. Two of the $2p$ orbitals, say $2p_x$ and $2p_y$, are half-filled while the remaining one, $2p_z$, is empty. How can then a C atom make 4 bonds with hydrogen atom to form methane? The C atom promotes one electron from $2s$ to $2p_z$. The $2s$ and the three $2p$ orbitals are then

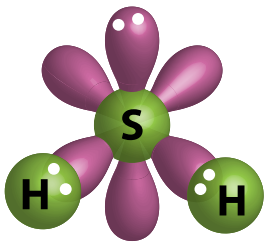


Figure 3.2: Orbital diagram of H_2S, depicting orbitals $3s$, $3p_x$, $3p_y$ and $3p_z$ of atom S. Orbitals $3p_x$ and $3p_y$ overlap with $1s$ orbitals of H atoms, whilst $3p_z$ contains a lone pair.

.

"mixed together" and four identical *hybridised*[4] *orbitals* are formed, called sp^3 orbitals (see Figure 3.1 which shows the shape of a hybridised sp^3 orbital). This notation indicates the number and type of the building blocks that mix together to form the hybrid. The four hybrid sp^3 orbitals of C then host one electron each. They are also dumbbell shaped, but with one lobe much larger than the other. They point in four different directions in the space, from the centre of a tetrahedron towards its apices (see Example 2.11). Each of the four orbitals make an overlap with $1s$ orbitals of a H atom to form the CH_4 molecule. Its orbital diagram is shown in Figure 3.3. *Hybridisation* costs energy that is then paid back

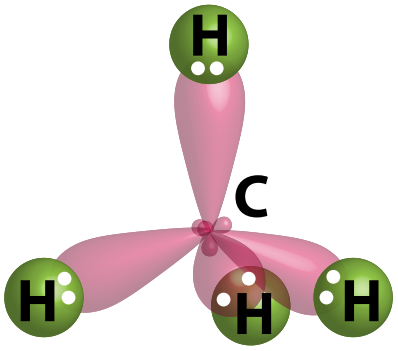


Figure 3.3: Orbital diagram of CH_4 with 4 sp^3 orbitals.

by formation of bonds.

Now, consider ethene with molecular formula C_2H_4 (its structural formula is drawn in Figure 3.9). In this case, the $2s$, $2p_x$ and $2p_y$ orbitals (on each C atom) mix together to form three sp^2 hybridised orbitals and the $2p_z$ orbital remain in its original form. The shape of a sp^2 orbital resembles a sp^3 orbital. The three sp^2 hybridised orbitals lie in the same plane and point in three different directions (from the centre of a triangle towards its vertices). The remaining p_z orbital is perpendicular to that plane. Two of the three hybridised orbitals of a C atom overlap with a $1s$ orbital of H to form a C–H bond. The

[4]They are called hybridised because they are hybrids of the original orbitals.

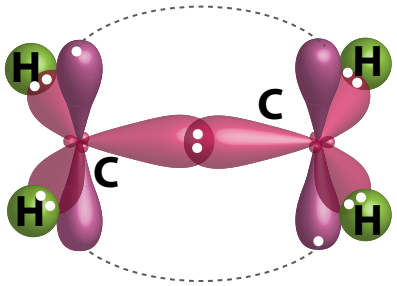


Figure 3.4: Orbital diagram of C_2H_4 with an overlap of 2 sp^2 orbitals.

remaining sp^2 hybridised orbital overlaps with another sp^2 of a different C atom to form the C−C bond. The two C atoms are therefore bonded with a double bond. The second bond comes from the side-by-side overlap of the p_z orbitals. We see that two kind of bonds can be formed: when the overlap occurs between the two bonding nuclei we speak of a σ (*sigma*) bond; when the overlap occurs above and below the nuclei and not between them we speak of a π (*pi*) bond. So, the double bond between the two C atoms is one σ bond and one π bond. Note that the C_2H_4 molecule is planar and that rotation is restricted around the C−C bond. In the ethane molecule (with molecular formula C_2H_6), there is only one σ bond between the two C atoms. There is (almost) free rotation around that bond which implies an infinite number of different conformations of very similar energy.

Example 3.2. This example shows another possible type of hybridisation. Consider the *ethyne* molecule with molecular formula C_2H_2. The $2s$ and $2p_x$ orbitals in C can mix

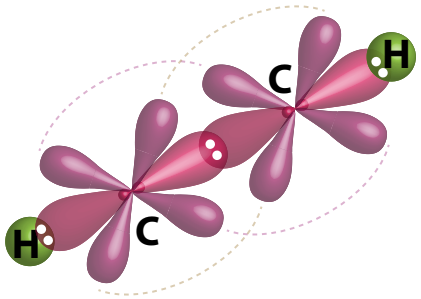


Figure 3.5: Orbital diagram of C_2H_2.

together to form two *sp* hybridised orbitals. They lie on the same line and point in the opposite directions. The remaining $2p_y$ and $2p_z$ orbitals are perpendicular to this line and to each other (see Figure 3.5). The bond formation is similar to the case of C_2H_4, except that there are now two pairs of $2p$ orbitals that overlap side-by-side. There is a triple bond between the two C atoms: one σ bond and two π bonds. □

3.2 Hückel Molecular-Orbital Theory

This section is a very brief introduction to Hückel Molecular-Orbital theory. We will mostly follow [44] which is the main reference. Here, we are not going to argue in favor or against the theory. We will accept it as it is. For an in-depth treatment of concepts from quantum chemistry the reader can consult any of the numerous references, such as [17, 157, 192, 195]. In theoretical chemistry, one can essentially do one of the two things: perform very accurate calculations (involving heavy-duty numerical computation) for a single molecule; develop general simplified models to obtain *insight* in a large class of molecules. *Hückel theory* is probably the simplest possible tool set that, despite of its simplicity, still gives meaningful results for unsaturated carbon frameworks and similar systems.

We will start in quantum mechanics and do a series of approximations until we finally end up with Hückel theory. The main goal of the quantum mechanics is to solve the *Schrödinger wave equation*

$$\mathcal{H}\Psi = E\Psi. \tag{3.1}$$

In this equation, $\mathcal{H}$ is a differential operator referred to as the *Hamiltonian operator* for a system of nuclei and electrons. It can be written as $\mathcal{H} = \mathcal{T} + \mathcal{V}$, i.e., it consists of two parts: the kinetic-energy operator $\mathcal{T}$ and the potential-energy operator $\mathcal{V}$. Let there be m nuclei and n electrons in the system and let $\mathbf{R}_i$ and $\mathbf{r}_j$ be the coordinates of the i-th nucleus and the j-th electron, respectively. Function $\Psi = \Psi(\mathbf{R}_1, \ldots, \mathbf{R}_m, \mathbf{r}_1, \ldots, \mathbf{r}_n)$ in (3.1) is the many-particle wave function.

The kinetic-energy operator in atomic units is

$$\mathcal{T} = -\sum_{i=1}^{n} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{m} \frac{1}{2M_i}\widetilde{\nabla}_i^2. \tag{3.2}$$

In the above expression, M_i is the ratio of the mass of i-th nucleus to the mass of an electron and ∇_i^2 is the Laplacian involving differentiation with respect to the coordinates of the i-th electron while $\widetilde{\nabla}_i^2$ is the Laplacian with respect to the coordinates of the i-th nucleus. In other words, the first term in (3.2) is the operator for the kinetic energy of the electrons and the second term is the operator for the kinetic energy of the nuclei.

The potential-energy operator is

$$\mathcal{V} = -\sum_{i=1}^{n}\sum_{j=1}^{m} \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m} \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|}. \tag{3.3}$$

In the above expression, Z_i is the atomic number of the i-th nucleus. The first term in (3.3) represents the Coulomb attraction between electrons and nuclei, whilst, the second and third term represent the repulsion between electrons and between nuclei, respectively.

In this process, we will do a series of approximations. First, we make the *Born-Oppenheimer approximation* which is central to quantum chemistry. Nuclei are much heavier than electrons and they move very slowly compared to the electrons. Therefore, we can clamp them (i.e., assign them fixed positions in space). Within this approximation the second term of (3.2) can be neglected and the third term of (3.3) is constant. It has no

effect on the operator eigenfunctions and will also be neglected. What remains is called the electronic Hamiltonian:

$$\mathcal{H}_{\text{elec}} = -\sum_{i=1}^{n} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{Z_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{3.4}$$

The Schrödinger equation becomes

$$\mathcal{H}_{\text{elec}}\Psi = E\Psi, \tag{3.5}$$

where $\Psi = \Psi(\mathbf{r}_1, \ldots, \mathbf{r}_n)$ is the many-electron wave function. One important principle of quantum mechanics is *antisymmetry*. This means that if we swap $\mathbf{r}_i$ and $\mathbf{r}_j$ we have:

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_n) = -\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_n). \tag{3.6}$$

We will limit our consideration to *conjugated π systems*, i.e., to planar carbon frameworks that have one p_π orbital at each carbon centre. That p_π orbital, called the atomic orbital, is perpendicular to the plane of the molecule[5]. In the molecular-orbital theory, electrons are not restrained to atomic orbitals, but reside in so-called *molecular orbitals*. Each π electron is allowed to move over the whole framework of the conjugated system. Only those π electrons will be considered.

In the LCAO (linear-combination-of-atomic-orbitals) approach, the *molecular orbital* is made out of *atomic orbitals*. Let there be n atoms in the framework and let

$$\{\chi_r\}_{r=1}^{n} \tag{3.7}$$

be the set of atomic orbitals called the *basis*. Molecular orbitals are of the form

$$\phi_i = \sum_{r=1}^{n} c_r^{(i)}\chi_r, \tag{3.8}$$

i.e., they are (weighted) linear combinations of the atomic orbitals. We assume that $\chi_1, \ldots, \chi_n$ are given in advance (they were obtained somewhere else). This approach is an approximation and is not expected to give the best possible results. When the best-possible molecular orbitals are attained, it is said that the *Hartree-Fock Limit* is reached. In order to get close to it, heavy-duty quantum mechanical calculations are required. The purpose of Hückel theory is not to determine properties of a single molecule very precisely, but to enable us to compare a large class of molecules one to another.

The Hamiltonian for a system with a large number of electrons (3.4) has a large number of terms. There is repulsion between every pair of electrons and the operator has to take this into account. Instead, one can use the effective Hamiltonian

$$\mathcal{H}_{\text{eff}} = \mathcal{T}_{\text{eff}} + \mathcal{V}_{\text{eff}} \tag{3.9}$$

for each π-electron. In this effective Hamiltonian, there is no reference to the other electrons. The interaction between them is "averaged". We will not bother to write down the $\mathcal{H}_{\text{eff}}$

[5]The subscript π in p_π indicates that this p orbital in involved in the formation of π bonds.

precisely. In Hückel theory this is not needed. Actually, this is the essence of the Hückel theory. Every result will be given in terms of two empirical parameters α and β. It is not even necessary to know what the numeric values of these parameters are. Our job now is to solve the

$$\mathcal{H}_{\text{eff}}\phi_i = \varepsilon_i\phi_i \tag{3.10}$$

equation, where ϕ_i is a molecular orbital and ε_i is the orbital energy. In quantum mechanics the expression

$$E = \frac{\int \Psi^*\mathcal{H}\Psi \,\mathrm{d}\tau}{\int \Psi^*\Psi \,\mathrm{d}\tau} \tag{3.11}$$

is the energy of the wave function Ψ. (In [44], it is called the *Rayleigh Ratio*.) The integration is done over all space (and over all variables). *Variation Principle* from quantum mechanics says that an approximate solution $\widetilde{\Psi}$ to the Schrödinger equation gives an approximate energy

$$\widetilde{E} = \frac{\int \widetilde{\Psi}^*\mathcal{H}\widetilde{\Psi} \,\mathrm{d}\tau}{\int \widetilde{\Psi}^*\widetilde{\Psi} \,\mathrm{d}\tau} \tag{3.12}$$

which is always larger than the exact energy E. The best wavefunction of a given form is the one that gives the lowest energy $\widetilde{E}$. It can be found by minimizing $\widetilde{E}$ with respect to the free parameters in the wavefunction.

In our case, $\phi = c_1\chi_1 + c_2\chi_2 + \cdots + c_n\chi_n$, where $\{\chi_r\}_{r=1}^n$ is a fixed basis, thus the energy ε is a function of $c_1, c_2, \ldots, c_n$. The minimum is obtained for those values of $c_1, c_2, \ldots, c_n$ where

$$\frac{\partial\varepsilon}{\partial c_r} = 0 \tag{3.13}$$

for every $r = 1, 2, \ldots, n$. We assume that the basis orbitals are normalised, i.e., for every $r = 1, 2, \ldots, n$:

$$\int \chi_r^*\chi_r \,\mathrm{d}\tau = 1. \tag{3.14}$$

Moreover, $\{\chi_r\}_{r=1}^n$ and ϕ can be chosen to be real and we may omit complex conjugation from the equations and write

$$\int \chi_r^2 \,\mathrm{d}\tau = 1. \tag{3.15}$$

We will also assume that atomic orbitals on different centres are orthogonal, i.e.,

$$\int \chi_r\chi_s \,\mathrm{d}\tau = 0 \tag{3.16}$$

for $r \neq s$. The energy ε can now be evaluated (note that the Hamiltonian operator is self-adjoint):

$$\begin{aligned}\int \phi^2 \,\mathrm{d}\tau &= \int (c_1\chi_1 + c_2\chi_2 + \cdots + c_n\chi_n)^2 \,\mathrm{d}\tau \\ &= \sum_{r=1}^{n} \int c_r^2\chi_r^2 \,\mathrm{d}\tau + \sum_{1\le r\ne s\le n} \int c_r c_s \chi_r \chi_s \,\mathrm{d}\tau \\ &= \sum_{r=1}^{n} c_r^2 \int \chi_r^2 \,\mathrm{d}\tau + 2 \sum_{1\le r<s\le n} c_r c_s \int \chi_r \chi_s \,\mathrm{d}\tau \\ &= \sum_{r=1}^{n} c_r^2\end{aligned} \tag{3.17}$$

and

$$\begin{aligned}\int \phi\mathcal{H}\phi \,\mathrm{d}\tau &= \int (c_1\chi_1 + \cdots + c_n\chi_n)\mathcal{H}(c_1\chi_1 + \cdots + c_n\chi_n) \,\mathrm{d}\tau \\ &= \sum_{r=1}^{n} \int c_r^2\chi_r\mathcal{H}\chi_r \,\mathrm{d}\tau + \sum_{1\le r\ne s\le n} \int c_r c_s \chi_r \mathcal{H}\chi_s \,\mathrm{d}\tau \\ &= \sum_{r=1}^{n} c_r^2 \int \chi_r\mathcal{H}\chi_r \,\mathrm{d}\tau + 2 \sum_{1\le r<s\le n} c_r c_s \int \chi_r \mathcal{H}\chi_s \,\mathrm{d}\tau.\end{aligned} \tag{3.18}$$

Let us define $H_{ij} \coloneqq \int \chi_i\mathcal{H}\chi_j \,\mathrm{d}\tau$. This enables us to write

$$\int \phi\mathcal{H}\phi \,\mathrm{d}\tau = \sum_{r=1}^{n} c_r^2 H_{rr} + 2 \sum_{1\le r<s\le n} c_r c_s H_{rs}. \tag{3.19}$$

Integrals H_{ii} and H_{ij}, $i \ne j$, have a physical interpretation. H_{ii} term is the energy of an electron in an orbital χ_i, i.e., it measures the attraction of i-th atom for electrons, whilst H_{ij}, $i \ne j$, is the energy of interactions between orbitals χ_i and χ_j on centres i and j, i.e., it measures the attraction of the bond region between i-th and j-th atom. Plainly speaking, H_{ii} tell us something about atoms and H_{ij}, $i \ne j$, tell us something about bonds. The energy can be expressed as

$$\varepsilon = \frac{\sum_{r=1}^{n} c_r^2 H_{rr} + 2\sum_{1\le r<s\le n} c_r c_s H_{rs}}{\sum_{r=1}^{n} c_r^2}. \tag{3.20}$$

Now, let us differentiate (3.20) with respect to c_i:

$$\begin{aligned}\frac{\partial\varepsilon}{\partial c_i} &= \left(\sum_{r=1}^{n} c_r^2\right)^{-2} \left(\left(2c_iH_{ii} + 2\sum_{r\ne i} c_r H_{ri}\right)\sum_{r=1}^{n} c_r^2 - \left(\sum_{r=1}^{n} c_r^2 H_{rr} + 2\sum_{1\le r<s\le n} c_r c_s H_{rs}\right) 2c_i\right) \\ &= \frac{2c_iH_{ii} + 2\sum_{r\ne i} c_r H_{ri}}{\sum_{r=1}^{n} c_r^2} - \frac{2\varepsilon c_i}{\sum_{r=1}^{n} c_r^2}\end{aligned} \tag{3.21}$$

If $\frac{\partial\varepsilon}{\partial c_i} = 0$ then the enumerator of the above expression should be 0. We may assume that $\sum_{r=1}^{n} c_r^2 \ne 0$ (or else, we would obtain the trivial solution $c_1 = c_2 = \cdots = c_n = 0$ which does not make physical sense and we are not interested in it). We obtain the following system of equations

$$2\sum_{r=1}^{n} c_r H_{ri} - 2\varepsilon c_i = 0 \qquad (i = 1, \ldots, n) \tag{3.22}$$

which are called the *secular equations* of the system. We need to solve the system for c_i and ε. The system can be rewritten in the matrix form (we also divided the equations by 2 to get rid of that constant):

$$\begin{bmatrix} H_{11}-\varepsilon & H_{12} & \cdots & H_{1n} \\ H_{21} & H_{22}-\varepsilon & \cdots & H_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n1} & H_{n2} & \cdots & H_{nn}-\varepsilon \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = 0_{n\times 1}. \tag{3.23}$$

Let us define the matrix $\mathbf{H} \coloneqq [H_{ij}]_{i,j=1}^{n}$, called the *Hückel Hamiltonian matrix* and let $\mathbf{c} = [c_1\ c_2\ \ldots\ c_n]^\intercal$. The above system can be written in a more "compact form":

$$(\mathbf{H} - \varepsilon I_{n\times n})\mathbf{c} = 0_{n\times 1}. \tag{3.24}$$

The matrix $\mathbf{H}$ is a Hermitian matrix (it is real and symmetric). In order to obtain a non-trivial solution, $\det(\mathbf{H} - \varepsilon I_{n\times n}) = 0$. This is called the *secular determinant*. Expansion of this determinant is a polynomial of order n in variable ε. This is nothing but the *characteristic polynomial* of $\mathbf{H}$. Its roots are eigenvalues of the matrix and they are all real. This means that there are n molecular orbitals associated with the system. Thus, we need to add an index to the molecular orbital:

$$\phi_i = \sum_{r=1}^{n} c_r^{(i)} \chi_r. \tag{3.25}$$

It only remains to show how to obtain the coefficients $\{c_r^{(i)}\}_{r=1}^{n}$ associated with the i-th orbital energy ε_i. They have to satisfy the system of equations

$$\sum_{r=1}^{n} c_r^{(i)} H_{jr} - \varepsilon_i c_j^{(i)} = 0 \qquad (j = 1, \ldots, n). \tag{3.26}$$

Because $\det(\mathbf{H} - \varepsilon I_{n\times n}) = 0$, one of the equations is a linear combination of the others. There are $n-1$ independent equations and n unknowns. We will also require that ϕ_i is normalised, i.e., $\int \phi_i^2 \,\mathrm{d}\tau = \left(c_1^{(i)}\right)^2 + \left(c_2^{(i)}\right)^2 + \cdots + \left(c_n^{(i)}\right)^2 = 1$. This uniquely determines the coefficients. Note that they are the eigenvectors of the corresponding eigenvalues ε_i.

We assumed that the basis is normalised and orthogonal, i.e., $S_{rr} = \int \chi_r^2 \,\mathrm{d}\tau = 1$ and $S_{rs} = \int \chi_r \chi_s \,\mathrm{d}\tau = 1$ for $r \neq s$. This can be, in a more compact way using the Kronecker δ notation, written as $S_{rs} = \delta_{rs}$. Let us define $\mathbf{S} \coloneqq [S_{ij}]_{i,j=1}^{n}$. In our case, $\mathbf{S} = I_{n\times n}$. Had we not made the above assumptions, we would have obtained (in exactly the same way by the variation principle) an equation that slightly generalises (3.24):

$$(\mathbf{H} - \varepsilon\mathbf{S})\mathbf{c} = 0_{n\times 1}. \tag{3.27}$$

Number S_{ij} also has a physical interpretation. It is an overlap between orbitals χ_i and χ_j on centres i and j.

When Hückel developed the theory, he used notation α_r instead of H_{rr} and β_{rs} instead of H_{rs} $(r \neq s)$. This is called the α, β*-notation*. Number α_r is called the *Coulomb integral*

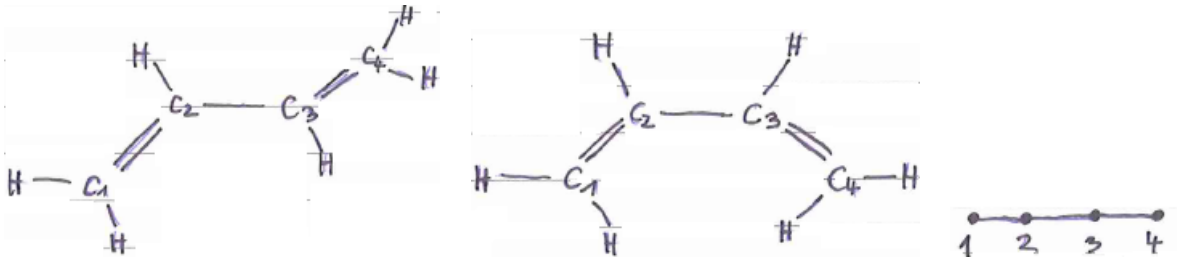
for atom r and β_{rs} is called the *resonance integral* for the bond between r and s. Chemical intuition tells us that β_{rs} is close to 0, except for atoms that are bonded neighbours. In the case of conjugated hydrocarbons, Hückel made the following assumptions (they are now known as the *Hückel approximations*):

(i) Integrals α_r are set equal to a common value $\alpha \coloneqq \alpha_1 = \cdots = \alpha_n$.

(ii) If r and s are not bonded neighbours then $\beta_{rs} = 0$. If they are bonded then they are set equal to a common value β.

Let G denote the Hückel graph, i.e., the carbon framework of the π system under consideration. Then $\beta_{rs} = \beta$ if $r \sim s$ in G, otherwise $\beta_{rs} = 0$.

Example 3.3. Let us work out an example in great detail. Take the *butadiene* molecule (see Figure 3.6). It comes in two isomers[6]. Trans isomer is on the left-hand side and cis isomer



(a) Trans butadiene (b) Cis butadiene (c) Hückel graph

Figure 3.6: Two isomers of butadiene and their Hückel graph.

is on the right-hand side. Hückel theory is only concerned with the σ-bond connectivity and not with molecular geometry. It makes no distinction between the two isomers. If we look at the skeleton graph associated with the molecules (see Figure 3.6(c)), we obtain the path P_4 in both cases. The Hückel Hamiltonian matrix (using the α, β-notation) is:

$$\mathbf{H} = \begin{bmatrix} \alpha & \beta & 0 & 0 \\ \beta & \alpha & \beta & 0 \\ 0 & \beta & \alpha & \beta \\ 0 & 0 & \beta & \alpha \end{bmatrix}. \tag{3.28}$$

First, we have to solve the following equation to obtain the energies of the molecular orbitals:

$$\det(\mathbf{H} - \varepsilon I_{4\times 4}) = \begin{vmatrix} \alpha - \varepsilon & \beta & 0 & 0 \\ \beta & \alpha - \varepsilon & \beta & 0 \\ 0 & \beta & \alpha - \varepsilon & \beta \\ 0 & 0 & \beta & \alpha - \varepsilon \end{vmatrix} = 0. \tag{3.29}$$

[6]The IUPAC "Gold Book" [138] defines an isomer as one of several molecular entities that have the same molecular formulae but different structural formulae or different stereochemical formulae (and hence different physical or chemical properties).

We may divide all elements of the determinant by β (since it is a zero determinant and $\beta \neq 0$). Also, let us introduce $\lambda = \frac{\varepsilon - \alpha}{\beta}$. The determinant becomes:

$$\begin{vmatrix} -\lambda & 1 & 0 & 0 \\ 1 & -\lambda & 1 & 0 \\ 0 & 1 & -\lambda & 1 \\ 0 & 0 & 1 & -\lambda \end{vmatrix} = \lambda^4 - 3\lambda^2 + 1 = 0. \tag{3.30}$$

□

Note that the solutions of (3.30) are exactly the eigenvalues of the skeleton graph G and Hückel theory eventually turns into the spectral graph theory. The energies of molecular graph, i.e., orbital energies of the molecular orbitals, can be expressed with the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the Hückel graph in the following way:

$$\varepsilon_i = \alpha + \lambda_i \beta. \tag{3.31}$$

As α is simply the origin of the energy scale, and $|\beta|$ is a unit of energy difference from α, some authors prefer to use λ_i as a proxy for orbital energy of the i-th molecular orbital.

Example 3.4. Let us solve equation (3.30) from Example 3.3. The four roots of the characteristic polynomial that is associated with the butadiene molecule (see Figure 3.6) are:

$$\begin{aligned} \lambda_1 &= \tfrac{\sqrt{5}+1}{2} \approx 1.61803, \\ \lambda_2 &= \tfrac{\sqrt{5}-1}{2} \approx 0.61803, \\ \lambda_3 &= \tfrac{-\sqrt{5}+1}{2} \approx -0.61803, \\ \lambda_4 &= \tfrac{-\sqrt{5}-1}{2} \approx -1.61803. \end{aligned}$$

□

The eigenvalues will always be given in non-increasing order, i.e., $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$. Recall that the eigenvalue λ_1 is non-degenerate for a connected graph, and $\lambda_n = -\lambda_1$ for bipartite graphs. The values of α and β are negative. Because α measures the energy of a π electron bound within a carbon atom, it has to be negative. In order to remove the electron, energy has to be expanded. A similar argument suggests that β also has to be negative since the bond attracts an electron. At this point we do not know anything about their magnitudes. The energy is measured relatively to α which is regarded as zero energy level. The four energy levels are visualised in the Figure 3.7. This is called a *molecular orbital diagram*. Because β is negative, the lowest energy level is the one associated with the largest eigenvalue λ_1. If $\lambda_i > 0$, the orbital is referred to as a *bonding orbital*, because an electron in such orbital decreases the total π electron energy (we will shortly see how the total π electron energy is defined) of the system. Orbitals for which $\lambda_i < 0$ are called *anti-bonding*. If the orbital is occupied by an electron, the total π electron energy is increased. If $\lambda_i = 0$, the orbital is called *non-bonding*. Note that the Perron-Frobenius theorem (Theorem 2.16) implies that there always exists a bonding molecular orbital in a π conjugated system.

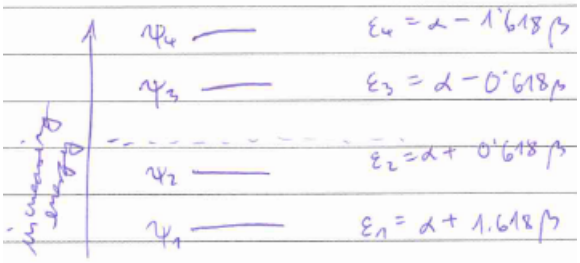


Figure 3.7: Visualisation of the energy levels of the molecular orbitals.

Note that the vector $\mathbf{c}^{(i)} = [c_1^{(i)}\ c_2^{(i)}\ \dots\ c_n^{(i)}]^\intercal$ that contains coefficients associated with the eigenvalue λ_i is the (normalised) eigenvector which corresponds to that eigenvalue. Entries in this eigenvector specify the molecular-orbital and are called *LCAO coefficients* or *MO coefficients*.

Example 3.5. Let us continue our calculations on the budatiene molecule (see Figure 3.6). We have already determined the eigenvalues of its Hückel graph (see Example 3.4). The normalised eigenvectors are (we only give numerical values):

$$\begin{aligned}
\mathbf{c}^{(1)} &\approx \begin{bmatrix} 0.371748 & 0.601501 & 0.601501 & 0.371748 \end{bmatrix}^\intercal, \\
\mathbf{c}^{(2)} &\approx \begin{bmatrix} -0.601501 & -0.371748 & 0.371748 & 0.601501 \end{bmatrix}^\intercal, \\
\mathbf{c}^{(3)} &\approx \begin{bmatrix} 0.601501 & -0.371748 & -0.371748 & 0.601501 \end{bmatrix}^\intercal, \\
\mathbf{c}^{(4)} &\approx \begin{bmatrix} -0.371748 & 0.601501 & -0.601501 & 0.371748 \end{bmatrix}^\intercal.
\end{aligned}$$

We can now write down the molecular orbitals expressed in the basis:

$$\begin{aligned}
\phi_1 &= 0.37\chi_1 + 0.60\chi_2 + 0.60\chi_3 + 0.37\chi_4, \\
\phi_2 &= -0.60\chi_1 - 0.37\chi_2 + 0.37\chi_3 + 0.60\chi_4, \\
\phi_3 &= 0.60\chi_1 - 0.37\chi_2 - 0.37\chi_3 + 0.60\chi_4, \\
\phi_4 &= -0.37\chi_1 + 0.60\chi_2 - 0.60\chi_3 + 0.37\chi_4.
\end{aligned}$$

□

Orbital ϕ_1 is the one with the lowest energy. We are going to assign electrons to molecular orbitals using three fundamental principles. By the *Aufbau Principle*, available orbitals of lowest energy are filled first. By the *Pauli Exclusion-Principle*, they contain at most 2 electrons (with opposite spins). If there are multiple orbitals with the same energy then by the *Hund's Rule* one electron is assigned into each of them before they start to double up. There are four electrons in the electron cloud and they are assigned to the two orbitals of lowest energy (see Figure 3.8). This *electron configuration* (assignment of electrons to orbitals) is the so-called *ground state* configuration for the system in consideration. Let v_i (for $i = 1, \dots, n$) denote the number of electrons assigned to the i-th molecular orbital. Note that

$$v_i = \begin{cases} 2, & \text{if } \phi_i \text{ is doubly occupied;} \\ 1, & \text{if } \phi_i \text{ is singly occupied;} \\ 0, & \text{if } \phi_i \text{ is unoccupied.} \end{cases} \tag{3.32}$$

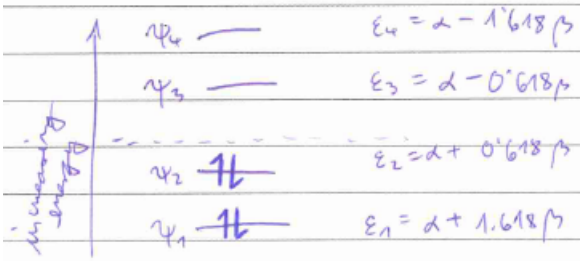


Figure 3.8: The assignment of electrons to the orbitals.

The electron configuration can be written down in the form of $(\phi_1)^{v_1}(\phi_2)^{v_2}\dots(\phi_n)^{v_n}$. The ground state for the butadiene in this notation is $(\phi_1)^2(\phi_2)^2$. The vector $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$ is another (equivalent) way of representing the electron configuration.

Certain orbitals of a given electron configuration deserve a special name (and acronym):

(i) highest (fully or partly) occupied molecular-orbital (HOMO);

(ii) lowest unoccupied (or partly unoccupied) molecular-orbital (LUMO);

(iii) singly occupied molecular-orbital (SOMO);

(iv) doubly occupied molecular-orbital (DOMO).

Their meaning is self-explanatory. With other words, HOMO is the last orbital with $v_i > 0$ and LUMO is the first orbital with $v_i < 2$. If there are SOMOs, HOMO and LUMO coincide. We will use λ_{HOMO} and λ_{LUMO} to denote the eigenvalue of HOMO and LUMO, respectively. Here are some more definitions. An electron configuration is said to be:

(i) *closed shell* if all orbitals are either doubly occupied or unoccupied;

(ii) *open shell* if there are some SOMOs (forced by either odd number of electrons or Hund's rule);

(iii) *properly closed shell* if the HOMO is bonding and LUMO is non- or antibonding, i.e., $\lambda_{\text{HOMO}} > 0 \geq \lambda_{\text{LUMO}}$;

(iv) *pseudo closed shell* if it is closed and some bonding orbitals are empty, i.e., $\lambda_{\text{LUMO}} > 0$;

(v) *meta closed shell* if it is closed and some antibonding orbitals are occupied, i.e., $\lambda_{\text{HOMO}} < 0$.

The *HOMO-LUMO gap* is defined as $|\lambda_{\text{HOMO}} - \lambda_{\text{LUMO}}|$. Recently, the HOMO-LUMO gap has gained the attention of mathematicians [115, 142, 144, 143, 160]. Note that larger HOMO-LUMO gaps indicate greater stability of systems [74, 75]. The *total π electron energy*, denoted E^π, is the sum of energies of electrons in the occupied molecular orbitals:

$$E^\pi = \sum_{i=1}^{n} v_i \varepsilon_i = \sum_{i=1}^{n} v_i(\alpha + \lambda_i \beta). \tag{3.33}$$

For the (ground state) butadiene from the above example, the total π electron energy is

$$\begin{aligned} E^{\pi} &= 2(\alpha + x_1\beta) + 2(\alpha + x_2\beta) = 2(\alpha + \tfrac{\sqrt{5}+1}{2}\beta) + 2(\alpha + \tfrac{\sqrt{5}-1}{2}\beta) \\ &= 4\alpha + 2\sqrt{5}\beta \approx 4\alpha + 4.472\beta. \end{aligned} \tag{3.34}$$

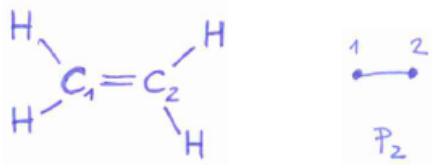
Chemists define the total π energy of the molecule as

$$E_{\pi} = \sum_{i=1}^{n} v_i \lambda_i. \tag{3.35}$$

(We put π as the lower index to distinguish between E_{π} and E^{π}.) A related quantity is the *graph energy* [97] which is defined as

$$E_G = \sum_{i=1}^{n} |\lambda_i|. \tag{3.36}$$

The reader can work out the total π electron energy for ethene (ethylene) which is shown in Figure 3.9. Its total π electron energy is $E^{\pi}(\text{ethene}) = 2\alpha + 2\beta$. Ethene has only one

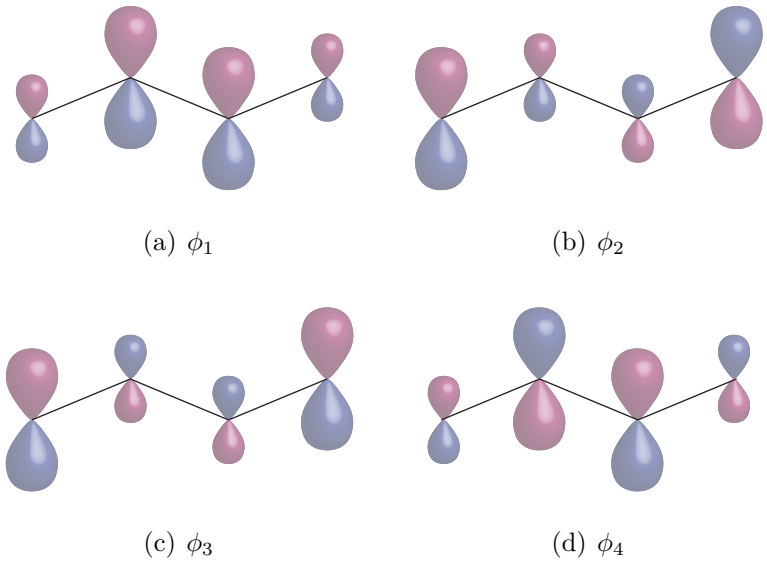


(a) Structural formula (b) Hückel graph

Figure 3.9: Structural formula of ethene and its skeleton graph, P_2.

C=C bond, so this value can be regarded as the π electron energy of a double bond. Now, suppose that the system in consideration would have, instead of delocalised π electrons, a group of localised double bonds such that every electron would participate in one of them. Then the energy of those electrons would be $2\nu(\alpha + \beta)$ where ν is the number of double bonds. The difference between the energy of a group of localised double bonds and the total π electron energy of the system is called the *delocalisation energy* (also the *resonance energy*). For the butadiene molecule from the above example, this value is 0.472β. It is the result of allowing electrons to delocalise. When it is large it suggests stability of the system. When the delocalisation energy is small or negative it suggest that the system is unstable or even non-existent.

The four molecular orbitals of the butadiene are visualised with diagrams in Figure 3.10. When the coefficient at the atomic orbital is positive, the red lobe is pointing up. When the coefficient is negative, it is pointing down. The sizes of those lobes in the figure are proportional to their corresponding coefficients. There is a physical significance in this. Think of solutions of the Schrödinger equation for a *particle in a box*. It does not take a lot of imagination to see waves in the above figures. When the sign between two consecutive atomic orbitals changes, we think of this as a "node" on the bond between the two centres. When an electron is added to an orbital, the bonds with nodes weaken and the bonds without nodes get stronger. Suppose that butadiene from the above example is in ground



(a) ϕ_1 (b) ϕ_2

(c) ϕ_3 (d) ϕ_4

Figure 3.10: Visualisation of molecular orbitals of butadiene molecule.

state. If another electron is added, it is assigned to the ϕ_3 orbital. The middle bond is then stronger, but the 1–2 and the 3–4 bonds are weakened. The same happens if one electrons is removed from the ground state. In the ground state, it is expected that the 1–2 and 3–4 bonds are stronger than the middle one; the ϕ_1 orbital adds to strength of all bonds while the ϕ_2 orbital weakens the middle bond and contributes to the strength of the 1–2 and 3–4 bonds.

Let us point out that there are certain ways to estimate numerical values of α and β. However, most of the properties that can be calculated via Hückel molecular-orbital theory (atomic charge, bond order, free valence etc.) are calculated from the eigenvectors and the numeric values of α and β are not really needed.

3.2.1 Atomic Charge, Bond Order and Free Valence

Recall that the molecular orbital $\phi_i = \sum_{r=1}^{n} c_r^{(i)} \chi_r$, where $c_r^{(i)}$ are the MO coefficients and χ_r are the atomic orbitals. In quantum mechanics, charge densities are given by the square of the amplitude of a wave function. Let us integrate $|\phi_i|^2$ over the whole space (recall that basis is normalised and orthogonal):

$$\int |\phi_i|^2 \, \mathrm{d}\tau = \int \left(c_1^{(i)} \chi_1 + \cdots + c_n^{(i)} \chi_n \right)^2 \mathrm{d}\tau = \sum_{r=1}^{n} \left(c_r^{(i)} \right)^2 = 1. \tag{3.37}$$

This makes perfect sense and comes as no surprise (the probability that an electron is somewhere in space is 1). Individual term $q_r^{(i)} := \left(c_r^{(i)} \right)^2$, called *partial π charge*, represents the probability that an electron in the i-th MO is found on atom r. Instead of probability, chemists use the term *charge density*. The *total π charge* on atom r, denoted Q_r, is the

sum of contributions from all individual electrons in all orbitals:

$$Q_r = \sum_{i=1}^{n} v_i q_r^{(i)} = \sum_{i=1}^{n} v_i \left(c_r^{(i)}\right)^2, \tag{3.38}$$

where v_i is the number of electrons assigned to the orbital ϕ_i. Total π charges $Q_1, \ldots, Q_n$ are easily calculated for an arbitrary conjugated system.

Example 3.6. Let us determine charge densities of (ground state) butadiene (in Figure 3.3):

$$\begin{aligned} Q_1 &= 2 \cdot 0.371748^2 + 2 \cdot (-0.601501)^2 = 1, \\ Q_2 &= 2 \cdot 0.601501^2 + 2 \cdot (-0.371748)^2 = 1, \\ Q_3 &= 2 \cdot 0.601501^2 + 2 \cdot 0.371748^2 = 1, \\ Q_4 &= 2 \cdot 0.371748^2 + 2 \cdot 0.601501^2 = 1. \end{aligned}$$

Surprisingly, charge density is equal to 1 on each atom. Later, we will see that this is true for all the so-called *alternant hydrocarbons* (defined in Section 3.2.2). □

When $|\phi_i^2|$ in (3.37) is expanded, it has terms of type $c_r^{(i)} c_s^{(i)} \chi_r \chi_s$. In our approximation, the orbitals χ_r and χ_s, $r \neq s$, were assumed to be orthogonal. Let us put this neglect of overlaps aside for a moment. Atomic orbital χ_r has large value near atom r. Similarly, χ_s has large value near s. Thus, the product $\chi_r \chi_s$ can be of significant importance only in the part of space near both r and s, i.e., in the region of the bond between r and s. It is natural to define the concept of bond order. An electron in ϕ_i is considered to contribute $p_{rs}^{(i)} := c_r^{(i)} c_s^{(i)}$ to the π bond order between atoms r and s. This contribution, called the *partial π bond order* between atoms r and s, is significant only if r and s are neighbours in the Hückel graph (i.e., they are joined by a σ bond). The *total π bond order* (also called the *Coulson π bond order*) between atoms r and s is defined as [43]:

$$P_{rs}^{\pi} = \sum_{i=1}^{n} v_i p_{rs}^{(i)} = \sum_{i=1}^{n} v_i c_r^{(i)} c_s^{(i)}, \tag{3.39}$$

and the *total bond order* is defined as

$$P_{rs}^{\sigma+\pi} = P_{rs}^{\sigma} + P_{rs}^{\pi}. \tag{3.40}$$

The bond order of a σ bond always equals 1, i.e., $P_{rs}^{\sigma} = 1$, because the two electrons that constitute the σ bond are considered to be constrained to the region between the two bonded atoms.

Example 3.7. In ethene (see Figure 3.9) the coefficients are:

$$c_1^{(1)} = c_2^{(1)} = \tfrac{1}{\sqrt{2}}, \quad c_1^{(2)} = \tfrac{1}{\sqrt{2}} \quad \text{and} \quad c_2^{(2)} = -\tfrac{1}{\sqrt{2}}.$$

In the ground state, $v_1 = 2$ and $v_2 = 0$. Therefore,

$$P_{1,2}^{\pi} = 2 \cdot \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} + 0 \cdot \tfrac{1}{\sqrt{2}} \cdot \left(-\tfrac{1}{\sqrt{2}}\right) = 1.$$

This is only the π electron contribution to the bond order. The total bond order is

$$P_{1,2}^{\sigma+\pi} = P_{1,2}^{\sigma} + P_{1,2}^{\pi} = 2.$$

□

Example 3.8. In the ground state butadiene (see Figure 3.6) the contributions from individual obitals and the total bond orders are shown in Table 3.3. Suppose we excited an

	bond		
orbitals	1–2	2–3	3–4
ϕ_1	0.2236	0.3618	0.2236
ϕ_2	0.2236	−0.1382	0.2236
$(\phi_1)^2(\phi_2)^2$	0.8944	0.4472	0.8944
total bond order	1.8944	1.4472	1.8944

Table 3.3: Contributions to bond order from individual orbitals and the total bond orders.

electron from ϕ_2 to ϕ_3. Bond orders for this excited state of the butadiene are shown in Table 3.4. If an electron is added to ϕ_3 without taking anything from ϕ_1 or ϕ_2 this results

	bond		
orbitals	1–2	2–3	3–4
ϕ_3	−0.2236	0.1382	−0.2236
$(\phi_1)^2(\phi_2)^1(\phi_3)^1$	0.4472	0.7236	0.4472
total bond order	1.4472	1.7236	1.4472
$(\phi_1)^2(\phi_2)^2(\phi_3)^1$	0.6708	0.5854	0.6708
total bond order	1.6708	1.5854	1.6708

Table 3.4: Bond orders in the first excited state of the butadiene and in the ionized state.

in $(\phi_1)^2(\phi_2)^2(\phi_3)^1$ electron configuration. Bond orders for this configuration are also in Table 3.4. □

Experimentally, it was observed that bond order is negatively correlated with bond length (i.e., bonds with higher bond order are shorter). This confirms the physical significance of the definition of bond order. (For more details see the plot in Figure 4-5 in reference [44] and the discussion that accompanies the figure.)

In chemistry, there exists an old concept of *partial valence* (or *resudial affinity*). In Hückel theory this corresponds to the quantitatively defined notion of *free valence*. Let us first define the *total bond number* at r, denoted $N_r^{\sigma+\pi}$, as the sum of total bond orders for all bonds that are incident to this atom:

$$N_r^{\sigma+\pi} \coloneqq \sum_{s \in G(r)} P_{rs}^{\sigma+\pi} = \sum_{s \in G(r)} (P_{rs}^{\sigma} + P_{rs}^{\pi}) = d_r + \sum_{s \in G(r)} \sum_{i=1}^{n} v_i c_r^{(i)} c_s^{(i)}. \tag{3.41}$$

The π *bond number* (also called the *Coulson bond number*), N_r^π, is the sum of π bond orders for the bonds rs for all carbon atoms s adjacent to r:

$$N_r^\pi := \sum_{s \in G(r)} P_{rs}^\pi = \sum_{s \in G(r)} \sum_{i=1}^{n} v_i c_r^{(i)} c_s^{(i)}. \tag{3.42}$$

This formula for N_r^π emphasises the rôle of neighbouring atoms: a large bond number for atom r requires large MO coefficients on atom r but also on its neighbours s.

Example 3.9. In (ground state) butadiene:

$$N_1^{\sigma+\pi} = N_4^{\sigma+\pi} = 1.8944 \qquad \text{and} \qquad N_2^{\sigma+\pi} = N_3^{\sigma+\pi} = 1.8944 + 1.4472 = 3.3416.$$

This implies that atoms 1 and 4 are less deeply involved in bonding than atoms 2 and 3. □

Intuitively, lower bond number means that an atom has larger potential to participate in reactions than an atom with a higher bond number. There exists a maximal bond number that can be theoretically obtained. Let us denote it by $N_{\max}^{\sigma+\pi}$. This values is widely stated to be $N_{\max}^{\sigma+\pi} = 3 + \sqrt{3} \approx 4.7321$. This seems a well established result, but its provenance is very hard to track down. Hereinafter, we will devote a special chapter to this matter and investigate it deeply. For now, let us assume that this values is indeed $3 + \sqrt{3}$. The *free valence*, denoted F_r, of atom r is defined as

$$F_r = N_{\max}^{\sigma+\pi} - N_r^{\sigma+\pi}. \tag{3.43}$$

This is regarded as "unused bonding" of the r-th atom.

Example 3.10. For (ground state) butadiene (in Figure 3.6):

$$F_1 = F_4 = N_{\max}^{\sigma+\pi} - N_1^{\sigma+\pi} = 4.7321 - 1.8944 = 2.8377,$$
$$F_2 = F_3 = N_{\max}^{\sigma+\pi} - N_2^{\sigma+\pi} = 4.7321 - 3.3416 = 1.3904.$$

In [44], Coulson also considered C–H bonds. Note that we only consider C–C bonds here, i.e., bonds in the carbon skeleton. □

It was experimentally observed that there is a positive correlation between reaction rate and free valence. (For more details see the plot in Figure 4-11 in reference [44].)

3.2.2 The Hückel "$4p + 2$" Rule of Aromaticity and the Coulson-Rushbrooke Theorem

Chemists of the early 20th century had noticed that ring systems with six π electrons are a very stable species and they coined the term "aromatic sextet". According to a legend, the structure of benzene (see Figure 3.11) came to August Kekulé in a dream. Kekulé claimed that in his dream he saw a snake gobbling its own tail. The name *aromatic compound* comes from the fact that many of the compounds that were first isolated were highly fragrant (they smelled pleasantly like vanilla, almond etc.). However, many of the presently known aromatic compounds smell unpleasantly or have no odor at all. The "aromatic sextet"

is just a special case of a more general rule which is called the *Hückel "$4p + 2$" rule of aromaticity*. Hückel observed [111, 112, 113, 114] that planar cyclic π electron systems with $4p + 2$ electrons, where $p \in \mathbb{N}$, show pronounced stability.

Here, we explore planar monocyclic systems with chemical formula C_nH_n. The n carbon atoms form a cycle, i.e., its Hückel graph is C_n. We have determined eigenvalues of C_n in Section 2.6.1. Recall that the eigenvalues of C_n are $\lambda_k = 2\cos\left(\frac{2k\pi}{n}\right)$ for $k = 0, \ldots, n-1$. Thus, energies of molecular orbitals are given by:

$$\varepsilon_k = \alpha + 2\beta\cos\left(\frac{2k\pi}{n}\right) \tag{3.44}$$

for $k = 0, 1, \ldots, n-1$. The lowest bonding orbital is obtained for $k = 0$ and its energy is $\alpha + 2\beta$. Non-bonding orbitals will arise exactly in case n is divisible by 4. Moreover, they will appear in degenerate pairs. Figure 3.12 shows how n electrons are assigned to the n molecular orbitals of the annulene C_nH_n.

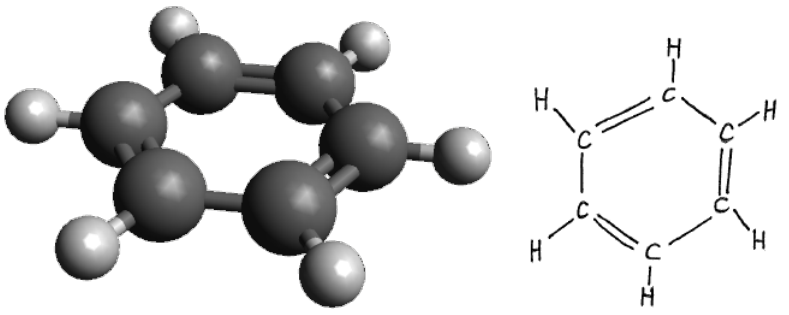
The neutral *cyclopropenyl* radical $C_3H_3^{\bullet}$ has one electron in one of the two degenerate anti-bonding orbitals. The system is expected to be more stable as the *cyclopropenyl cation* $C_3H_3{}^+$ with just two π electrons.

Let us consider *cyclobutadiene* C_4H_4. In our idealised model there are two non-bonding degenerate orbitals. It is known that in this case the Jahn-Teller type of distortion occurs [35, 36]. The Jahn-Teller theorem predicts that degenerate electronic states in (non-linear) molecules give rise to (symmetry-lowering) distortions which remove the intial degeneracy. Under this distortion one of the orbitals has slightly higher energy than α, and the other one has slightly lower energy than α. Note that the HOMO-LUMO gap is small in this case. One way or another (if we stick with our model), C_4H_4 is unstable.

Consider the neutral *cyclopentadienyl* radical $C_5H_5^{\bullet}$. In Figure 3.12, we can see that there is still room for one more electron in the boding molecular orbitals. Therefore, the system would be more stable as *cyclopentadienyl anion* $C_5H_5{}^-$, i.e., the stable number of electrons in this system is 6.

System C_6H_6 is the classic *benzene* example (see Figure 3.11). In the neutral configuration the bonding molecular orbitals are completely filled.

In the neutral *cycloheptatrienyl* radical $C_7H_7^{\bullet}$, one electron goes to an anti-bonding level (as in the case of $C_3H_3^{\bullet}$). This electron can easily be removed to obtain *cycloheptatrienyl*



(a) (b)

Figure 3.11: The benzene molecule and its structural formula.

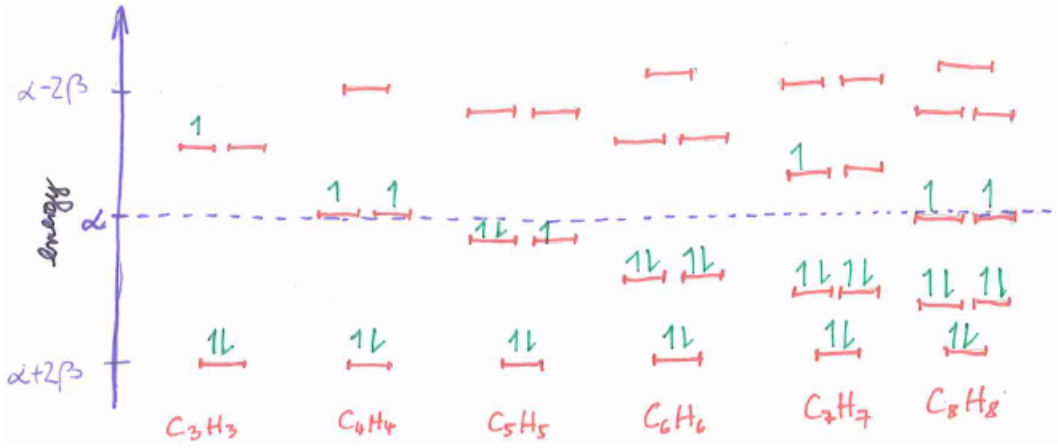


Figure 3.12: π electron energy levels in C_nH_n for $n = 3, 4, \ldots, 8$.

cation $C_7H_7^+$ which has 6 electrons.

The *cyclooctatetraene* system with molecular formula C_8H_8 is similar to C_4H_4. We can now see how this consideration can be generalised to $[n]$annulene C_nH_n for any $n \geq 3$. There are four cases:

(i) $n = 4p$: this system is unstable due to two degenerate non-bonding orbitals;
(ii) $n = 4p + 1$: this system has room for one more electron in the bonding MOs and is more stable as $C_nH_n^-$;
(iii) $n = 4p + 2$: the bonding MOs are completely filled;
(iv) $n = 4p + 3$: one electron is in an anti-bonding molecular orbital and the system is more stable as $C_nH_n^+$.

One of the foundations of the Hückel theory is the dichotomy between alternant and non-alternant conjugated hydrocarbons. The *starring process* in chemistry asks us to place stars on certain carbon-centres. The atoms are divided into two groups, *starred* and *unstarred*, such that no atom of one group is adjacent to another atom from the same group. This process is precisely *2-colouring* of the Hückel graph. Bipartite hydrocarbons are called *alternant* hydrocarbons and non-bipartite hydrocarbons are called *non-alternant*.

Theorem 3.1 (Coulson-Rushbrooke Theorem, [44])**.** *In an alternant hydrocarbon the following statements hold:*

(i) Molecular orbital energy levels are symmetrically paired about the zero energy level α, i.e., if $\varepsilon_i = \alpha + \lambda\beta$ then $\varepsilon_{n-i+1} = \alpha - \lambda\beta$.
(ii) The LCAO coefficients of any pair of orbitals (orbitals of energy $\alpha + \lambda\beta$ and $\alpha - \lambda\beta$) are identical apart from a change of sign in the coefficients of the atomic orbitals on the unstarred atoms.
(iii) In a neutral hydrocarbon, atomic charges Q_r on carbon-atoms are, in the ground state, all precisely unity.

Proof. Statements (i) and (ii) from Theorem 3.1 are nothing more than Proposition 2.13, rewritten in the languege of the Hückel theory.

Let us prove the third statement. Rewrite equation (3.24) as

$$\mathbf{H}\mathbf{c}^{(i)} = \varepsilon_i \mathbf{c}^{(i)} \qquad (i = 1, \ldots, n). \tag{3.45}$$

We added index i since there are n molecular orbitals. Equations (3.45) can be combined in one matrix equation

$$\mathbf{HC} = \mathbf{EC}, \tag{3.46}$$

where $\mathbf{C} = \left[\mathbf{c}^{(1)}\ \mathbf{c}^{(2)}\ \ldots\ \mathbf{c}^{(n)}\right]$ and

$$\mathbf{E} = \begin{bmatrix} \varepsilon_1 & 0 & \cdots & 0 \\ 0 & \varepsilon_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \varepsilon_n \end{bmatrix}.$$

Recall that LCAO coefficients are normalised and orthogonal, i.e.,

$$\left(\mathbf{c}^{(i)}\right)^{\intercal} \mathbf{c}^{(i)} = 1 \qquad \text{and} \qquad \left(\mathbf{c}^{(i)}\right)^{\intercal} \mathbf{c}^{(j)} = 0 \quad (i \neq j). \tag{3.47}$$

These two equations can be combined into (using the Kronecker δ notation):

$$\left(\mathbf{c}^{(i)}\right)^{\intercal} \mathbf{c}^{(j)} = \delta_{ij} \qquad (1 \leq i, j \leq n). \tag{3.48}$$

In the matrix form this is $\mathbf{C}^{\intercal}\mathbf{C} = I_{n\times n}$. It follows that $\mathbf{C}^{\intercal} = \mathbf{C}^{-1}$, i.e., $\mathbf{C}$ is an orthogonal matrix. From $\mathbf{C}\mathbf{C}^{\intercal} = I_{n\times n}$ we obtain:

$$\sum_{i=1}^{n} \left(c_r^{(i)}\right)^2 = 1 \qquad (r = 1, \ldots, n). \tag{3.49}$$

Note that (3.49) holds in general (for both alternant and non-alternant hydrocarbons).

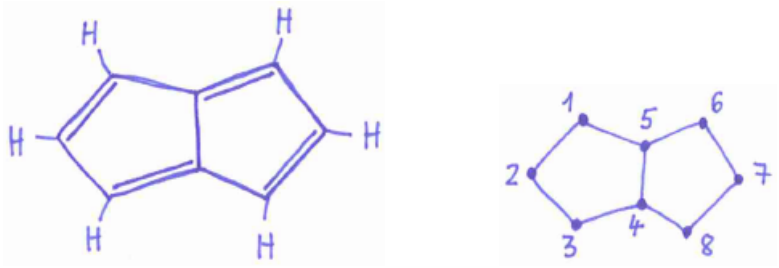
Let us now assume that the hydrocarbon under consideration is alternant. Suppose that it has k bonding orbitals. By (i), there are k corresponding anti-bonding orbitals. It is clear that $k \leq \frac{n}{2}$. There are also $n - 2k$ non-bonding orbitals. In a neutral alternant hydrocarbon, in the ground state, all its bonding orbitals are doubly occupied and the $n-2k$ degenerate non-bonding orbitals are singly occupied. From (3.38) it follows that (we know from (ii) that $c_r^{(i)} = c_r^{(n-i+1)}$ or $c_r^{(i)} = -c_r^{(n-i+1)}$):

$$\begin{aligned} Q_r &= 2\sum_{i=1}^{k} \left(c_r^{(i)}\right)^2 + \sum_{i=k+1}^{n-k} \left(c_r^{(i)}\right)^2 \\ &= \sum_{i=1}^{k} \left(c_r^{(i)}\right)^2 + \sum_{i=1}^{k} \left(c_r^{(i)}\right)^2 + \sum_{i=k+1}^{n-k} \left(c_r^{(i)}\right)^2 \\ &= \sum_{i=1}^{k} \left(c_r^{(i)}\right)^2 + \sum_{i=n-k+1}^{n} \left(c_r^{(i)}\right)^2 + \sum_{i=k+1}^{n-k} \left(c_r^{(i)}\right)^2 = \sum_{i=1}^{n} \left(c_r^{(i)}\right)^2 = 1. \end{aligned} \tag{3.50}$$

□

Theorem 3.1 is also known as the *Pairing Theorem* [132]. Example 3.11 shows that Theorem 3.1 does not apply to non-alternant hydrocarbons. It does not apply to anions and cations. Moreover, it does not apply to excited states, except if excitation takes place from one orbital to its complementary counterpart. Note that, as a consequence of Theorem 3.1, E_π and E_G are equal for the ground state configurations of bipartite graphs and $\lambda_{\text{HOMO}} = -\lambda_{\text{LUMO}}$.

Example 3.11. Consider the *pentalene* molecule. Its structural formula is in Figure 3.13. It is clearly a non-alternant hydrocarbon. Its eigenvalues are:



(a) Structural formula (b) Hückel graph

Figure 3.13: Structural formula of pentalene and its skeleton graph.

$$\lambda_1 \approx 2.3429, \quad \lambda_2 \approx 1.4142, \quad \lambda_3 = 1, \quad \lambda_4 \approx 0.4707,$$
$$\lambda_5 = 0, \quad \lambda_6 \approx -1.4142, \quad \lambda_7 \approx -1.8136, \quad \lambda_8 = -2.$$

We can see that $\lambda_1 \neq -\lambda_8$, $\lambda_2 \neq -\lambda_7$, $\lambda_3 \neq -\lambda_6$ and $\lambda_4 \neq -\lambda_5$, so statement (i) of Theorem 3.1 clearly does not hold. The LCAO coefficients that correspond to the orbitals of the lowest and the highest energy are:

$$\mathbf{c}^{(1)} \approx \begin{bmatrix} 0.3179 & 0.2714 & 0.3179 & 0.4735 & 0.4735 & 0.3179 & 0.2714 & 0.3179 \end{bmatrix},$$
$$\mathbf{c}^{(8)} \approx \begin{bmatrix} 0.2887 & 0.0000 & -0.2887 & 0.5774 & -0.5774 & 0.2887 & 0.0000 & -0.2887 \end{bmatrix}.$$

Those coefficients are clearly *not* identical up to change of sign. There is no way to apply statement (ii) of Theorem 3.1, because the starring process cannot be performed. In the ground state, the electron configuration of pentalene is $(\phi_1)^2(\phi_2)^2(\phi_3)^2(\phi_4)^2$. Atomic charges on carbon atoms are:

$$Q_1 \approx 0.8146, \quad Q_2 \approx 1.1732, \quad Q_3 \approx 0.8146, \quad Q_4 \approx 1.1976,$$
$$Q_5 \approx 1.1976, \quad Q_6 \approx 0.8146, \quad Q_7 \approx 1.1732, \quad Q_8 \approx 0.8146.$$

None of the atomic charges equals 1, so statement (iii) of Theorem 3.1 also does not hold. Note that $E_\pi = E_G$ in this example. This follows from the general fact of spectral graph theory that $\sum_{i=1}^{n} \lambda_i = 0$, because in the case of pentalene bonding molecular orbitals are doubly occupied and all the other orbitals are unoccupied, i.e.,

$$\lambda_{\text{HOMO}} \approx 0.4707 \quad \text{and} \quad \lambda_{\text{LUMO}} = 0.$$

□

3.3 Fullerenes

Fullerene graphs are an important class of chemical graphs. Here, we will only give some of their basic properties. For an in-depth treatment of fullerenes the reader is referred to monograph [70]. For an overview of the recent developments in this field, the reader can consult the survey paper by Schwerdtfeger et al. [186].

Carbon can form many *allotropes*[7] among which *diamond* and *graphite* are known for a long time. Recently, another allotrope called *graphene* has received a lot of attention; it has the form of the infinite hexagonal lattice. *Fullerenes* are closed carbon-cage molecules that contain only pentagonal and hexagonal rings. Each C atom is bonded to exactly 3 other C atoms. Kroto, Heath, O'Brien, Curl and Smalley performed the laser vaporisation experiment on graphite. There was a pronounced mass peak at C_{60} and a slightly less pronounced mass peak at C_{70}. The paper by Kroto et al. [127] which appeared in 1985 proposed the structure of the C_{60} molecule with icosahedral symmetry which is known as *Buckminsterfullerene* (also known as the *buckyball*). It is the most famous fullerene. Their paper initiated the wide reasearch of fullerenes. Kroto, Smalley and Curl were awarded the 1996 Nobel Prize in Chemistry for discovery of fullerenes. In 1984 Rohlfing et al. [180] also performed the laser vaporisation on graphite. In this experiment, they obtained a mass spectrum (carbon cluster mass distribution) which showed that for $n \geq 40$ there are only C_n clusters for even values of n. This was another empirical evidence that was in favour of the *fullerene hypothesis*, which predicts that these carbon clusters have the structure of fullerenes.

The first mathematical model that one may consider to describe a fullerene is a *convex polyhedron*: atoms correspond to vertices of the polyhedron, bonds correspond to edges of the polyhedron and rings correspond to its faces. Fullerenes can be modelled with those trivalent polyhedra that contain only pentagonal and hexagonal faces. An equivalent approach is via graph theory:

Definition 3.1. *A* fullerene graph *is a planar cubic 3-connected graph that contains only pentagonal and hexagonal faces (including the outer face).*

The following theorem tells us that the two representations are essentially equivalent (if we are only concerned with the combinatorial structure and not with geometry):

Theorem 3.2 (Steinitz's Theorem, Theorem 4.1 in [162])**.** *The 1-skeleton of an arbitrary convex polyhedron in $\mathbb{R}^3$ is a planar 3-connected graph, and each planar 3-connected graph is polyhedral, i.e., it is a 1-skeleton of convex polyhedron in $\mathbb{R}^3$.*

The combinatorial structure of the polyhedron is recoverable from the graph by Whitney's Theorem (Theorem 2.9).

Let the number of vertices, edges and faces of a fullerene graph be denoted by n, e and f, respectively. Equation (2.7) (i.e., Euler's formula) in this notation is

$$n - e + f = 2. \tag{3.51}$$

[7]The IUPAC "Gold Book" [138] defines allotropes as different structural modifications of an element, i.e., the atoms of the element are bonded together in a different way (in the same state of matter).

As a fullerene graph is 3-regular it follows that

$$e = \frac{3n}{2}. \tag{3.52}$$

We immediately see that fullerene graphs can exist only for even values of n. From (3.51) and (3.52) it also follows that

$$f = \frac{n}{2} + 2. \tag{3.53}$$

Let f_5 and f_6 denote the number of pentagonal and hexagonal faces, respectively. Note that

$$f = f_5 + f_6 = \frac{n}{2} + 2. \tag{3.54}$$

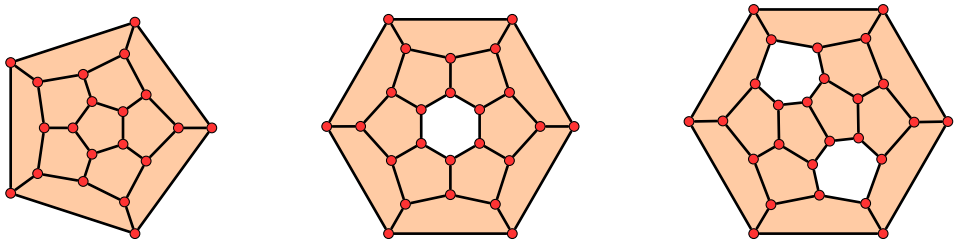
From equation (2.5) it follows that

$$5f_5 + 6f_6 = 2e = 3n. \tag{3.55}$$

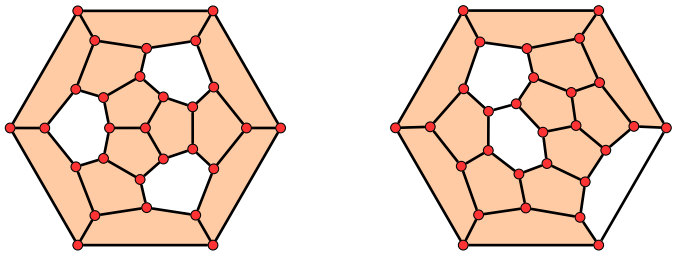
From equations (3.54) and (3.55) it immediately follows that

$$f_5 = 12 \qquad \text{and} \qquad f_6 = \frac{n}{2} - 10. \tag{3.56}$$

Note that the plane dual of a fullerene graph is a *triangulation* (i.e., a plane graph where all faces are triangles) with 12 degree-5 vertices and $\frac{n}{2} - 10$ degree-6 vertices. There exists at least one fullerene graph for each even number $n \geq 20$ with the exception of $n = 22$. The smallest fullerene graph is the skeleton of the dodecahedron (see Figure 3.14(a)); it is



(a) $\mathbf{I}_h$ symmetry, $n = 20$ (b) $\mathbf{D}_{6d}$ symmetry, $n = 24$ (c) $\mathbf{D}_{3h}$ symmetry, $n = 26$



(d) $\mathbf{T}_d$ symmetry, $n = 28$ (e) $\mathbf{D}_2$ symmetry, $n = 28$

Figure 3.14: All fullerene graphs on $n \leq 28$ vertices. Inner pentagonal faces are shaded.

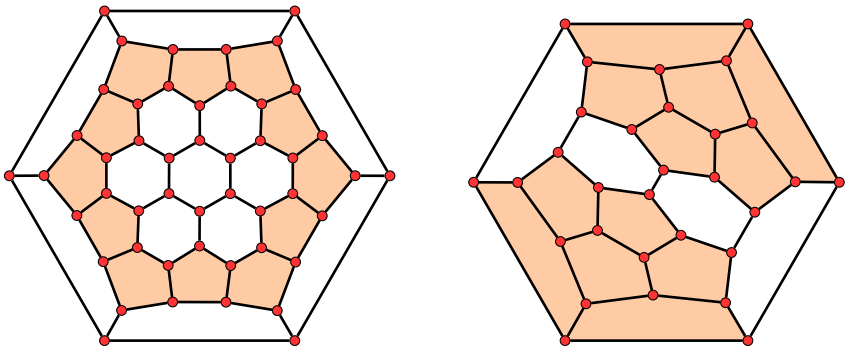
the only one with 20 vertices. There is also only one fullerene graph with 24 vertices (see Figure 3.14(b)) and only one with 26 vertices (see Figure 3.14(c)). There are two fullerene

graphs with 28 vertices (see Figure 3.14(d) and Figure 3.14(e)). Their number increases rapidly beyond that point. More precisely, the number of non-isomorphic fullerene graphs on n vertices is of the order of n^9 [183].

A special class of fullerenes are the IPR fullerenes. In an IPR (Isolated Pentagon Rule) fullerene no two pentagons share an edge. Using various methods from quantum mechanics, Albertazzi et al. [2] provided evidence that IPR fullerenes are chemically very stable. Let us slightly refine this concept:

Definition 3.2. *A* PIP *(*Pentagonal Incidence Partition*) of a fullerene $\mathcal{F}$, denoted $PIP(\mathcal{F})$, is a partition of the number 12 which can be obtained in the following way. Take the subgraph of the plane dual of the fullerene $\mathcal{F}$ that is induced on the degree-5 vertices. The orders of connected components of the induced subgraph are the parts of the partition.*

In other words, define a relation $\equiv$ on the pentagonal faces of fullerene $\mathcal{F}$ such that $p \equiv p$ for every pentagonal face p and $p \equiv q$ if pentagonal faces p and q share an edge. The transitive closure of $\equiv$ is an equivalence relation on the pentagonal faces of $\mathcal{F}$. The sizes of its equivalence classes are parts of the PIP. For every fullerene $\mathcal{F}$ with $n \leq 28$ vertices $PIP(\mathcal{F}) = 12$. There are only 41 such fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 12$. The largest of them has 48 vertices (see Figure 3.15(a)). All fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 12$ are listed in Appendix A. The smallest fullerene $\mathcal{F}$ with $PIP(\mathcal{F}) = 6 + 6$ is the fullerene on 30 vertices with $\mathbf{D}_{5h}$ symmetry group (see Figure 3.15(b)). It is also the smallest fullerene



(a) $\mathbf{D}_{6d}$ symmetry, $n = 48$ (b) $\mathbf{D}_{5h}$ symmetry, $n = 30$

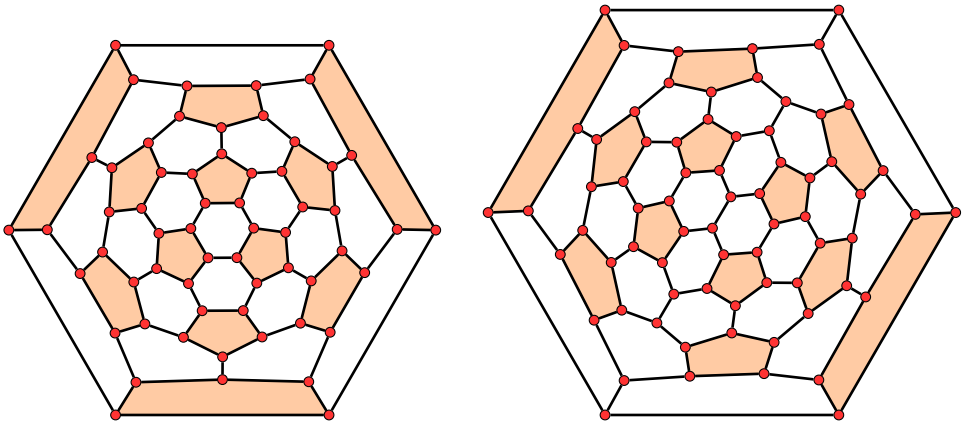
Figure 3.15: The figure on the left shows the largest fullerene $\mathcal{F}$ with $PIP(\mathcal{F}) = 12$. The figure on the right shows the smallest fullerene $\mathcal{F}$ with $PIP(\mathcal{F}) \neq 12$.

with $PIP(\mathcal{F}) \neq 12$. We can define IPR fullerenes using this notion:

Definition 3.3. *Fullerene $\mathcal{F}$ is an* IPR fullerene *if*

$$PIP(\mathcal{F}) = \underbrace{1 + 1 + \cdots + 1}_{12}. \tag{3.57}$$

The buckyball (see Figure 3.16(a)) is the smallest IPR fullerene. It has the structure of truncated icosahedron (soccer ball with black pentagons and white hexagons). The second smallest IPR fullerene has 70 vertices (see Figure 3.16(b)). Their claimed stability would



(a) $\mathbf{I}_h$ symmetry, $n = 60$ (b) $\mathbf{D}_{5h}$ symmetry, $n = 70$

Figure 3.16: Buckyball (on the left) is the smallest IPR fullerene. The fullerene on the right has 70 vertices and is the second smallest IPR fullerene.

explain why in the laser vaporisation experiment the two pronounced peaks in the mass spectrum appeared at $n = 60$ and $n = 70$.

Let us consider symmetries of fullerenes. In principle, there are four groups that can be associated with a fullerene.

The *combinatorial symmetry group* is the automorphism group of the fullerene graph. Recall that by our definition (Definition 3.1), the fullerene graph is a planar graph and not a plane graph (i.e., the embedding is not given in advance).

Since a fullerene can be embedded into a sphere, we may consider it as a combinatorial surface and model it with a map. The *topological symmetry group* is the symmetry group of the map (an automorphism of a map is a permutation of vertices, edges and faces of the map which preserves adjacency). In general, a map may have lower symmetry than its 1-skeleton.

As we have already seen, a fullerene can also be modeled as a convex polyhedron. This structure posesses even more information than the map alone. The *geometrical symmetry group* is the point group of the polyhedron with the highest symmetry among all polyhedra that represent a given fullerene, i.e., the maximally symmetric 3D embedding of the fullerene graph. Such a polyhedron always exists (see Theorem 3.3). This does not hold for maps in general. There exist maps which can not be realised in the 3D space such that the 3D object would exhibit all symmetries of the corresponding map. By Mani's Theorem, which we will present shortly, all three types of symmetry groups discussed so far are isomorphic (as abstract groups) in the realm of fullerenes. Among the three, we will only consider the last one and call it the *ideal symmetry group*.

The fourth is the *physical symmetry group* which is the point group of the physical molecule. The shape of the actual molecule is often different from the ideal one due to the Jahn-Teller effect and other distortions of physical nature.

We will now consider the ideal symmetry group. Because fullerenes are trivalent and comprise of pentagonal and hexagonal rings, they can have n-fold rotational symmetry axis only if $n \in \{1, 2, 3, 5, 6\}$ [68]. On the other hand a fullerene is a finite non-linear

molecule and hence only finite subgroups of the $O(3)$ are admissible ideal symmetry groups. Therefore, among all point groups discussed in Section 2.4 only 36 of them are possible for fullerenes, namely:

$$\mathbf{C}_1, \mathbf{C}_i, \mathbf{C}_s, \mathbf{C}_n, \mathbf{C}_{nh}, \mathbf{C}_{nv}, \mathbf{S}_{2n}, \mathbf{D}_n, \mathbf{D}_{nh}, \mathbf{D}_{nd}, \mathbf{T}, \mathbf{T}_h, \mathbf{T}_d, \mathbf{I} \text{ and } \mathbf{I}_h, \tag{3.58}$$

where $n \in \{2, 3, 5, 6\}$. A subgroup of the symmetry group of an object is also referred to as symmetry group of that object. In this context, the largest symmetry group is called the *full symmetry group*. It was shown in [71] that $\mathbf{C}_5$, $\mathbf{C}_{5h}$, $\mathbf{C}_{5v}$ and $\mathbf{S}_{10}$ occur only as subgroups of $\mathbf{D}_{5h}$, $\mathbf{D}_{5d}$, $\mathbf{D}_5$, $\mathbf{I}$ or $\mathbf{I}_h$ and never as full symmetry groups. Similarly, groups $\mathbf{C}_6, \mathbf{C}_{6h}, \mathbf{C}_{6v}$ and $\mathbf{S}_{12}$ occur only as subgroups of $\mathbf{D}_6$, $\mathbf{D}_{6h}$ or $\mathbf{D}_{6d}$. The list of possible symmetries is thus reduced to 28 point groups. They are all listed in Table 3.5 together with their orders.

Point group	Order	Point group	Order
$\mathbf{C}_1$	1	$\mathbf{D}_{2d}$	8
$\mathbf{C}_i$	2	$\mathbf{D}_5$	10
$\mathbf{C}_s$	2	$\mathbf{D}_6$	12
$\mathbf{C}_2$	2	$\mathbf{D}_{3h}$	12
$\mathbf{C}_3$	3	$\mathbf{D}_{3d}$	12
$\mathbf{C}_{2h}$	4	$\mathbf{T}$	12
$\mathbf{C}_{2v}$	4	$\mathbf{D}_{5h}$	20
$\mathbf{S}_4$	4	$\mathbf{D}_{5d}$	20
$\mathbf{D}_2$	4	$\mathbf{D}_{6h}$	24
$\mathbf{C}_{3h}$	6	$\mathbf{D}_{6d}$	24
$\mathbf{C}_{3v}$	6	$\mathbf{T}_h$	24
$\mathbf{S}_6$	6	$\mathbf{T}_d$	24
$\mathbf{D}_3$	6	$\mathbf{I}$	60
$\mathbf{D}_{2h}$	8	$\mathbf{I}_h$	120

Table 3.5: All 28 possible symmetry groups for fullerenes.

One way to determine the symmetry group is by finding a maximally symmetric 3D embedding. A nice embedding of a fullerene graph into the space $\mathbb{R}^3$ can be obtained by using the algorithm described in the paper by Manolopoulos and Fowler [69]. First, one has to find three molecular orbitals which have only one nodal plane (i.e., one negative and one positive lobe). Let us describe this in graph-theoretical terms. Let $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of the fullerene graph under consideration and let $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \ldots, \mathbf{c}^{(n)}$ be their corresponding eigenvectors. Consider an eigenvector $\mathbf{c}^{(k)}$. Let us partition the vertices of the fullerene graph into three sets, black, gray and white, according to whether their corresponding coefficients in $\mathbf{c}^{(k)}$ are positive, zero or negative, respectively. The molecular orbital has only one nodal plane if it is not the orbital with the lowest energy and the following two conditions hold:

(i) The subgraph induced on white vertices is connected and the subgraph induced on black vertices is connected.

(ii) After all black vertices with at least one white neighbour and all white vertices with at least one black neighbour are removed, the two subgraphs induced on white and black vertices are both connected.

In [69], Manolopoulos and Fowler noticed that there exist exactly three such molecular orbitals. Let $1 < k_x < k_y < k_z \leq n$ be the indices of the three such orbitals. Define $s_x = (\lambda_1 - \lambda_{k_x})^{-\frac{1}{2}}$, $s_y = (\lambda_1 - \lambda_{k_y})^{-\frac{1}{2}}$ and $s_z = (\lambda_1 - \lambda_{k_z})^{-\frac{1}{2}}$. Then the coordinates of the i-th vertex can be obtained as:

$$(x_i, y_i, z_i) = \left(c_i^{(k_x)} s_x, c_i^{(k_y)} s_y, c_i^{(k_z)} s_z\right). \tag{3.59}$$

By considering symmetry operations of this embedding, one can assign a unique point group to a fullerene graph. Graph drawing methods based on eigenvectors were further studied by Pisanski et al. [76, 128, 163]. See also the paper [50] by Cvetković et al.

Theorem 3.2 shows us how 3-connected planar graphs on one side and 1-skeleta of convex polyhedra on the other side are related. It is clear that a symmetry operation of a polyhedron induces an automorphism of the graph that corresponds to its 1-skeleton. By the following theorem of Mani [133], the converse is also true:

Theorem 3.3 (Mani's Theorem, Theorem 3.5 in [193])**.** *Let G be a 3-connected planar graph. Then there exists a convex polyhedron P in $\mathbb{R}^3$ such that G is the 1-skeleton of P and $\mathrm{Aut}(G)$, i.e., the full automorphism group of graph G, is displayed by P (i.e., every automorphism of G induces a symmetry operation on P).*

Theorem 3.3 tells us that a fullerene graph can not posesses a "hidden symmetry". Thus, the point group can also be determined in a purely combinatorial manner without having to obtain a 3D embedding of the fullerene. This algorithm is described in a paper by Myrvold, Fowler et al. [153].

To obtain a physically accurate embedding, time consuming quantum mechanical calculations can be used. In practice, force field optimisation techniques are preferred, which are computationally much less demanding and still yield good results. The first force-field for fullerenes with four parameters was designed by Wu et al. [200]. In following years several refinements were published by various authors. In the software package named **Fullerene** [185], a 22-parameter force field is implemented. This software can be obtained from `http://ctcp.massey.ac.nz/index.php?page=fullerenes` free of charge (for non-commercial use).

3.3.1 Generating fullerenes

The first algorithm for generating fullerenes was the spiral algorithm which is described in [70]. The idea resides in the paper by Manolopoulos, May and Down [135]. They noticed that many fullerene structures can be "peeled like an orange". More precisely, they conjectured the following:

Conjecture 3.1 (The Spiral Conjecture, Conjecture 2.1 in [70])**.** *The surface of a fullerene polyhedron may be unwound in a continuous spiral strip of edge-sharing pentagons and hexagons such that each new face in the spiral after the second shares an edge with both:*

(a) its immediate predecessor in the spiral and

(b) the first face in the preceding spiral that still has an open edge.

□

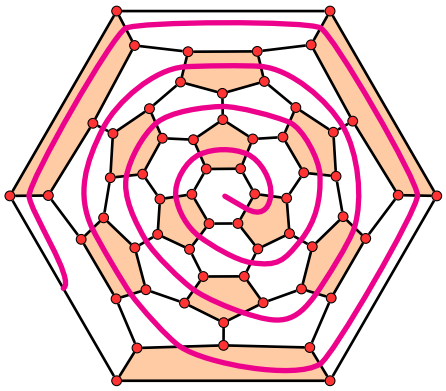


Figure 3.17: A spiral of the buckyball.

An example of a spiral is shown in Figure 3.17. Later, it turned out that the spiral conjecture is false [134, 202, 203, 27]. The smallest unspiralable fullerene has 380 vertices [29]. In 1995 Brinkmann and Dress developed an efficient generator which is called **fullgen** [26]. In 2012 Brinkmann, Goedgebeur and McKay developed a faster generator which is called **buckygen** [28]. In a recent paper by Goedgebeur and McKay [81] an even faster algorithm was developed for generating the class of IPR fullerenes. It was incorporated in the **buckygen** program. Both **fullgen** and **buckygen** are part of the **CaGe** package [24] which can be obtained free of charge from `https://caagt.ugent.be/CaGe/` or `https://www.math.uni-bielefeld.de/~CaGe/`. For our needs the **fullgen** program is sufficient. Planar embeddings of all fullerene graphs that are shown on figures in this section were obtained by the **embed** program, which can be found in the **CaGe** package.

3.4 Kekulé structures

In the early considerations of electronic structure of conjugated π systems, a molecule was thought of as having localised double bonds. A classic example is the naphthalene molecule whose structural formula is in Figure 2.12. Note that every C atom in naphthalene is involved in precisely one double bond. But this single valence-bond structure does not accurately represent the state of the molecule. In the so called "resonance theory", the electronic structure is represented by a combination of several distinct valence-bond structures. The molecule is then said to resonate among these structures or to have a structure that is a hybrid of these valence-bond structures. The naphthalene molecule has three distinct such structures (see Figure 3.18).

A *Kekulé structure* is a valence-bond structure in which every C atom is involved in exactly one double bond. A Kekulé structure is a synonym for a *perfect matching* alias 1-*factor* in

the Hückel graph of the molecule. Our approach to this subject is purely mathematical. It should give the necessary mathematical rigor to the treated concepts in order to clarify to what extent the mathematical model reflects reality. However, we give sufficient translation to the language of mathematical chemistry. The theory of matchings in graphs has been widely studied. The book by Lovász and Plummer [166] is an excellent reference.

Definition 3.4. *A* matching F *in a graph* G *is any subset of mutually non-incident edges. A matching* F *is called a* perfect matching *if every vertex of* G *is an end vertex of some member of* F*. The cardinality of set* F *is called the* size *of the matching.*

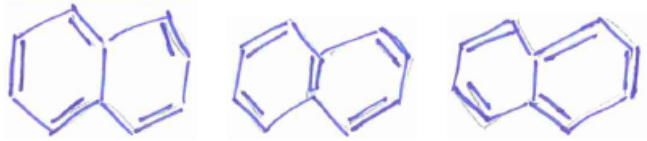
Even though a perfect matching is a formal mathematical concept and a Kekulé structure is a chemical notion, we often refer to perfect matching as Kekulé structures. A *maximum matching* is a matching that contains the largest possible number of edges. In other words, F is a maximum matching in G if $|F'| \leq |F|$ for every matching F' in G. The size of the maximum matching in a graph G is known as the *matching number*, denoted $\nu(G)$. Note that only graphs of even order may possess a perfect matching and that $|F| = \nu(G) = \frac{|V(G)|}{2}$ if F is a perfect matching of G. Also note that every cycle graph of an even order has exactly 2 perfect matchings.

Let F be a perfect matching of G and $e = uv \in E(G)$. If $e \in F$ then $F \setminus \{e\}$ is a perfect matching in $G - uv$. If $e \notin F$ then F is a perfect matching in $G - e$.

Definition 3.5. *The* Kekulé number *of a graph* G*, denoted* $K = K(G)$*, is the number of distinct perfect matchings in* G*.*

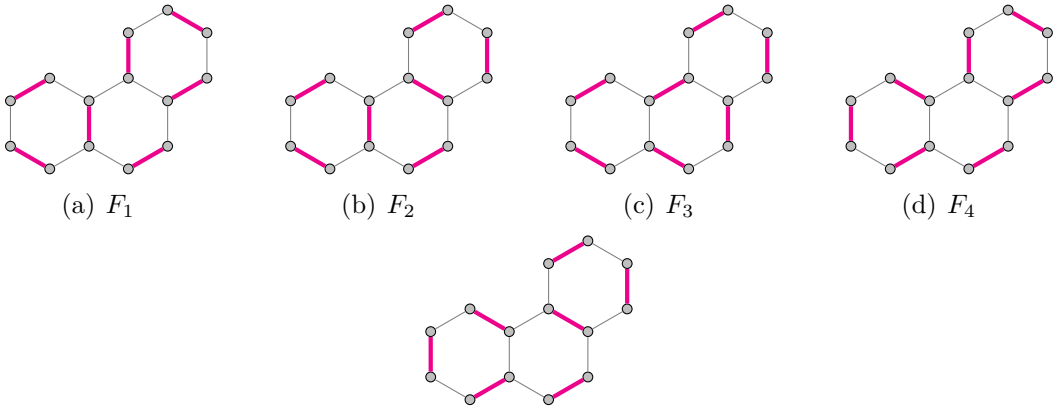
The set of all perfect matchings will be denoted as $\{F_1, F_2, \ldots, F_K\}$. A graph possessing a perfect matching is called *Kekulean*. The phenantherene graph is Kekulean; its Kekulé number is five and the five of its Kekulé structures are shown in Figure 3.19.

Determining whether a given graph G is Kekulean can be done in polynomial time. The algorithm for determining the matching number $\nu(G)$ of a bipartite graph G can be found in any standard reference on algorithms (see e.g. [38]). Let $\vec{G}$ be a directed graph associated with G, such that every arc is directed from a black vertex to a white vertex (where black and white vertices are the bipartition of the graph G). Let us add another two vertices labeled s and t to the digraph $\vec{G}$. They are called the *source* and the *sink*, respectively. For every black vertex v add an arc from s to v and for every white vertex w add an arc from w to t. The resulting digraph $\vec{G}_n$ for the naphthalene graph G_n is in Figure 3.20. We say that an arc $(u, w) \in A(\vec{G})$ *leaves* u and *enters* w. Let $\delta^+(u)$ denote all arcs that enter vertex u and let $\delta^-(u)$ denote all arcs that leave vertex u. An (s, t)-*flow* $\mathcal{F}$ in $\vec{G}$ is a subset



(a) (b) (c)

Figure 3.18: The three distinct valence-bond structures of naphthalene.



(a) F_1 (b) F_2 (c) F_3 (d) F_4

(e) F_5

Figure 3.19: The five Kekulé structures F_i, $1 \leq i \leq 5$, of the phenanthrene graph G_p. Its Kekulé number $K(G_p)$ is 5.

of arcs, such that

$$|\delta^+(u) \cap \mathcal{F}| = |\delta^-(u) \cap \mathcal{F}| \tag{3.60}$$

for all $u \in V(\vec{G}) \setminus \{s, t\}$. The number $|\delta^-(s) \cap \mathcal{F}| = |\delta^+(t) \cap \mathcal{F}|$ is called the *value* of flow $\mathcal{F}$. The flow of maximum value can be found in polynomial time by using the *Ford-Fulkerson method*. The *Push-relabel algorithm* due to Goldberg and Tarjan can be implemented in $O(|V(G)|^2\sqrt{|E(G)|})$ time. There exist more sophisticated algorithms with even better time complexity; for an overview see reference [124]. The value of the maximum flow in $\vec{G}$ is precisely the size of the maximum matching in G. A maximum flow in the digraph that corresponds to the naphthalene graph G_n is in Figure 3.21. Maximum matching in a general (not necessarily bipartite) graph can also be found in polynomial time by using the *Edmonds' matching algorithm* [124] which is among the more involved algorithms in combinatorial optimisation. In book [124] by Korte and Vygen one can find an implementation which runs in $O(|V(G)|^3)$ time. In mathematical chemistry, one often needs to determine the Kekulé number of a graph or compile a list of all Kekulé structures. However, the enumeration problem (i.e., determining the Kekulé number) for perfect matchings in general graphs (even in bipartite) is NP-hard. In [154] Pauling writes: "A few minutes suffice to draw the four unexcited structures for anthracene, the five for phenanthrene, or the six for

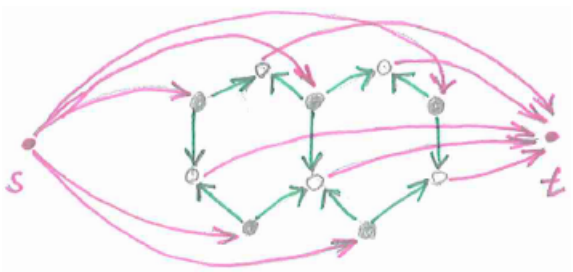


Figure 3.20: Digraph $\vec{G}_n$ that is associated to the naphthalene graph G_n.

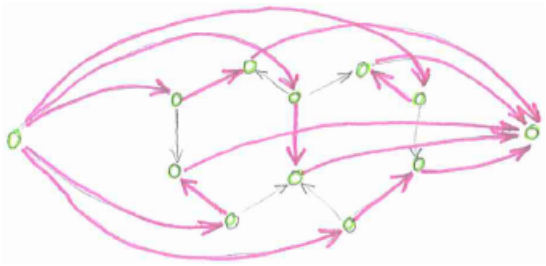


Figure 3.21: The maximum flow in $\vec{G}_n$. The arcs that belong to the flow are coloured magenta. The associated matching corresponds to the Kekulé structure in Figure 3.18(b).

pyrene and to determine the bond numbers of the various bonds; an hour or two might be needed for the 110 structures of tetrabenzoheptacene." For pictures of the mentioned benzenoids, see Figures 5.4 and 5.5. Of course, Pauling was referring to the pen and paper approach. But even a modern computer would not suffice to count (in a reasonable amount of time) the Kekulé structures of a fullerene on, say, 100 vertices by a brute force search. It therefore makes sense:

(i) To seek *special classes of graphs* for which the enumeration problem can be solved in polynomial time.

(ii) To seek good *lower* and *upper bounds* on the Kekulé number.

(iii) To seek *faster* (or *simpler*) algorithms for special classes of graphs to decide if a graph is Kekulean.

3.4.1 Determinants and Kekulé structures

Determinants are an essential tool in linear algebra. Recall that the determinant of an $n \times n$ matrix $A = [a_{ij}]_{i,j=1}^{n}$ is defined as

$$\det A = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i,\pi(i)}, \tag{3.61}$$

where S_n is the set of all permutations on $\{1, 2, \ldots, n\}$ and $\operatorname{sgn}(\pi)$ is the *sign* of permutation π. The *permanent* of an $n \times n$ matrix A is defined as

$$\operatorname{per} A = \sum_{\pi \in S_n} \prod_{i=1}^{n} a_{i,\pi(i)}. \tag{3.62}$$

Permanents are less known than determinants. Note that the definition of a permanent is very similar to that of a determinant. The only difference is that the $\operatorname{sgn}(\pi)$ factor is missing in the equation (3.62). From the computational point of view, there is a huge difference between permanents and determinants, because the permanent is much more difficult to evaluate than determinant. The determinant of an $n \times n$ matrix can be evaluated in $O(n^3)$ time by simple Gaussian elimination or the LU decomposition in which the diagonal

elements of L are all 1, while the value of the determinant is the product of the diagonal elements of matrix U (see reference [55] for details on the LU decomposition). However, the evaluation of a permanent is a NP-hard problem.

Let G be a bipartite graph and let $V(G) = V_1 \sqcup V_2$ be the bipartition. It is clear that a Kekulé structure in graph G may exist only if $|V_1| = |V_2|$. Let $V_1 = \{v_1, v_2, \ldots, v_m\}$ and $V_2 = \{w_1, w_2, \ldots, w_m\}$. The matrix $B(G) = [b_{ij}]_{i,j=1}^m$ where

$$b_{ij} = \begin{cases} 1, & \text{if } v_i \sim w_j; \\ 0, & \text{otherwise.} \end{cases} \tag{3.63}$$

is called the *biadjacency matrix* of the bipartite graph G. (We have already encountered the biadjacency matrix in Section 2.6.) The determinant of $B(G)$ is

$$\det B(G) = \sum_{\pi \in S_m} \operatorname{sgn}(\pi) \prod_{i=1}^{m} b_{i,\pi(i)}. \tag{3.64}$$

Every non-zero term in the expansion of $\det B(G)$ corresponds to a Kekulé structure in G. If G is non-Kekulean then $\det B(G) = 0$, but the converse is not true. The non-zero terms in the expansion are either 1 or -1 and they may cancel each other (as in the case of complete bipartite graphs $K_{n,n}$ for $n \geq 2$). On the other hand, it is clear that

$$K(G) = \operatorname{per} B(G) \tag{3.65}$$

for a bipartite graph G.

Let x_e be a formal variable for each edge $e \in E(G)$. Define the *variable biadjacency matrix* $\widetilde{B}(G) = [b_{ij}]_{i,j=1}^m$ where

$$b_{ij} = \begin{cases} x_e, & \text{if } e = v_i w_j; \\ 0, & \text{otherwise.} \end{cases} \tag{3.66}$$

The determinant $\det \widetilde{B}(G)$ is a polynomial in variables x_e, $e \in E(G)$. It is clear that the terms in the expansion of $\det \widetilde{B}(G)$ are in one-to-one correspondence with Kekulé structures of G and we obtain:

Proposition 3.4 (Theorem 8.2.1 in [166])**.** *A bipartite graph G is Kekulean if and only if*

$$\det \widetilde{B}(G) \neq 0. \tag{3.67}$$

□

Determinants with formal variables can also be used for non-bipartite graphs. An *orientation* $\vec{G}$ of a graph G is a digraph that can be obtained from G by choosing a direction on every edge of G (and thus turning it into an arc). Let $\vec{G}$ be any orientation of G and let x_e be a formal variable associated with $e \in E(G)$. Let $V(G) = \{v_1, v_2, \ldots, v_n\}$. Define the *variable skew adjacency matrix* $\widetilde{A}(G) = [a_{ij}]_{i,j=1}^n$ where

$$a_{ij} = \begin{cases} x_e, & \text{if } e = (v_i, v_j); \\ -x_e, & \text{if } e = (v_j, v_i); \\ 0, & \text{otherwise.} \end{cases} \tag{3.68}$$

If graph G is bipartite such that $V_1 = \{v_1, v_2, \ldots, v_m\}$ and $V_2 = \{v_{m+1}, v_{m+2}, \ldots, v_{2m}\}$, where $V(G) = V_1 \sqcup V_2$ is the bipartition, then the matrix $\widetilde{A}(G)$ can be expressed as

$$\widetilde{A}(G) = \begin{bmatrix} 0_{m\times m} & \widetilde{B}(G) \\ -\widetilde{B}(G) & 0_{m\times m} \end{bmatrix}. \tag{3.69}$$

From equation (3.69) it follows that

$$\det \widetilde{A}(G) = (\det \widetilde{B}(G))^2. \tag{3.70}$$

Note that equation (3.70) does not make sense if graph G is not bipartite (the matrix $\widetilde{B}(G)$ is not defined).

Matrix A is a *skew symmetric* matrix if $-A = A^\intercal$. Note that the variable skew adjacency matrix $\widetilde{A}(G)$ defined above is a skew symmetric matrix. Let A be a skew symmetric $2m\times 2m$ matrix. Let $P = \{\{i_1, j_1\}, \{i_2, j_2\}, \ldots, \{i_m, j_m\}\}$ be a partition of $\{1, 2, \ldots, 2\mathrm{m}\}$ into blocks of size 2. Define

$$a_P = \operatorname{sgn} \begin{pmatrix} 1 & 2 & \ldots & 2m-1 & 2m \\ i_1 & j_1 & \ldots & i_m & j_m \end{pmatrix} \prod_{k=1}^{m} b_{i_k, j_k}.$$

The *Pfaffian* of matrix A, denoted $\operatorname{pf} A$, is defined as

$$\operatorname{pf} A = \sum_{P\in\mathcal{P}_m} a_P, \tag{3.71}$$

where $\mathcal{P}_m$ is the family of all partitions of $\{1, 2, \ldots, 2\mathrm{m}\}$ into blocks of size 2. From linear algebra it is known [152] that if A is a skew symmetric matrix then

$$\det A = (\operatorname{pf} A)^2. \tag{3.72}$$

Consider the Pfaffian $\operatorname{pf} \widetilde{A}(G)$. Note that every non-zero term a_P in $\operatorname{pf} \widetilde{A}(G)$ corresponds to a Kekulé structure of G. The following proposition is a generalisation of Proposition 3.4 for general graphs:

Proposition 3.5 (Theorem 8.2.3 in [166])**.** *A graph G is Kekulean if and only if*

$$\det \widetilde{A}(G) \neq 0. \tag{3.73}$$

□

It turns out that this theory can be used to count the number of Kekulé structures in graphs of certain families. Let $V(G) = \{v_1, v_2, \ldots, v_n\}$ and let $\vec{G}$ be any orientation of G. The *skew adjacency matrix* of $\vec{G}$ is an $n \times n$ matrix $S(\vec{G}) = [s_{ij}]_{i,j=1}^{n}$ where

$$s_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in A(\vec{G}); \\ -1, & \text{if } (v_j, v_i) \in A(\vec{G}); \\ 0, & \text{otherwise.} \end{cases} \tag{3.74}$$

Each term in $\operatorname{pf} S(\vec{G})$ is either 1, -1 or 0. The non-zero terms are in one-to-one correspondence with the Kekulé structures of G. If F is a Kekulé structure then the *sign* of F, denoted $\operatorname{sgn} F$, is the corresponding term in the Pfaffian. It is clear that

$$|\operatorname{pf} S(\vec{G})| \leq K(G). \tag{3.75}$$

The terms in $\operatorname{pf} S(\vec{G})$ may cancel each other, but if all the terms have the same sign then the equality in (3.75) is obtained. If $|\operatorname{pf} S(\vec{G})| = K(G)$ then the digraph $\vec{G}$ is called *Pfaffian*. If there exist an orientation $\vec{G}$ then the graph G is also called *Pfaffian*. If we can find a Pfaffian orientation then we will also be able to count the number of Kekulé structures. The following theorem is crucial for this purpose. Before we can state it we need a few definitions.

Definition 3.6. *Let F be a matching in G. An* F-alternating cycle *is a cycle on which edges that belong to matching F and edges that do not belong to matching F alternate.*

Definition 3.7. *A cycle C in G is* nice *if $G - V(C)$ contains a Kekulé structure.*

Example 3.12. Let G be the graph in Figure 3.22 and let F be one of its Kekulé structures (edges of F are coloured magenta in the figure). Let

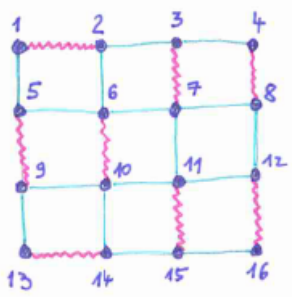


Figure 3.22: A Kekulean graph G and one of its Kekulé structures F (coloured magenta).

$$\begin{aligned} C^{(1)} &= (1,2,6,10,14,13,9,5,1), \\ C^{(2)} &= (3,4,8,12,16,15,11,7,3), \\ C^{(3)} &= (2,3,7,11,15,14,10,6,2). \end{aligned}$$

Cycles $C^{(1)}$ and $C^{(2)}$ are both F-alternating, whilst, the cycle $C^{(3)}$ is not. All the cycles listed above are nice. An example of a cycle which is not nice is

$$C^{(4)} = (1,5,6,10,11,15,16,12,8,4,3,2,1).$$

Definition 3.8. *Let C be a cycle of even length in G. Choose an arbitrary direction and while traversing the cycle C, count the number of arcs in $\vec{G}$ that agree with the direction of the traversal. The cycle C is* evenly oriented *with respect to $\vec{G}$ if it has an even number of arcs oriented in the direction of the traversal, otherwise it is called* oddly oriented.

Now we are ready to state the theorem:

Theorem 3.6 (Theorem 8.3.2 in [166]). *Let G be a graph with an even number of vertices and let $\vec{G}$ be an orientation of G. Then the following statements are equivalent:*

(a) $\vec{G}$ is a Pfaffian orientation of G.
(b) Every Kekulé structure of G has the same sign with respect to $\vec{G}$.
(c) Every nice cycle in G is oddly oriented with respect to $\vec{G}$.
(d) If G has a Kekulé structure then for some Kekulé structure F, every F-alternating cycle is oddly oriented with respect to $\vec{G}$.

□

Just as graph, digraphs can also be embedded in the plane (see Figure 3.23 for an example). A *clockwise facial walk* of a face f in a plane digraph is a cyclic sequence of arcs that are

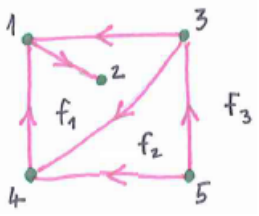


Figure 3.23: A plane digraph with 5 vertices, 6 arcs and 3 faces.

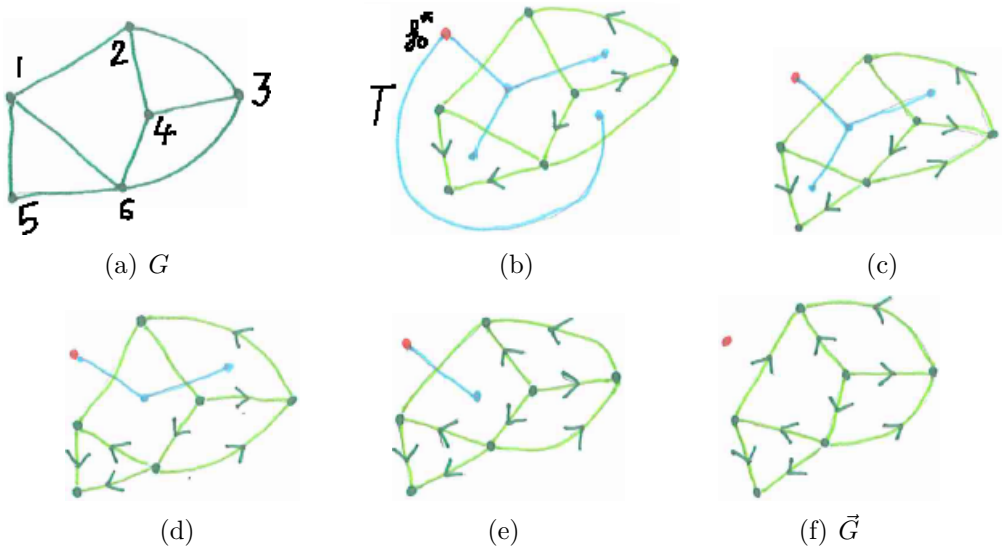
encountered when the boundary of the face is traversed in the clockwise direction. For example, the clockwise facial walk of face f_2 of the digraph in Figure 3.23 is the sequence $((3,4),(5,3),(5,4))$. The clockwise facial walk of face f_1 is the same digraph is the sequence $((4,1),(1,2),(1,2),(3,1),(3,4))$. An arc can be traversed in its own direction or in the opposite direction. Those arcs in the sequence that are traversed in their own direction are called *properly oriented* with respect to the face. In the clockwise facial walk of face f_2 only the arc $(5,4)$ is properly oriented. Note that an arc can be traversed more than once in a facial walk. In the above facial walk of f_1 only the first occurrence of $(1,2)$ is properly oriented, whilst the second is not. We can now state the following lemma:

Lemma 3.7 (Lemma 8.3.3 in [166]). *Let $\vec{G}$ be a connected plane digraph such that every bounded face has an odd number of properly oriented arcs in its clockwise facial walk. Then in every cycle C (which is not necessarily directed) the number of arcs that are oriented clockwise is of the opposite parity to the number of vertices of $\vec{G}$ inside the cycle and, consequently, the digraph $\vec{G}$ is Pfaffian.* □

The next theorem is an important result, because it implies that the Kekulé number of planar graphs (e.g. coronoids and fullerenes) can be found in polynomial time:

Theorem 3.8 (Kasteleyn's Theorem, Theorem 8.3.4 in [166]). *Every plane graph has a Pfaffian orientation. Such an orientation can be found in polynomial time.* □

Kasteleyn's Theorem can be proved by induction on the number of edges with the help of Lemma 3.7. The proof is constructive and it therefore gives rise to the algorithm for obtaining such an orientation. We will describe it with an example.



(a) G (b) (c)

(d) (e) (f) $\vec{G}$

Figure 3.24: Steps in the algorithm for finding a Pfaffian orientation of a plane graph.

Example 3.13. Let G be the graph in Figure 3.24(a). Pick a rooted tree T in G^* (i.e., in the plane dual of G) with the root in vertex f_0^* which corresponds to the outer face f_0 of G. Those edges e whose dual counterparts e^* are not in $E(T)$ comprise a spanning tree in G. Choose any orientation on those edges (see Figure 3.24(b)). Note that Lemma 3.7 holds trivially for trees (because trees have no bounded face).

While the tree T has at least one edge, keep doing the following procedure: Select a leaf f^* of the tree T and remove f^* together with its incident edge e^* from T. Convert edge e (that corresponds to the edge e^* that was just removed from T) to an arc in such a way that the face f will have an odd number of properly oriented edges.

In the end, a Pfaffian orientation of G will be obtained. This procedure is illustrated in Figure 3.24. Now that we have obtained $\vec{G}$, we can write down the matrix $S(\vec{G})$:

$$S(\vec{G}) = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & -1 \\ -1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 1 & -1 & 1 & 0 \end{bmatrix}.$$

As matrix $S(\vec{G})$ is skew symmetric we have that $(\operatorname{pf} S(\vec{G}))^2 = \det S(\vec{G}) = 9$ by equation (3.72). Since $\vec{G}$ is a Pfaffian orientation it follows that $K(G) = |\operatorname{pf} S(\vec{G})| = 3$. The reader can verify that there are indeed three distinct Kekulé structures in graph G. □

3.4.2 Matching polynomials

Let G be a graph and let $p(G, k)$ denote the number of matchings of size k in G. Note that $p(G, 0) = 1$ because an empty set of edges is a matching of size 0 by our definition. It is

clear that $p(G,1) = |E(G)|$. If graph G has even order then $p(G, \frac{|V(G)|}{2}) = K(G)$. It is also clear that $p(G,k) = 0$ for $k > \nu(G)$.

Definition 3.9. *Let G be a graph and let $n = |V(G)|$. The* matching polynomial *of G, denoted $\mu(G,x)$, is defined as*

$$\mu(G,x) = \sum_{k=0}^{\lfloor n/2 \rfloor} (-1)^k p(G,k) x^{n-2k}. \tag{3.76}$$

Let F be a matching in graph G. The number $\mathrm{def}(F) = |V(G)| - 2|F|$ is called the *defect* of matching G. It is the number of vertices of G not covered by the matching F. The matching polynomial can be expressed as

$$\mu(G,x) = \sum_{F} (-1)^{|F|} x^{\mathrm{def}(F)}, \tag{3.77}$$

where F in the above summation runs over all matching in G. Polynomial $\mu(G,x)$ is also called *acyclic polynomial* [196] and *matching defect polynomial* [166] in the literature. Another related polynomial is the *matching generating polynomial*, denoted $g(G,x)$, which is defined as

$$g(G,x) = \sum_{k=0}^{\nu(G)} p(G,k) x^k. \tag{3.78}$$

The two polynomials are related by the identity

$$\mu(G,x) = x^n g(G, -x^{-2}). \tag{3.79}$$

The next proposition gives a simple recurrence relation for the numbers $p(G,k)$:

Proposition 3.9 (Lemma 8.5.1 in [166])**.** *Let G be a graph and let $uv \in E(G)$. Then*

$$p(G,k) = p(G-uv,k) + p(G-u-v,k-1) \tag{3.80}$$

and

$$\mu(G,x) = \mu(G-uv,x) - \mu(G-u-v,x). \tag{3.81}$$

Proof. The matchings of size k can be partitioned into two disjoint subsets: (a) those that include the edge uv and (b) those that do not. If uv does not belong to the matching then all edges of the matching are in $E(G-uv)$. The number of such matchings is exactly $p(G-uv,k)$. If uv does belong to the matching then the edges that are incident with uv do not belong to the matching and it has the remaining $k-1$ edges in $E(G-u-v)$. The matchings of size $k-1$ in $G-u-v$ are in one-to-one correspondence with those matchings of size k in G that include the edge uv, thus there are $p(G-u-v,k-1)$ such matchings and the results follows. By the use of recurrence (3.80), obtaining equality (3.81) is a simple exercise. □

In the chemical literature, one can find the so-called *method of fragmentation* which is attributed to Randić [174]. It is essentially a brute-force algorithm which directly follow from Proposition 3.9. The next proposition is a generalisation of Proposition 3.9:

Proposition 3.10 (Lemma 8.5.2 in [166]). *Let G be a graph and let $u \in V(G)$. Let $G(u) = \{v_1, v_2, \ldots, v_k\}$ be the neighbourhood of vertex u. Then*

$$\mu(G, x) = x\mu(G - u, x) - \sum_{i=1}^{k} \mu(G - u - v_i, x). \tag{3.82}$$

□

The proof is very similar to that of Proposition 3.9. The matchings are first divided into: (a) those that cover vertex u and (b) those that do not. Matchings from (a) are further divided into k disjoint subsets depending on which of the edges uv_i, $i = 1, \ldots, k$, they contain. The following theorem tell us that the matching polynomial is equal to the characteristic polynomial in the case of a forest:

Theorem 3.11 (Mowshowitz [151], Theorem 8.5.3 in [166]). *If G is a forest then*

$$\mu(G, x) = p_G(x), \tag{3.83}$$

where $p_G(x)$ is the characteristic polynomial of graph G.

Unfortunately, Theorem 3.11 does not hold for general graphs. In Section 2.6 we have seen that roots of the characteristic polynomial of an undirected graph are real numbers. The roots of the matching polynomial could, in principle, be any complex numbers. The following proposition tells us that this is not the case:

Proposition 3.12 (Corollary 8.5.7 in [166]). *Let G be a graph. All roots of the matching polynomial $\mu(G, x)$ are real. Moreover, they are symmetrically placed with respect to 0.* □

The matching polynomial has direct applications in mathematical chemistry. Many measures of stability of chemical compounds have been introduced and studied in the past. One of the most significant is the *resonance energy*. For some families of compounds it has been observed that the number of different Kekulé structures is positively correlated with the stability of the compound. One way to precisely formulate the resonance energy is as follows.

Definition 3.10. *Let G be a graph and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be the eigenvalues of G. Let $\lambda_1^{(m)} \geq \lambda_2^{(m)} \geq \cdots \geq \lambda_n^{(m)}$ be the roots of the matching polynomial of graph G. The* topological resonance energy *of G, denoted $TRE(G)$, is defined as*

$$TRE(G) = \sum_{i=1}^{n} v_i \left(\lambda_i - \lambda_i^{(m)}\right), \tag{3.84}$$

where v_i is the occupation number of the i-th molecular orbital.

Topological resonance energy was defined by Gutman, Milun and Trinajstić [103]. It was further studied by Gutman and Mohar [104]. Note that the roots of the matching polynomial can be ordered because they are all real numbers by Proposition 3.12.

3.4.3 Conjugated circuits

This section is devoted to the *conjugated circuits* approach to π-electron currents which is a generalised notion of *ring currents*. A ring current is a physical effect in aromatic molecules that are placed in a magnetic field. A ring current is induced in the delocalized π electrons by the perpendicular component of the field, which are thereby caused to move freely across the molecule, rather than being localized in bonds. The study of conjugated circuits was initiated in 1976 by Milan Randić [173] in order to study *resonance energy* of polycyclic conjugated hydrocarbons. The concept of conjugated circuits has since found several applications in theoretical chemistry and in discrete mathematics. The 1979 paper by Gutman and Randić [105], among other uses, mentions clarification of the concept of aromaticity and generalisation of the Hückel rule to polycyclic conjugated systems. The 157-page review paper by Randić [175] on aromaticity indicates in a short aside that conjugated circuits can also be used to determine π-electron currents. This application of conjugated curcuits is further elaborated in [176].

Our first model is purely combinatorial and has limited application to actual π-electron currents. The model depends only on the graph and not on its position in space. This gives a reasonable approximation to the case when the molecule is planar or nearly planar (e.g. a benzenoid) and the magnetic field is directed perpendicularly to the plane of the molecule. One could refine the model to spatial structures. In the refined model the values of ring currents should be multiplied by a cosine of the angle between the normal of the plane and the direction of magnetic field. For a molecule that changes shape if viewed from different angles the ring currents cannot be computed independently of the direction of the magnetic field.

We may define the *symmetric difference* of two perfect matchings, denoted by $C_{i,j}$, as follows:

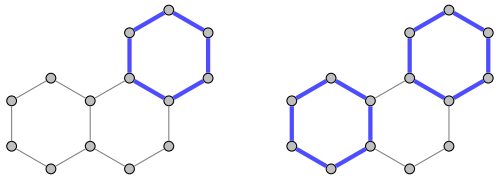
$$C_{i,j} \coloneqq F_i \oplus F_j. \tag{3.85}$$

Gutman and Cyvin [102] use k_i to denote i-th perfect matching and Δ to denote symmetric difference operation. Expression $k_i \,\Delta\, k_j$ in their notation corresponds to $F_i \oplus F_j$ in ours.

Before we go any further, we have to decide how exactly to interpret the operation $\oplus$. Formally, a perfect matching F_i is a subset of edges of graph G. Therefore, $C_{i,j} = F_i \oplus F_j$ is also a subset of edges. However, we can also view F_i and $C_{i,j}$ as subgraphs of G. What is the vertex set of the subgraph $C_{i,j}$? Since both F_i and F_j are spanning subgraphs of G it would make sense to also consider $C_{i,j}$ a spanning subgraph of G. But we take a different approach and omit all isolated vertices. Thus, the *support* (i.e., the vertex set) of $C_{i,j}$ consists of all vertices of G that belong to at least one edge of $C_{i,j}$.

Example 3.14. In this section, we consider the phenanthrene graph G_p as a generic example for clarifying the introduced structures. Figure 3.19 depicts the five Kekulé structures F_i, $1 \leq i \leq 5$, of the phenanthrene graph G_p. Figure 3.25 depicts the symmetric difference of F_1 and F_2 and the symmetric difference of F_1 and F_5. □

Let H be a subgraph of G. Recall that $G - H$ denotes the subgraph of G induced on the set $V(G) \setminus V(H)$, i.e., on the set of vertices of G that do not belong to H. Graph $G - H$ will be called *residual* of H.



(a) $C_{1,2} = F_1 \oplus F_2$ (b) $C_{1,5} = F_1 \oplus F_5$

Figure 3.25: Examples of symmetric difference of perfect matchings of G_p.

Let $F_{i,j} = F_i \cap F_j$ denote the *intersection of matchings*, i.e., the graph induced by the set of edges common to both F_i and F_j. We give a simple lemma:

Lemma 3.13. *Let G be a graph with perfect matchings F_1 and F_2. Then*

$$V(G) = V(C_{1,2}) \sqcup V(F_{1,2}),$$

where $C_{1,2}$ is the symmetric difference and $F_{1,2}$ is the intersection of perfect matchings.

Proof. A vertex $v \in V(G)$ is endpoint of an edge $e_{v,1} \in F_1$ and endpoint of an edge $e_{v,2} \in F_2$. If $e_{v,1} = e_{v,2}$ then $v \in V(F_{1,2})$. If $e_{v,1} \neq e_{v,2}$ then $v \in V(C_{1,2})$. It is clear that v can not belong to both $V(F_{1,2})$ and $V(C_{1,2})$. □

Proposition 3.14. *Let G be a graph and let $\{F_1, F_2, \ldots, F_K\}$ be the set of its perfect matchings. For any i, j, $1 \leq i, j \leq K$, the graph $C_{i,j}$ is either:*

(a) empty, when $i = j$, or

(b) vertex-disjoint union of one or more even cycles, when $i \neq j$.

Moreover, the residual graph $G - C_{i,j}$ is Kekulean.

Proof. Part (a) is clear, because $X \oplus Y = \emptyset$ if and only if $X = Y$ for any sets X and Y.

Graph $C_{i,j}$ is finite and each vertex of $C_{i,j}$ is incident to one edge from F_i and one other edge from F_j, i.e., $C_{i,j}$ is a 2-regular graph where edges are colourable with two colours (use colour i for edges which come from F_i). Hence, $C_{i,j}$ is a union of vertex-disjoint even cycles.

By Lemma 3.13, $F_{i,j}$ is a perfect matching of $G - C_{i,j}$, hence $G - C_{i,j}$ is Kekulean. □

The number $K(G - H)$ will be called the *residual Kekulé number of H*. Motivated by Proposition 3.14 we define *conjugated circuits* as follows:

Definition 3.11. *Let G be a graph and C its subgraph. If C is a vertex-disjoint union of (zero or more) even cycles and the residual graph $G - C$ is Kekulean then C is called a* conjugated circuit.

Note that the empty graph is a conjugated circuit of every Kekulean graph. We will call it the *empty conjugated circuit*. When Randić first defined conjugated circuits [174] he considered only the case when the circuit is composed of a single cycle. For this reason we give to such structure a special name.

Definition 3.12. *Let G be a graph and C one of its conjugated circuits. If C is a connected non-empty graph (i.e., a single cycle) then C is called a* conjugated cycle.

Using this terminology one could say that in [174] Randić considered conjugated cycles and not conjugated circuits in general. Later, in 1979, Gutman and Randić [105] treated conjugated circuits in general (they called such structures *disjoint conjugated circuits*). Note that the notion of F-alternating cycles (see Definition 3.6) is closely related to conjugated circuits. In the special case when F is a perfect matching, alternating cycles are exactly conjugated cycles.

We define the restriction of a Kekulé structure to a subgraph:

Definition 3.13. *Let G be a graph, let $H \subseteq G$ be its subgraphs and let $F \subseteq E(G)$ be a perfect matching in G. The set of edges $F \cap E(H)$ is called the* restriction of F to H.

Example 3.15. Look at the graph Q_3 in Figure 3.26. Let $F = \{(1,4),(2,6),(3,7),(5,8)\}$ be one of its perfect matchings (it is emphasized on Figure 3.26). Let H_1 be the subgraph

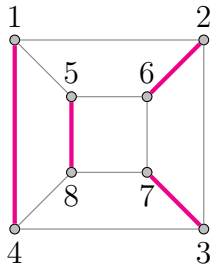


Figure 3.26: Graph Q_3 with one of its perfect matchings.

of Q_3 that is induced on vertices $\{5,6,7,8\}$ and let H_2 be the subgraph that is induced on vertices $\{2,3,6,7\}$. We can see that the restriction of F to H_2 is a perfect matching in H_2, whilst, the restriction of F to H_1 is not a perfect matching in H_1. □

Conjugated circuits of a graph are closely related to its Kekulé structures.

Definition 3.14. *Let G be a graph, C one of its conjugated circuits and F one of its perfect matchings. We say that C and F are* compatible *if the restriction of F to C is a perfect matching in C.*

Note that in the Definition 3.14, the restriction of F to $G - C$ is also a perfect matching in $G - C$. The following theorem by Gutman and Randić gives us a characterisation of conjugated circuits:

Theorem 3.15 (Characterisation of conjugated circuits, [105])**.** *Let C be a subgraph of G. Then C is a conjugated circuit if and only if there exists a pair of Kekulé structures of G, say F_i and F_j, such that $C = C_{i,j} = F_i \oplus F_j$.*

Proof. For the reverse direction, the proof is immediate. By Proposition 3.14, for any pair of perfect matchings F_i and F_j the graph $C_{i,j}$ is a conjugated circuit.

In the other direction the proof is slightly more involved. Let C be a conjugated circuit in G. Since $G - C$ is Kekulean, we may choose a perfect matching of $G - C$, says E_0. Assume C is composed of s vertex-disjoint even cycles $C_1, C_2, \ldots, C_s$. Let E_i and E'_i be the two perfect matchings in C_i. Then the union $F = E_0 \cup E_1 \cup \ldots \cup E_s$ determines a perfect matching in G. Analogously, $F' = E_0 \cup E'_1 \cup \ldots \cup E'_s$ is also a perfect matching in G. Since $F \oplus F' = C$, the result follows. □

Note that in Theorem 3.15 the Kekulé structures F_i and F_j are both compatible with conjugated circuit C. The theorem tells us that each ordered pair (i, j), $i \neq j$, $1 \leq i, j \leq K$, determines a conjugated circuit. However, the same conjugated circuit is counted more than once. The following result gives exact numbers:

Theorem 3.16. *Let C be a conjugated circuit of G, having s connected components, and let $k = K(G - C)$ be the residual Kekulé number of C. Then the number of ordered pairs (i, j), such that $C = C_{i,j}$, is equal to $2^s k$.*

Proof. Let (i, j) be a pair that determines C. The restrictions of matchings F_i and F_j to $G - C$ coincide. Each of the matchings F_i and F_j can be assembled from a matching in C and a matching in $G - C$. The shared part (matching in $G - C$) can be chosen in k different ways. The two matchings must be complementary in each cycle of C. Since each cycle of C has two matchings, one of them goes to F_i and the other one goes to F_j. Thus, there are 2^s ways to do that. Since the two parts (matching in C and matching in $G - C$) can be chosen independently and this covers all possibilities the product of the counts gives the result. □

Note that pairs (i, j) and (j, i) give the same conjugated circuits due to symmetry of operation $\oplus$. The number of unordered pairs $\{i, j\}$ such that $C = C_{i,j}$ is therefore equal to $2^{s-1}k$. In [72] a different counting has been performed for $s = 1$. Here, we give a slight generalisation:

Theorem 3.17. *Let C be a conjugated circuit of G, having s connected components, and let $k = K(G-C)$ be the number of perfect matchings of $G-C$. Then the number of ordered pairs (i, j), such that C is contained in $C_{i,j}$, is equal to $2^s k^2$.*

Proof. Let (i, j) be a pair that determines C. The restrictions of perfect matchings F_i and F_j to $G - C$ may be any pair of matchings in $G - C$. This can be arranged in k^2 ways. On the other hand, they must be complementary in each cycle of C. This can be done in 2^s ways. Since the two parts can be chosen independently and this covers all possibilities the product of the counts gives the result. □

Example 3.16. Some conjugated circuits can be written as a symmetric difference in more than one way. For example, the circuit in Figure 3.25(a) can be written as $F_1 \oplus F_2$ or $F_4 \oplus F_5$ and the circuit in Figure 3.25(b) can be written as $F_1 \oplus F_5$ or $F_2 \oplus F_4$. □

Lemma 3.18. *Let G be a graph, C a conjugated circuit and F a perfect matching in G such that C and F are compatible. There exists a unique matching F' such that $C = F \oplus F'$.*

Proof. If C and F are compatible then the restriction of F to C is a perfect matching in C and the restriction of F to $G-C$ is a perfect matching in $G-C$. Because none of the edges incident with vertices $V(G-C)$ belongs to C, matchings F and F' must coincide on $G-C$, i.e., $F \cap E(G-C) = F' \cap E(G-C)$. Recall that cycles of even length have exactly 2 different perfect matchings. One of them is $F \cap E(C)$. Let M denote the other perfect matching in C. It is easy to see that $C = F \oplus F'$ implies $F' \cap E(C) = M$. The matching F' must be equal to $M \cup (F \cap E(G-C))$ and is therefore uniquely determined. □

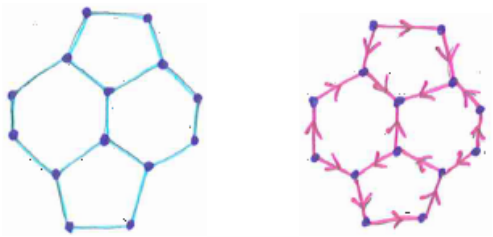
The original approach of Randić to conjugated circuits was different. He was looking at the ways to extend a given Kekulé structure to a conjugated circuit, i.e., he was looking for conjugated circuits that are compatible with a given Kekulé structure. In the paper with Gutman [105] they announced what Randić later calls 'Gutman's important theorem' [176], and can be phrased in our terminology in the following way:

Theorem 3.19 (Gutman and Randić, [105])**.** *Let G be a graph with K perfect matchings and let F be any of its perfect matchings. There are exactly $K-1$ distinct non-empty conjugated circuits that can be written as a symmetric difference between F and some other perfect matching of G.*

In other words, there are exactly $K-1$ distinct non-empty conjugated circuits that are compatible with F.

Proof of Theorem 3.19. Let $\mathcal{C}(F)$ be the set of conjugated circuits that are compatible with F. By Theorem 3.15, this means that $C \in \mathcal{C}(F)$ if and only if there exists a perfect matching F' such that $C = F \oplus F'$. Clearly, F' is different from F and is uniquely determined by C and F by Lemma 3.18. The bijection between the set of perfect matchings different from F and the set $\mathcal{C}(F)$ establishes that fact that $|\mathcal{C}(F)| = K-1$. □

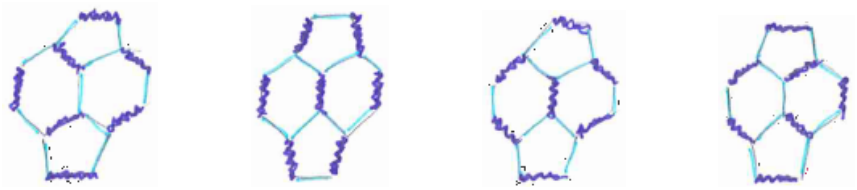
Let us finish this section with a chemical application of conjugated circuits. We describe the conjugated circuits model of π-electron currents [176]. Let G be a plane Kekulean graph with a distinguished outer face. Let us convert it to a directed graph by choosing an arbitrary direction for each edge (which converts the edge to an arc). For an example see



(a) G (b) $\vec{G}$

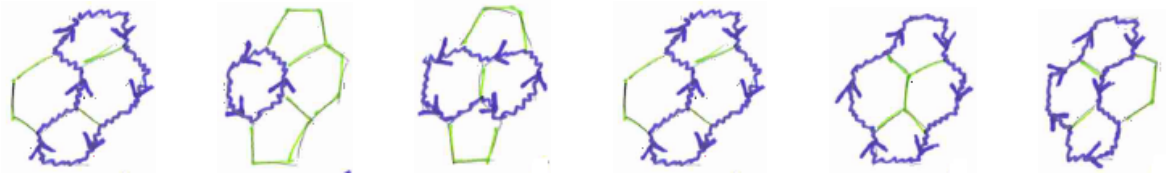
Figure 3.27: A plane Kekulean graph G and its corresponding digraph $\vec{G}$.

Figure 3.27. Then determine all of its Kekulé structures. The graph G in Figure 3.27 has Kekulé number $K(G) = 4$. All of its Kekulé structures are shown in Figure 3.28. Then find all the conjugated circuits in graph G by using Theorem 3.15. All symmetric differences

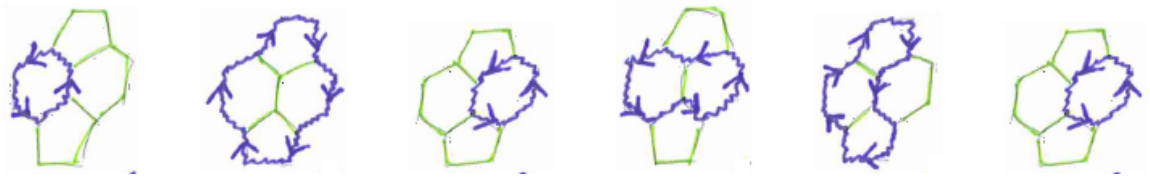


(a) F_1 (b) F_2 (c) F_3 (d) F_4

Figure 3.28: The four Kekulé structures F_i, $1 \leq i \leq 4$, of the graph G.



(a) $\vec{C}_{1,2}$ (b) $\vec{C}_{1,3}$ (c) $\vec{C}_{1,4}$ (d) $\vec{C}_{2,1}$ (e) $\vec{C}_{2,3}$ (f) $\vec{C}_{2,4}$



(g) $\vec{C}_{3,1}$ (h) $\vec{C}_{3,2}$ (i) $\vec{C}_{3,4}$ (j) $\vec{C}_{4,1}$ (k) $\vec{C}_{4,2}$ (l) $\vec{C}_{4,3}$

Figure 3.29: The complete list of conjugated circuits of the graph G.

$F_i \oplus F_j$ of Kekulé structures F_i and F_j for all $i \neq j$ are shown in Figure 3.29. Assign an orientation to each conjugated cycle of each conjugated circuit. If the cycle is aromatic (i.e., its length is congruent to 2 modulo 4) then use anti-clockwise orientation. If it is anti-aromatic (i.e., its length is congruent to 0 modulo 4) then use clockwise orientation. We obtain a directed conjugated circuit $\vec{C}_{i,j}$ from conjugated circuit $C_{i,j}$. Note that all conjugated circuits in Figure 3.29 have already been assigned an orientation.

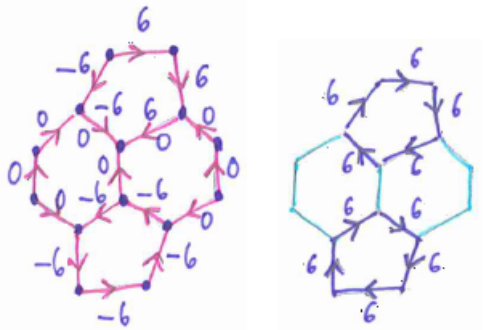
We say that an arc $e \in A(\vec{G})$ agrees with the direction of a directed subgraph $\vec{H} \subseteq \vec{G}$ if e and the corresponding arc in $\vec{H}$ have the same orientation. For each arc $e \in A(\vec{G})$ and for each conjugated circuit $C_{i,j}$ define

$$s(C_{i,j}, e) = \begin{cases} 1, & \text{if } e \in C_{i,j} \text{ and } e \text{ agrees with the direction of } \vec{C}_{i,j}; \\ -1, & \text{if } e \in C_{i,j} \text{ and } e \text{ doest not agree with the direction of } \vec{C}_{i,j}; \\ 0, & \text{otherwise.} \end{cases} \tag{3.86}$$

Now we can define the *π-electron current distribution* on $\vec{G}$:

$$w(\vec{G}, e) = \sum_{i=1}^{K(G)} \sum_{j=1}^{K(G)} s(C_{i,j}, e). \tag{3.87}$$

For the digraph $\vec{G}$ in Figure 3.27 we obtain the π-electron current distribution which is visualised by Figure 3.30(a). The digraph $\vec{G}$ can always be chosen in such way that $w(\vec{G}, e) \geq 0$



(a) (b)

Figure 3.30: The π-electron current distribution on $\vec{G}$.

for every arc $e \in A(\vec{G})$. Such representation is in Figure 3.30(b). The direction and values on arcs e for which $w(\vec{G}, e) = 0$ are not displayed in the figure. This represents the flow of delocalised π electrons when the molecule is placed in a magnetic field which is perpendicular to the plane of the molecule.

3.4.4 Pauling bond order

Let G be a Kekulean graph and let $P(e)$ be the number of Kekulé structures passing through the edge $e \in E(G)$. The number $P(e)$ is called the *matching number of the edge e*.

Definition 3.15. *The number*

$$p(e) = \frac{P(e)}{K(G)}, \tag{3.88}$$

i.e., the fraction of those Kekulé structures that pass through e, is called the Pauling bond order *of e.*

This concept was introduced in 1935 by L. Pauling [156].

Example 3.17. Consider the phenanthrene graph G_p. Recall that $K(G_p) = 5$. Its five Kekulé structures are shown in Figure 3.19. It is not hard to determine the Pauling bond

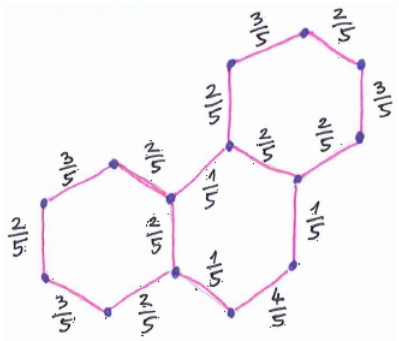


Figure 3.31: Pauling bond orders of the phenanthrene graph G_p.

orders for graph G_p. One can determine the numbers $P(e)$ for each $e \in E(G_p)$ from the Figure 3.19. Pauling bond orders for all edges of G_p are shown in Figure 3.31. □

Note that $p(e)$ is not defined if the graph G is not Kekulean. It is easy to see that

$$\sum_{v \in G(u)} p(uv) = 1 \tag{3.89}$$

for all $u \in V(G)$. It is also obvious that $p(e) \in [0,1] \cap \mathbb{Q}$ for all $e \in E(G)$. For a given number $p \in [0,1] \cap \mathbb{Q}$ let $E_p(G) = \{e \in E(G) \mid p(e) = p\}$, i.e., $E_p(G)$ is the set of edges of G which have Pauling bond order p. Let us define a mapping $\eta_G \colon [0,1] \cap \mathbb{Q} \to \mathbb{N}_0$ by

$$\eta_G(p) = |E_p(G)|. \tag{3.90}$$

Symbol $\mathbb{N}_0$ denotes the set of all non-negative integers. It is easy to see that

$$\sum_{p \in [0,1] \cap \mathbb{Q}} \eta_G(p) = |E(G)|. \tag{3.91}$$

Example 3.18. For the cycle graph $G = C_{2n}$ of even length we have $\eta_G\left(\frac{1}{2}\right) = 2n$ and $\eta_G(p) = 0$ for all $p \neq \frac{1}{2}$.

The edges from $E_0(G)$ have the property that they do not belongs to any Kekulé structures. Similarly, each edge from $E_1(G)$ belongs to all Kekulé structures of G. The Pauling bond order is a pre-quantum analogue of the Coulson π bond order.

3.5 Polyhedral self-assembly

In 2013 Gradišar et al. [84] successfully designed a self-assembly tetrahedral *polypeptide* called TET12. They designed a linear chain of twelve *peptides*, separated by flexible links, such that certain pairs of peptides "glued" together and formed *coiled coil dimers*. The end result was a stable tetrahedron in which each of its six edges was a coiled coil dimer. Every peptide is a chain composed of several *amino acids*. There exist 20 *standard* amino acids, each of which has its name, a three-letter code and a one-letter code (see Table 3.6). The TET12 is composed of 476 amino acids and can be encoded as a string of length 476 using one-letter codes:

$$\begin{aligned}\text{TET12} = {} & \text{START} + \text{APH} + \text{LINK} + \text{P3} + \text{LINK} + \text{BCR} + \text{LINK} + \text{GCN}_{\text{sh}} + \\ & \text{LINK} + \text{APH} + \text{LINK} + \text{P7} + \text{LINK} + \text{GCN}_{\text{sh}} + \text{LINK} + \text{P4} + \\ & \text{LINK} + \text{P5} + \text{LINK} + \text{P8} + \text{LINK} + \text{BCR} + \text{LINK} + \text{P6} + \text{STOP},\end{aligned} \tag{3.92}$$

where

$$\begin{aligned}\text{APH} &= \texttt{"MKQLEKELKQLEKELQAIEKQLAQLQWKAQARKKKLAQLKKKLQA"},\\ \text{BCR} &= \texttt{"DIEQELERAKASIRRLEQEVNQERSRMAYLQTLLAK"},\\ \text{GCN}_{\text{sh}} &= \texttt{"QLEDKVEELLSKNYHLENEVARLKKLVG"},\\ \text{P3} &= \texttt{"SPEDEIQQLEEEIAQLEQKNAALKEKNQALKYG"},\\ \text{P4} &= \texttt{"SPEDKIAQLKQKIQALKQENQQLEEENAALEYG"},\\ \text{P5} &= \texttt{"SPEDENAALEEKIAQLKQKNAALKEEIQALEYG"},\\ \text{P6} &= \texttt{"SPEDKNAALKEEIQALEEENQALEEKIAQLKYG"},\\ \text{P7} &= \texttt{"SPEDEIQALEEKNAQLKQEIAALEEKNQALKYG"},\\ \text{P8} &= \texttt{"SPEDKIAQLKEENQQLEQKIQALKEENAALEYG"},\\ \text{START} &= \texttt{"MYHHHHHHSRAG"},\\ \text{LINK} &= \texttt{"SGPG"} \text{ and}\\ \text{STOP} &= \texttt{"SGTS"}.\end{aligned}$$

In the above expression (3.92), operation + denotes concatenation of strings. Some of the peptide pairs in this chain are "compatible", i.e., they will interlock and form stable coiled coil dimers. On the other hand, those pairs that do not have strong affinity to each other will not form a coiled coil dimer. Let $\mathcal{P}$ be a multiset of $2m$ peptides. In the case of TET12 the multiset $\mathcal{P}$ is

$$\{\text{APH}, \text{APH}, \text{BCR}, \text{BCR}, \text{GCN}_{\text{sh}}, \text{GCN}_{\text{sh}}, \text{P3}, \text{P4}, \text{P5}, \text{P6}, \text{P7}, \text{P8}\}. \tag{3.93}$$

We can imagine a graph with a vertex for every peptide in $\mathcal{P}$ and an edge between two of them if and only if they are compatible. If this graph is a disjoint union of m copies of K_2 it induces a partition on $\mathcal{P}$ into pairs. Then $\mathcal{P}$ is called a multiset of *orthogonal peptide pairs*. Even though the problem of determining whether a pair of peptides forms a stable coiled coil dimer and the problem of finding a large multiset of orthogonal peptide pairs

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic acid	Asp	D
Glutamic acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

Table 3.6: List of the 20 standard amino acids.

are both interesting from a chemical point of view, we will not address them here. This is described in the paper by Gradišar and Jerala [85] and in references cited therein. In our model this information will be given in advance. We know from experimental evidence that the following pairs are orthogonal peptide pairs:

$$(\text{P3}, \text{P4}), (\text{P5}, \text{P6}), (\text{P7}, \text{P8}), (\text{GCN}_{\text{sh}}, \text{GCN}_{\text{sh}}), (\text{APH}, \text{APH}) \text{ and } (\text{BCR}, \text{BCR}). \tag{3.94}$$

Certain pairs consist of two copies of the same peptide. They are called *homodimers*. Otherwise they are called *heterodimers*. This information alone is not sufficient to fully describe the *glueing process* that leads to the tetrahedron. The polypeptide chain TET12 is directed (from START to STOP) and so are all peptides along it. Two peptides may be glued together in such way that they both point in the same direction. In this case they form a *parallel* dimer. If they point in opposite directions, they form an *anti-parallel* dimer. The problem of determining whether a given dimer is parallel or anti-parallel resides outside our current model. This information is provided as input data. For the case of TET12, we know that (APH, APH) and (BCR, BCR) are antiparallel dimers, whilst, all other pairs from (3.94) are parallel dimers.

Before we demonstrate the glueing process, we will describe the mathematical model that we use. The polypeptide chain is modeled with a labeled directed path of length $2m$, where m is the number of peptide pairs. Instead of names APH, P3, BCR, ... we will use letter $a, b, c, \ldots$

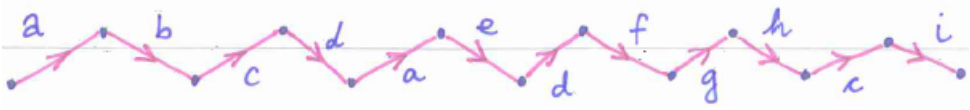


Figure 3.32: Representation of TET12 with a labeled directed path $\vec{P}_{13}$.

Let $\vec{P}_{2m+1}$ be the directed path on $2m+1$ vertices. (It contains $2m$ arcs which represent peptides.) Let Σ be an alphabet with one symbol for each different peptide. For the case of TET12,

$$\Sigma = \{a, b, c, d, e, f, g, h, i\}. \tag{3.95}$$

The map $w\colon A(\vec{P}_{2m+1}) \to \Sigma$ assigns a symbol to each arc. The TET12 can be modeled by the digraph in Figure 3.32. This labeled digraph can be represented as the sequence

$$(a, b, c, d, a, e, d, f, g, h, c, i). \tag{3.96}$$

The information on which peptides glue together is encoded in the *glueing mapping*

$$gl\colon \Sigma \to \Sigma \sqcup \Sigma^{-1}, \tag{3.97}$$

where $\Sigma^{-1} = \{x^{-1} \mid x \in \Sigma\}$. The mapping gl for TET12 is given in Table 3.7. If $gl(x) \in \Sigma$

x	a	b	c	d	e	f	g	h	i
$gl(x)$	a^{-1}	f	c^{-1}	d	h	b	i	e	g

Table 3.7: The glueing mapping for TET12.

then x and $gl(x)$ glue together in a parallel way. If $gl(x) \in \Sigma^{-1}$ then x and $gl(x)$ glue together in an anti-parallel way. Note that if $gl(x) \in \{x, x^{-1}\}$ then x is a homodimer.

The mapping gl can be extended to $gl\colon \Sigma \sqcup \Sigma^{-1} \to \Sigma \sqcup \Sigma^{-1}$ by defining

$$gl(x^{-1}) = gl(x)^{-1} \quad \text{and} \quad (x^{-1})^{-1} = x. \tag{3.98}$$

Note that gl is an involution on $\Sigma \sqcup \Sigma^{-1}$. Let $r\colon \Sigma \sqcup \Sigma^{-1} \to \Sigma \sqcup \Sigma^{-1}$ such that $r(x) = x^{-1}$. Then $\langle gl, r\rangle$ generates a group that is a subgroup of bijections on $\Sigma \sqcup \Sigma^{-1}$. The group $\langle gl, r\rangle$ acts on $\Sigma \sqcup \Sigma^{-1}$. The set $x^{\langle gl,r\rangle} = \{\alpha(x) \mid \alpha \in \langle gl, r\rangle\}$ is called the *orbit* of x. The elements of the orbit space $(\Sigma \sqcup \Sigma^{-1})/\langle gl, r\rangle = \{x^{\langle gl,r\rangle} \mid x \in \Sigma\}$ correspond to peptide pairs.

A *glueing sequence* is a bijection $s\colon \{1, 2, \ldots, m\} \to (\Sigma \sqcup \Sigma^{-1})/\langle gl, r\rangle$. The mapping s is usually given as a vector with m elements from Σ which are representatives of each orbit. This mapping tells us that the peptide pair $s(1)$ glues first, followed by $s(2)$ and so on.

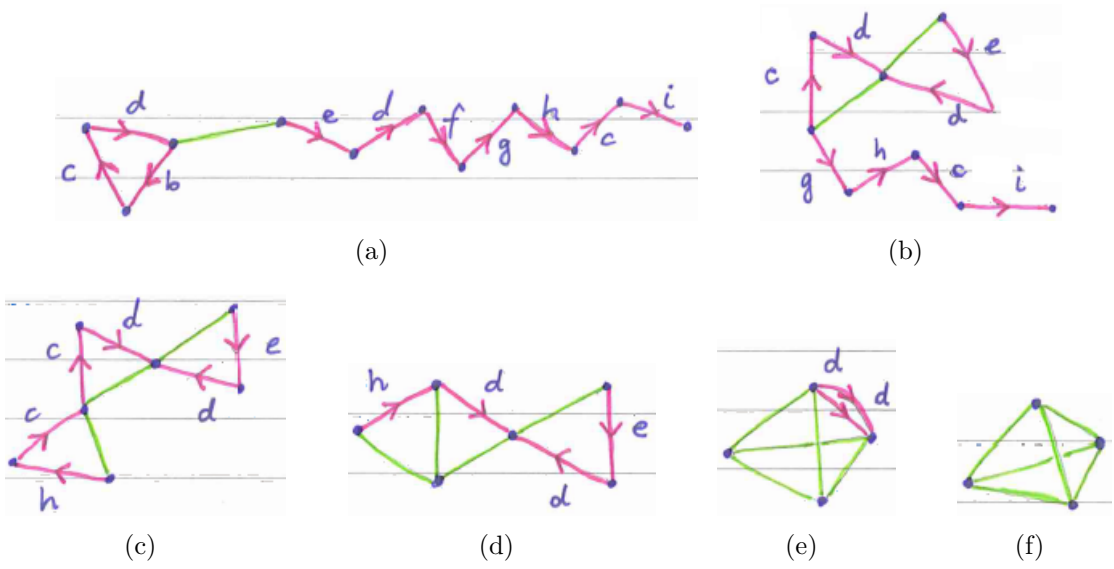
The question of determining the glueing sequence lies outside this model. The glueing sequence is crucial from the chemical viewpoint and was considered in the papers by Kočar et al. [123, 126]. If the sequence is wisely chosen then the polypeptide chain has a greater chance to form the desired polyhedron (note that in a real chemical experiment, a range of malformed byproducts may occur). Besides high folding yield, a good glueing sequence

also folds rapidly on temperature quenching. Note that our model is general enough to also cover the DNA self-assembly.

We know that the pair (APH, APH) is the most stable and will be the first to glue. It is followed respectively by (P3, P4), (P5, P6), (BCR, BCR), (P7, P8) and finally (GCN_{sh}, GCN_{sh}). In our abstract model we use letters $a, b, c, \ldots$ to represent peptides. The glueing sequence of TET12 is

$$(a, b, g, c, e, d). \tag{3.99}$$

Figure 3.33 demonstrates the glueing process of TET12. Because $s(1) = a$, the arcs labeled with a are identified and the graph in Figure 3.33(a) is obtained. The edge that occured by this identification is coloured green in the figure and its direction is not displayed. Because $gl(a) = a^{-1}$, one of the arcs has to be reversed before the identification is carried out. We can arbitrarily choose any of the two, because the end result is essentially the same. Once the two peptides are glued together their orientation does not matter anymore. Because $s(2) = b$, the arcs labeled with b and f are next to be identified. Since $gl(b) = f$ they are glued in the parallel way and the graph on Figure 3.33(b) is obtained. In the end, we obtain the tetrahedron (see Figure 3.33(f)). All the intermediate steps in this process are shown in Figure 3.33.



(a) (b) (c) (d) (e) (f)

Figure 3.33: Glueing process of the TET12.

Note that the labeling of the path $\vec{P}_{13}$ cannot be chosen arbitrarily. Example 3.19 provides a path (see Figure 3.34(a)) which does not result in the tetrahedron even though its labels are a permutation of the original ones.

Example 3.19. Take the same glueing mapping and glueing sequence as in the glueing process of the TET12. The path in Figure 3.34(a) is different from the one in Figure 3.32. The end result in this case is the *bouquet* graph with 6 loops. All the intermediate steps are shown in Figure 3.34. □

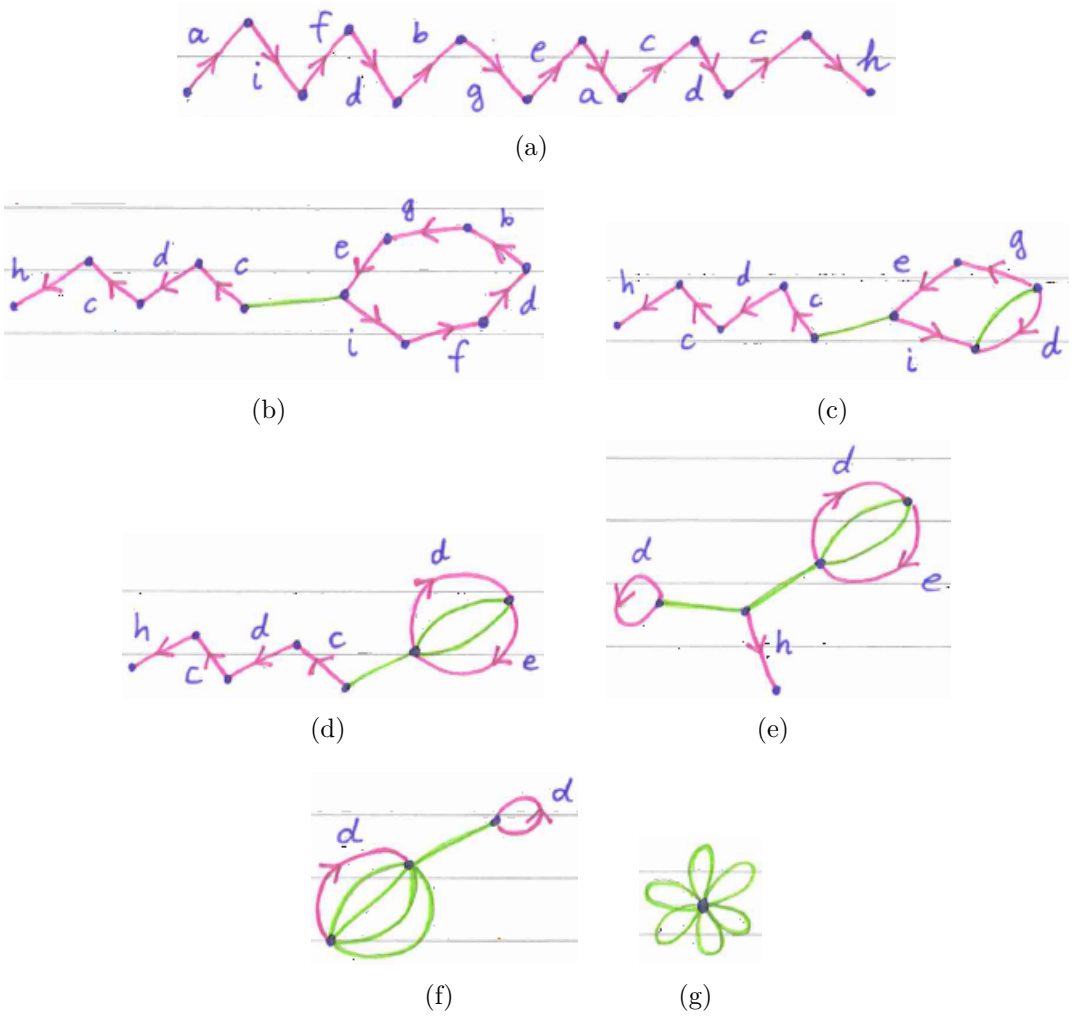


Figure 3.34: A glueing process that does not result in a tetrahedron.

The above example shows us that it is not trivial to find a labeled path that will result in the desired polyhedron. The first mathematical model for this problem, which was developed with trivalent polyhedra in mind, was described by Klavžar and Rus [120]. A year later, the model was refurbished by Fijavž, Pisanski and Rus [67] to include all polyhedra.

Definition 3.16. *A* double trace *in a graph G is a walk which traverses every edge exactly twice.*

A double trace in a simple graph can be given as a sequence of vertices

$$W = w_0\, w_1\, w_2 \ldots w_{2m}, \tag{3.100}$$

where indices are taken modulo $2m$. It is easy to see that every graph G admits a double trace. If we replace every edge of G by a digon, we obtain an *Eulerian multigraph* G'. An *Eulerian circuit* in G' corresponds to a double trace in G. Because all vertices in G' are of even degree the following proposition follows:

Proposition 3.20 (Proposition 2.1 in [67]). *Every graph G has a double trace.* □

Example 3.20. This example demonstrates that a double trace in a graph G does not in general give rise to a directed path (which represents a polypeptide) that will result in the graph G after the glueing process is performed. The skeleton of the tetrahedron is the graph K_4. Let $V(K_4) = \{1, 2, 3, 4\}$. The walk

$$W = 1\ 2\ 4\ 2\ 3\ 1\ 4\ 3\ 1\ 2\ 3\ 4\ 1 \tag{3.101}$$

is a double trace in K_4. The polypeptide chain can be obtained in the following way. Take the directed path $\vec{P}_{13}$ and label its arcs consecutively with letters $a, b, c, \ldots, l$. Simultaneously traverse the path $\vec{P}_{13}$ and W. When an edge of K_4 is traversed for the first time, label it with the label of the corresponding arc of $\vec{P}_{13}$ and direct the edge of K_4 in the direction of its traversal. Now suppose that it is traversed for the second time. Let x be the label on the arc of K_4 and y the label on the arc of $\vec{P}_{13}$. If the arc of K_4 is traversed in the same direction, define $gl(x) = y$ and $gl(y) = x$. If it is traversed in the opposite direction, define $gl(x) = y^{-1}$ and $gl(y) = x^{-1}$. The resulting glueing mapping is given in Table 3.8 and the labeled digraph that was obtained from K_4 during this procedure is in Figure 3.35.

x	a	b	c	d	e	f	g	h	i	j	k	l
$gl(x)$	i	c^{-1}	b^{-1}	j	h	l^{-1}	k^{-1}	e	a	d	g^{-1}	f^{-1}

Table 3.8: The glueing mapping from Example 3.20.

Take an arbitrary glueing sequence and perform the glueing process on the $\vec{P}_{13}$ using the

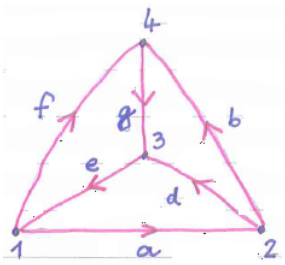


Figure 3.35: The labeled digraph that was obtained from K_4 during the procedure of determining the polypeptide chain from a given double trace.

glueing mapping that was obtained in the above procedure. (Note that the end result is independent of the glueing sequence.) The resulting graph is shown in Figure 3.36. Note that if vertices 4 and $4'$ were identified then we would have obtained the K_4. □

Definition 3.17. *Let $e = uv \in E(G)$ and let $W = w_0\, w_1\, w_2 \ldots w_{2m}$ be a double trace in G. If there exists an integer i such that $(w_{i-1}, w_i, w_{i+1}) = (v, u, v)$ then W has a* retracing.

Let $u \in V(G)$ and let $v, v' \in G(u)$ such that $v \neq v'$. If there exist integers i and j, $i \neq j$, such that $w_i = w_j = u$ and $\{w_{i-1}, w_{i+1}\} = \{w_{j-1}, w_{j+1}\} = \{v, v'\}$ then W has a repetition *through vertex u.*

Klavžar and Rus defined the notion of a stable trace:

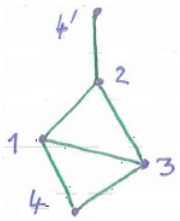


Figure 3.36: The result of the glueing process on the polypeptide obtained in Example 3.20.

Definition 3.18. *A* stable trace *is a double trace that has no retracing and no repetition through its vertices.*

A stable trace of a graph G gives rise to a polypeptide that results in the graph G when the glueing process is performed. Note that the double trace from Example 3.20 has both a retracing and a repetition. In both cases the vertex labeled with 4 is involved. They proved the following theorem:

Theorem 3.21 (Klavžar and Rus, Theorem 3.1 in [120])**.** *A graph G admits a stable trace if and only if $\delta(G) \geq 3$.* □

Fijavž, Pisanski and Rus generalised the notion of a stable trace:

Definition 3.19. *Let G be a graph, $v \in V(G)$ and $N \subseteq G(v)$. Let W be a double trace in G. Then W has a* N-repetition *at v if for all integers i such that $w_i = v$ it holds that the pair $\{w_{i-1}, w_{i+1}\}$ is either contained in N or disjoint from N.*

Definition 3.20. *A* n-stable trace *is a double trace where for all $v \in V(G)$ and for all $N \subseteq G(v)$ such that $1 \leq |N| \leq n$ it holds that W has no N-repetition.*

They defined the important notion of a strong trace:

Definition 3.21. *Let G be a graph and let W be a double trace of G. If for every vertex $v \in V(G)$ it holds that W has a N-repetition at v if and only if $N = G(v)$ or $N = \emptyset$ then W is called a* strong trace*.*

They also proved the following theorem which is based on a deep result from topological graph theory [93]:

Theorem 3.22 (Fijavž, Pisanski and Rus, Theorem 3.5 in [67])**.** *Every graph admits a strong trace.* □

Let W be a double trace of a graph G. The double trace W traverses each edge $e \in E(G)$ exactly two times. If W traverses an edge e in the same direction both times then e is *parallel* with respect to W. Otherwise, it is *antiparallel* with respect to W. If all edges of G are parallel with respect to W then W itself is called a *parallel trace*. Similarly, if all edges of G are antiparallel with respect to W then W is called an *antiparallel trace*. It is easy to see that a parallel / antiparallel edge of W gives rise to a parallel / antiparallel dimer in the self-assembled polypeptide.

It is possible to obtain a new double trace from existing one. We can change the direction of tracing (reverse the trace W) or start at a different vertex (shift the trace W). If the graph G posesses a symmetry, we can obtain a new trace by acting by a graph automorphism $\alpha \in \mathrm{Aut}(G)$ on W.

Definition 3.22. *Double traces W and W' are* equivalent *if W' can be obtained from W by using any combination of the following operations:*

(i) reversion of W;

(ii) shifting W;

(iii) applying a permutation on W induced by an automorphism of G.

Otherwise, traces W and W' are called non-equivalent.

Let $\mathcal{T}(G)$ denote the set of all double traces of a graph G. The equivalence of double traces is an equivalence relation on the set $\mathcal{T}(G)$. This is clearly also an equivalence relation on any subset of $\mathcal{T}(G)$, such as stable traces, strong traces and so on. Assume that vertices $V(G) = \{v_0, \ldots, v_{n-1}\}$ of the graph G are linearly ordered as

$$v_0 < v_1 < \cdots < v_{n-1}. \tag{3.102}$$

This ordering induces a lexicographic ordering on the set of double traces of G, i.e., $W \leq W'$ if and only if $W = W'$ or there exists an index i, $0 \leq i \leq 2m$, such that $w_i < w'_i$ and $w_j = w'_j$ for all $j < i$. Every subset $\mathcal{S} \subseteq \mathcal{T}(G)$ has a lexicographically smallest member which is called the *canonical representative* of $\mathcal{S}$.

Let us define mappings $\rho, \sigma_i \colon \mathcal{T}(G) \to \mathcal{T}(G)$ as

$$\rho(w_0 \ldots w_{2m}) = w_{2m} \ldots w_0 \qquad \text{and} \qquad \sigma_i(w_0 \ldots w_{2m}) = w_i \ldots w_{2m+i}. \tag{3.103}$$

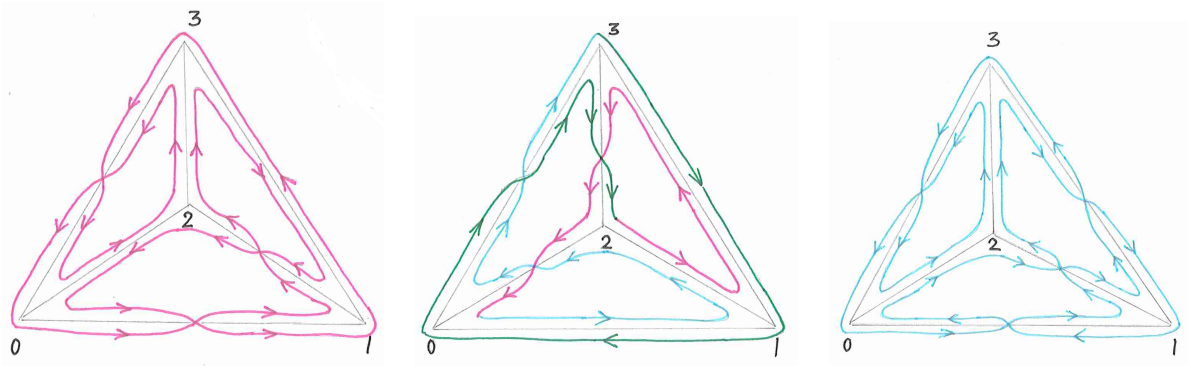
Note that $\sigma_0 = \sigma_{2m} = \mathrm{id}_{\mathcal{T}(G)}$. Let $\alpha \in \mathrm{Aut}(G)$. The automorphism α acts on $\mathcal{T}(G)$ in the following way:

$$\alpha(w_0 \ldots w_{2m}) = \alpha(w_0) \ldots \alpha(w_{2m}). \tag{3.104}$$

Then $\mathrm{Aut}(G)$, $R = \{\mathrm{id}, \rho\}$ and $S = \{\sigma_i \mid i = 0, \ldots, 2m-1\}$ are three groups acting on $\mathcal{T}(G)$. Elements of the orbit space $\mathcal{T}(G)/(\mathrm{Aut}(G) \times R \times S)$ are precisely equivalence classes of double traces for the equivalence relation from Definition 3.103.

The group $\mathrm{Aut}(G) \times R \times S$ partitions the 672 different strong traces of the tetrahedron graph into 3 equivalence classes of sizes 288, 288 and 96 [12]. Representatives of those three equivalence classes are visualised in Figure 3.37. Note that the traces are represented by closed curves, because the position of the starting vertex can be arbitrarily chosen (since traces that can be obtained from each other by shifting operations σ_i are equivalent). The curve that represents the trace is drawn in such a way that it intersects itself and crosses the drawing of graph G only at the midpoint of an edge.

A branch-and-bound algorithm that outputs each canonical strong trace of G is described in [12]. Here, we introduce *map traces* and describe an algorithm for enumerating them that uses dynamic programming.



(a) (b) (c)

Figure 3.37: The three non-equivalent strong traces of the tetrahedron graph.

We may view the strong traces in Figure 3.37 as walks traversing all flags of a map M which is a combinatorial representation of a cellular embedding of the graph G. The trace can be therefore interpreted as a 2-regular graph whose vertices are flags of M. This motivates us to define the following:

Definition 3.23. *Let $M = (\Phi, \tau_0, \tau_1, \tau_2)$ be a map. Choose a mapping $\lambda\colon E(M) \to \{-1, 1\}$. An* (undirected) map trace*, denoted Q_λ, is a 2-regular graph whose set of vertices is Φ and its edges are defined as follows:*

a) $\phi \sim \tau_1(\phi)$;

b) $\phi \sim \tau_0(\phi)$ when $\lambda(e_\phi) = 1$ and $\phi \sim \tau_2\tau_0(\phi)$ when $\lambda(e_\phi) = -1$.

Map trace Q_λ in general consists of one or more even cycles. We are only interested in those traces that are connected:

Definition 3.24. *A* connected (undirected) map trace *Q_λ is a map trace which has a single connected component.*

If we choose an initial flag ϕ_0 and a direction, we can assign a strong trace to a connected map trace Q_λ. Traverse the map trace Q_λ in the given direction starting at ϕ_0 to obtain the closed walk $(\phi_0, \phi_1, \phi_2, \ldots, \phi_{4m} = \phi_0)$. Then

$$W(Q_\lambda) = (v_{\phi_0}, v_{\phi_2}, v_{\phi_4}, \ldots, v_{\phi_{4m}}) \tag{3.105}$$

is a strong trace. Most of the time, we are interested in non-equivalent strong traces. Because by shifting and reversing a strong trace W we obtain an equivalent strong trace, the intial flag ϕ_0 and the direction on the connected map trace Q_λ can be arbitrarily chosen.

Example 3.21. The concepts introduced above are illustrated in Figure 3.38 which shows an embedding of the (trigonal) bipyramid graph in the plane. Its vertices are labeled with numbers $1, 2, \ldots 5$, its 9 edges are labeled with lowercase letters $a, b, \ldots, i$ and its 6 faces are labeled with uppercase letters $A, B, \ldots, F$. Green dots in Figure 3.38 represent flags. This particular map is flag-simple, so we can uniquely label the flags with vertex-edge-face triples; two of those labels, namely $(5, g, F)$ and $(5, i, F)$, are indicated in Figure 3.38. It

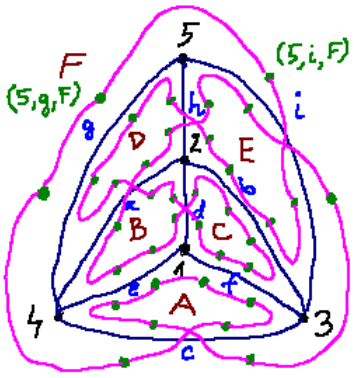


Figure 3.38: Embedding of the bipyramid in the plane and one of its connected map traces.

should be obvious to the reader how to label the rest. The magenta line represents the map trace.

Note that in Figure 3.38 the edges of the map trace Q_λ (magenta segments) can be drawn in such a way that they do not intersect the lines that represent edges of the skeleton graph $G = \mathrm{Skel}(M)$, except in the middle of some edges where the map trace also crosses itself. Such a crossing on edge e occurs precisely when $\lambda(e) = -1$. Note that the drawing of the map trace does not intersect itself near any vertex of the skeleton graph. This property makes perfect sense from a chemical point of view [15]. □

Definition 3.25. *Map M is* compatible *with a strong trace W if there exists a map trace Q such that $W = W(Q)$.*

Not every strong trace is compatible with every map. For example, if a strong trace of the bipyramid contains the subsequence $(1, 2, 5)$ then it is obviously not compatible with the map in Figure 3.38.

Note that different map traces may yield the same strong trace as in the case of the cycle graph C_5 embedded in the plane as shown in Figure 3.39. However, if minimum degree in

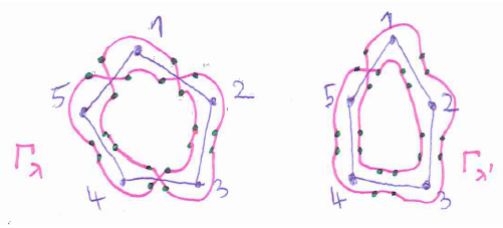


Figure 3.39: Two different connected map traces Q_λ and $Q_{\lambda'}$ yielding the same strong trace, i.e., $W(Q_\lambda) = W(Q_{\lambda'})$.

the graph $G = \mathrm{Skel}(M)$ is at least 3, the map trace is either:

(a) uniquely determined from the strong trace or

(b) it is not compatible with the given map M.

For most applications it is sufficient to find strong traces that are compatible with a fixed embedding of the graph G. We can immediately obtain a simple brute-force algorithm with the time complexity that is exponential in m, where m is the number of edges of the skeleton graph $\mathrm{Skel}(M)$. The algorithm enumerates all strong traces that are compatible with a given map M:

Let M be a map with $4m$ flags and let $G = \mathrm{Skel}(M)$. Let $L = \emptyset$. In the end, the set L will contain all canonical strong traces that are compatible with M. Iterate over all subsets $\mathcal{E} \subseteq E(G)$. For a given set $\mathcal{E}$ define the mapping $\lambda_{\mathcal{E}} \colon E(G) \to \{-1, 1\}$, such that

$$\lambda_{\mathcal{E}}(e) = \begin{cases} 1, & e \in \mathcal{E}; \\ -1, & \text{otherwise.} \end{cases} \tag{3.106}$$

Check whether $Q_{\lambda_{\mathcal{E}}}$ is a connected map trace. If it is connected, take $W = W(Q_{\lambda_{\mathcal{E}}})$. Then determine the canonical strong trace W' that is equivalent to W and add W' to the set L. After we process all subsets of $E(G)$, the set L will contain all canonical strong traces that are compatible with M.

Now imagine that we cut the map (that has a strong trace drawn on it) along its edges. From the tetrahedron we obtain four 'jigsaw puzzle pieces' (see Figure 3.40). Each piece corresponds to one face. The strong trace that was drawn on the map can be obtained

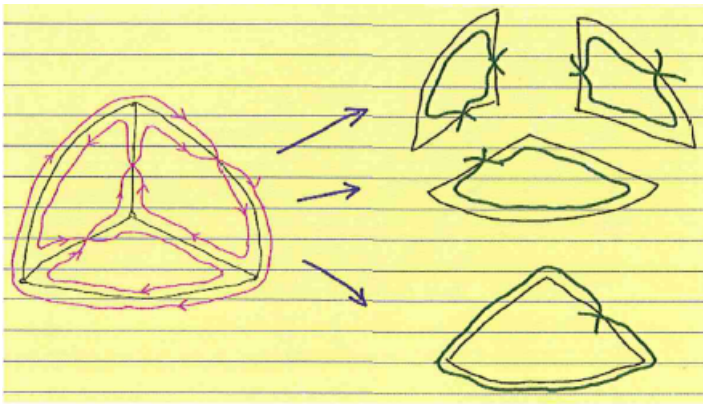


Figure 3.40: Jigsaw puzzle pieces.

from these jigsaw pieces by glueing them back together. This is the main idea behind the *dynamic programming* algorithm. In order to provide a formal description of the algorithm, we first define:

Definition 3.26. *Let Q_λ be a map trace. A* partial map trace of a face $f \in F(M)$ *is a subgraph of Q_λ induced on the set of flags $\{\phi \mid f_\phi = f\}$.*

A partial map trace of a subset of faces $\mathcal{F} \subseteq F(M)$ *is a subgraph of Q_λ induced on the set of flags $\{\phi \mid f_\phi \in \mathcal{F}\}$.*

Clearly, a partial trace is a disjoint union of cycles and paths. We will call it *admissible* if:

(a) it does not contain cycles and $\mathcal{F}$ is a proper subset of $F(M)$ or

(b) it is a connected trace and $\mathcal{F} = F(M)$.

By choosing a proper subset of faces of the map, we obtain a surface with boundary. This boundary is a disjoint union of cycles:

$$\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_r. \tag{3.107}$$

On each of those cycles we choose a reference vertex $v_i \in \mathcal{C}_i$ and a direction. The *signature*, denoted by $\sigma(\widetilde{Q})$, of a partial map trace $\widetilde{Q}$ can be obtained in the following way:

Traverse, consecutively, the cycles $\mathcal{C}_1, \mathcal{C}_2, \ldots$ of the boundary by starting at the reference vertex and travelling in the chosen direction. For each flag ϕ that you encounter along the walk, write down one symbol:

(a) If the degree of vertex ϕ in the partial map trace is 2, write down the symbol $\star$.

(b) If the degree of vertex ϕ is 1, find the other end of this path. If it was already labeled, write down its label. If it was not labeled yet, label it using the smallest positive integer that was not used so far.

Example 3.22. Figure 3.41 shows an example where 5 faces were chosen from a larger map.

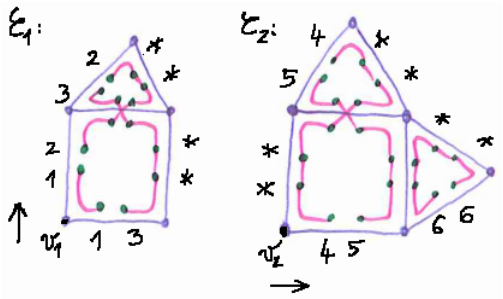


Figure 3.41: A partial map trace together with its signature.

The boundary consists of two cycles denoted $\mathcal{C}_1$ and $\mathcal{C}_2$. Reference vertices and directions are also indicated. The signature of this partial map trace is

$$(1, 2, 3, 2, \star, \star, \star, \star, 3, 1, 4, 5, 6, 6, \star, \star, \star, \star, 4, 5, \star, \star). \tag{3.108}$$

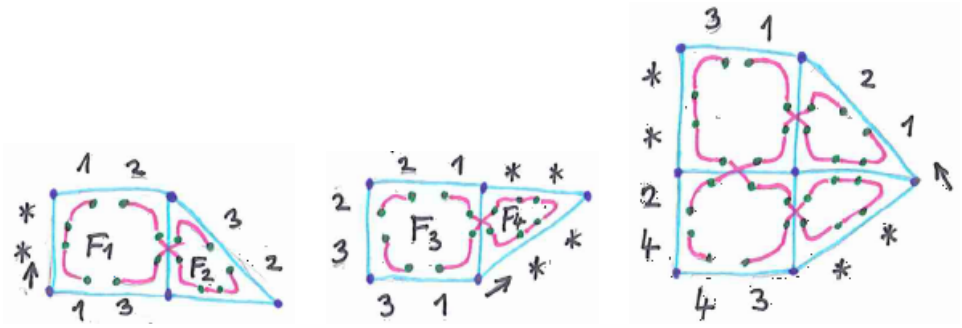
□

If the partial map trace $\widetilde{Q}$ is admissible then we say that $\sigma(\widetilde{Q})$ is also admissible.

Definition 3.27. *Let $\mathcal{F}, \mathcal{F}' \subset F(M)$ such that $\mathcal{F} \cap \mathcal{F}' = \emptyset$. Let $\widetilde{Q}$ and $\widetilde{Q}'$ be their corresponding partial map traces. We say that $\widetilde{Q}$ and $\widetilde{Q}'$ are* compatible *if for each edge $e \in E(M)$ that is shared between the boundaries of $\mathcal{F}$ and $\mathcal{F}'$ it holds that all vertices $\{\phi \mid e_\phi = e\}$ of the disjoint union of partial map traces $\widetilde{Q}$ and $\widetilde{Q}'$ have the same degree (which is either 1 or 2).*

We can *merge* partial map traces $\widetilde{Q}$ and $\widetilde{Q}'$ if they are compatible. While there exists an edge $e \in E(M)$ that is shared between $\mathcal{F}$ and $\mathcal{F}'$ and a degree-1 vertex ϕ such that $e_\phi = e$, connect the vertices ϕ and $\tau_2\tau_0(\phi)$. In this way we obtain the merged partial map trace $\widetilde{Q} \circ \widetilde{Q}'$. If two partial map traces $\widetilde{Q}_1$ and $\widetilde{Q}_2$ are compatible then we say that $\sigma(\widetilde{Q}_1)$ and $\sigma(\widetilde{Q}_2)$ are also compatible.

Example 3.23. Let $\widetilde{Q}_1$ and $\widetilde{Q}_2$ be two partial map traces (see Figures 3.42(a) and 3.42(b)). Suppose that faces $F_1, \ldots, F_4$ are part of a larger map in which faces F_1 and F_3 as well as faces F_2 and F_4 are adjacent. The partial map traces $\widetilde{Q}_1$ and $\widetilde{Q}_2$ are compatible. We can



(a) $\widetilde{Q}_1$ (b) $\widetilde{Q}_2$ (c) $\widetilde{Q}_1 \circ \widetilde{Q}_2$

Figure 3.42: A merged partial map trace $\widetilde{Q}_1 \circ \widetilde{Q}_2$.

therefore merge them and obtain the partial map trace $\widetilde{Q}_1 \circ \widetilde{Q}_2$ in Figure 3.42(c). □

The key observation is that in order to obtain $\sigma(\widetilde{Q} \circ \widetilde{Q}')$, we only need $\sigma(\widetilde{Q})$ and $\sigma(\widetilde{Q}')$. To do so, first change the labels in one of the two signatures to obtain distinct labels. Then look at pairs of vertices $(\phi, \tau_0\tau_2(\phi))$ where ϕ belongs to an edge that is shared between $\mathcal{F}$ and $\mathcal{F}'$. The corresponding labels in the signature can be one of the following (if the two partial map traces are compatible):

(a) Both are labeled by $\star$ in which case we ignore them.

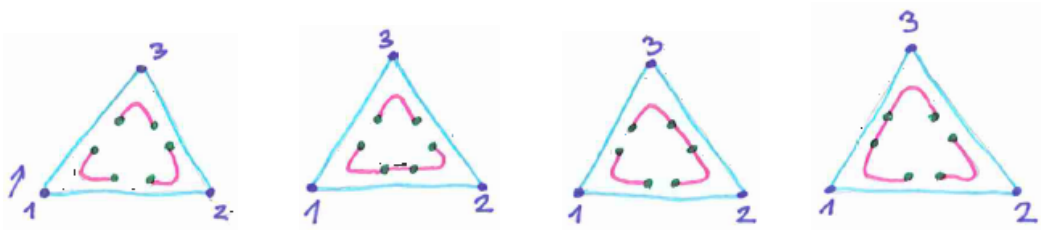
(b) The labels on ϕ and $\tau_0\tau_2(\phi)$ may be two different numbers (in that case relabel the signature using the smallest of the two numbers).

(c) If the two labels are equal, ignore them (in that case the merged partial trace will contain a cycle and will not be admissible, unless $\mathcal{F} \cup \mathcal{F}' = F(M)$).

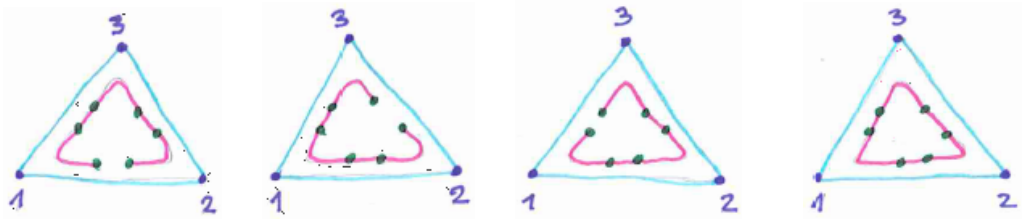
Then traverse the boundaries of $\mathcal{F} \cup \mathcal{F}'$ with respect to chosen reference vertices and directions. Relabel the signature in such a way that the first occurances of numerical labels will appear in increasing order within the trace and consecutive integers (starting with 1) will be used as labels.

The dynamic programming algorithm has three phases which we call the *preparation phase*, the *main phase* and the *final phase*.

In the preparation phase of the algorithm, we prepare a list of signatures of all admissible partial map traces of each face $f_i \in F(M)$.



(a) $(1, 2, 2, 3, 3, 1)$ (b) $(1, 2, 2, 1, \star, \star)$ (c) $(1, 2, \star, \star, 2, 1)$ (d) $(\star, \star, 1, 2, 2, 1)$



(e) $(\star, \star, \star, \star, 1, 1)$ (f) $(\star, \star, 1, 1, \star, \star)$ (g) $(1, 1, \star, \star, \star, \star)$ (h) Non-admissible.

Figure 3.43: Partial map traces of a triangular face.

Example 3.24. A triangular face has 7 admissible partial map traces and 1 non-admissible partial map trace. They are shown in Figure 3.43. Their signatures were also determined. The partial map trace in Figure 3.43(h) is non-admissible. □

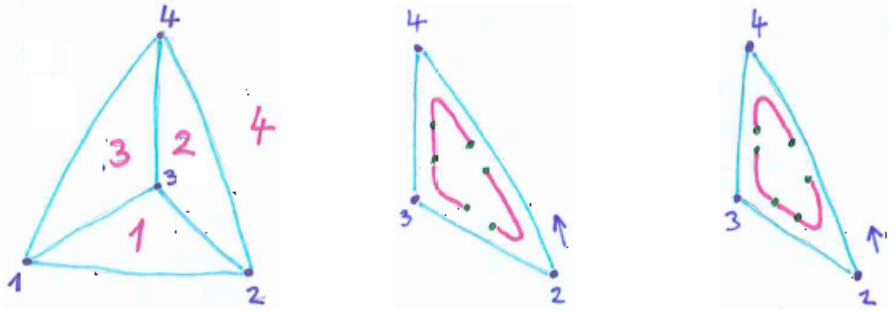
In the preparation phase we also choose a linear ordering $f_1 < f_2 < \cdots < f_{|F(M)|}$ on the faces of $F(M)$.

In the main phase, we begin with a single face f_1 and iteratively add new faces to it, one by one. Formally, define $\mathcal{F}_i = \{f_1, \ldots, f_i\}$ for $i = 1, \ldots, |F(M)|$. For each $\mathcal{F}_i$, choose an ordering on the cycles that comprise the boundary of $\mathcal{F}_i$ and choose a reference vertex and a direction on each of the cycles. Those reference vertices and directions will be common to all signatures of $\mathcal{F}_i$.

For each $i = 1, 2, \ldots, |F(M)| - 1$ we will create a dictionary $\mathcal{D}_i$ whose keys will be signatures of all admissible partial map traces of $\mathcal{F}_i$. The value in the dictionary that corresponds to a signature will be the number of all partial map traces with the given signature. The initial dictionary $\mathcal{D}_1$ contains admissible signatures of the face f_1. Note that all values in this dictionary are equal to 1.

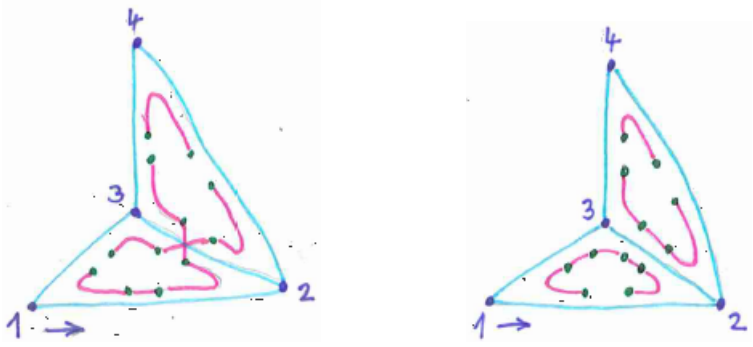
The dictionary $\mathcal{D}_{i+1}$ can be obtain from the dictionary $\mathcal{D}_i$ and the list of admissible signatures of the face f_{i+1}. Combine each signature in $\mathcal{D}_i$ with each signature of f_{i+1}. If they are compatible and the resulting signature is admissible, increase the count of the resulting admissible signature in $\mathcal{D}_{i+1}$ by the corresponding value stored in $\mathcal{D}_i$.

Example 3.25. Order the faces of the tetrahedral map as indicated by red numbers in Figure 3.44(a). The dictionary $\mathcal{D}_1$ contains all admissible signatures from Figure 3.43. The signature in Figure 3.44(b) cannot be merged with the one in Figure 3.43(e), because they are not compatible. But it can be merged with the signature in Figure 3.43(b) and the partial map trace in Figure 3.45(a) is obtained. On the other hand, the signature in Figure 3.44(c) is not compatible with the one in Figure 3.43(b), but is is compatible with the one in Figure 3.43(e) and they merge into a partial map trace in Figure 3.45(b). □



(a) (b) $(1, 2, \star, \star, 2, 1)$ (c) $(1, 2, 2, 1, \star, \star)$

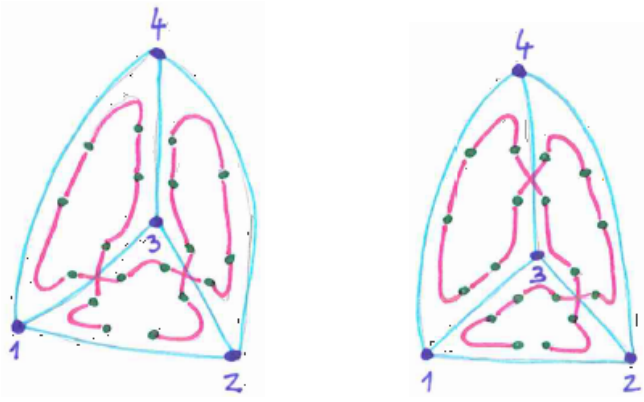
Figure 3.44: Ordering on the faces of the tetrahedron and two (admissible) partial map traces of the face f_2.



(a) $(\star, \star, 1, 2, 2, 3, 1, 3)$ (b) $(1, 1, 2, 3, 3, 2, \star, \star)$

Figure 3.45: Two examples of signatures that are stored in dictionary $\mathcal{D}_2$.

Example 3.26. Note that the same signature can be obtained in many different ways. The signature $(1, 1, \star, \star, \star, \star)$ from dictionary $\mathcal{D}_3$ belongs to more than one partial map trace. Two of them are shown in Figure 3.46.



(a) (b)

Figure 3.46: The signature $(1, 1, \star, \star, \star, \star)$ from dictionary $\mathcal{D}_3$ belongs to more than one partial map trace.

In the final phase of the algorithm, we build the dictionary $\mathcal{D}_{|F(M)|}$. When the last face $f_{|M(G)|}$ is added, the boundary disappears. Again, combine each key of $\mathcal{D}_{|F(M)|-1}$ with each admissible signature of $f_{|M(G)|}$. If they are compatible and merge into a connected map trace, increase the total count of connected map traces by the corresponding value in $\mathcal{D}_{|F(M)|-1}$.

In this algorithm, the number of different connected map trace is obtained. It is possible to reconstruct the i-th map trace from the data structures that were used in the algorithm. Note that the symmetries of the map M were not taken into account here. For more details on this dynamic programming algorithm, see reference [13].

Chapter 4

The Coulson conjecture on maximum bond number

When we defined the free valence in Section 3.2.1, we assumed that the maximum π *bond number* (i.e., the total π bond order around a sp^2 carbon atom) that can be theoretically obtained (on any centre in any sp^2 π system) is no larger than $\sqrt{3}$. This statement does not appear to have been formally proved. We will provide empirical evidence on the behaviour of maximum π bond number as a function of vertex count, n, of chemical graphs and describe the family of graphs that realises local maxima for small n. This research was initiated by Henry Zubaida in his M.Chem research project at The University of Sheffield.

Let

$$N^{\pi}(G) = \max_{r} N_r^{\pi}(G), \tag{4.1}$$

where G is a given graph. If $\mathcal{G}$ is a family of graphs, then

$$N^{\pi}(\mathcal{G}) = \max_{G \in \mathcal{G}} N^{\pi}(G). \tag{4.2}$$

Let us state the conjecture on maximum π bond number.

Conjecture 4.1. *For all chemical graphs (i.e., connected subcubic graphs) G it holds that*

$$N^{\pi}(G) \leq \sqrt{3}. \tag{4.3}$$

□

Conjecture 4.2. *The star on 4 vertices $K_{1,3}$ is the only chemical graph for which the conjectured maximum bond number $\sqrt{3}$ is attained, i.e., $N^{\pi}(K_{1,3}) = \sqrt{3}$ and $N^{\pi}(G) < \sqrt{3}$ if $G \neq K_{1,3}$ is a chemical graph.* □

Let us first give a brief historical overview. This story goes back to the early days of molecular-orbital theory. How strong, in total, can the π bonds involving a single sp^2 carbon centre be? The folklore answer, based on publications of the Coulson school from the late 1940's and early 1950's, is that in the Hückel model the π bond number of an atom cannot exceed $\sqrt{3}$ (see Conjecture 4.1), and further that this maximum is uniquely realised by the *trimethylenemethane* carbon skeleton (see Conjecture 4.2) in neutral, cation and

anionic states. How did this belief in what is effectively a graph-theoretical conjecture came about? Unfortunately, the main actors in the story are no longer available for comment, and we are left with indirect inference from clues in the literature.

A discussion of the historical context is given in the book by Gavroglu and Simões [80]. In the early days of applications of quantum mechanics in chemistry, in the 1940's radical reactivity was studied by the French school of Daudel, Pullman and Pullman [168] and in England and Scotland by Coulson and co-workers [40]. The general principle was that the reactivity at a given centre would depend on the *free valence* remaining after the electrons around the atom had contributed to existing local and delocalised bonds.

Evolution of Coulson's thinking on this topic can be followed in the literature. In comments contributed to the published proceedings of the 1945 Faraday Discussion on Oxidation [119], he notes that free valence is reduced whenever a carbon atom is involved in making bonds, and he characterizes the degree of that involvement by the *bond number*, N. Discussion of bond number can be complicated by the existence of different conventions. N may include both σ and π bonds, or π bonds alone; the two values differ trivially, as $N_r^{\sigma+\pi} = d_r + N_r^{\pi}$, where d_r is the degree of vertex r in the graph representing the carbon skeleton. To avoid ambiguity, we explicitly write either $N^{\sigma+\pi}$ or N^{π}, e.g., $3+\sqrt{3}$ for $N^{\sigma+\pi}$ in trimethylenemethane, but $\sqrt{3}$ for N^{π} in the same molecule. If both σ and π bonds are included in the count, the bond number of atom r

$$N_r^{\sigma+\pi} = \sum_{s\sim r}(1 + P_{rs}^{\pi}) = d_r + \sum_{s\sim r}\sum_{i=1}^{n} v_i \left(c_r^{(i)}\right)^{\star} c_s^{(i)}, \tag{4.4}$$

where d_r is the degree of r in the carbon skeleton of the molecule, the notation $s \sim r$ implies that s runs over neighbours σ-bonded to r, v_i is the occupation number (2, 1 or 0) of π molecular orbital i, and $c_r^{(i)}$ is the coefficient of the p_π atomic orbital on centre s in that molecular orbital. If complex orbitals are used, bond orders remain real, provided that complete eigenspaces (full sets of degenerate orbitals) are included in the sum.

By 1947, Coulson is noting in a paper contributed to the Faraday Discussion on The Labile Molecule, that free valence can be quantified as the gap between the actual value of N_r and some theoretical maximum value $N_{\max}$ [40]. Initially, he assumed that $N_{\max}^{\sigma+\pi} = 3 + 1.680$, equating to a maximum π bond number of 1.680. Although not explicitly stated in the paper, this value ("... the maximum ... found so far for carbon in any molecule ...") is consistent with the bond number for the degree-3 carbon centre in the benzyl *radical* ($N^{\pi} = 1.6801416$), a molecule that features in the paper. Almost immediately, this value was revised. In a stop-press footnote to his 1948 paper on Free Valence in Organic Reactions, contributed at a French meeting, he corrects this number, writing [41] "(*) Note added in proof. Mr W. E. Moffitt has just shown that the greatest possible value of $N_{\max}$ is $3 + \sqrt{3} = 4.73$. This only occurs in the hypothetical molecule $C(CH_2)_3$. 4.68 seems to be the largest value for molecules which are not free radicals." Given the context of radical reactivity, it is perhaps surprising that Coulson should have excluded radicals; it is not clear whether he intended also to exclude closed-shell ionic states of molecules that would be radicals as neutrals. The circumstantial evidence indicates that he was aware that benzyl cation would give $N_r^{\pi} \approx 1.68$, as would the neutral benzyl radical. However, if benzyl cation is to be included, why not include trimethylenemethane cation as a non-radical too?

The footnote already implies at least two mathematical conjectures, one about the value of $N_{\max}$ for chemical graphs (see the above Conjecture 4.1), and another for the restriction to *non-radical chemical graphs*.

Definition 4.1. *A chemical graph G is a* non-radical *if it has an even number of vertices and a non-zero gap between eigenvalues $\lambda_{n/2}$ and $\lambda_{n/2+1}$.*

Longuet-Higgins [129] remarks on this change from 1.68 to $\sqrt{3}$ for maximum π bond order. (He cites the 1948 footnote for Moffitt's contribution and adds a citation of his own unpublished work.) Moffitt was a student of Coulson, who submitted his D. Phil. thesis on 'On the Electronic Structure of Molecules' in Oxford in October 1948 [140]. It includes (Chapter III) a theoretical investigation of the trimethylenemethane radical, which inspired experimental investigation and characterisation of this species [62, 66, 199].

Moffitt's thesis contains a brief discussion of $N_{\max}$ in the context of a spin-coupled model of electronic structure, and offers a plausibility argument (rather than a mathematical proof) for the limiting value, asserting "... Now the strength of the π bonding at atom i will be a maximum, and therefore the residual affinity will be a minimum, when the spins of the electrons on neighbouring atoms j are free to align with s_i in the strongest way – i.e., when the spins are not pre-emptively aligned owing to the presence of other conjugated parts of the molecule..." [140]. In the published version of the chapter [141] he adds "Coulson has recently suggested that the molecular orbital method may also be used to predict reactivity. Analogous quantities $r_i = 1 - \sum_j^{\S} p_{ij}$ may also be defined for these calculations, and arguments similar to those used above lead to minimum affinities at $r = 0$, $(1 - \sqrt{2})$ and $(1 - \sqrt{3})$ respectively."

This seems to be the only statement that Moffitt made in print on the value of $N_{\max}$. Mathematically, it amounts to a conjecture that $N_{\max}$ is equal to $\sqrt{d}$ and is realised by the vertices of degree $d = 1$, 2 and 3 in the graphs $K_{1,1}$, $K_{1,2}$ and $K_{1,3}$, respectively (see Figure 4.1). We will call the bound $\sqrt{d}$ for graphs G with $\Delta(G) \leq d$ the *Moffitt bound*. The justification of the $\sqrt{3}$ maximum is repeated, more or less verbatim, in later

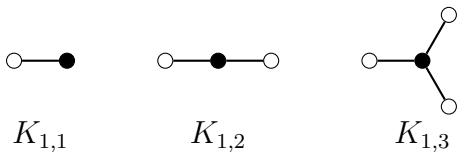


Figure 4.1: The star graphs $K_{1,1}$ to $K_{1,3}$, with maximum π bond number $\sqrt{1}$, $\sqrt{2}$ and $\sqrt{3}$, respectively, on the black centres. $K_{1,1}$ is the complete graph on two vertices (molecular graph of ethene) and $K_{1,2}$ is the path on three vertices (molecular graph of the allyl radical). $K_{1,3}$ is the molecular graph of the trimethylenemethane radical.

literature, e.g., in Pullman and Pullman's textbook on 'Quantum Biochemistry' [169]. In other textbooks [44, 191], the value of $N_{\max}^{\pi} = \sqrt{3}$ for sp^2 carbon atoms is simply stated as a fact. For a time, it was enshrined in the official IUPAC definition of free valence [138].

The $\sqrt{d}$ values for $N_{\max}^{\pi}$ were used for the Hückel problem in 1951 in the paper by Burkitt [33]. Adoption of $\sqrt{3}$ as a global maximum for atoms of arbitrary degree was

discussed in the contemporary literature, as were the values appropriate for heteroatoms, and atoms in cumulated bonds.

Returning to the earliest statement of the claim: if Moffitt had a formal proof of his statement, he does not appear to have published it. He died prematurely in 1958 [42]. Coulson himself died in 1974 [3] and does not seem to have elaborated in print on the demonstration referred to in his 1948 footnote. The reconstruction of his lectures to Oxford undergraduate classes, published some years after his death by two former students O'Leary and Mallion [44] includes a statement of the $\sqrt{3}$ maximum, without comment on the history.

The notification of a proof in a marginal note has obvious echoes in the great mathematical detective stories, and it may never be possible to reconstruct the putative demonstration of the result $N^{\pi}_{\max} = \sqrt{3}$. We are not aware of a formal proof, but computers now allow the empirical investigation of the π bond number on a scale that was not possible for the pioneers in this area. Here, we report on the behaviour of the function $N^{\pi}_{\max}(n)$, the maximum π bond number over the set of chemical graphs of order $2 \leq n \leq 20$. Although no example exceeding the conjectured global maximum of $\sqrt{3}$ is found, some interesting behaviour is noted, and a family of locally maximal graphs is identified for $4 \leq n \leq 20$ and conjectured to give the highest values of $N^{\pi}_{\max}(n)$ for all n. Perhaps of most interest is the empirically established existence of a gap between the value of $N^{\pi}_{\max}(4) = \sqrt{3}$, realized by the star graph $K_{1,3}$, apparently uniquely, and the limit derived from a family of graphs that correspond to local maxima. We propose to call this limit the *Coulson-Moffitt number*.

4.1 Further Properties of Bond Number

Recall that MO coefficients are object to normalisation conditions within an molecular orbital (see Equation (3.37)) and for each vertex (see Equation 3.49) which act to limit the maximum achievable value for N^{π}_{r}.

An alternative formula for N^{π}_{r} can be derived from the adjacency eigenvalue equation

$$A(G)\mathbf{c}^{(i)} = \lambda_i \mathbf{c}^{(i)}, \tag{4.5}$$

from which a local condition for each vertex r is

$$\lambda_i c^{(i)}_r = \sum_{s \in G(r)} c^{(i)}_s. \tag{4.6}$$

Note that this is precisely (2.21). Thus,

$$N^{\pi}_{r} = \sum_{s \in G(r)} P^{\pi}_{rs} = \sum_{i=1}^{n} \sum_{s \in G(r)} v_i c^{(i)}_r c^{(i)}_s = \sum_{i=1}^{n} v_i c^{(i)}_r \sum_{s \in G(r)} c^{(i)}_s = \sum_{i=1}^{n} v_i \lambda_i \left(c^{(i)}_r \right)^2. \tag{4.7}$$

This formula gives useful chemical information about the rôles of orbitals of different energies: occupation of a bonding orbital typically increases bond number, occupation of a non-bonding orbital has no effect on bond number, and occupation of a anti-bonding orbital typically reduces bond number. This implies an alternative definition of 'non-bonding': not only does occupation of a non-bonding orbital make no contribution to the overall π energy, it also makes no contribution to the bond number around atom r in a π system.

In contrast, an individual bond order may increase, decrease or remain unchanged on occupation of a non-bonding orbital (see Example 4.1).

Example 4.1. Figure 4.2 shows the non-bonding π molecular orbital of the smallest chem-

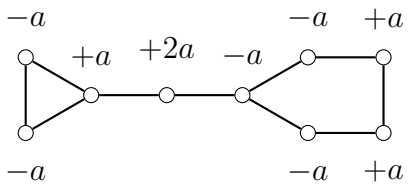


Figure 4.2: The non-bonding π molecular orbital of the smallest chemical nut graph.

ical nut graph. Recall that a nut graph has $n_0 = 1$ and only core vertices, i.e., non-zero entries in the eigenvector for the unique zero eigenvalue. On addition of an electron to the non-bonding orbital, five bonds would weaken, and five strengthen in this hypothetical planar π system. In Figure 4.2, a is $\frac{1}{\sqrt{12}}$. □

N_r^π is implicitly a function of the structure of the molecular graph G, the number of vertices, the specific vertex r and the number of π electrons. When we look for the maximum, $N_{\max}^\pi$, to be used in calculation of free valence, we need in principle to consider all possible electron counts for all vertices in all possible chemical graphs. Luckily, the equation (4.7) gives a simple way to avoid this extra complication. For a given molecule, there is an *ideal* π electron count compatible with Aufbau, Pauli and Hund's rules for electron configurations, that maximises the bond numbers of *all* centres: the occupation number v_i should be 2 for all eigenvectors corresponding to positive λ_i, and zero for all others. Hence, when searching for $N_{\max}^\pi$ we may use

$$N_r^\pi = \sum_{i=1}^{n_+} 2\lambda_i \left(c_r^{(i)}\right)^2, \tag{4.8}$$

where (n_+, n_0, n_-) is the inertia of the graph G. As the eigenvector corresponding to the maximum eigenvalue (i.e., the Perron–Frobenius eigenvalue) of a connected graph has strictly positive entries (by Theorem 2.16), we have that $N_{\max}^\pi$ is strictly positive.

We can reformulate the definition of the maximum bond number for a given molecule by noting that the π bond order for every bond rs, and therefore the bond number for every centre r, is reduced to zero when *all* π orbitals, bonding, non-bonding and anti-bonding, are fully occupied. This chemical fact follows from the orthonormality condition

$$\sum_{i=1}^{n} c_r^{(i)} c_s^{(i)} = \delta_{rs}, \tag{4.9}$$

where δ_{rs} is the Kronecker delta. Hence, the maximum bond number can be converted from a sum over bonding to a sum over all molecular orbitals:

$$N_r^\pi = \sum_{i:\ \lambda_i > 0} 2\lambda_i \left(c_r^{(i)}\right)^2 = -\sum_{i:\ \lambda_i \leq 0} 2\lambda_i \left(c_r^{(i)}\right)^2 = \sum_{i=1}^{n} |\lambda_i| \left(c_r^{(i)}\right)^2. \tag{4.10}$$

The expression on the right-hand size of (4.10) removes the dependence of maximum bond number on occupied versus empty orbitals. It is similar in form to the definition of graph energy [97]:

$$E_G = \sum_{i=1}^{n} |\lambda_i| \,, \tag{4.11}$$

which is the maximum possible Hückel π energy for a given molecule (corresponding to full occupation of the n_+ bonding orbitals). In both (4.10) and (4.11) all eigenvectors and eigenvalues are treated on an equal footing.

The formulation (4.8) in terms of positive eigenvalues also has an implication for the participation of given vertices in each eigenvector/molecular orbital. Recall that in the language of graph theory as developed by Sciriha (see Section 2.6.3), a vertex r is *core* for the eigenspace $\mathcal{E}(\lambda_i)$ (i.e., the set of orbitals corresponding to energy $\alpha + \lambda_i\beta$) if there is a non-zero entry $c_r^{(i)}$ for *some* vector within the eigenspace; otherwise the vertex is *core-forbidden*. The formulation (4.8) shows that eigenvectors contribute to N_r^π if and only if $\lambda_i > 0$ and r is core in eigenspace $\mathcal{E}(\lambda_i)$.

An easy upper bound on N_r^π follows from the equation (4.8):

Proposition 4.1. *Let G be a Hückel graph. Then*

$$N_r^\pi \leq 2\lambda_1(G). \tag{4.12}$$

Proof. As $\lambda_1 > \lambda_i$ for all $i \neq 1$ and $\sum_{k=1}^{n} \left(c_r^{(k)}\right)^2 = 1$:

$$\begin{aligned} N_r^\pi &= \sum_{i=1}^{n_+} 2\lambda_i \left(c_r^{(i)}\right)^2 \leq \sum_{i=1}^{n_+} 2\lambda_1 \left(c_r^{(i)}\right)^2 \\ &\leq \sum_{i=1}^{n} 2\lambda_1 \left(c_r^{(i)}\right)^2 = 2\lambda_1. \end{aligned} \tag{4.13}$$

□

This is typically a very loose upper bound. It can be improved for bipartite graphs (alternant hydrocarbons). The eigenvalues of bipartite graphs obey Theorem 3.1: if λ is an eigenvalue, then so is $-\lambda$, and an eigenvector for λ can be converted to an eigenvector for $-\lambda$ by reversing the signs of the entries for one partite set of vertices. Zero eigenvalues may be self-paired.

An implication for eigenvector entries is that the normalisation condition can be refined, since

$$\sum_{i=1}^{n} \left(c_r^{(i)}\right)^2 = 1 = \sum_{k\in\mathcal{K}_+} \left(c_r^{(k)}\right)^2 + \sum_{k\in\mathcal{K}_0} \left(c_r^{(k)}\right)^2 + \sum_{k\in\mathcal{K}_-} \left(c_r^{(k)}\right)^2 \tag{4.14}$$

and

$$\sum_{k\in\mathcal{K}_+} \left(c_r^{(k)}\right)^2 = \sum_{k\in\mathcal{K}_-} \left(c_r^{(k)}\right)^2 \leq \frac{1}{2}, \tag{4.15}$$

where $\mathcal{K}_+ = \{k \mid \lambda_k > 0\}$, $\mathcal{K}_- = \{k \mid \lambda_k < 0\}$ and $\mathcal{K}_0 = \{k \mid \lambda_k = 0\}$. The upper bound (4.12) is then improved for bipartite graphs:

Proposition 4.2. *Let G be a Hückel graph of an* alternant hydrocarbon. *Then*

$$N_r^\pi \leq \lambda_1(G). \tag{4.16}$$

□

This bound is sharp in the case that G is bipartite with *exactly one positive eigenvalue*, and where vertex r is core forbidden in the eigenspace $\mathcal{E}(0)$. In Section 2.6.1 we have described the spectra of bipartite graphs. Star graphs $K_{1,n-1}$ on n vertices are bipartite, have unique positive eigenvalues $\lambda_1(K_{1,n-1}) = \sqrt{n-1}$ (note that $n-1$ is the degree of the central vertex) and when $n_0 > 0$ ($n_0 > 0$ if and only if $n > 2$) that central vertex is core-forbidden for eigenvalue 0. Therefore, all star graphs have a vertex (the central vertex) with $N_r^\pi = \lambda_1 = \sqrt{n-1}$, providing the *lower* bounds (and conjectured exact values) to the best $N_{\max}$ for sp^2 carbon atom of $\sqrt{1}$, $\sqrt{2}$ and $\sqrt{3}$ (see Table 4.1), and in general $\sqrt{n-1}$ for the set of all graphs, chemical and otherwise, on n vertices.
Smith has characterised graphs with precisely one positive eigenvalue [189]:

Theorem 4.3 (Theorem 6.7 in [49])**.** *A graph has exactly one positive eigenvalue if and only if its non-isolated vertices form a complete multipartite graph.* □

There are only 9 complete multipartite subcubic graphs (with at least one edge). All of them are shown in Figure 4.3. Vertices that realize the maximum bond number are coloured red. Table 4.2 displays the maximum bond number for each of those graphs. They are sorted

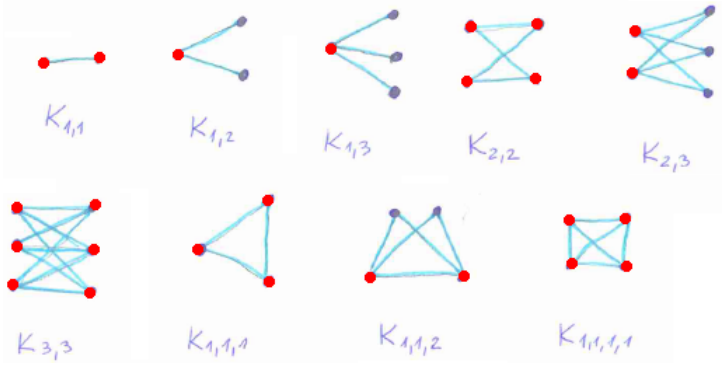


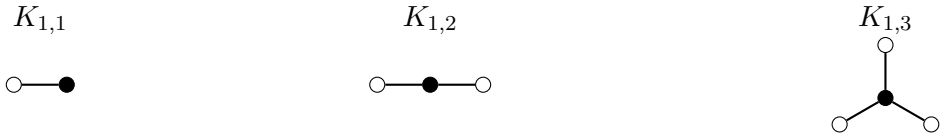
Figure 4.3: All complete multipartite chemical graphs.

by $N^\pi(G)$ in decreasing order. Hence, we can state the following:

Proposition 4.4. *The Moffitt value of $\sqrt{3}$ for the maximum π bond order is the best over the set of all chemical graphs with exactly one positive eigenvalue, i.e., of all π systems with exactly one bonding orbital.* □

There are infinitely many graphs with exactly one positive eigenvalue, but only a finite number of such graphs exist when the vertex degree is bounded. We will show that there are finitely many graphs with bounded vertex degree and bounded number of positive eigenvalues.

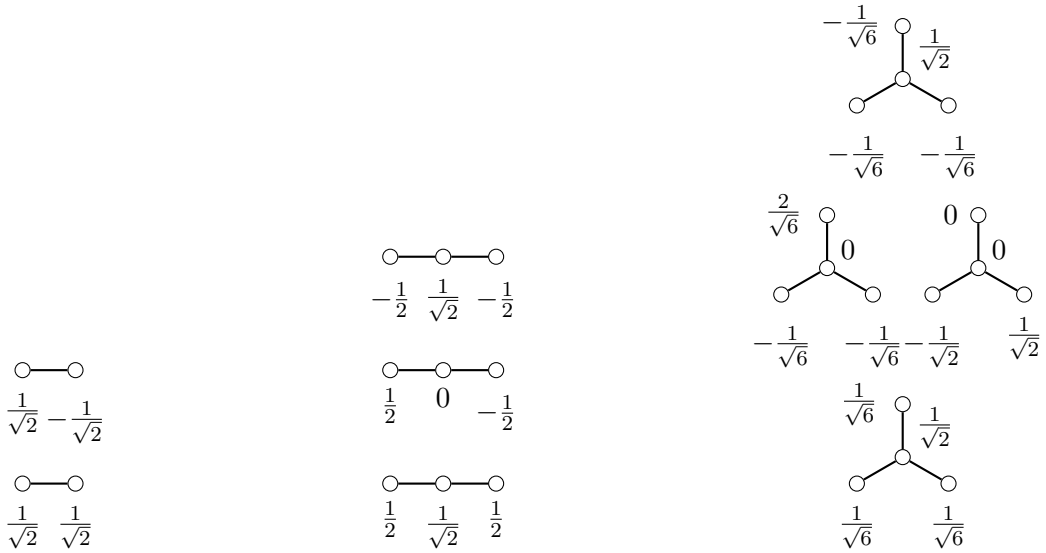
Graphs:



Eigenvalues:

	$\lambda_3 = -\sqrt{2}$	$\lambda_4 = -\sqrt{3}$
$\lambda_2 = -1$	$\lambda_2 = 0$	$\lambda_2 = \lambda_3 = 0$
$\lambda_1 = +1$	$\lambda_1 = \sqrt{2}$	$\lambda_1 = \sqrt{3}$

Eigenvectors:



Bond number $N_\bullet(\pi)$:

$2(1)\left(\frac{1}{\sqrt{2}}\right)^2 = 1$	$2(\sqrt{2})\left(\frac{1}{\sqrt{2}}\right)^2 = \sqrt{2}$	$2(\sqrt{3})\left(\frac{1}{\sqrt{2}}\right)^2 = \sqrt{3}$

Table 4.1: Bond number of $K_{1,1}$, $K_{1,2}$ and $K_{1,3}$

Lemma 4.5. *The path P_n has $\lfloor \frac{n}{2} \rfloor$ positive eigenvalues.*

Proof. In Section 2.6.1 we showed that the spectrum of P_n contains values $2\cos\left(\frac{\pi j}{n+1}\right)$ for $j = 1, \ldots, n$. Note that $0 < \frac{\pi j}{n+1} < \pi$ for $1 \leq j \leq n$. The value $2\cos\left(\frac{\pi j}{n+1}\right)$ is greater than zero if $\frac{\pi j}{n+1} < \frac{\pi}{2}$, i.e., if $j < \frac{n+1}{2}$. This is equivalent to $j \leq \lfloor \frac{n}{2} \rfloor$. □

There is a straightforward upper bound on the largest possible order of a connected graph G of maximum degree $\Delta(G)$ and diameter $\mathrm{diam}(G)$. If $\Delta(G) = 1$ then $G \cong K_2$ and $\mathrm{diam}(G) = 1$ which is a trivial observation.

G	$N^\pi(G)$	
$K_{1,3}$	$\sqrt{3}$	1.732051
$K_{1,1,2}$	$\frac{17+9\sqrt{17}}{34}$	1.591410
$K_{1,1,1,1}$	$\frac{3}{2}$	1.500000
$K_{1,2}$	$\sqrt{3}$	1.414214
$K_{1,1,1}$	$\frac{4}{3}$	1.333333
$K_{2,3}$	$\frac{\sqrt{6}}{2}$	1.224745
$K_{1,1}$	1	1.000000
$K_{2,2}$	1	1.000000
$K_{3,3}$	1	1.000000

Table 4.2: Bond numbers of chemical graphs with one bonding orbital.

Lemma 4.6 ([139])**.** *Let G be a graph with $\Delta(G) \geq 2$. Then*

$$|V(G)| \leq \begin{cases} 1 + \Delta(G)\frac{(\Delta(G)-1)^{\mathrm{diam}(G)}-1}{\Delta(G)-2} & \text{if } \Delta(G) > 2; \\ 2\,\mathrm{diam}(G) + 1 & \text{if } \Delta(G) = 2. \end{cases} \tag{4.17}$$

□

Lemma 4.6 tells us that there are only finitely many graphs with a given maximum degree and diameter. Equation (4.17) is called the *Moore bound*. We are ready to prove the following theorem:

Theorem 4.7. *There exist finitely many graphs with a prescribed maximum degree Δ which have n_+ positive eigenvalues.*

Proof. Let us assume that there are infinitely many such graphs, i.e., the set

$$\mathcal{G} = \{G \mid G \text{ has } n_+ \text{ positive eigenvalues and maximum degree } \Delta\}$$

is infinite. Then we can choose an arbitrarily large graph (i.e., of orbitrary large order) from $\mathcal{G}$. From Lemma 4.6 it follows that we can choose a graph with arbitrary large diameter from the set $\mathcal{G}$.

Let $G \in \mathcal{G}$ be a graph with $\mathrm{diam}(G) \geq 2n_+ + 1$. This means that there exist two vertices $u, v \in V(G)$ such that $d(u, v) = 2n_+ + 1$. Let $u = u_0, u_1, u_2, \ldots, u_{2n_+}, u_{2n_+1} = v$ be a shortest (u, v)-path and let H be a subgraph of G induced on vertices $\{u_0, u_1, \ldots, u_{2n_+1}\}$. Clearly, $H \cong P_{2n_++2}$. From Lemma 4.5 and Corollary 2.15 it follows that G has at least $\lfloor \frac{2n_++2}{2} \rfloor = n_+ + 1$ positive eigenvalues. We have arrived at a contradiction which means that the set $\mathcal{G}$ must be finite. □

From Theorem 4.7 we immediately obtain:

Corollary 4.8. *There are finitely many chemical graph with a prescribed number, n_+, of positive eigenvalues.* □

Corollary 4.8 implies that we can enumerate all chemical graphs with a given number of positive eigenvalues, just as we did in the case of $n_+ = 1$ above. Efforts to obtain a general proof of a statement equivalent to Proposition 4.4 for all $n_+ > 1$ have not succeeded so far, and we turn now to computer-aided search for higher values of $N^{\pi}_{\max}$.

4.2 Computer Searches

Complete sets of chemical graphs on $2 \leq n \leq 20$ vertices were generated using the **geng** program. In each search, the bond number for each graph maximised with respect to electron count and choice of vertex was calculated by using the **numpy.linalg.eigh** function [118, 190] which in turn uses the **syevd** function from LAPACK [4]. This routine uses a divide and conquer algorithm to compute eigenvalues and eigenvectors. A specialised library for arbitrary precision arithmetic called **mpmath** [117] was used where necessary to separate contenders for best graph.

Definition 4.2. *A* crippled caterpillar *on n vertices is a graph G with*

$$V(G) = \{v_0, v_1, \ldots, v_{n-1}\}$$

and

$$E(G) = \{v_{i-1}v_i \mid 1 \leq i \leq n-1 \wedge i \bmod 3 \neq 0\} \cup \{v_{i-2}v_i \mid 2 \leq i \leq n-1 \wedge i \bmod 3 = 0\}.$$

Definition 4.3. *A* kite graph *on $n \geq 5$ vertices is a graph G with*

$$V(G) = \{v_0, v_1, \ldots, v_{n-1}\}$$

and

$$E(G) = \{v_0v_2, v_0v_3, v_1v_2, v_1v_3, v_2v_4, v_3v_4\} \cup \{v_{i-1}v_i \mid 5 \leq i \leq n-1 \wedge i \bmod 3 \neq 1\} \cup \{v_{i-2}v_i \mid 7 \leq i \leq n-1 \wedge i \bmod 3 = 1\}.$$

We denote the crippled caterpillar on n vertices by CriCat(n) and the kite graph on n vertices by Kite(n).

Chemical graphs that attain the maximal bond number are listed in Table 4.3 and drawn on Figure 4.4. We can state the following:

Proposition 4.9. *No chemical graph with a π bond number exceeding the conjectured $\sqrt{3}$ global maximum exists in the range $2 \leq n \leq 20$.* □

Let us examine the graphs that realize the maximal bond number for a given n. By observing Table 4.3 we anticipate the following:

Conjecture 4.3. *Among all chemical graphs on n, $n \geq 2$, vertices, the graph that attains the maximum π bond number is the kite graph for the values of $n \geq 14$ where $n \bmod 3 = 2$, and the crippled caterpillar for all other values of n.* □

n	$\max_r N_r^\pi$	Graph
2	1.00000000	CriCat(2) $\cong K_2$
3	1.41421356	CriCat(3) $\cong P_3$
4	1.73205081	CriCat(4) $\cong K_{1,3}$
5	1.68924640	CriCat(5)
6	1.70130162	CriCat(6)
7	1.70710678	CriCat(7)
8	1.70481999	CriCat(8)
9	1.70573213	CriCat(9)
10	1.70609618	CriCat(10)
11	1.70590690	CriCat(11)
12	1.70599266	CriCat(12)
13	1.70602352	CriCat(13)
14	1.70601452	Kite(14)
15	1.70601400	CriCat(15)
16	1.70601701	CriCat(16)
17	1.70601642	Kite(17)
18	1.70601604	CriCat(18)
19	1.70601636	CriCat(19)
20	1.70601631	Kite(20)

Table 4.3: Chemical graph on n vertices that attains the maximal bond number.

n	$\max_r N_r^\pi$	Graph
14	1.70601452	CriCat(14)
17	1.70601642	CriCat(17)
20	1.70601631	CriCat(20)

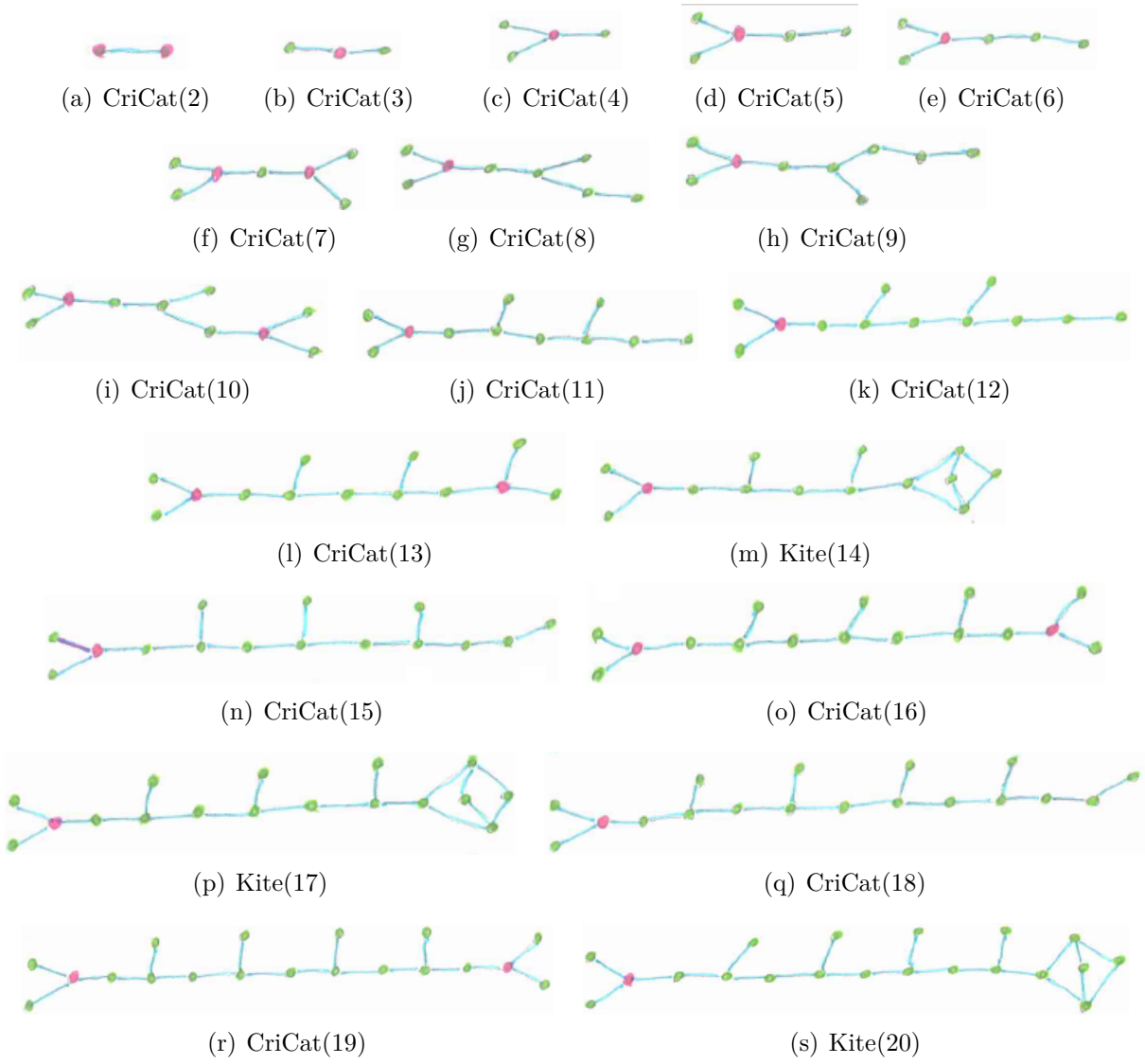
Table 4.4: Chemical trees on 14, 17 and 20 vertices that attain the maximal bond number.

Let us restrict our attention to chemical trees. Almost all graphs in Table 4.3, except for $n \in \{14, 17, 20\}$, are trees. If a tree attains the maximal π bond number among all chemical graphs then it clearly attains the maximal value within the subfamily of chemical trees. In Table 4.4 only those trees are listed which are not already included in the list of chemical graphs. See also Figure 4.5 for their drawings. We see that all chemical trees up to 20 vertices are crippled caterpillars and we can pose the following conjecture:

Conjecture 4.4. *Among all chemical trees on n, $n \geq 2$, vertices, the crippled caterpillar obtains the maximum π bond number.* □

In Section 3.3 we presented fullerenes which are an important class of chemical graphs. The search for graphs that attain the maximal bond number was done in the class of fullerenes on $20 \leq n \leq 100$ vertices. The results are in Table 4.5. The naming convention for fullerenes in this table is n:id, where:

(i) n is the number of its vertices and



(a) CriCat(2) (b) CriCat(3) (c) CriCat(4) (d) CriCat(5) (e) CriCat(6)

(f) CriCat(7) (g) CriCat(8) (h) CriCat(9)

(i) CriCat(10) (j) CriCat(11) (k) CriCat(12)

(l) CriCat(13) (m) Kite(14)

(n) CriCat(15) (o) CriCat(16)

(p) Kite(17) (q) CriCat(18)

(r) CriCat(19) (s) Kite(20)

Figure 4.4: Chemical graph on n vertices that attains the maximal bond number.

(ii) *id* is the place in which this fullerene appears when all fullerenes on n vertices are generated by the **fullgen** program.

The symmetry group of each fullerene is also given in the Table 4.5. Their drawings can be found in Appendix B. We can state the following:

Proposition 4.10. *No fullerene graph with a π bond number exceeding the conjectured $\sqrt{3}$ global maximum exists in the range $20 \leq n \leq 100$.* □

We also searched for fullerenes that attain the minimal number $\max_r N_r^{\pi}$ among all fullerenes on n vertices. Results are presented in Table 4.6. Their drawings can also be found in Appendix B.

When we examine the graphs in Table 4.5 we see that the vertex that maximizes the bond number is always incident to hexagonal faces only, unless the pentagonal faces are

(a) CriCat(14)

(b) CriCat(17)

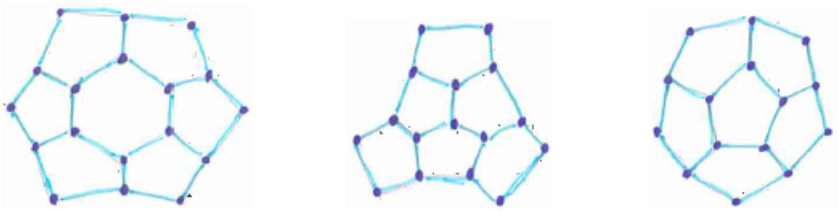
(c) CriCat(20)

Figure 4.5: Chemical trees on 14, 17 and 20 vertices that attain the maximal bond number.

distributed in such a way that there do not exist vertices that are incident to three hexagonal faces. We pose the following conjecture:

Conjecture 4.5. *If a fullerene graph $\mathcal{F}$ has a non-empty set of vertices $\mathcal{V} \subset V(\mathcal{F})$ that are incident to three hexagonal faces then any vertex that maximises the Coulson bond number in $\mathcal{F}$ will be contained in $\mathcal{V}$.* □

Let us compare fullerenes in Table 4.5 and Table 4.6. The order of the symmetry group of the fullerene that minimises the maximal bond number is in most cases (but not always) larger than the order of the symmetry group of the fullerene that attains the maximal bond number. Symmetry is apparently not an advantage if we are trying to maximise the maximal bond number. Intuitively speaking, many vertices have to sacrifice in order for one chosen vertex to succeed and accumulate enough charge. If there are large vertex orbits, this prevents the possibility of the Musketeers' "All for one, and one for all!" principle. We also see that certain types of capped nanotubes occur frequently in the Table 4.6. If $n \equiv 0 \pmod{12}$, the fullerene that minimizes the maximal Coulson bond order is the nanotube that is capped with the patch in Figure 4.6(a), except for $n = 60$ when the buckyball overtakes the tube. However, the capped nanotube is the third best fullerene



(a) (b) (c)

Figure 4.6: Caps of certain nanotubes that appear frequently in the Table 4.6.

that minimizes the maximal Coulson bond number among all fullerenes on $n = 60$ vertices. That nanotube has the $\mathbf{D}_{6d}$ symmetry if $n \equiv 0 \pmod{24}$ and the $\mathbf{D}_{6h}$ symmetry if $n \equiv 12 \pmod{24}$. If $n \geq 60$ and $n \equiv 2 \pmod{12}$ then the fullerene that minimizes the maximal

n	$\max_r N_r^\pi$	Graph
20	1.47082039	20:1 ($\mathbf{I}_h$ symmetry)
24	1.50852610	24:1 ($\mathbf{D}_{6d}$ symmetry)
26	1.53611943	26:1 ($\mathbf{D}_{3h}$ symmetry)
28	1.55439908	28:2 ($\mathbf{D}_2$ symmetry)
30	1.56881522	30:2 ($\mathbf{C}_{2v}$ symmetry)
32	1.58365220	32:2 ($\mathbf{D}_{3h}$ symmetry)
34	1.58171525	34:5 ($\mathbf{C}_{3v}$ symmetry)
36	1.58154470	36:7 ($\mathbf{C}_1$ symmetry)
38	1.58450304	38:12 ($\mathbf{C}_{2v}$ symmetry)
40	1.58370654	40:31 ($\mathbf{C}_s$ symmetry)
42	1.58452376	42:21 ($\mathbf{C}_1$ symmetry)
44	1.58445872	44:1 ($\mathbf{C}_{2v}$ symmetry)
46	1.58386557	46:50 ($\mathbf{C}_s$ symmetry)
48	1.58397604	48:137 ($\mathbf{C}_s$ symmetry)
50	1.58474274	50:55 ($\mathbf{C}_{3v}$ symmetry)
52	1.58421130	52:128 ($\mathbf{C}_1$ symmetry)
54	1.58425237	54:194 ($\mathbf{C}_1$ symmetry)
56	1.58418419	56:293 ($\mathbf{C}_s$ symmetry)
58	1.58436088	58:255 ($\mathbf{C}_s$ symmetry)
60	1.58451202	60:359 ($\mathbf{C}_s$ symmetry)
62	1.58414014	62:493 ($\mathbf{C}_1$ symmetry)
64	1.58436765	64:1481 ($\mathbf{C}_s$ symmetry)
66	1.58408912	66:3652 ($\mathbf{C}_1$ symmetry)
68	1.58428693	68:1959 ($\mathbf{C}_1$ symmetry)
70	1.58442267	70:4087 ($\mathbf{C}_s$ symmetry)
72	1.58442937	72:5633 ($\mathbf{C}_1$ symmetry)
74	1.58445980	74:5814 ($\mathbf{D}_{3h}$ symmetry)
76	1.58458610	76:9132 ($\mathbf{T}_d$ symmetry)
78	1.58439360	78:10288 ($\mathbf{C}_s$ symmetry)
80	1.58444297	80:11391 ($\mathbf{C}_{3v}$ symmetry)
82	1.58438720	82:13243 ($\mathbf{C}_1$ symmetry)
84	1.58434577	84:14510 ($\mathbf{C}_1$ symmetry)
86	1.58434048	86:15156 ($\mathbf{C}_1$ symmetry)
88	1.58436022	88:14729 ($\mathbf{C}_1$ symmetry)
90	1.58438207	90:16030 ($\mathbf{C}_1$ symmetry)
92	1.58433751	92:16163 ($\mathbf{C}_1$ symmetry)
94	1.58421952	94:47431 ($\mathbf{C}_s$ symmetry)
96	1.58424765	96:126731 ($\mathbf{C}_s$ symmetry)
98	1.58422793	98:101379 ($\mathbf{C}_1$ symmetry)
100	1.58421996	100:186008 ($\mathbf{C}_1$ symmetry)

Table 4.5: Fullerenes that attain the maximal bond number.

n	$\max_r N_r^\pi$	Graph
20	1.47082039	20:1 ($\mathbf{I}_h$ symmetry)
24	1.50852610	24:1 ($\mathbf{D}_{6d}$ symmetry)
26	1.53611943	26:1 ($\mathbf{D}_{3h}$ symmetry)
28	1.54956698	28:1 ($\mathbf{T}_d$ symmetry)
30	1.54522586	30:1 ($\mathbf{D}_{5h}$ symmetry)
32	1.55434763	32:4 ($\mathbf{D}_{3d}$ symmetry)
34	1.56340035	34:1 ($\mathbf{C}_s$ symmetry)
36	1.55033775	36:1 ($\mathbf{D}_{6h}$ symmetry)
38	1.56217235	38:8 ($\mathbf{C}_2$ symmetry)
40	1.55432207	40:34 ($\mathbf{T}_d$ symmetry)
42	1.56572204	42:39 ($\mathbf{D}_3$ symmetry)
44	1.55676795	44:52 ($\mathbf{D}_{3h}$ symmetry)
46	1.57356433	46:44 ($\mathbf{C}_2$ symmetry)
48	1.56538961	48:1 ($\mathbf{D}_{6d}$ symmetry)
50	1.56089506	50:105 ($\mathbf{D}_{5h}$ symmetry)
52	1.57387513	52:140 ($\mathbf{C}_2$ symmetry)
54	1.57297777	54:30 ($\mathbf{C}_{2v}$ symmetry)
56	1.57396484	56:311 ($\mathbf{D}_3$ symmetry)
58	1.57540560	58:794 ($\mathbf{C}_1$ symmetry)
60	1.55269340	60:936 ($\mathbf{I}_h$ symmetry)
62	1.57426202	62:1612 ($\mathbf{D}_{3h}$ symmetry)
64	1.57411256	64:150 ($\mathbf{C}_s$ symmetry)
66	1.57491926	66:56 ($\mathbf{C}_{2v}$ symmetry)
68	1.57424416	68:4656 ($\mathbf{D}_3$ symmetry)
70	1.57382683	70:1 ($\mathbf{D}_{5h}$ symmetry)
72	1.57046136	72:1 ($\mathbf{D}_{6d}$ symmetry)
74	1.57491346	74:9249 ($\mathbf{D}_{3h}$ symmetry)
76	1.57468957	76:7956 ($\mathbf{D}_{2d}$ symmetry)
78	1.57257030	78:1992 ($\mathbf{D}_{3h}$ symmetry)
80	1.57425326	80:18087 ($\mathbf{D}_{5d}$ symmetry)
82	1.57544373	82:36884 ($\mathbf{C}_2$ symmetry)
84	1.57118716	84:1 ($\mathbf{D}_{6h}$ symmetry)
86	1.57456053	86:40357 ($\mathbf{D}_{3h}$ symmetry)
88	1.57479103	88:50148 ($\mathbf{C}_{2v}$ symmetry)
90	1.57319886	90:54468 ($\mathbf{D}_{5h}$ symmetry)
92	1.57465797	92:43314 ($\mathbf{D}_3$ symmetry)
94	1.57568328	94:75372 ($\mathbf{C}_1$ symmetry)
96	1.57162575	96:1 ($\mathbf{D}_{6d}$ symmetry)
98	1.57458570	98:142141 ($\mathbf{D}_{3h}$ symmetry)
100	1.57510187	100:1 ($\mathbf{D}_{5d}$ symmetry)

Table 4.6: Fullerenes that minimize the maximal bond number.

Coulson bond number is the nanotube with the $\mathbf{D}_{3h}$ symmetry that is capped with the patch in Figure 4.6(b). For $n = 70$ and $n = 100$ the nanotube capped with the patch in Figure 4.6(c) emerges. It has the $\mathbf{D}_{5h}$ symmetry for $n = 70$ and the $\mathbf{D}_{5d}$ symmetry for $n = 100$.

We also see that fullerene graphs are not good candidates as families of chemical graphs that would maximise the Coulson bond number.

The next family under consideration are benzenoid graphs. They are extensively treated in Chapter 5. The results are in Table 4.7. Benzenoids on up to $n = 32$ vertices were considered. The second column in the table gives the number of non-equivalent benzenoids on n vertices. Benzenoids are given via their boundary-edges codes (see Definition 5.14). For each $n \leq 32$, the table lists both the benzenoid that obtains the maximal Coulson bond number and the one that minimizes it. Figures can be found in Appendix B. Benzenoids were generated by our own computer program. The algorithm is described in Section 5.5.

n	Count	$\max_G N^\pi(G)$	$BEC(G)$	$\min_G N^\pi(G)$	$BEC(G)$
6	1	1.33333333	6	1.33333333	6
10	1	1.62763379	55	1.62763379	55
13	1	1.63054753	444	1.63054753	444
14	2	1.62631158	5252	1.62334845	5351
16	1	1.62132013	4343	1.62132013	4343
17	1	1.63220884	52441	1.63220884	52441
18	5	1.62715855	532521	1.59242502	515151
19	1	1.63164451	43342	1.63164451	43342
20	3	1.63130593	523431	1.62228423	514341
21	6	1.63182435	5241521	1.63060923	5314421
22	15	1.63238336	424242	1.62191039	53215321
23	4	1.63206671	5142431	1.63032477	4413431
24	15	1.63672514	44224411	1.59253929	51415141
25	27	1.63252587	4323422	1.62993398	521521521
26	46	1.63273787	52242421	1.59980247	5151151511
27	26	1.63584712	442234311	1.62969256	521513421
28	77	1.63635309	5211442241	1.61604295	51333331
29	127	1.63223591	514224321	1.62602800	53141351221
30	189	1.63620345	4422243311	1.59179608	515121515121
31	159	1.63575386	52311442231	1.62362874	433233411
32	385	1.63644271	532114422411	1.59969316	515114151411

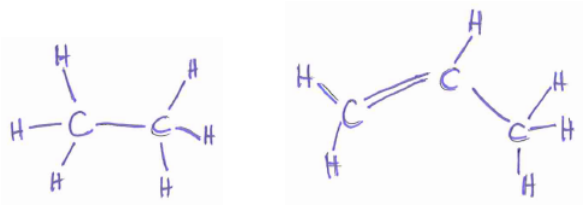
Table 4.7: Benzenoids and the maximum Coulson bond number.

Chapter 5

Hexagonal Systems

5.1 Hexagonal systems from a chemical viewpoint

Benzenoid hydrocarbons are important well studied compounds [58, 57], important on their own right, and now as mimics of graphene. In their monograph from 1989 [102], Gutman and Cyvin provide an in-depth treatment of this subject. There, they define *benzenoid hydrocarbons* as condensed polycyclic unsaturated fully conjugated hydrocarbons composed exclusively of six-membered rings. While the above definition tells a lot to a chemist, it might cause discomfort to pure mathematicians and computer scientists. We will therefore explain the above definition word by word. Hydrocarbons are, as the name suggests, chemical compounds consisting of carbon and hydrogen atoms only. An example is *ethane* with chemical formula C_2H_6 whose skeletal formula is depicted in Figure 5.1(a). Another example is *propene* with chemical formula C_3H_6, depicted in Figure 5.1(b). The



(a) Ethane (b) Propene

Figure 5.1: Two examples of simple hydrocarbons.

carbon framework of a hydrocarbon is interconnected with strong σ-bonds (think of it as a simple graph). Each hydrogen is attached to a certain carbon (with a σ-bond). There may be only one σ-bond between a pair of carbon atoms and each carbon atom may form at most four σ-bonds. Some bonded pairs of carbons may be connected with π-bonds (interaction between p orbitals) which count towards the total valency of four. The π-bonds are less strong than σ-bonds, giving rise to double and triple bonds. *Unsaturated* hydrocarbons are those which contain at least one double or triple bond. Whilst, ethane is saturated, propene is not. According to [138], a π-conjugated system is a molecule whose carbon framework may be represented as a system of alternating single and double bonds, as in the case of

propene. However, the distribution of the π-electrons is not accurately represented by a simple structural formula involving full double bonds. Instead, we should think of a π-conjugated hydrocarbon as a carbon framework with one p orbital at each carbon atom contributing one electron to the shared electron cloud. *Polycyclic* means that one or more cycles may be present in the carbon framework (see the example in Figure 5.2). Note that this compound possesses 5-membered rings and is therefore not a benzenoid hydrocarbon. In the chemical literature, one often encounters *polycyclic aromatic hydrocarbons*, often

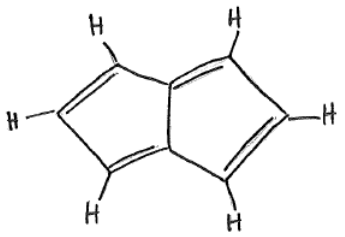


Figure 5.2: Pentalene

abbreviated as PAHs, which is a large class of organic molecules. PAHs are aromatic by definition. Because benzenoid hydrocarbons include highly reactive compounds such as heptacene or triangulene, they are not a proper subclass of PAHs. Nevertheless, the majority of benzenoid hydrocarbons that were experimentally studied so far are a fortiori stable compounds.

Another restriction is that the molecules of benzenoid hydrocarbons are planar (or reasonably close to that). Helicenic (e.g. heptahelicene) compounds are thus not considered as benzenoids. Gutman and Cyvin [102] state that in the early 1990s around 300 benzenoid hydrocarbons were known. PAHs and benzenoid hydrocarbons can be found among products of incomplete combustion of organic materials. Evidence suggest that they can be found in tar and in small amounts everywhere in the environment, even in food that has been processed by frying and roasting. A number of benzenoid hydrocarbons have been discovered to be carcinogenic. They can be found (amongst other lethal substances) in tobacco smoke.

Our purpose is to study a purely mathematical model for these compounds. One possible route to construction of such a model is via benzenoid graphs. Some authors introduce them as surfaces with boundary and call them *benzenoid systems* (or *hexagonal systems*). We will also take the latter approach and call these objects *benzenoids*.

5.2 The infinite hexagonal grid

To define the infinite hexagonal grid, we will use a different definition of a (multi)graph. In this definition, edges have their own identity:

Definition 5.1. *A (multi)graph is an ordered triple $G = (V, E, r)$, where V is a set of vertices, E is a set of edges and $r\colon E \to \binom{V}{2} \cup V$ is a mapping that assigns end vertices to each edge.*

Note that this definition also allows parallel edges and loops. The set

$$\binom{V}{2} \cup V = \{\{u, v\} \mid u, v \in V\}, \tag{5.1}$$

i.e., an edge may only have one end vertex in which case the edge is called a loop. Now we can define the infinite hexagonal lattice [106]:

Definition 5.2. *The* infinite hexagonal lattice $G(\mathcal{H})$ *is an infinite graph where*

$$V = \mathbb{Z} \times \mathbb{Z} \times \{0, 1\},$$
$$E = \mathbb{Z} \times \mathbb{Z} \times \{0, 1, 2\},$$

and $r\colon E \to \binom{V}{2} \cup V$ *is defined as*

$$r(\xi, \eta, \nu) = \begin{cases} \{(\xi, \eta, 0), (\xi - 1, \eta, 1)\}, & \text{if } \nu = 0; \\ \{(\xi, \eta, 0), (\xi, \eta, 1)\}, & \text{if } \nu = 1; \\ \{(\xi + 1, \eta - 1, 0), (\xi, \eta, 1)\}, & \text{if } \nu = 2. \end{cases} \tag{5.2}$$

The graph $G(\mathcal{H})$ does not contain loops or parallel edges. It is clearly a cubic graph. Note that $G(\mathcal{H})$ does not possess any geometric information. To obtain the infinite hexagonal grid, we have to embed it into the Euclidean plane $\mathbb{R}^2$. Let $\varphi\colon V(G(\mathcal{H})) \to \mathbb{R}^2$ be a mapping that assigns coordinates to each vertex of $G(\mathcal{H})$ that is defined as:

$$\varphi(\xi, \eta, \nu) = \begin{cases} \left(\sqrt{3}\xi + \frac{\sqrt{3}}{2}\eta, \frac{3}{2}\eta + 1\right), & \text{if } \nu = 0; \\ \left(\sqrt{3}\xi + \frac{\sqrt{3}}{2}\eta + \frac{\sqrt{3}}{2}, \frac{3}{2}\eta + \frac{1}{2}\right), & \text{if } \nu = 1. \end{cases} \tag{5.3}$$

Each edge $e \in E(G(\mathcal{H}))$ is embedded as a straight line segment between the two points that correspond to the end vertices of e. The infinite hexagonal lattice with the above embedding is a plane graph which is called the *infinite hexagonal grid* and is denoted with $\mathcal{H}$ (see Figure 5.3). The $G(\mathcal{H})$ is a bipartite graph with the bipartition $V_1 = \mathbb{Z} \times \mathbb{Z} \times \{0\}$ and $V_2 = \mathbb{Z} \times \mathbb{Z} \times \{1\}$. The vertices of V_1 and V_2 are coloured orange and green, respectively, in Figure 5.3.

Every face of $\mathcal{H}$ is a regular hexagon with a unit-length side. Those hexagons can be labeled with elements of the set $\mathbb{Z} \times \mathbb{Z}$. The hexagon with the label $(0, 0)$ has its centre in coordinates $(0, 0)$ of the plane $\mathbb{R}^2$. The hexagon labeled with (ξ, η) has its centre in coordinates $\left(\sqrt{3}\xi + \frac{\sqrt{3}}{2}\eta, \frac{3}{2}\eta\right)$ and is incident with the vertices

$$(\xi, \eta, 0), (\xi, \eta, 1), (\xi + 1, \eta - 1, 0), (\xi, \eta - 1, 1), (\xi, \eta - 1, 0) \text{ and } (\xi - 1, \eta, 1). \tag{5.4}$$

Hexagon (ξ, η) is incident with the edges

$$(\xi, \eta, 0), (\xi, \eta, 1), (\xi, \eta, 2), (\xi + 1, \eta - 1, 0), (\xi, \eta - 1, 1) \text{ and } (\xi - 1, \eta, 2). \tag{5.5}$$

The faces of $\mathcal{H}$ are vertices of its plane dual $\mathcal{H}^*$ which is a 6-regular plane triangulation. The vertex $(\xi, \eta) \in V(\mathcal{H}^*)$ is adjacent to vertices

$$(\xi, \eta + 1), (\xi + 1, \eta), (\xi + 1, \eta - 1), (\xi, \eta - 1), (\xi - 1, \eta) \text{ and } (\xi - 1, \eta + 1). \tag{5.6}$$

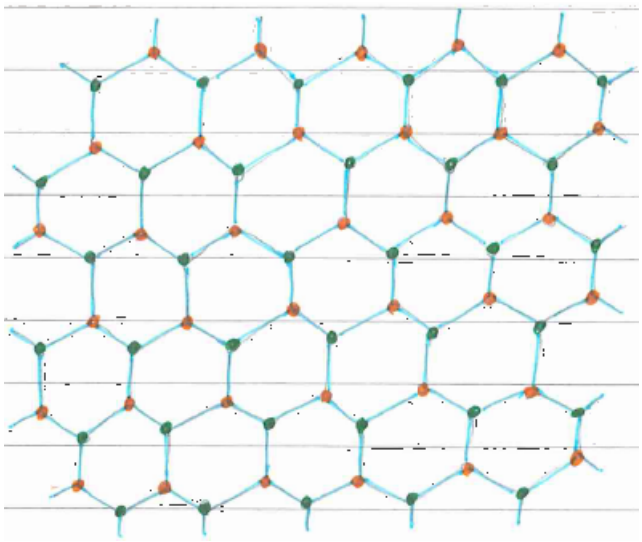


Figure 5.3: The infinite hexagonal grid $\mathcal{H}$.

5.3 Mathematical treatment of hexagonal systems

Traditionally, a *benzenoid system* $\mathcal{B}$ is a collection of hexagons that constitute a simply connected bounded region of the infinite hexagonal grid $\mathcal{H}$ in the Euclidean plane. Other equivalent definitions are also possible. For a complete treatment of this topic, see [102]. Many authors consider benzenoids as plane graphs. For instance, Hammack, Imrich and Klavžar [106] define them as follows (also Definition C in [102]):

Definition 5.3. *Let Z be a cycle of the infinite hexagonal grid $\mathcal{H}$. A* benzenoid graph *is formed by the vertices and edges of $\mathcal{H}$ lying on and in the interior of Z.*

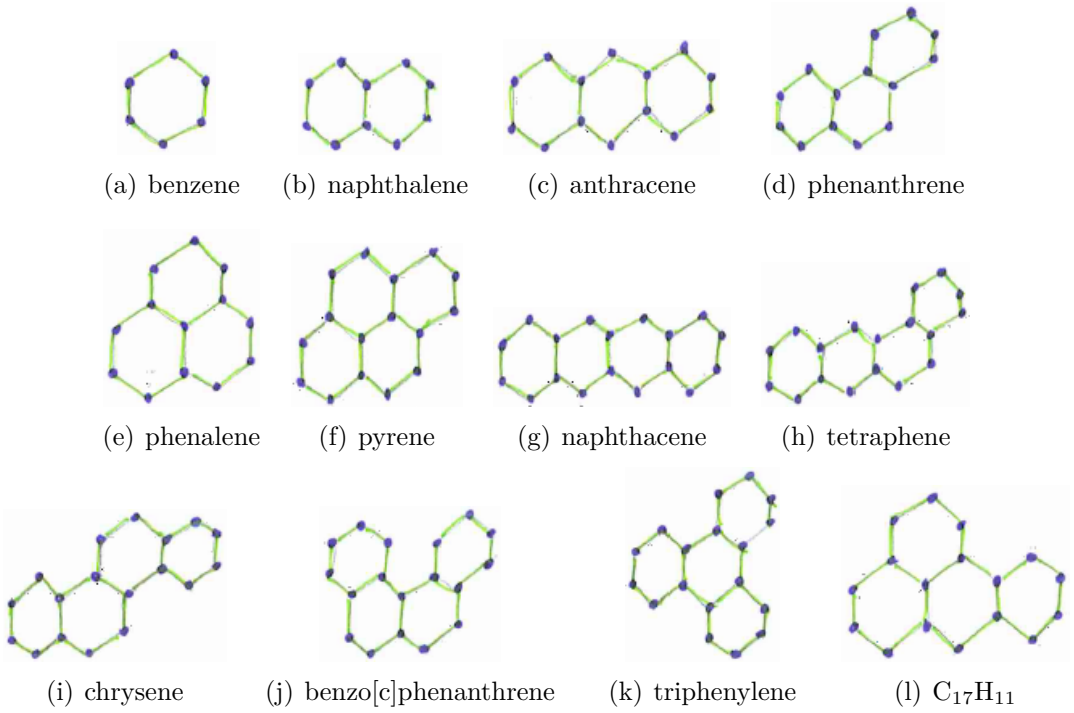
Take a benzenoid. Note that one of its faces, called the *outer face*, is unbounded. Moreover, every edge is incident to exactly two distinct faces (of which one may be the outer face). Every vertex is incident to 2 or 3 edges. Therefore, a benzenoid graph is 2-connected and 2-edge-connected. In a chemical context, we use terms *atom* and *bond* as synonyms for vertex and edge, respectively. Note that a benzenoid graph is bipartite, as it is a subgraph of $G(\mathcal{H})$.

Another way to define benzenoids is via *polygonal surfaces* (see Section 2.2):

Definition 5.4. *A benzenoid is a polygonal surface consisting of hexagons that can be embedded in the Euclidean plane such that all hexagons are regular.*

Note that if all hexagons of a polygonal surface are regular then their sides have equal length. Without this condition, we would obtain a much larger class of surfaces with boundary that also includes *helicenes* amongst other creatures. If we also dropped the condition that the polygonal surface can be embedded in the Euclidean plane then the class of such surfaces would be even larger.

Example 5.1. Figure 5.4 shows all (non-equivalent) benzenoids with up to 4 hexagons. Figure 5.5 shows a benzenoid with 11 hexagons. □



(a) benzene (b) naphthalene (c) anthracene (d) phenanthrene

(e) phenalene (f) pyrene (g) naphthacene (h) tetraphene

(i) chrysene (j) benzo[c]phenanthrene (k) triphenylene (l) $C_{17}H_{11}$

Figure 5.4: All non-equivalent benzenoids with up to 4 hexagons. (Their names are taken from reference [102].)

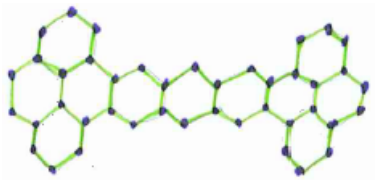
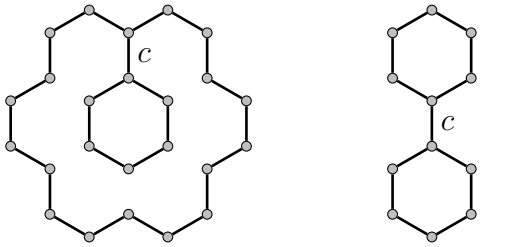


Figure 5.5: A benzenoid with 11 hexagons, called tetrabenzoheptacene [154].

Due to their importance in chemistry several benzenoid families have been studied in the past in addition to cata- and peri-condensed dichotomy (see Definition 5.12). Their Kekulé structures [31], symmetry groups [161] and other properties have been studied [37, 46, 48, 102, 171, 172]. We will consider some of them in the remainder of this chapter.

Another important class of polycyclic hydrocarbons that was widely studied in the past are coronoids [51, 52]. Intuitively speaking, a *coronoid* is a "benzenoid with holes", i.e., a benzenoid with some internal bonds and atoms removed. To define the class of coronoids precisely, some additional restrictions are needed. Normally, the resulting structure must be composed entirely of hexagons and connected if it is to be of interest in the theory of conjugated carbon frameworks. Although the two plane graphs in Figure 5.6 can be obtained from benzenoids by removing vertices and edges, they are not coronoids. This motivates the following mathematical formalisation of coronoids and benzenoids.



(a) (b) Biphenyl

Figure 5.6: These two plane graphs are not coronoids. The edge denoted by c (in both examples) is not incident to two distinct faces.

5.4 New approach towards hexagonal systems

We will consider the infinite planar hexagonal grid $\mathcal{H}$ as a collection of hexagons. Let a and b be two hexagons from $\mathcal{H}$. We say that a and b are *adjacent*, and denote this by $a \sim b$, if and only if they are different and share an edge. If two distinct hexagons share an edge we call them *neighbours*. The set of all neighbours of a will be denoted $N(a)$. Note that each hexagon of $\mathcal{H}$ has exactly 6 neighbours. Let $\mathcal{K} \subseteq \mathcal{H}$ be a *hexagonal system*, i.e., an arbitrary collection of hexagons from $\mathcal{H}$. $N_{\mathcal{K}}(a)$ will denote the set of those hexagons of $\mathcal{K}$ that belong to $N(a)$, i.e., $N_{\mathcal{K}}(a) \coloneqq N(a) \cap \mathcal{K}$. We call them the *neighbours of a in $\mathcal{K}$*. Define equivalence relation $\equiv_{\mathcal{K}}$ as follows: for $a, b \in \mathcal{K}$ it holds that $a \equiv_{\mathcal{K}} b$ if there exists a sequence $c_0 = a, c_1, c_2, \ldots, c_m = b$ such that $c_{i-1} \sim c_i$ for $i = 1, 2, \ldots, m$ and $c_i \in \mathcal{K}$ for $i = 0, 1, \ldots, m$. In particular, this means that it is possible to move from hexagon a to hexagon b along a pathway composed of adjacent hexagons that all belong to $\mathcal{K}$ (see Figure 5.7). Note that $a \equiv_{\mathcal{K}} b$ if $a \in N_{\mathcal{K}}(b)$ and $b \in \mathcal{K}$. Also, it is easy to see that $\mathcal{K} \subseteq \mathcal{L}$ and $a \equiv_{\mathcal{K}} b$ imply $a \equiv_{\mathcal{L}} b$. Hexagonal system $\mathcal{K}$ is naturally decomposed into equivalence

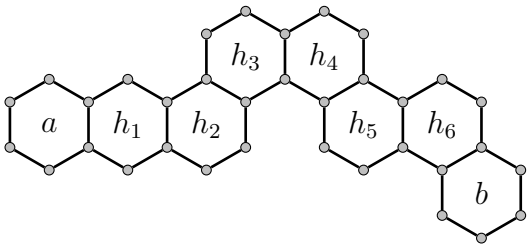


Figure 5.7: A path of hexagons $h_1, h_2, \ldots, h_6$ joining hexagon a to hexagon b.

classes $\{\mathcal{C}_i\}_{i \in C(\mathcal{K})}$, called *connected components*. Of course, $\mathcal{K} = \cup_{i \in C(\mathcal{K})} \mathcal{C}_i$. If $\mathcal{K}$ is finite, the number of its connected components, i.e., the cardinality of $C(\mathcal{K})$, is also finite. A hexagonal system is *connected* if it comprises only one connected component.

With respect to the number and positions of its neighbours, a hexagon of a connected hexagonal system $\mathcal{K}$ (with $h \geq 2$ hexagons) can be in one of 12 possible modes [102]. For example, a hexagon is in L_1 mode if it has only one neighbour. The complete list of modes is given in Figure 5.8.

Lemma 5.1. *Let $\mathcal{K}$ be a hexagonal system and $\{\mathcal{C}_i\}_{i \in C(\mathcal{K})}$ its decomposition into equivalence*

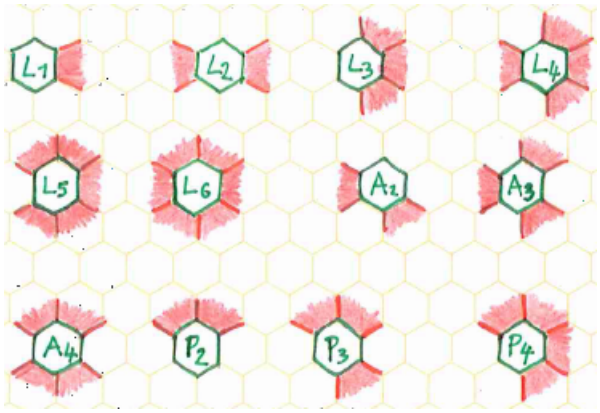


Figure 5.8: Modes of hexagons.

classes. Let $\mathcal{L}$ be a connected hexagonal system. If $\mathcal{L} \subseteq \mathcal{K}$ then $\mathcal{L} \subseteq \mathcal{C}_i$ for some $i \in C(\mathcal{K})$.

Proof. Suppose there exist hexagons $h_i \in \mathcal{L} \cap \mathcal{C}_i$ and $h_j \in \mathcal{L} \cap \mathcal{C}_j$. Connectedness of $\mathcal{L}$ implies $h_i \equiv_{\mathcal{L}} h_j$. From $\mathcal{L} \subseteq \mathcal{K}$ it follows that $h_i \equiv_{\mathcal{K}} h_j$. A contradiction. □

Lemma 5.2. *Let $\mathcal{K} \neq \emptyset$ be an arbitrary hexagonal system and $a \in \mathcal{K}$ any of its hexagons. Then hexagon a belongs to some connected component $\mathcal{C}_i$ of $\mathcal{K}$. Let $b \in N(a)$ be any of the neighbours of a in $\mathcal{H}$. Then either $b \in \mathcal{C}_i$ or $b \in \mathcal{K}^{\complement}$. In other words, no hexagon of $\mathcal{C}_i$ is adjacent to a hexagon of $\mathcal{C}_j$ if $i \neq j$.*

Proof. Suppose there is a hexagon $b \in \mathcal{C}_j$, $i \neq j$, such that $b \in N(a)$. Then a and b are in the same equivalence class by the definition of $\equiv_{\mathcal{K}}$. A contradiction. □

Lemma 5.3. *Let $\mathcal{K}$ be an arbitrary non-empty hexagonal system that is a proper subset of $\mathcal{H}$, i.e., $\emptyset \neq \mathcal{K} \subset \mathcal{H}$. Let $\mathcal{C}_a$ be any connected component of the complement $\mathcal{K}^{\complement}$. Then there exists a hexagon $\tilde{a} \in \mathcal{C}_a$ that is adjacent to some hexagon in $\mathcal{K}$.*

Proof. Let $N(\mathcal{R})$ denote the set of all hexagons that are not contained in $\mathcal{R}$ and are adjacent to some member of $\mathcal{R}$, i.e., $N(\mathcal{R}) = (\cup_{r \in \mathcal{R}} N(r)) \setminus \mathcal{R}$. Let $a \in \mathcal{C}_a$ and let $\mathcal{P}_0 = \{a\}$. Define $\mathcal{P}_k = \mathcal{P}_{k-1} \cup N(\mathcal{P}_{k-1})$ for $k \geq 1$.

It is clear that $\mathcal{P}_0 \subseteq \mathcal{C}_a$. Let n be the smallest number such that $\mathcal{P}_n \nsubseteq \mathcal{C}_a$. Suppose that such a number n exists. Note that $\mathcal{P}_{n-1} \subseteq \mathcal{C}_a$. Then $\mathcal{P}_n$ contains some hexagon $h \notin \mathcal{C}_a$. By Lemma 5.2, $h \in \mathcal{K}$. By construction of family $\{\mathcal{P}_k\}_k$, there exists a hexagon $\tilde{a} \in \mathcal{P}_{n-1}$, such that $h \in N(\tilde{a})$.

Now suppose that the desired number n does not exist. This means that $\mathcal{P}_n \subseteq \mathcal{C}_a$ for all n. But $\cup_{n=0}^{\infty} \mathcal{P}_n = \mathcal{H}$, i.e., for every $h \in \mathcal{H}$ there exists some number m, s.t. $h \in \mathcal{P}_m$. Since $\mathcal{K}$ is non-empty there is some hexagon $k \in \mathcal{K} \subset \mathcal{H}$ and therefore $k \in \mathcal{P}_l$ for some number l, which is a contradiction. □

Definition 5.5. *A finite connected hexagonal system $\mathcal{K}$ is called a* (general) coronoid.

Definition 5.6. *A finite connected hexagonal system $\mathcal{K}$ whose complement $\mathcal{K}^{\complement}$ is also connected is called a* benzenoid.

We prove a useful lemma:

Lemma 5.4. *Let $\mathcal{K}$ be any finite hexagonal system. Then the complement $\mathcal{K}^{\complement}$ of $\mathcal{K}$ consists of finitely many, say $d+1$, $d \geq 0$, connected components:*

$$\mathcal{K}^{\complement} = \mathcal{C}_\infty \sqcup \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_d. \tag{5.7}$$

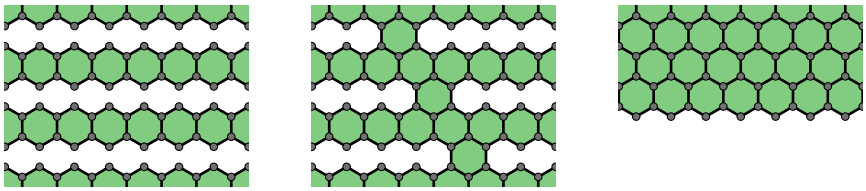
Each of these components but one, denoted by $\mathcal{C}_\infty$, is finite. Each of the finite components $\mathcal{C}_i$, $1 \leq i \leq d$, is a coronoid.

Component $\mathcal{C}_\infty$ is called the *exterior* of $\mathcal{K}$ and each $\mathcal{C}_i$, $1 \leq i \leq d$, is called a *corona hole*. In the above expression $\sqcup$ stands for disjoint union. The size of a coronoid $\mathcal{K}$, denoted $|\mathcal{K}|$, is the cardinality of the set $\mathcal{K}$, i.e., the number of hexagons it consists.

Proof. Let a be an arbitrary hexagon of $\mathcal{H}$. Let the family $\{\mathcal{P}_i\}_{i=0}^{\infty}$ be as defined in the proof of Lemma 5.3. Since $\mathcal{K}$ is finite, there exists $n \in \mathbb{N}$ such that $\mathcal{K} \subseteq \mathcal{P}_n$. (More precisely, for every $k \in \mathcal{K}$ there exists n_k such that $k \in \mathcal{P}_{n_k}$. Take $n := \max_{k \in \mathcal{K}} n_k$.) The complement of $\mathcal{P} := \mathcal{P}_n$ is contained in $\mathcal{K}^{\complement}$. Because $\mathcal{P}^{\complement}$ is connected, it is contained in a connected components of $\mathcal{K}^{\complement}$. Denote this component by $\mathcal{C}_\infty$. Note that this is the infinite component. In addition to $\mathcal{C}_\infty$, $\mathcal{K}^{\complement}$ may have more connected components. All of them (if there are any) are contained in $\mathcal{P}$. Each is finite, because $\mathcal{P}$ is finite. Their number is bounded by the number of hexagons in $\mathcal{P}$. Therefore, $\mathcal{K}^{\complement}$ has finitely many connected components.

The fact that all finite components are coronoids is clear from the definition of coronoids. □

The above lemma does not apply to infinite hexagonal systems. (See Figure 5.9 for examples.)



$\mathcal{K}_1$ $\mathcal{K}_2$ $\mathcal{K}_3$

Figure 5.9: $\mathcal{K}_1$ consists of infinitely many disjoint infinite lines of hexagons. Its complement has infinitely many connected components that are themselves infinite. $\mathcal{K}_2$ is obtained from $\mathcal{K}_1$ by adding another line of hexagons with a different slope. The hexagonal system $\mathcal{K}_2$ is a connected example. Hexagonal system $\mathcal{K}_3$ (half-plane) is infinite and its complement $\mathcal{K}_3^{\complement}$ has a single connected component. In fact, a connected hexagonal system with arbitrary many finite and arbitrary many infinite connected components can be obtained.

Lemma 5.5. *Let $\mathcal{K}$ be a coronoid and let $\mathcal{C}_a$ and $\mathcal{C}_b$ be two connected components of its complement $\mathcal{K}^{\complement}$. Let $a \in \mathcal{C}_a$ and $b \in \mathcal{C}_b$. Then $a \equiv_{\mathcal{K} \cup \mathcal{C}_a \cup \mathcal{C}_b} b$.*

Components $\mathcal{C}_a$ and $\mathcal{C}_b$ in the above lemma may be either two corona holes or one corona hole and the exterior of $\mathcal{K}$.

Proof. By Lemma 5.3 there exists $\tilde{a} \in \mathcal{C}_a$, such that $\tilde{a} \in N(k_a)$ for some $k_a \in \mathcal{K}$. By the same lemma, there exists $\tilde{b} \in \mathcal{C}_b$, such that $\tilde{b} \in N(k_b)$ for some $k_b \in \mathcal{K}$. Since $\mathcal{C}_a$ is a connected component, $a \equiv_{\mathcal{C}_a} \tilde{a}$. Similarly, $b \equiv_{\mathcal{C}_b} \tilde{b}$. From $\tilde{a} \equiv_{\mathcal{K}\cup\mathcal{C}_a} k_a$, $\tilde{b} \equiv_{\mathcal{K}\cup\mathcal{C}_b} k_b$ and $k_a \equiv_{\mathcal{K}} k_b$, it follows that $a \equiv_{\mathcal{K}\cup\mathcal{C}_a\cup\mathcal{C}_b} b$. □

Corollary 5.6. *Let $\mathcal{K}$ be a coronoid and let $\mathcal{C}_a$ be a connected component of its complement $\mathcal{K}^{\complement}$. Let $a \in \mathcal{C}_a$ and $k \in \mathcal{K}$. Then $a \equiv_{\mathcal{K}\cup\mathcal{C}_a} k$.*

Proof. In the proof of Lemma 5.5 we have already shown the existence of $\tilde{a} \in \mathcal{C}_a$, such that $\tilde{a} \in N(k_a)$ for some $k_a \in \mathcal{K}$. From $a \equiv_{\mathcal{C}_a} \tilde{a}$, $\tilde{a} \equiv_{\mathcal{K}\cup\mathcal{C}_a} k_a$ and $k_a \equiv_{\mathcal{K}} k$, it follows that $a \equiv_{\mathcal{K}\cup\mathcal{C}_a} k$. □

Theorem 5.7. *Let $\mathcal{K}$ be a coronoid. The complement $\mathcal{K}^{\complement}$ of $\mathcal{K}$ has a finite number $b(\mathcal{K}) := d + 1$, $d \geq 0$, of connected components $\mathcal{B}_\infty, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_d$ such that*

$$\mathcal{K}^{\complement} = \mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_d. \tag{5.8}$$

Exactly one component, denoted $\mathcal{B}_\infty$, is infinite and the other d components are finite, each being a benzenoid. Moreover, $\mathcal{B}_\infty^{\complement}$ is also a benzenoid.

Proof. From Lemma 5.4, it immediately follows that $\mathcal{K}^{\complement} = \mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_d$, where $\mathcal{B}_\infty$ is an infinite and $\mathcal{B}_i$, $1 \leq i \leq d$, are finite connected components. We need to show that each $\mathcal{B}_i$, $1 \leq i \leq d$, is a benzenoid.

It only remains to show that the complement $\mathcal{B}_i^{\complement}$ of $\mathcal{B}_i$, $1 \leq i \leq d$, is connected. From $\mathcal{H} = \mathcal{K} \sqcup \mathcal{K}^{\complement} = \mathcal{K} \sqcup (\mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_d)$ it follow that

$$\mathcal{B}_i^{\complement} = \mathcal{K} \sqcup \mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_{i-1} \sqcup \mathcal{B}_{i+1} \sqcup \cdots \sqcup \mathcal{B}_d. \tag{5.9}$$

Hexagonal systems $\mathcal{K}, \mathcal{B}_\infty, \mathcal{B}_1, \ldots, \mathcal{B}_{i-1}, \mathcal{B}_{i+1}, \ldots, \mathcal{B}_d$ are all connected. By Lemma 5.5 and Corollary 5.6, their union is also connected.

To show that $\mathcal{B}_\infty^{\complement}$ is a benzenoid, we need to show that $\mathcal{B}_\infty^{\complement}$ and $(\mathcal{B}_\infty^{\complement})^{\complement}$ are connected and that $\mathcal{B}_\infty^{\complement}$ is finite. Since $(\mathcal{B}_\infty^{\complement})^{\complement} = \mathcal{B}_\infty$, it is clearly connected. Since

$$\mathcal{B}_\infty^{\complement} = \mathcal{K} \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_d, \tag{5.10}$$

it is connected by the same argument that worked for $\mathcal{B}_i^{\complement}$ above. Hexagonal systems $\mathcal{K}, \mathcal{B}_1, \ldots, \mathcal{B}_d$ are all finite and therefore their union $\mathcal{B}_\infty^{\complement}$ is also finite. □

Definition 5.7. *Let $\mathcal{K}$ be a coronoid. The* benzenoid closure *of $\mathcal{K}$, denoted $\overline{\mathcal{K}}$, is the intersection of all those benzenoids that include $\mathcal{K}$ as a subset, i.e.,*

$$\overline{\mathcal{K}} = \bigcap \{\mathcal{B} \mid \mathcal{B} \text{ is benzenoid} \wedge \mathcal{K} \subseteq \mathcal{B}\}. \tag{5.11}$$

For a coronoid $\mathcal{K}$ we define $Benz(\mathcal{K}) = \{\mathcal{B} \mid \mathcal{B} \text{ is benzenoid} \wedge \mathcal{K} \subseteq \mathcal{B}\}$. This set will be repeatedly used in several proofs that follow. Using this terminology, (5.11) from the above definition can be written in a shorter form as $\overline{\mathcal{K}} = \bigcap Benz(\mathcal{K})$.

Lemma 5.8. *The benzenoid closure $\overline{\mathcal{K}}$ of a (general) coronoid $\mathcal{K}$ is a benzenoid. Moreover, $\overline{\mathcal{K}} = \mathcal{B}_\infty^{\complement} = \mathcal{K} \sqcup \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_d$, where $\mathcal{K}^{\complement} = \mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_d$ as in Theorem 5.7.*

Proof. By Theorem 5.7, $\mathcal{K}^{\complement} = \mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \cdots \sqcup \mathcal{B}_d$, where $\mathcal{B}_\infty$ is infinite and $\mathcal{B}_i$, $1 \leq i \leq d$, are benzenoids. We will show that $\overline{\mathcal{K}} = \mathcal{K} \cup \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_d = \mathcal{B}_\infty^{\complement}$.

Let $\mathcal{L}$ be an arbitrary benzenoid such that $\mathcal{K} \subseteq \mathcal{L}$. Then $\mathcal{L}^{\complement} \subseteq \mathcal{K}^{\complement}$. Because $\mathcal{L}^{\complement}$ is connected, it is contained in at most one of components $\mathcal{B}_\infty, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_d$. Since $\mathcal{L}^{\complement}$ is infinite, $\mathcal{L}^{\complement} \subseteq \mathcal{B}_\infty$. Therefore,

$$\mathcal{K} \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_d = \mathcal{B}_\infty^{\complement} \subseteq (\mathcal{L}^{\complement})^{\complement} = \mathcal{L}. \tag{5.12}$$

This implies $\mathcal{K} \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_d \subseteq \overline{\mathcal{K}}$.

By Theorem 5.7, $\mathcal{B}_\infty^{\complement} = \mathcal{K} \cup \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_d$ is a benzenoid. Thus $\overline{\mathcal{K}} \subseteq \mathcal{K} \cup \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_d$. We have proved that $\overline{\mathcal{K}} = \mathcal{K} \cup \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_d$, where $\mathcal{K} \cup \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_d$ is a benzenoid. This proves existence, and also uniqueness, of $\overline{\mathcal{K}}$. □

In the above proof of Lemma 5.8 we have also shown how to construct $\overline{\mathcal{K}}$ for a given $\mathcal{K}$. The reader should note that in general case the intersection of two benzenoids is not necessarily a benzenoid (see Figure 5.10).

Proposition 5.9. *The benzenoid closure $\mathcal{K} \mapsto \overline{\mathcal{K}}$ is an operation on the set of all coronoids that satisfies the following three conditions:*

(a) $\forall \mathcal{K}\colon \mathcal{K} \subseteq \overline{\mathcal{K}}$,

(b) $\forall \mathcal{K}, \mathcal{L}\colon \mathcal{K} \subseteq \mathcal{L} \implies \overline{\mathcal{K}} \subseteq \overline{\mathcal{L}}$, *and*

(c) $\forall \mathcal{K}\colon \overline{\overline{\mathcal{K}}} = \overline{\mathcal{K}}$.

Note that the co-domain of the mapping $\mathcal{K} \mapsto \overline{\mathcal{K}}$ can be restricted to the set of all benzenoids. This mapping is surjective and the preimage of every benzenoid is a finite set of coronoids.

Proof. By definition, $\overline{\mathcal{K}} = \bigcap Benz(\mathcal{K})$. From the definition it follows directly that $\mathcal{K} \subseteq \overline{\mathcal{K}}$.

Let us now show that $\overline{\overline{\mathcal{K}}} = \overline{\mathcal{K}}$. By definition, $\overline{\overline{\mathcal{K}}} = \bigcap Benz(\overline{\mathcal{K}})$. Therefore $\overline{\mathcal{K}} \subseteq \overline{\overline{\mathcal{K}}}$. Since $\overline{\mathcal{K}}$ is a benzenoid, $\overline{\mathcal{K}} \in Benz(\overline{\mathcal{K}})$. Therefore, $\overline{\overline{\mathcal{K}}} \subseteq \overline{\mathcal{K}}$.

Finally, we will show that $\mathcal{K} \subseteq \mathcal{L} \Rightarrow \overline{\mathcal{K}} \subseteq \overline{\mathcal{L}}$. By definition, $\overline{\mathcal{K}} = \bigcap Benz(\mathcal{K})$ and $\overline{\mathcal{L}} = \bigcap Benz(\mathcal{L})$. From $\mathcal{K} \subseteq \mathcal{L}$ it follows that every element of $Benz(\mathcal{L})$ is also an element of $Benz(\mathcal{K})$, i.e., $Benz(\mathcal{L}) \subseteq Benz(\mathcal{K})$. Therefore $\overline{\mathcal{K}} \subseteq \overline{\mathcal{L}}$. □

We define another closure operation.

Definition 5.8. *An* alternative benzenoid closure *$\mathcal{K} \mapsto \widetilde{\mathcal{K}}$ is a mapping on the set of all coronoids that satisfies the following three conditions:*

(i) $\forall \mathcal{K}\colon \widetilde{\mathcal{K}}$ *is benzenoid,*

(ii) $\forall \mathcal{K}\colon \mathcal{K} \subseteq \widetilde{\mathcal{K}}$, *and*

(iii) $\forall \mathcal{K}\colon \forall$*benzenoid* $\mathcal{L}\colon \mathcal{K} \subseteq \mathcal{L} \implies \widetilde{\mathcal{K}} \subseteq \mathcal{L}$.

The following lemma tells us that this corresponds to an alternative definition of the benzenoid closure operation.

Lemma 5.10. *Let $\mathcal{K}$ be a general coronoid. Then $\widetilde{\mathcal{K}} = \overline{\mathcal{K}}$. With other words, benzenoid closure operation and the alternative benzenoid closure operation coincide.*

Proof. First, we will prove existence of $\widetilde{\mathcal{K}}$ by proving that $\overline{\mathcal{K}}$ satisfies the three conditions of Definition 5.8.

By Lemma 5.8, $\overline{\mathcal{K}}$ is a benzenoid, so (i) holds. Proposition 5.9 tells us that $\mathcal{K} \subseteq \overline{\mathcal{K}}$, so (ii) also holds. It only remains to see that $\mathcal{K} \subseteq \mathcal{L} \implies \overline{\mathcal{K}} \subseteq \mathcal{L}$ for every benzenoid $\mathcal{L}$.

Let $\mathcal{L}$ be any benzenoid such that $\mathcal{K} \subseteq \mathcal{L}$. By definition, $\mathcal{L} \in \mathit{Benz}(\mathcal{K})$. It is clear that $\mathcal{L} \in \mathit{Benz}(\mathcal{K})$ implies $\overline{\mathcal{K}} = \bigcap \mathit{Benz}(\mathcal{K}) \subseteq \mathcal{L}$.

In principle, there could exist some other benzenoid, different from $\overline{\mathcal{K}}$, which would also satisfy the three conditions in Definition 5.8. We will show that this is not the case by proving that every $\widetilde{\mathcal{K}}$ from Definition 5.8 is $\overline{\mathcal{K}}$.

Let $\mathcal{L}$ be any element of $\mathit{Benz}(\mathcal{K})$. Condition (iii) implies that $\widetilde{\mathcal{K}} \subseteq \mathcal{L}$. Therefore $\widetilde{\mathcal{K}} \subseteq \bigcap \mathit{Benz}(\mathcal{K}) = \overline{\mathcal{K}}$. By condition (i) and (ii), $\widetilde{\mathcal{K}}$ is a benzenoid such that $\mathcal{K} \subseteq \widetilde{\mathcal{K}}$. This means that $\widetilde{\mathcal{K}}$ is a member of $\mathit{Benz}(\mathcal{K})$. Clearly, $\overline{\mathcal{K}} = \bigcap \mathit{Benz}(\mathcal{K}) \subseteq \widetilde{\mathcal{K}}$. From $\widetilde{\mathcal{K}} \subseteq \overline{\mathcal{K}}$ and $\overline{\mathcal{K}} \subseteq \widetilde{\mathcal{K}}$ it follows that $\widetilde{\mathcal{K}} = \overline{\mathcal{K}}$ which completes the proof. □

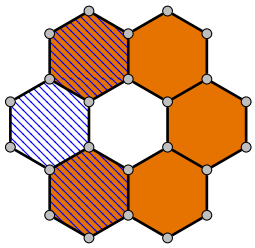


Figure 5.10: The intersection of two benzenoids is not necessarily a benzenoid.

Lemma 5.11. *The intersection of two benzenoids is a finite hexagonal system. Each of its connected components is a benzenoid.*

Proof. Let $\mathcal{L}_a$ and $\mathcal{L}_b$ be two benzenoids and let $\mathcal{L} = \mathcal{L}_a \cap \mathcal{L}_b$. Because $\mathcal{L}_a$ and $\mathcal{L}_b$ are finite, $\mathcal{L}$ is also finite. It consists of finitely many (possibly 0) finite connected components. That allows us to write $\mathcal{L} = \mathcal{K}_1 \sqcup \mathcal{K}_2 \sqcup \cdots \sqcup \mathcal{K}_d$. To complete the proof, we have to show that each $\mathcal{K}_i^{\complement}$ is connected.

By Lemma 5.4, $\mathcal{K}_i^{\complement} = \mathcal{C}_\infty \sqcup \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_m$. From $\mathcal{K}_i \subseteq \mathcal{L} \subseteq \mathcal{L}_a$ we obtain $\mathcal{L}_a^{\complement} \subseteq \mathcal{K}_i^{\complement}$. Since $\mathcal{L}_a^{\complement}$ is connected and infinite, $\mathcal{L}_a^{\complement} \subseteq \mathcal{C}_\infty$. In addition to $\mathcal{C}_\infty$, $\mathcal{K}_i^{\complement}$ may have 0 or more other connected components. Suppose $m \geq 1$.

By Lemma 5.3, there exists $h \in \mathcal{C}_1$ which is adjacent to some $k \in \mathcal{K}_i$. Then $h \notin \mathcal{L}_a$ or $h \notin \mathcal{L}_b$. If h belonged to both $\mathcal{L}_a$ and $\mathcal{L}_b$, then $h \in \mathcal{K}_j$ for some j. From $h \in \mathcal{C}_1 \subseteq \mathcal{K}_i^{\complement}$, it follows that $j \neq i$. But this contradicts Lemma 5.2, so h indeed belongs to at most one of $\mathcal{L}_a$ or $\mathcal{L}_b$. Without loss of generality, we can assume $h \notin \mathcal{L}_a$. In other words, $h \in \mathcal{L}_a^{\complement} \subseteq \mathcal{C}_\infty$. This contradicts the fact that $h \in \mathcal{C}_1$.

This means that $m = 0$, implying $\mathcal{K}_i^{\complement} = \mathcal{C}_\infty$, and the proof is complete. □

Look at Figure 5.11. Benzenoids $\mathcal{K}_1$ and $\mathcal{K}_2$ are not equal, i.e., $\mathcal{K}_1 \neq \mathcal{K}_2$. For instance, $h_1 \in \mathcal{K}_1$ but $h_1 \notin \mathcal{K}_2$. If one would draw them on a piece of paper and cut them out, they

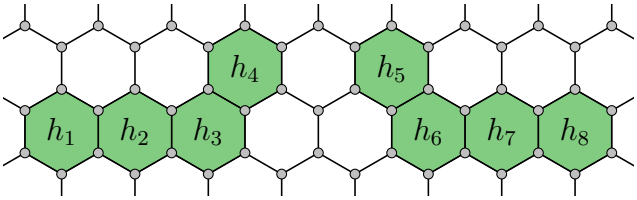


Figure 5.11: Benzenoids $\mathcal{K}_1 = \{h_1, h_2, h_3, h_4\}$ and $\mathcal{K}_2 = \{h_5, h_6, h_7, h_8\}$ are equivalent.

would coincide. This notion of coincidence can be precisely defined. One possible approach is via isometries of the plane that map hexagons to hexagons. (For more details on this topic see [94].) If the coordinate system is placed so that the origin is in the centre of a chosen hexagon and if hexagons have sides of unit length then isometries $\phi_1, \phi_2, \phi_3, \phi_4 \colon \mathbb{R}^2 \to \mathbb{R}^2$, where

$$\begin{aligned}
\phi_1(x, y) &= \left(x + \sqrt{3}, y\right), \\
\phi_2(x, y) &= \left(x + \tfrac{\sqrt{3}}{2}, y + \tfrac{3}{2}\right), \\
\phi_3(x, y) &= \left(-x, y\right), \text{ and} \\
\phi_4(x, y) &= \left(x \cdot \cos\tfrac{\pi}{3} - y \cdot \sin\tfrac{\pi}{3}, x \cdot \sin\tfrac{\pi}{3} + y \cdot \cos\tfrac{\pi}{3}\right)
\end{aligned}$$

generate a subgroup of the group of isometries of $\mathbb{R}^2$. Elements of this subgroup are precisely those isometries that map hexagons to hexagons. Note that this group is not a subgroup of the $O(2)$, becuase there exist elements which do not fix any point of $\mathbb{R}^2$. However, there is another approach which is purely combinatorial. Let $\mathrm{Aut}(\mathcal{H})$ be the group of symmetries of the hexagonal grid, i.e.,

$$\mathrm{Aut}(\mathcal{H}) = \{\phi \colon \mathcal{H} \to \mathcal{H} \mid \phi \text{ is bijection} \wedge \forall a, b \in \mathcal{H} \colon \phi(a) \sim \phi(b) \Leftrightarrow a \sim b\}. \tag{5.13}$$

Now, we can define:

Definition 5.9. *Hexagonal systems $\mathcal{K}$ and $\mathcal{L}$ are* equivalent*, denoted $\mathcal{K} \cong \mathcal{L}$, if there exists a symmetry $\psi \in \mathrm{Aut}(\mathcal{H})$, such that $\psi(\mathcal{K}) = \mathcal{L}$.*

Let $\mathcal{K}$ be a (general) coronoid. From Lemma 5.8 it follows that

$$\overline{\mathcal{K}} \setminus \mathcal{K} = \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_d \quad (d \geq 0), \tag{5.14}$$

where each $\mathcal{B}_i$ is a benzenoid. The benzenoids $\mathcal{B}_i$ are called *corona holes*. They consist of one or more hexagons. The (general) coronoids (a) and (b) in Figure 5.12 have two corona holes each, whilst coronoid (c) has only one. Examples (b) and (c) are special in a sense. A general coronoid is called *degenerate* if one of its corona holes is a single hexagon. This is because such a corona hole has no interpretation in chemistry. (Recall that in a benzenoid molecule, as opposed to a benzenoid graph, all vertices of degree 2 represent CH groups and all of degree 3 represent C atoms. There is no notion of the infilling of each hexagon having a physical meaning. Larger holes, with their internal H atoms or degree 2 vertices, have a chemical meaning.) Otherwise it is called *non-degenerate*. Corona holes of size 1 will be called *degenerate corona holes*.

Definition 5.10. *Let $\mathcal{K}$ be any coronoid. The non-degenerate closure of $\mathcal{K}$, NonDeg($\mathcal{K}$), is the intersection of all those non-degenerate coronoids that include $\mathcal{K}$ as a subset, i.e.,*

$$\overline{\mathcal{K}} = \bigcap \{\mathcal{N} \mid \mathcal{N} \text{ is non-degenerate coronoid} \wedge \mathcal{K} \subseteq \mathcal{N}\}. \tag{5.15}$$

Plainly speaking, it is the smallest non-degenerate coronoid which includes $\mathcal{K}$. It is not hard to see that *NonDeg*($\mathcal{K}$) can be obtained from $\mathcal{K}$ by adding to it exactly its degenerate corona holes. Hereinafter, where we consider applications of this theory in chemistry, by the word *coronoid* we will always mean a non-degenerate coronoid.

The definition of coronoids includes benzenoids as a special case.

Definition 5.11. *A* proper coronoid *is a coronoid that is not also a benzenoid.*

Using the terminology of topology one may say that the only difference between benzenoids and proper coronoids is that both are connected, but the former are also simply connected. To give a precise meaning to this, one has to consider benzenoids as surfaces with a boundary. Plane graphs on their own are not sufficient model for coronoids. Some faces have to be labelled as "not present". We will call them *holes*. In this context, the outer face of a coronoid is merely one of its holes. Chemists do not always distinguish between the two models (coronoids as plane graphs and coronoids as surfaces with boundary); for them, exactly those faces of length strictly greater than 6 are holes. In other words, they do not recognise degenerate corona holes. As we will see in Chapter 6, when we generalise coronoids to perforated patches, some care should be taken over this distinction.

We defined coronoids (and benzenoids) as subsets of hexagons in the Euclidean plane. We may consider them as 2-dimensional *cell complexes* [109]. Cell complexes are a generalisation of polygonal surfaces (defined in Section 2.2). In what follows, we will adopt some topological terminology and notation. Recall that in Section 5.2 we obtained the infinite hexagonal grid $\mathcal{H}$ by embedding the infinite cubic graph called *hexagonal lattice* in the Euclidean plane. Vertices of the hexagonal lattice are 0-cells, edges are 1-cells and faces are 2-cells. Every edge of the hexagonal lattice is incident to exactly two distinct hexagonal faces. Any two distinct faces can either share a single edge or nothing at all.

Let $\mathcal{K}$ be a coronoid. By our definition it is a collection of 2-cells. We can assign to $\mathcal{K}$ the smallest *subcomplex* of the hexagonal grid which includes all hexagons (2-cells) of $\mathcal{K}$. Note that a subcomplex contains the closure of each of its cells. Its 1-skeleton is a graph that is a subgraph of the hexagonal lattice consisting of those vertices and edges which are incident with at least one hexagon of $\mathcal{K}$. We will denote this graph by $G(\mathcal{K})$ and call it a *skeleton* in the general setting and a *coronoid graph* when we deal with coronoids. Note that if $\mathcal{H}$ denotes the hexagonal grid, then $G(\mathcal{H})$ is the hexagonal lattice. Note that $\mathrm{Aut}(G(\mathcal{H}))$ acts transitively on vertices of $G(\mathcal{H})$. Elements of $\mathrm{Aut}(G(\mathcal{H}))$ are graph automorphisms.

Consider the three examples of coronoids on Figure 5.12. As in the case of benzenoids, every edge is incident to exactly two distinct faces of which one may be a hole, but not both. Every vertex is incident to either 2 or 3 edges. A coronoid graph is a 2-connected and 2-edge-connected graph.

The edges of a coronoid graph are naturally divided into two types: internal and boundary. An edge belonging to two (adjacent) hexagons is *internal* and an edge belonging to only one hexagon is a *boundary edge*. Vertices of a coronoid graph can also be divided

into internal and boundary. *Internal* vertices are incident with 3 internal edges. All other vertices of $G(\mathcal{K})$ are called *boundary*. Boundary vertices can be further divided into two types: those of degree 2 and those of degree 3. A hexagon h of a coronoid $\mathcal{K}$ is *internal* if $|N_{\mathcal{K}}(h)| = 6$. Otherwise it is a *boundary* hexagon. In other words, an internal hexagon is surrounded by 6 internal edges. The subgraph of $G(\mathcal{K})$ that consists of all boundary vertices and edges will be called *border* of $\mathcal{K}$ and denoted $\partial\mathcal{K}$. When $\mathcal{K} = \{k\}$ is a singleton, we will write ∂k.

Definition 5.12. *A benzenoid is called* catacondensed *if it has no internal vertices, and* pericondensed *otherwise.*

Pericondensed benzenoids can be further divided into two groups:

Definition 5.13. *Pericondensed benzenoids which have internal hexagons are called* corpulent *benzenoids. Those without internal hexagons are called* gaunt *benzenoids.*

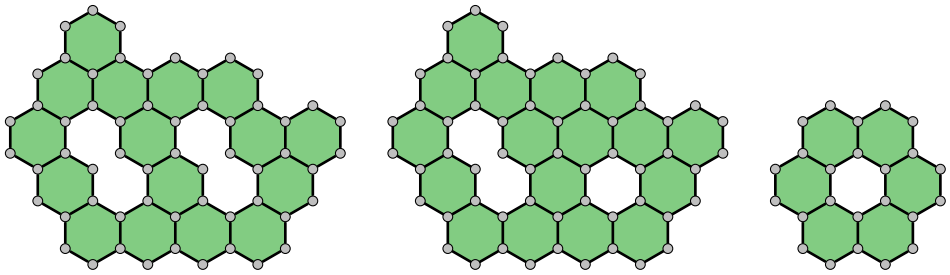
Example 5.2. The benzenoid $\mathcal{B}_1$ in Figure 5.13 is catacondensed, whilst, benzenoids $\mathcal{B}_2$ and $\mathcal{B}_3$ are pericondensed. Benzenoid $\mathcal{B}_2$ is gaunt and $\mathcal{B}_3$ is corpulent. □

Let us prove the following property of the infinite hexagonal grid:

Proposition 5.12. *The girth of the infinite hexagonal lattice is 6, i.e.,*

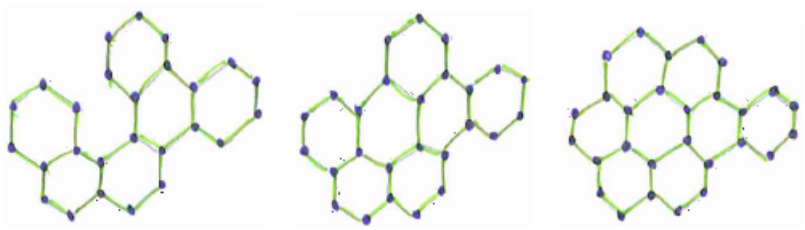
$$\text{girth}(G(\mathcal{H})) = 6. \tag{5.16}$$

Let C be an arbitrary cycle of $G(\mathcal{H})$. Then $|C| = 6$ if and only if there exists a hexagon $h \in \mathcal{H}$ such that $C = \partial h$.



(a) (b) (c)

Figure 5.12: Three examples of general coronoids. The first is non-degenerate whilst the second and third are degenerate.



(a) $\mathcal{B}_1$ (b) $\mathcal{B}_2$ (c) $\mathcal{B}_3$

Figure 5.13: A catacondensed, a gaunt and a corpulent benzenoid.

Proof. Let $h \in \mathcal{H}$ be any hexagon. Then $\partial h \cong C_6$. We will show that there are no shorter cycles and also that all 6-cycles are the borders of hexagonal faces of the hexagonal grid.

The distance between vertices u and v of a graph G is the length of the shortest path between u and v in G. We will denote it by $d_G(u, v)$.

Owing to symmetry, to investigate the structural properties of cycles of $G(\mathcal{H})$ it is enough to investigate cycles that contain a given vertex u. Up to symmeties of $G(\mathcal{H})$ there is only one path of length 2. See Figure 5.14. Vertices u and v are not adjacent, so there

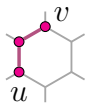


Figure 5.14: The only type of path of length 2 in $G(\mathcal{H})$.

are no triangles in $G(\mathcal{H})$. There are two possible paths of length 3 (see Figure 5.15). In

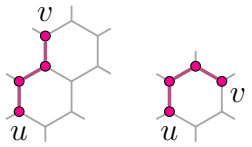


Figure 5.15: The two types of paths of length 3 in $G(\mathcal{H})$.

both the endpoints are not adjacent, so there are no rectangles in $G(\mathcal{H})$. Both paths of length 3 can be extended in two ways yielding four paths of length 4 (see Figure 5.16). In

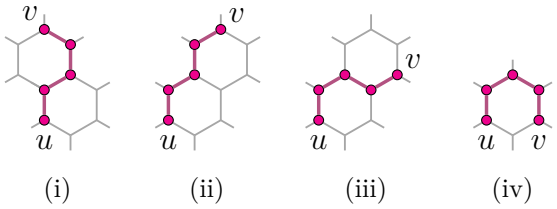
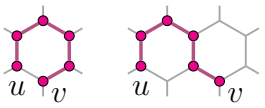


Figure 5.16: The four types of paths of length 4 in $G(\mathcal{H})$.

all cases their endpoints are non-adjacent, so there are no 5-cycles in $G(\mathcal{H})$. In cases (i), (ii) and (iii) we have that $d(u, v) = 4$. Therefore, these paths cannot form 6-cycles. The case (iv) can be extended to a 5-path in two ways (see Figure 5.17). In case (b), we have that $d(u, v) = 3$, which means that no 6-cycle can be formed. In case (a), vertices u and v are adjacent and therefore form a 6-cycle. We have examined all options and no other 6-cycles arise, which means that all of them can be obtained as borders of hexagons. □

Note that every vertex of $G(\mathcal{K})$ belongs to one, two or three 6-cycles and that every edge of $G(\mathcal{K})$ belongs to one or two 6-cycles. Because $G(\mathcal{K}) \subseteq G(\mathcal{H})$, every 6-cycle of $G(\mathcal{K})$ is also a 6-cycle of $G(\mathcal{H})$. Therefore, for every 6-cycle C in $G(\mathcal{K})$ there exists some $h \in \mathcal{H}$,



(a) (b)

Figure 5.17: Paths of length 5 in $G(\mathcal{H})$.

such that $C = \partial h$. Let u be an arbitrary vertex of $G(\mathcal{K})$. It is incident with 3 hexagons, say h_1, h_2 and h_3, of $\mathcal{H}$. By definition of skeleton, at least one of those hexagons must also be in $\mathcal{K}$. Without loss of generality $h_1 \in \mathcal{K}$. Therefore, u belongs to 6-cycle ∂h_1. It may also happen that $h_2 \in \mathcal{K}$ and/or $h_3 \in \mathcal{K}$. Then vertex u also belongs to cycle ∂h_2 and/or ∂h_3. Let e be an arbitrary edge of $G(\mathcal{K})$. There exist $h_1 \in \mathcal{H}$ and $h_2 \in \mathcal{H}$ such that $e \in \partial h_1$ and $e \in \partial h_2$. By definition of skeleton at least one of them also belongs to $\mathcal{K}$. It may also happen that both of them belong to $\mathcal{K}$. We can now state the following proposition:

Proposition 5.13. *Let $\mathcal{K} \subseteq \mathcal{H}$ be an arbitrary coronoid. Then the following statements are true:*

(a) The graph $G(\mathcal{K})$ is connected.

(b) For every edge $e \in E(G(\mathcal{K}))$ there exists $h \in \mathcal{K}$ such that $e \in \partial h$.

(c) For every cycle $C \subseteq G(\mathcal{K})$ the inequality $|C| \geq 6$ holds. The equality is attained if and only if there exists a hexagon $h \in \mathcal{H}$ such that $C = \partial h$. Moreover, if $\mathcal{K}$ is non-degenerate then $h \in \mathcal{K}$.

Proof. To show that $G(\mathcal{K})$ is connected, we will find a (u, v)-path for any pair of vertices $u, v \in V(G(\mathcal{K}))$. We already know that there exist h_u and h_v such that $u \in \partial h_u$ and $v \in \partial h_v$. There exists a sequence of hexagon $h_0 = h_v, h_1, \ldots, h_n = h_u$ such that h_i and h_{i+1} are adjacent for all i. If two hexagons h and k are adjacent, there exists a path from any vertex of ∂h to any vertex of ∂k. On every hexagon h_i, $1 \leq i < n$, choose a vertex v_i. Let $v_0 = v$ and $v_n = u$. There exist paths P_i with endvertices v_i and v_{i+1}. Concatenation of paths $P_0, P_1, \ldots, P_{n-1}$ is a (u, v)-walk.

Statement (b) was already proved in the discussion preceding Proposition 5.13.

If $H \subseteq G$, then $\text{girth}(H) \geq \text{girth}(G)$. The fact that $G(\mathcal{K}) \subseteq G(\mathcal{H})$ and Proposition 5.12 give us the inequality. Those cycles which attain the equality are characterized by Proposition 5.12. Now suppose that $h \notin \mathcal{K}$ and that $\mathcal{K}$ is non-degenerate. Hexagon h is sorrounded by 6 hexagons $h_1, h_2, \ldots, h_6$ of $\mathcal{H}$. By definition of $G(\mathcal{K})$, at least one of h or h_i must be in $\mathcal{K}$ for all i. This means that $h_1, \ldots, h_6 \in \mathcal{K}$. But then $\{h\}$ is a corona hole of size 1, a contradiction. □

Proposition 5.14. *For any coronoid $\mathcal{K}$, the subgraph $\partial\mathcal{K}$ forms a collection of $b(\mathcal{K})$, $b(\mathcal{K}) \geq 1$, cycles. A coronoid is a benzenoid if $b(\mathcal{K}) = 1$ and is a proper coronoid with $b(\mathcal{K}) - 1$ corona holes if $b(\mathcal{K}) > 1$.*

Proof. Look at Figure 5.18. There are two types of boundary vertex: the one in Figure 5.18(a) is incident with only one hexagon of $\mathcal{K}$; the one in Figure 5.18(b) is incident

with 2 hexagons of $\mathcal{K}$. We consider the subgraph containing boundary vertices and edges. In the first case vertex v is incident with edges e and e' and has degree 2. In the second case, v is incident with e and e'' and also has degree 2. Therefore, the subgraph containing boundary vertices and edges is a 2-regular graph which is a union of vertex-disjoint cycles. By the Theorem 2.3 (Jordan curve theorem), every cycle splits the plane into two discon-

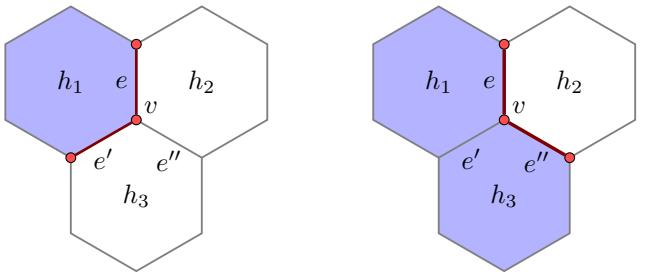


Figure 5.18: Two types of boundary vertices. Shaded hexagons are present in the coronoid. Vertex v is of degree 2 in the first case and of degree 3 in the second case. Boundary vertices and edges are emphasized.

nected regions. Consider a hexagon $h \in \mathcal{K}^{\complement}$. It is contained in a region that is surrounded by one of the cycles. No member of $\mathcal{K}$ is inside this region, because a pathway of hexagons cannot enter the region elsewhere but through the perimeter. Therefore, perimeters separate corona holes from the coronoid $\mathcal{K}$. By Theorem 5.7, there must be the same number of cycles as there are corona holes (including the outer face). □

For two graphs G and H by $H \hookrightarrow G$ we denote the embedding (injective homomorphism) of graph H into G. Note that graph $G(\mathcal{K})$ is not a plane graph and does not possess any geometric information. Let $\mathcal{A}$ be anthracene. From Figure 5.19, it is clear that $G(\mathcal{A})$ can be drawn in the plane in many different ways.

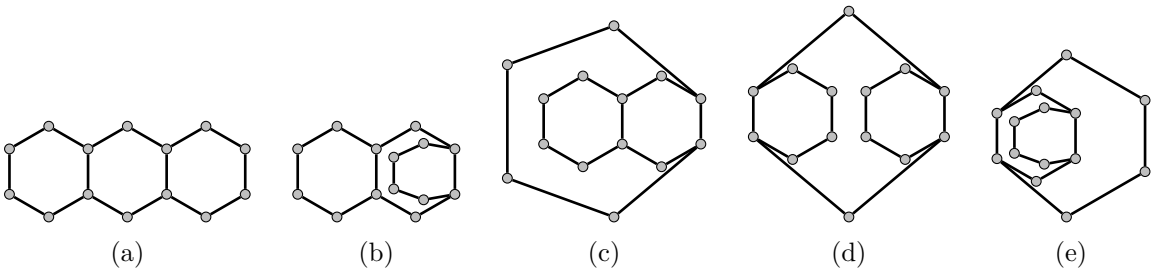


Figure 5.19: Some different embeddings of the anthracene graph $G(\mathcal{A})$ in the Euclidean plane. Only (a) represents a part of the hexagonal grid $\mathcal{H}$.

The most appropriate drawing of anthracene for chemical purposes is the left-most. It is the only one that can be obtained from the hexagonal grid $\mathcal{H}$. There is an embedding of a coronoid graph into the hexagonal grid $\mathcal{H}$ called the *natural embedding*. Recall that $G(\mathcal{K})$ was obtained from $\mathcal{K}$ by taking all 0-cells and 1-cells of the corresponding subcomplex. The

natural embedding just sends the graph back to its place of birth. We have the following theorem to tell us that there is only one drawing up to symmetries of the hexagonal grid:

Theorem 5.15. *Let $\mathcal{K}$ be a coronoid and let $C \subseteq G(\mathcal{K})$ be a cycle of length 6 ($|C| = 6$). Then $C \hookrightarrow G(\mathcal{H})$ can be extended to $G(\mathcal{K}) \hookrightarrow G(\mathcal{H})$ in an unique way.*

Proof. Let $e = uv \in E(C_6)$. Then $e \hookrightarrow G(\mathcal{H})$ can be extended to $C_6 \hookrightarrow G(\mathcal{H})$ in two different ways.

The distance between hexagons h and k in $\mathcal{K}$, $d(h, k)$, is the smallest n for which a sequence $h_0 = h, h_1, \ldots, h_n = k$ of sequentially adjacent hexagons of $\mathcal{K}$ exist. Let $\mathcal{K} = \{h_1, h_2, \ldots, h_d\}$. Suppose that $C = \partial h_1$ and that $d(h_1, h_i) \leq d(h_1, d_{i+1})$. We start with $C \hookrightarrow G(\mathcal{H})$ and extend it step by step. On i-th step we find images of those vertices of ∂h_i that are not already embedded. For h_i there exists some h_j, $j < i$, such that ∂h_j was already embedded and h_i shares an edge e_i with h_j. If vertices of ∂h_i were already embedded, there is nothing left to do. Otherwise, $e_i \hookrightarrow G(\mathcal{H})$ can be extended to $\partial h_i \hookrightarrow G(\mathcal{H})$ in two ways. As we are constructing an injective homomorphism, images of ∂h_j and ∂h_i may not overlap. Therefore, only one option remains. □

This constructive proof gives rise to an algorithm for embedding an arbitrary coronoid graph G into the hexagonal lattice. If anything goes wrong during this procedure, that means that the input graph G was not a valid coronoid graph. If the graph G is a valid coronoid graph, the algorithm will always finish successfully.

Theorem 5.15 no longer holds, if we permit arbitrary subgraphs of $G(\mathcal{H})$. (See the examples given in Figure 5.20, Figure 5.21 and Figure 5.22.)

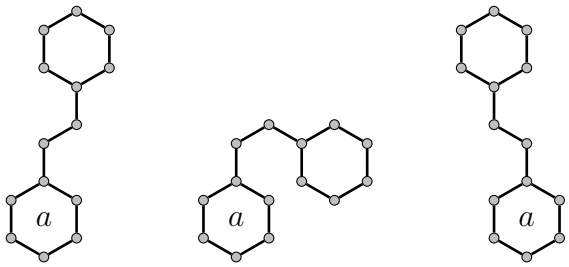


Figure 5.20: A graph consisting of two hexagons connected by a path of length 3 can be embedded in the hexagonal lattice in more than one way, even when the hexagon denoted by a is fixed.

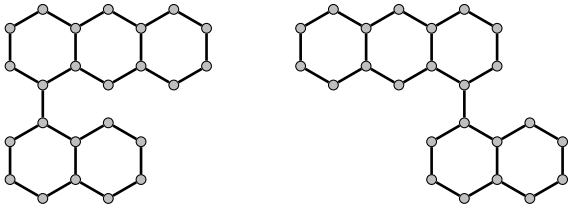


Figure 5.21: Another example of a non-coronoid subgraph of the lattice that can be embedded in more than one way.

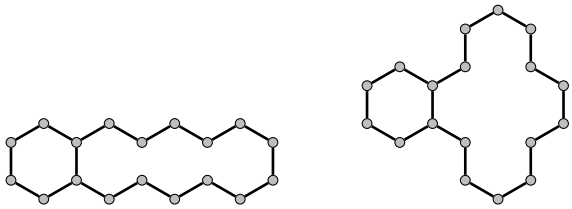


Figure 5.22: An example of a 2-connected non-coronoid graph that can be embedded in more than one way. (Not all possible embeddings are listed.)

Lemma 5.16. *Let $\mathcal{K}$ be a coronoid and let $G \coloneqq G(\mathcal{K})$. Then there exists (up to symmetry of $\mathcal{H}$) exactly one non-degenerate coronoid $\mathcal{N}$ such that $G = G(\mathcal{N})$. Moreover, $\mathcal{N} \cong NonDeg(\mathcal{K})$.*

Proof. Choose some hexagon $h \in \mathcal{H}$ and choose an arbitrary 6-cycle C in G. Let $\phi\colon C \hookrightarrow G(\mathcal{H})$ be an embedding such that $\phi(C) = \partial h$. By Theorem 5.15, ϕ can be extended to $\Phi\colon G \hookrightarrow G(\mathcal{H})$ in a unique way. Let $\mathcal{N} = \{h \in \mathcal{H} \mid \partial h \subseteq \Phi(G)\}$. For every 6-cycle C in $\Phi(G)$ there exists some $h_C \in \mathcal{N}$ such that $\Phi(C) = \partial h_C$.

First we show that $\mathcal{N}$ is non-degenerate. Suppose that it has a corona hole of size 1, i.e., there exists $\widetilde{h} \in \mathcal{N}^{\complement}$ which is sorrounded by $h_1, h_2, \ldots, h_6 \in \mathcal{N}$. Every edge of cycle $\partial\widetilde{h}$ is present in $\Phi(G)$ because all hexagons $h_1, \ldots, h_6$ are present in $\mathcal{N}$. But then, by definition of $\mathcal{N}$, also $\widetilde{h} \in \mathcal{N}$. A contradiction.

Now, we show that $\mathcal{N} \cong NonDeg(\mathcal{K})$. Without loss of generality, we can assume that Φ is the natural embedding. If $k \in \mathcal{K}$ then from the definition of $G(\mathcal{K})$ we conclude that $\partial k \in G(\mathcal{K})$. Therefore, $k \in \mathcal{N}$. This means that $\mathcal{K} \subseteq \mathcal{N}$.

Let $h \in \mathcal{N} \setminus \mathcal{K}$. Let $\tilde{h} \in N(h)$ and suppose that $\tilde{h} \notin \mathcal{K}$. We know that h and $\tilde{h}$ share an edge e which does not belong to $G(\mathcal{K})$, because neither h nor $\tilde{h}$ belongs to $\mathcal{K}$. But $e \in G(\mathcal{K})$ by definition of $\mathcal{N}$. This is a contradiction. Therefore, $\tilde{h} \in \mathcal{K}$ for every $\tilde{h} \in N(h)$. With other words, h is a degenerate corona hole inside $\mathcal{K}$. That means that $\mathcal{N}$ is obtained from $\mathcal{K}$ by adding degenerate corona holes. □

We use the terminology of Gutman and Cyvin [102]. Each boundary cycle is called a *perimeter.* In a proper coronoid there is one *outer perimeter* and one or more *inner perimeters.* It may happen that one of the inner perimeters is longer than the outer perimeter (see Figure 5.23). Nevertheless, the cycle of a coronoid graph that belong to the outer perimeter can be

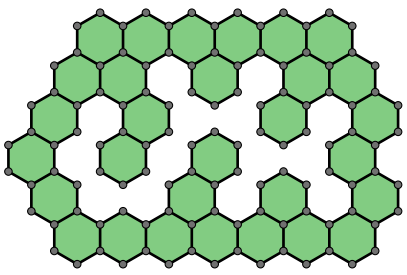


Figure 5.23: The outer perimeter of this single coronoid is of length 48 whilst the inner perimeter is of length 58.

easily recognised. By Theorem 5.15, a coronoid can be embedded in the hexagonal lattice in a unique way (up to symmetry). Then the left-most vertex of the coronoid belongs to the outer perimeter.

A *corona hole* is a unique benzenoid that can fill the interior of an inner perimeter. An inner perimeter of a coronoid may be viewed as the (outer) perimeter of the corresponding corona hole. The roles of boundary vertices of degrees 2 and 3 are interchanged when we make this change of viewpoint. A corona hole of a non-degenerate coronoid has at least 2 hexagons and the corresponding inner perimeter has at least two vertices of degree 2. A typical coronoid is schematically illustrated in Figure 5.24.

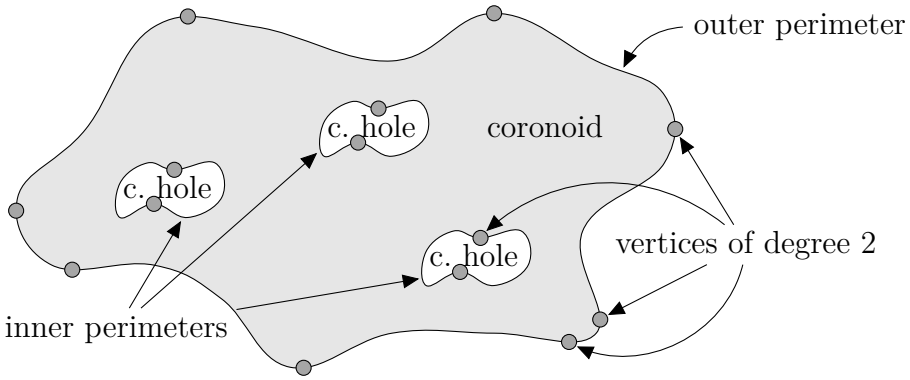


Figure 5.24: Schematic illustration of a typical coronoid. The outer perimeter has at least 6 vertices of degree 2 and any inner perimeter has at least 2 vertices of degree 2.

There are exactly two vertices of degree 2 on the inner perimeter in the case of a naphthalene corona hole. For all other corona holes the number of such vertices is strictly greater than 2. A lower bound can easily be obtained from the equations in [102]. Let us prove a simple lemma first:

Lemma 5.17. *Let h, n, m, n_i and ν denote, respectively, the number of hexagons, vertices, edges, internal vertices and boundary vertices of degree 3 in a benzenoid $\mathcal{B}$. Then:*

$$\nu = 2h - 2 - n_i \tag{5.17}$$

and

$$n = 4h + 2 - n_i. \tag{5.18}$$

Proof. Both equations can be proved by the mathematical induction on the number of hexagons h. Every benzenoid $\mathcal{B}$ with h hexagons can be obtained from a benzenoid $\mathcal{B}'$ with $h-1$ hexagons by adding a hexagon a to the perimeter of $\mathcal{B}'$. The hexagon $a \in \mathcal{B}$ can be in L_1, L_3, L_5, P_2 or P_4 mode.

The only benzenoid with a single hexagon is the benzene for which $h = 1$, $n = 6$, $m = 6$, $n_i = 0$ and $\nu = 0$. Equations (5.17) and (5.18) clearly hold for the benzene.

Let h', n', m', n'_i and ν' be the parameters of the benzenoid $\mathcal{B}'$ (from which $\mathcal{B}$ is obtained by a single addition). By the induction hypothesis it holds that

$$\nu' = 2h' - 2 - n'_i$$

and

$$n' = 4h' + 2 - n'_i.$$

Clearly, $h = h' + 1$. Suppose that the hexagon a is in L_3 mode. By this addition the number of vertices increases by 2 and the number of edges increases by 3, i.e., $n = n' + 2$ and $m = m' + 3$. The number of internal vertices increases by 2, i.e., $n_i = n'_i + 2$. The number of boundary vertices of degree 3 remains the same, i.e., $\nu' = \nu$. It follows that

$$\nu = \nu' = 2h' - 2 - n'_i = 2(h-1) - 2 - (n_i - 2) = 2h - 2 - n_i$$

and

$$n = n' + 2 = 4h' + 2 - n'_i + 2 = 4(h-1) + 2 - (n_i - 2) + 2 = 4h + 2 - n_i.$$

The cases when a is in L_1, L_5, P_2 or P_4 mode can be proved by the same technique and are left as an exercise to the reader. □

Proposition 5.18. *Let h, n, m and n_i denote, respectively, the number of hexagons, vertices, edges and internal vertices in a corona hole $\mathcal{B}$ of a coronoid $\mathcal{C}$. The number of boundary vertices of degree 3 in $\mathcal{B}$ (which correspond to vertices of degree 2 on the corresponding inner perimeter of $\mathcal{C}$), is*

$$2h - 2 - n_i = n - 2h - 4 \geq \sqrt{12h - 3} - 3. \tag{5.19}$$

Proof. Let ν denote the number of boundary vertices of degree 3 that belong to the corona hole $\mathcal{B}$ (those vertices are exactly degree-2 vertices of the corresponding inner perimeter). From equations (5.17) and (5.18) we can eliminate variable n_i to obtain $\nu = n - 2h - 4$. Then apply $2h + 1 + \sqrt{12h - 3} \leq n$ [108] to get $\nu \geq \sqrt{12h - 3} - 3$. □

5.5 Description via BEC and enumeration

There are several combinatorial descriptions of a benzenoid. A benzenoid can be, for instance, represented as a set of hexagons of the hexgonal grid, i.e., as a subset of $\mathbb{Z} \times \mathbb{Z}$. For example, *chrysene* in Figure 5.4(i) can be represented by the set

$$\{(-1, 0), (0, 0), (0, 1), (1, 1)\} \tag{5.20}$$

or by the set

$$\{(1, 1), (1, 2), (2, 2), (2, 3)\}. \tag{5.21}$$

There are infinitely many such representations of any given benzenoid. By appying an isometry of the hexagonal grid we obtain another representation. Assume that a benzenoid $\mathcal{B}$ with h hexagons is represented by the set

$$\{(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_h, \eta_h)\}. \tag{5.22}$$

We may assume that these labels are lexicographically ordered, i.e.,

(a) $\xi_i < \xi_{i+1}$ or
(b) $\xi_i = \xi_{i+1}$ and $\eta_i < \eta_{i+1}$

for all $1 \le i < h$. If we also require that $\min_{1\le i\le h} \xi_i = 0$ and $\min_{1\le i\le h} \eta_i = 0$ then we obtain a finite list of possible representations. The one that is lexicographically the smallest is called the *canonical representation*. The canonical representation of the chrysene is

$$\{(0,0),(0,1),(1,1),(1,2)\}. \tag{5.23}$$

Another possible description, which is particularly user-friendly, is the so-called *boundary-edges code* popularized by Hansen et al. [107]. See also [125]. The reader should be aware that helicenes are not uniquely determined by their boundary-edges codes.

Definition 5.14. *The* boundary-edges code *of a benzenoid $\mathcal{B}$, denoted $BEC(\mathcal{B})$ is a circular sequence of numbers, where each of them counts the number of edges on the boundary between two consecutive degree-3 boundary vertices. The exception is benzene with the boundary-edges code* `6`.

Benzene is an exception because it has no degree-3 vertices on its boundary. Any other benzenoid can be described by a sequence of numbers chosen from the alphabet $\{1,2,3,4,5\}$.

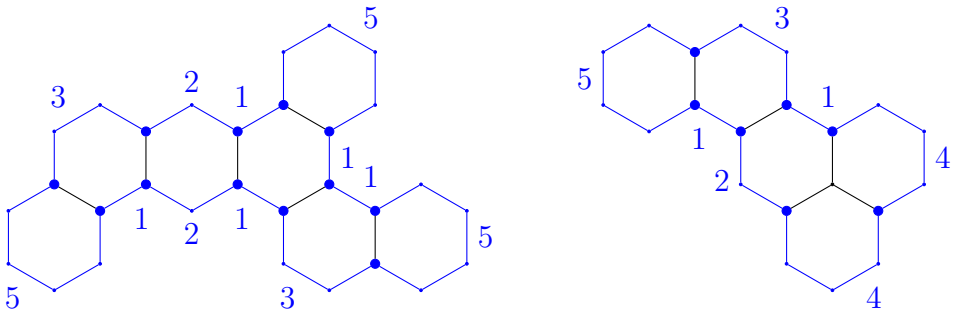
Example 5.3. Figure 5.25(a) depicts a catacondensed benzenoid $\mathcal{B}_1$ of 7 hexagons with

$$BEC(\mathcal{B}_1) = \texttt{532151153121}. \tag{5.24}$$

A pericondensed benzenoid $\mathcal{B}_2$ of 5 hexagons with

$$BEC(\mathcal{B}_2) = \texttt{5314421} \tag{5.25}$$

is in Figure 5.25(b). □



(a) $\mathcal{B}_1$ (b) $\mathcal{B}_2$

Figure 5.25: Examples of benzenoids together with corresponding numbers in boundary-edges codes.

We may shift and reverse a boundary-edges code. For example, from the code `5314421` we obtain `3144215`, `1442153`, `1244135`, ... The string that is lexicographically the largest among all cyclic shifts and their reverses is called the *canonical boundary-edges code*. The codes (5.24) and (5.25) are both canonical. This also gives us an easy method to test whether two benzenoids are equivalent. The following proposition is clear:

Proposition 5.19. *Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two benzenoids and let $BEC(\mathcal{B}_1)$ and $BEC(\mathcal{B}_2)$ be their canonical boundary-edges codes. If $\mathcal{B}_1 \cong \mathcal{B}_2$ then $BEC(\mathcal{B}_1) = BEC(\mathcal{B}_2)$.* □

The problem of enumeration of benzenoids was widely studied in the past. An overview of algorithms for generating benzenoids is given in [150] and in references cited therein. See also papers [22, 23] by Brinkmann, Caporossi and Hansen. An efficient generator for benzenoids is implemented in the **CaGe** package [24]. Here, we will describe a simple algorithm that exploits the canonical boundary-edges code. This algorithm was used in Section 4.2.

We create a sequence of sets $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \ldots$ The set $\mathcal{D}_h$ will contain canonical boundary-edges codes of all benzenoids with h hexagons. We start with $\mathcal{D}_1 = \{\texttt{6}\}$, i.e., the set $\mathcal{D}_1$ contains benzene which is the only benzenoid with a single hexagon. We can obtain $\mathcal{D}_{h+1}$ from the set $\mathcal{D}_h$. Iterate over all members of $\mathcal{D}_h$. For each $\mathcal{B} \in \mathcal{D}_h$, we attach a new hexagon to $\mathcal{B}$ in all possible ways. This produces a finite list of coronoids $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$. We can immediately discard all proper coronoids. The remaining hexagonal systems are all benzenoid. We determine their canonical boundary-edges codes and add them to the set $\mathcal{D}_{h+1}$.

From $\mathcal{D}_1$ we obtain $\mathcal{D}_2 = \{\texttt{55}\}$. We can attach a new hexagon to benzene at 6 different places, but in each case we obtain the naphthalene whose boundary-edges code is `55`. From $\mathcal{D}_2$ we obtain $\mathcal{D}_3 = \{\texttt{5252}, \texttt{5351}, \texttt{444}\}$ and so on.

The time that is needed to obtain $\mathcal{D}_{h+1}$ from $\mathcal{D}_h$ depends on the size of the set $\mathcal{D}_h$ and on lengths of the boundary-edges codes stored in $\mathcal{D}_h$. Let $\ell_1, \ell_2, \ldots, \ell_{|\mathcal{D}_h|}$ be lengths of boundary-edges codes in $\mathcal{D}_h$. For the benzenoid that corresponds to the code ℓ_i there are $O(\ell_i)$ places at which a new hexagon may be attached and the new boundary-edges code has length $O(\ell_i)$. The canonical boundary-edges code can be determined in linear time with respect to the length of the code [20]. If the set data structure is implemented as a has table, the insertion costs on average $O(1)$ time and the hashing function takes $O(\ell_i)$ time. Therefore, it takes $O(\sum_{i=1}^{|\mathcal{D}_h|} \ell_i^2)$ time to create $\mathcal{D}_{h+1}$ from $\mathcal{D}_h$.

5.6 Convex benzenoids

Recently Cruz, Gutman and Rada [47] considered an interesting subfamily of (finite) benzenoids that they call *convex*. Independently, a webpage by Scott Reynolds (a.k.a. Nekura Ca) exists that enumerates convex benzenoids [178]. The sequence was considered, yet again independently, by Allan C. Wechsler and is recorded in the OEIS [1] as the sequence A116513 [198].

By a *walk* in $\mathcal{H}$ we mean a sequence of hexagons such that any two consecutive ones are either adjacent or the same. Recall that a *path* is a walk that consists of pairwise distinct hexagons. For any pair of hexagons a and b in the infinite hexagonal grid $\mathcal{H}$ define the *interval* $I_{\mathcal{H}}(a, b)$ to be the benzenoid (which turns out to be of rhombic shape) that is composed of all hexagons on any of the shortest paths in $\mathcal{H}$ from a to b.

Definition 5.15. *A benzenoid $\mathcal{B}$ is* convex *if for any pair of its hexagons a and b the whole interval $I_{\mathcal{H}}(a, b)$ is contained in $\mathcal{B}$.*

This definition follows readily from the fact that hexagons of $\mathcal{H}$ form a metric space and $\mathcal{B}$ can be viewed as a subspace of $\mathcal{H}$. We are therefore considering convex sets in metric spaces. This notion of convexity (also known as *geodesic convexity*) can be found in a survey on metric graph theory by Bandelt and Chepoi [10].

Let $a, b \in \mathcal{B}$. With $d_{\mathcal{B}}(a, b)$ we denote the distance between a and b inside $\mathcal{B}$ and $d_{\mathcal{H}}(a, b)$ is the distance between a and b with respect to $\mathcal{H}$. Note that

$$d_{\mathcal{B}}(a, b) \geq d_{\mathcal{H}}(a, b), \tag{5.26}$$

but they are not necessarily equal (see Figure 5.26). If, in addition, $\mathcal{K} \subseteq \mathcal{H}$, then

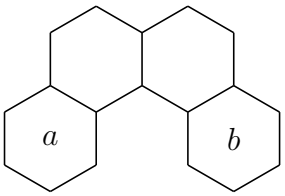


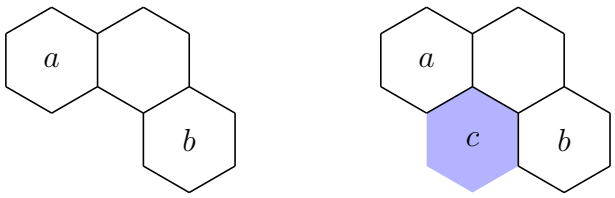
Figure 5.26: Benzo[c]phenanthrene $\mathcal{B}$, $d_{\mathcal{B}}(a, b) = 3 \neq d_{\mathcal{H}}(a, b) = 2$.

$$d_{\mathcal{H}}(a, \mathcal{K}) = \min_{k \in \mathcal{K}} d_{\mathcal{H}}(a, k). \tag{5.27}$$

Definition 5.16. *A benzenoid $\mathcal{B}$ obeys the* Small Parallelogram Rule *if for every $a, b \in \mathcal{B}$ such that $d_{\mathcal{H}}(a, b) = 2$ inclusion of some shortest path between a and b implies inclusion of the whole interval $I_{\mathcal{H}}(a, b)$, i.e.,*

$$\forall a, b \in \mathcal{B}\colon (d_{\mathcal{H}}(a, b) = 2 \wedge d_{\mathcal{B}}(a, b) = 2 \implies I_{\mathcal{H}}(a, b) \subseteq \mathcal{B}).$$

We call it the *Small Parallelogram Rule* for the following reason. If a sub-benzenoid consisting of 3 hexagons and defined by the boundary-edges code `5351` (phenanthrene) is present then the fourth hexagon that extends it to to pyrene (with the boundary-edges code `4343`) is also present (see Figure 5.27). The case when a and b lie on the same line is trivial (there is only one shortest path between a and b).



(a) $d_{\mathcal{B}}(a, b) = d_{\mathcal{H}}(a, b) = 2$ (b) $c \in I_{\mathcal{H}}(a, b) \subseteq \mathcal{B}$

Figure 5.27: Small Parallelogram Rule

Often it is convenient to introduce a coordinate system on $\mathcal{H}$. First, observe that the set of all edges of $\mathcal{H}$ can be partitioned into 3 classes with respect to their direction:

(a) v-edges (vertical edges),

(b) p-edges (positive slope edges), and

(c) n-edges (negative slope edges).

Every hexagon of $\mathcal{H}$ has two v-neighbours, two p-neighbours, and two n-neighbours. Recall that each $a \in \mathcal{H}$ can be assigned two integer coordinates. Consider the hexagon labeled by $(0, 0)$. Its right v-neighbour has its first coordinate increased by 1, and its left v-neighbour has its first coordinate decreased by 1. Its upper-right n-neighbour has its second coordinate increased by 1, and the opposite n-neighbour has its second coordinate decreased by 1. Figure 5.28 displays a subregion of $\mathcal{H}$ equipped with this coordinate system.

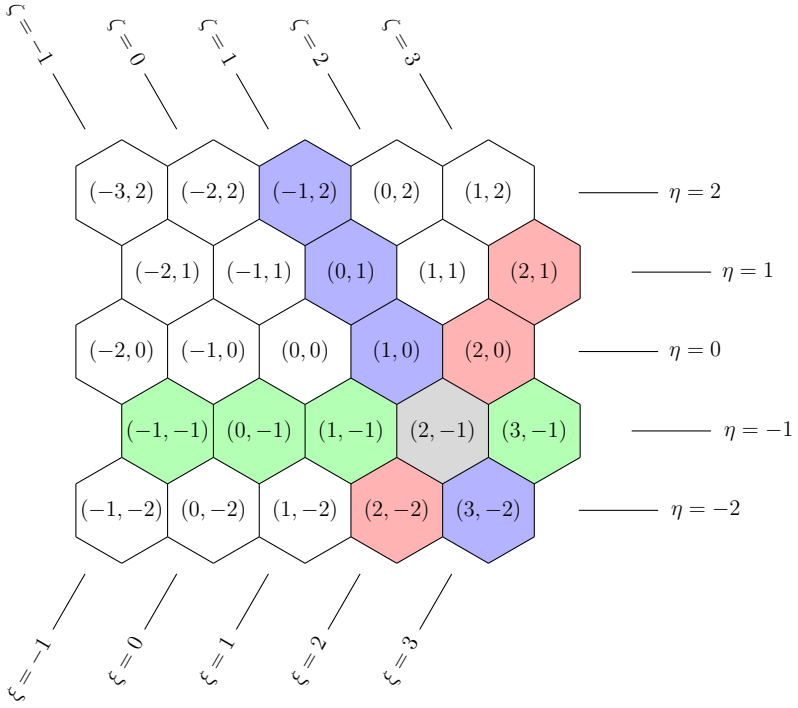


Figure 5.28: Coordinate system on $\mathcal{H}$.

Let $a \in \mathcal{H}$ be a hexagon of the infinite hexagonal grid $\mathcal{H}$. Let $\xi(a)$ and $\eta(a)$ denote the first and second coordinate of a. We can associate to the hexagon a another number $\zeta(a) = \xi(a) + \eta(a)$. It can be treated as yet another coordinate which increases/decreases with respect to the p-direction but is not independent of the other two.

Definition 5.17. *The sets*

$$\{a \in \mathcal{H} \mid \xi(a) = k\}, \{a \in \mathcal{H} \mid \eta(a) = k\}, \text{ and } \{a \in \mathcal{H} \mid \zeta(a) = k\}, \tag{5.28}$$

where $k \in \mathbb{Z}$ are called lines *in $\mathcal{H}$. We will distinguish them by their directions and call them* positive slope line, horizontal line *and* negative slope line, *respectively.*

According to Definition 5.17, hexagons $a, b \in \mathcal{H}$ belong to the same line if either:

(a) $\xi(a) = \xi(b)$,

(b) $\eta(a) = \eta(b)$, or

(c) $\zeta(a) = \zeta(b)$.

Examples of lines are shown on Figure 5.28. Green hexagons and the gray one constitute a horizontal line, red hexagons and the gray one constitute a positive slope line, and blue hexagons and the gray one constitute a negative slope line. Note that any two lines with different directions intersect in exactly one hexagon and, moreover, there alway exists exactly one line with the remaining direction that includes the intersection of the previous two.

Proposition 5.20. *A benzenoid system $\mathcal{B}$ is convex if and only if:*

(a) $\mathcal{B}$ is connected and

(b) $\mathcal{B}$ obeys the Small Paralellogram Rule.

Proposition 5.20 gives us much more concise conditions for convexity than the original Definition 5.15. To determine if a benzenoid is convex, one only needs to check whether it is connected and whether it obeys the Small Parallelogram Rule, which is a "local" condition. To prove the Proposition 5.20 we need the following lemma:

Lemma 5.21. *Let a benzenoid $\mathcal{B}$ satisfy conditions (a) and (b) from Proposition 5.20. If $a, b \in \mathcal{H}$ are on the same line then $I(a, b) \subseteq \mathcal{B}$.*

Proof. Let $a, b \in \mathcal{H}$ be on the same line. Without loss of generality we can assume that $a, b \in \mathcal{L} = \{h \in \mathcal{H} \mid \xi(h) = 0\}$. Since $\mathcal{B}$ is connected, there exists a path $P = p_1\, p_2 \ldots p_l$ between $a = p_1$ and $b = p_l$. (By the definition of a path, p_i and p_{i+1} are adjacent hexagons for all $1 \leq i < l$.) Let $n = \max_{1 \leq i \leq l} d_{\mathcal{H}}(p_i, \mathcal{L})$. The proof goes by induction on n.

In the case of $n = 0$ (base of induction) the image of path P is a subset of $\mathcal{L}$. Since the image of a path is connected it follow that $I(a, b) \subseteq \mathcal{B}$.

Now suppose that $n = \max_{1 \leq i \leq l} d_{\mathcal{H}}(p_i, \mathcal{L}) > 0$. By the induction hypothesis the existence of a path $Q = q_1\, q_2 \ldots q_k$ between a and b, such that $\max_{1 \leq i \leq k} d_{\mathcal{H}}(q_i, \mathcal{L}) < n$, would imply that $I(a, b) \subseteq \mathcal{B}$. Let r be the smallest index such that $d_{\mathcal{H}}(p_{r-1}, \mathcal{L}) = n - 1 \wedge d_{\mathcal{H}}(p_r, \mathcal{L}) = n$ and let s be the smallest index such that $d_{\mathcal{H}}(p_s, \mathcal{L}) = n \wedge d_{\mathcal{H}}(p_{s+1}, \mathcal{L}) = n - 1 \wedge s \geq r$. This situation is depicted on Figure 5.29. This subpath may look like $p'_{r-1}\, p_r \ldots p_s\, p'_{s+1}$,

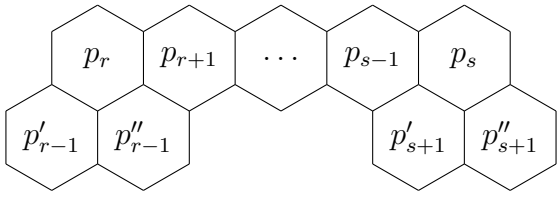


Figure 5.29: Subpath $p_{r-1}\, p_r \ldots p_{s+1}$ of path P.

$p''_{r-1}\, p_r \ldots p_s\, p'_{s+1}$, $p'_{r-1}\, p_r \ldots p_s\, p''_{s+1}$, or $p''_{r-1}\, p_r \ldots p_s\, p''_{s+1}$. If $r = s$ then by deletion of p_r

from the path P we obtain a walk (which can be further reduced to a proper path) that has one less hexagon at distance n from $\mathcal{L}$ than the original path P. If $r < s$ then we have two options:

(a) if $p_{r-1} = p''_{r-1}$ we can erase p_r from P since p_{r-1} and p_{r+1} are adjacent;

(b) if $p_{r-1} = p'_{r-1}$ then the Small Parallelogram Rule implies that $p''_{r-1} \in \mathcal{B}$ so we can replace p_r with p''_{r-1} in path P.

In both cases we obtain a path with one less hexagon at distance n from $\mathcal{L}$. If we iteratively apply this procedure, we obtain the path Q with desired property. □

Proof of Proposition 5.20. If $\mathcal{B}$ is convex then it is clear that $\mathcal{B}$ is connected and that $\mathcal{B}$ obeys the Small Parallelogram Rule.

Let us prove the other direction. Choose arbitrary hexagons $a, b \in \mathcal{B}$. We have to show that $I_{\mathcal{H}}(a, b) \subseteq \mathcal{B}$. Interval between two hexagons is always of rhombic shape. Without loss of generality (thanks to the symmetries of $\mathcal{H}$) we can assume that a is in its lower-left corner and b is in its upper-right corner (as depicted on Figure 5.30). Condition (a) implies the existence of a path P in $\mathcal{B}$ between a and b. An example of such a path is on Figure 5.30. First, we show that we can assume that the path $P = p_1\, p_2 \ldots p_l$ lies entirely in

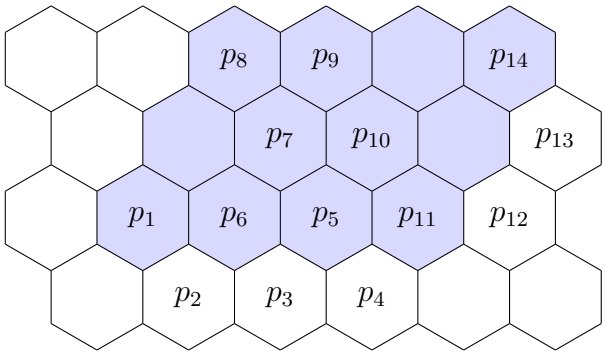


Figure 5.30: Interval $I(a = p_1, b = p_{14})$

the half-plane $\{h \in \mathcal{H} \mid \eta(h) \geq \eta(a)\}$. Let r be the smallest index such that $\eta(p_r) < \eta(a)$. Note that $r = 2$ in the example from Figure 5.30. Let s be the largest index such that $\eta(p_s) < \eta(a)$. Note that $s = 4$ in the example from Figure 5.30. If such indices do not exist, there is nothing left to prove. If such indices exist, then $1 < r \leq s < l$. Moreover, $\eta(p_{r-1}) = \eta(a) = \eta(p_{s+1})$. So p_{r-1} and p_{s+1} are on the same line and $I(p_{r-1}, p_{s+1}) \subseteq \mathcal{B}$ by Lemma 5.21. So the subpath $p_r \ldots p_s$ can be replaced with the shortest (p_{r-1}, p_{s+1})-path. Similarly, we can assume that P lies entirely in the half-plane $\{h \in \mathcal{H} \mid \eta(h) \leq \eta(b)\}$, in half-plane $\{h \in \mathcal{H} \mid \xi(h) \geq \xi(a)\}$, and in half-plane $\{h \in \mathcal{H} \mid \xi(h) \leq \xi(b)\}$. Therefore $P \subseteq I_{\mathcal{H}}(a, b)$.

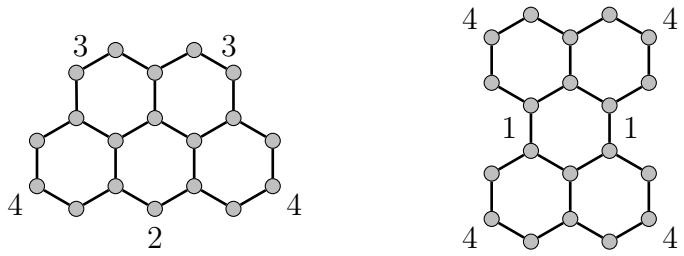
So far, we know that there is a path P in $\mathcal{B}$ connecting a to b that lies entirely in $I_{\mathcal{H}}(a, b)$. Recall our assumption that a is in its lower-left corner and b is in its upper-right corner. Now, we will show that there also exists a path $P' = p'_1 p'_2 \ldots p'_m$ in $\mathcal{B}$ with the special property that p'_{i+1} is the right v-neighbour or the upper-right n-neighbour of p'_i for each i, i.e., in each step either the η coordinate increases or the ξ coordinate increases.

The path P' can be obtained from the path P with the help of Lemma 5.21. Find the smallest index i, such that either $\xi(p_{i+1}) < \xi(p_i)$ or $\eta(p_{i+1}) < \eta(p_i)$. Without loss of generality we can assume that $\xi(p_{i+1}) < \xi(p_i)$. In each step of the path the ξ coordinate either increases by 1, decreases by 1 or stays the same. The path has to reach the upper-right corner eventually, so there exists an index $j > i$, such that $\xi(p_j) = \xi(p_i)$. By Lemma 5.21, the subpath $p_{i+1}, \ldots, p_{j-1}$ can be replaced with a straight line segment from p_i to p_j that lies in $\mathcal{B}$. We can iteratively use the above operation until we finally obtain the path P' with the desired property.

The path P' that we obtained can be described as a sequence of symbols $\rightarrow$ and $\nearrow$ which indicate the increase in the ξ and in the η coordinate, respectively. One of such paths in the example from Figure 5.30 is $(\rightarrow, \nearrow, \rightarrow, \rightarrow, \nearrow)$. Note that every possible shortest path in $I_{\mathcal{H}}(a, b)$ can be described as a permutation of these symbols. The Small Parallelogram Rule implies that we may swap two consecutive symbols in the sequence and thus obtain another path which also lies in $\mathcal{B}$. From $(\rightarrow, \nearrow, \rightarrow, \rightarrow, \nearrow)$ we can obtain, say, $(\rightarrow, \rightarrow, \nearrow, \rightarrow, \nearrow)$. It is clear that any permutation can be obtained in this way which implies that the whole interval $I_{\mathcal{H}}(a, b)$ is contained in $\mathcal{B}$. □

Proposition 5.22. *A benzenoid $\mathcal{B}$ is* convex *if and only if its boundary-edges code $BEC(\mathcal{B})$ does not contain symbol* `1`.

Example 5.4. The distinction between convex and non-convex benzenoids is visible from their boundary-edges codes. For instance, the boundary-edges code of the benzenoid $\mathcal{B}_1$ on Figure 5.31(a) is `24334` and does not contain symbol `1`. Therefore this benzenoid is convex,



(a) $BEC(\mathcal{B}_1) =$ `24334`, convex (b) $BEC(\mathcal{B}_2) =$ `144144`, non-convex

Figure 5.31: The distinction between convex and non-convex benzenoids is visible from their boundary-edges codes.

whilst, the boundary-edges code for the benzenoid in Figure 5.31(b) is `144144`, containing a `1`. Hence the benzenoid itself is non-convex. See Figure 5.31. □

Proof. Let $\mathcal{B}$ be a convex benzenoid. We will prove that there is no `1` in its boundary-edges code. Suppose that there is one. Then that part of the boundary of $\mathcal{B}$ that corresponds to the symbol `1` in its boundary-edges code locally looks as shown in Figure 5.32. There exist three hexagons a, b and c which are positioned as indicated in the Figure 5.32. This means that the Small Parallelogram Rule is not obeyed, which contradicts the fact that $\mathcal{B}$ is convex. Thus, there can be no `1` in its boundary-edges code.

Let $\mathcal{B}$ be a benzenoid such that there is no symbol `1` in the boundary-edges code of $\mathcal{B}$. Assume that $\mathcal{B}$ is not convex. By Proposition 5.20, it does not obey the Small Parallelogram

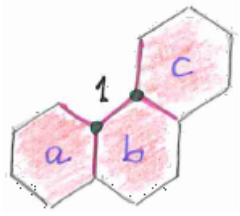


Figure 5.32: The part of boundary corresponding to symbol 1 in the boundary-edges code.

Rule. There must exist three hexagons in $\mathcal{B}$ which are positioned in the phenanthrene shape as shown in Figure 5.27(b) and where the hexagon labeled with c is not present (see Figure 5.27(b)). Then the corresponding part of the boundary contains two consecutive degree-3 vertices, i.e., there is a 1 in its boundary-edges code. This is a contradiction, thus benzenoid $\mathcal{B}$ is convex. □

We also propose the following definition:

Definition 5.18. *A benzenoid $\mathcal{B}$ is* pseudo-convex *if there are no two consecutive 1s in its boundary-edges code.*

In addition to finite benzenoids we will need (for the purpose of theoretical reasoning) some infinite ones:

Definition 5.19. *An infinite connected hexagonal system $\mathcal{K}$ whose complement $\mathcal{K}^{\complement}$ consists of infinite connected components is called an* infinite benzenoid.

Theorem 5.23. *There are uncountably many mutually non-equivalent infinite benzenoids.*

Proof. The interval $[0, 1)$ has the cardinality of the continuum. Each number $x \in [0, 1)$ can be written in its binary representation

$$0.x_1x_2x_3x_4\ldots \tag{5.29}$$

Note that $x_i \in \{0, 1\}$ for each $i \in \mathbb{N}$ and that

$$x = \sum_{i=1}^{\infty} x_i \cdot 2^{-i}. \tag{5.30}$$

The sequence $\{x_i\}_{i=1}^{\infty}$ is uniquely determined if we require that for each $n \in \mathbb{N}$ there exists an integer $m > n$ such that $x_m \neq 1$. For example, the number $\frac{11}{16}$ can be written in binary representation as 0.1011.

We will assign an infinite benzenoid to every such sequence $\{x_i\}_{i=1}^{\infty}$. Define

$$\mathcal{P}_x = \{h \in \mathcal{H} \mid \eta(h) \leq 0\} \setminus (\{(-2, 0), (-1, 0)\} \cup \{(2i - 1, 0) \mid i \in \mathbb{N} \wedge x_i = 1\}).$$

The infinite benzenoid $\mathcal{P}_x$ is obtained from the half-plane $\mathcal{P} = \{h \in \mathcal{H} \mid \eta(h) \leq 0\}$ by removing hexagons $(-2, 0)$, $(-1, 0)$ and all hexagon $(2i - 1, 0)$ where $i \in \mathbb{N}$ and $x_i = 1$. For example, the infinite benzenoid that corresponds to number $\frac{11}{16}$ is shown in Figure 5.33.

It is not hard to see that $\mathcal{P}_x \ncong \mathcal{P}_y$ if $x \neq y$. We constructed an injective mapping from the set $[0, 1)$ to the class of infinite benzenoids. Therefore, the class of all infinite benzenoids is uncountable. □

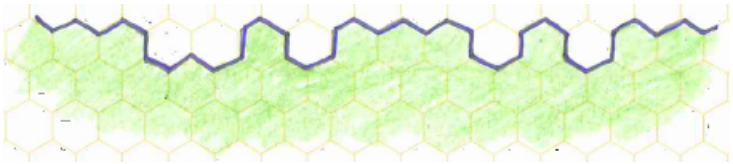
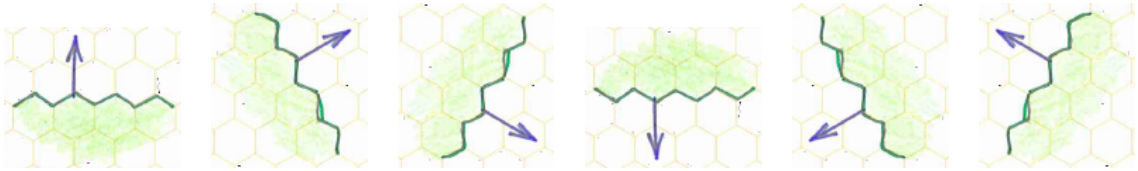


Figure 5.33: Infinite benzenoid $\mathcal{P}_{\frac{11}{16}}$.

Theorem 5.27 means that one cannot describe all of the infinite benzenoids algorithmically. This means that there exist infinite benzenoids for which there does not exist a finite program for constructing them (even if the program runs infinitely long).

Let us consider infinite convex benzenoids. The whole hexagonal grid $\mathcal{H}$ and the empty set $\emptyset$ are clearly convex. A hexagonal half-plane is also convex. Each half-plane has a



(a) $\mathcal{HP}^-_\eta(n)$ (b) $\mathcal{HP}^-_\zeta(n)$ (c) $\mathcal{HP}^-_\xi(n)$ (d) $\mathcal{HP}^+_\eta(n)$ (e) $\mathcal{HP}^+_\zeta(n)$ (f) $\mathcal{HP}^+_\xi(n)$

Figure 5.34: The normal of a hexagonal half-plane can point in 6 different directions.

normal which is pointing out of the half-plane. There are exactly 6 possible directions of half-planes (see Figure 5.34). We will denote them as follows:

$$\mathcal{HP}^+_\xi(n) = \{h \in \mathcal{H} \mid \xi(h) \geq n\}, \tag{5.31}$$
$$\mathcal{HP}^-_\xi(n) = \{h \in \mathcal{H} \mid \xi(h) \leq n\}, \tag{5.32}$$
$$\mathcal{HP}^+_\eta(n) = \{h \in \mathcal{H} \mid \eta(h) \geq n\}, \tag{5.33}$$
$$\mathcal{HP}^-_\eta(n) = \{h \in \mathcal{H} \mid \eta(h) \leq n\}, \tag{5.34}$$
$$\mathcal{HP}^+_\zeta(n) = \{h \in \mathcal{H} \mid \zeta(h) \geq n\}, \tag{5.35}$$
$$\mathcal{HP}^-_\zeta(n) = \{h \in \mathcal{H} \mid \zeta(h) \leq n\}. \tag{5.36}$$

Note that the intersection of two half-planes with the same normal is equal to one of the two. For example, $\mathcal{HP}^+_\xi(3) \cap \mathcal{HP}^+_\xi(5) = \mathcal{HP}^+_\xi(5)$. Therefore, the interection of any number of half-planes is equal to an intersection of at most 6 of those half-planes. For each direction of the normal that is present in the list, we select the half-plane that is contained in all other half-planes that have the same direction.

Definition 5.20. *Let $\mathcal{B}$ be a benzenoid (finite or infinite). The smallest convex benzenoid containing $\mathcal{B}$ is called the* convex closure *of $\mathcal{B}$ and is denoted $Conv(\mathcal{B})$.*

Proposition 5.24. *Any intersection of convex (finite or infinite) benzenoids is a convex benzenoid.*

Note that the empty set $\emptyset$ can also be considered as a convex benzenoid. To a chemist, this means nothing concrete. To a mathematician, it is clear that every two members of an empty set satisfy the condition in Definition 5.15, i.e., this condition is void.

Proof. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two convex benzenoids. If their intersection is $\emptyset$ then there is nothing to prove. Suppose that $\mathcal{B}_1 \cap \mathcal{B}_2 \neq \emptyset$. Let $a, b \in \mathcal{B}_1 \cap \mathcal{B}_2$. Because $\mathcal{B}_1$ is convex, we have that $I_{\mathcal{H}}(a,b) \subseteq \mathcal{B}_1$. Similarly, because $\mathcal{B}_2$ is convex we have that $I_{\mathcal{H}}(a,b) \subseteq \mathcal{B}_2$. But this means that $I_{\mathcal{H}}(a,b) \subseteq \mathcal{B}_1 \cap \mathcal{B}_2$, which completes the proof. □

Proposition 5.25. *A benzenoid $\mathcal{B}$ is convex if and only if it can be obtained as an intersection of half-planes, i.e., if there exist integers n_ξ^+, n_ξ^-, n_η^+, n_η^-, n_ζ^+ and n_ζ^-, such that*

$$\mathcal{B} = \mathcal{HP}_\xi^+(n_\xi^+) \cap \mathcal{HP}_\xi^-(n_\xi^-) \cap \mathcal{HP}_\eta^+(n_\eta^+) \cap \mathcal{HP}_\eta^-(n_\eta^-) \cap \mathcal{HP}_\zeta^+(n_\zeta^+) \cap \mathcal{HP}_\zeta^-(n_\zeta^-). \tag{5.37}$$

Proof. If a benzenoid $\mathcal{B}$ is obtained as an intersection of half-planes then it is convex by Proposition 5.24.

Now let $\mathcal{B}$ be a convex benzenoid. We will show that it is equal to the intersection of 6 half-planes. Let h_η^+ and h_η^- be hexagons of $\mathcal{B}$ (see Figure 5.35) such that

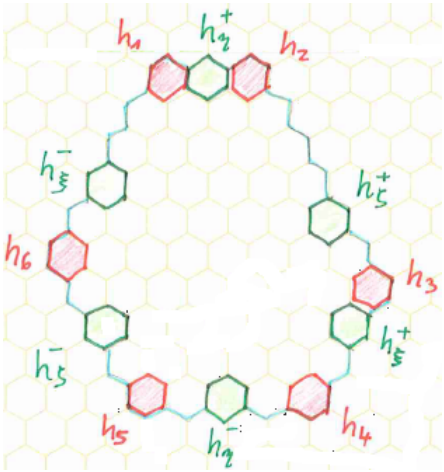


Figure 5.35: A convex benzenoid $\mathcal{B}$ from the proof of Proposition 5.25.

$$\eta(h_\eta^+) = \max_{h \in \mathcal{B}} \eta(h) \text{ and} \tag{5.38}$$

$$\eta(h_\eta^-) = \min_{h \in \mathcal{B}} \eta(h). \tag{5.39}$$

Similarly, let h_ξ^+, h_ξ^-, h_ζ^+ and h_ζ^- be hexagons of $\mathcal{B}$ such that

$$\xi(h_\xi^+) = \max_{h \in \mathcal{B}} \xi(h), \tag{5.40}$$

$$\xi(h_\xi^-) = \min_{h \in \mathcal{B}} \xi(h), \tag{5.41}$$

$$\eta(h_\zeta^+) = \max_{h \in \mathcal{B}} \zeta(h) \text{ and} \tag{5.42}$$

$$\eta(h_\zeta^-) = \min_{h \in \mathcal{B}} \zeta(h). \tag{5.43}$$

Let $h_1, h_2, \ldots, h_6$ be hexagons of $\mathcal{H}$ such that:

(i) $\eta(h_1) = \eta(h_\eta^+)$, $\xi(h_1) = \xi(h_\xi^-)$;

(ii) $\eta(h_2) = \eta(h_\eta^+)$, $\zeta(h_2) = \zeta(h_\zeta^+)$;

(iii) $\zeta(h_3) = \zeta(h_\zeta^+)$, $\xi(h_3) = \xi(h_\xi^+)$;

(iv) $\eta(h_4) = \eta(h_\eta^-)$, $\xi(h_4) = \xi(h_\xi^+)$;

(v) $\eta(h_5) = \eta(h_\eta^-)$, $\zeta(h_5) = \zeta(h_\zeta^-)$;

(vi) $\zeta(h_6) = \zeta(h_\zeta^-)$, $\xi(h_6) = \xi(h_\xi^-)$.

As $h_1 \in I_{\mathcal{H}}(h_\eta^+, h_\xi^-)$ it follows that $h_1 \in \mathcal{B}$. By analogy, $h_2, \ldots, h_6 \in \mathcal{B}$. Note that h_i and h_{i+1} (indices taken modulo 6) are on the same line, thus the line segment between h_i and h_{i+1} is also contained in $\mathcal{B}$. All hexagons inside the 'polygon' defined by $h_1, \ldots, h_6$ must belong to $\mathcal{B}$ or there would be a corona hole in $\mathcal{B}$. If any hexagon outside that polygon belonged to $\mathcal{B}$ that would contradict one of equations (5.38), (5.39), (5.40), (5.41), (5.42) or (5.43). It follows that

$$\mathcal{B} = \mathcal{HP}_\xi^+(n_\xi^-) \cap \mathcal{HP}_\xi^-(n_\xi^+) \cap \mathcal{HP}_\eta^+(n_\eta^-) \cap \mathcal{HP}_\eta^-(n_\eta^+) \cap \mathcal{HP}_\zeta^+(n_\zeta^-) \cap \mathcal{HP}_\zeta^-(n_\zeta^+). \tag{5.44}$$

□

Proposition 5.26. *An infinite benzenoid $\mathcal{B}$ is convex if and only if it can be obtained as an intersection of half-planes.* □

The proof of Proposition 5.26 is similar to the proof of Proposition 5.25 and is therefore omitted. The only difference is that some of the half-planes are not present. Therefore, for infinite convex benzenoids we have the following theorem:

Theorem 5.27. *There exist only countably many non-equivalent convex infinite benzenoids.*

Proof. By using Proposition 5.26 we can classify the infinite convex hexagons (up to symmetries of the hexagonal grid) into several sporadic cases and countable families (see Figure 5.36). The half-plane, the anvil and the wedge are sporadic cases. The strip is a family parametrized with an integer $n \geq 1$. The chomped wedge is parametrized with an integer $n \geq 2$. The knife is parametrized with an integer $n \geq 2$. The sword is parametrized with two integers n and m such that $m \geq n \geq 2$. □

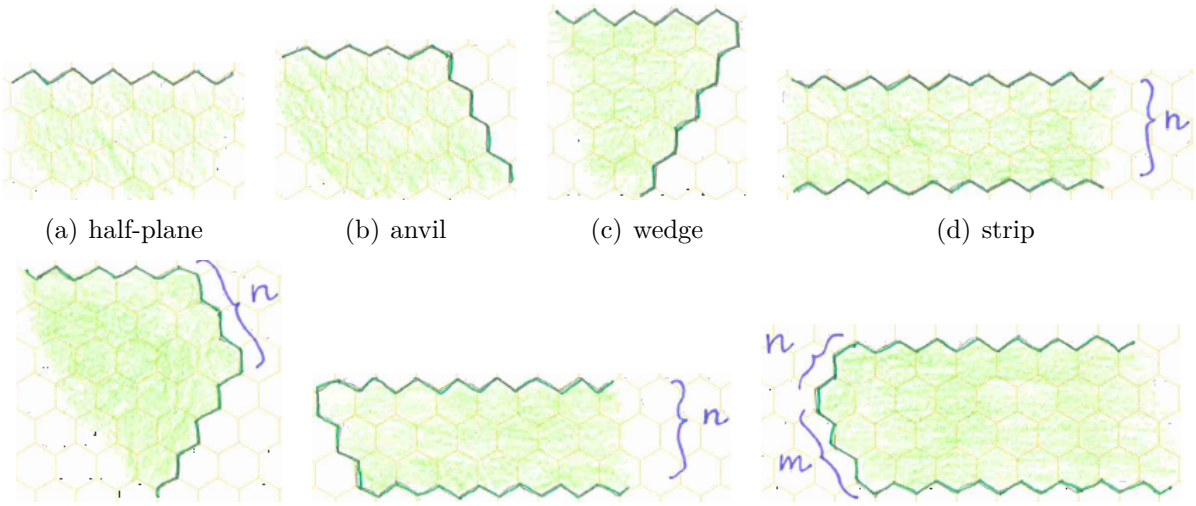
In a smiliar manner, by using Proposition 5.25, the finite convex benzenoids can be classified into six mutually disjoint families that we call *fundamental families of convex benzenoids*:

(a) The *linear chain* $L(n)$, $n \geq 2$, having n hexagons. The corresponding boundary-edges code is

$$52^{n-2}52^{n-2}.$$

(b) The *equilateral triangle* $T_3(n)$, $n \geq 2$, having n hexagons on each of its sides. The corresponding boundary-edges code is

$$42^{n-2}42^{n-2}42^{n-2}.$$



(a) half-plane (b) anvil (c) wedge (d) strip

(e) chomped wedge (f) knife (g) sword

Figure 5.36: The classification of infinite convex benzenoids.

(c) The *equilateral trapezoid* $T_4(n, m)$, $n > m \geq 2$, having n hexagons on the bottom base and m hexagons on the top base. The corresponding boundary-edges code is

$$42^{n-2}42^{n-m-1}32^{m-2}32^{n-m-1}.$$

(d) The *rhomboid* $R(n, m)$, $m \geq n \geq 2$, having n hexagons on its base and m hexagons on its side. The corresponding boundary-edges code is

$$42^{n-2}32^{m-2}42^{n-2}32^{m-2}.$$

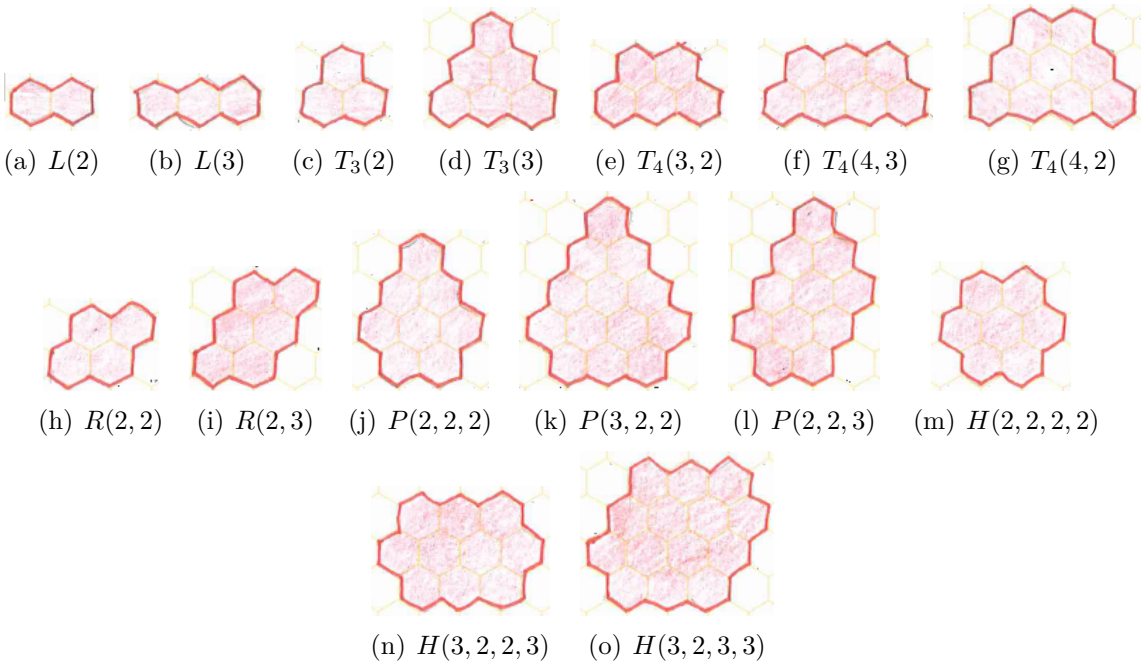
(e) The *pentagonal benzenoid* $P(n, m, k)$, $n \geq 2$, $k \geq m \geq 2$ having n hexagons on the base, m hexagons on its left side and k hexagons on its right side. The corresponding boundary-edges code is

$$32^{n-2}32^{k-1}32^{m+n-3}42^{n+k-3}32^{m-2}.$$

(f) The *hexagonal benzenoid* $H(n, m, k, t)$, $n \geq 2$, $k \geq m \geq 2$, $n + m - 2 \geq t \geq 2$, having n hexagons on the base, m hexagons on its left side, k hexagons on its right side and t hexagons on its top side. The corresponding boundary-edges code is

$$32^{n-2}32^{m-2}32^{n+k-t-2}32^{t-2}32^{n+m-t-2}32^{k-2}.$$

Note that an exponent n in the boundary-edges code means that the corresponding symbol repeats n times, e.g. $2^5 = 22222$. This is a compact way of writing the code. Benzene is also a convex benzenoid and is not included in any of the above families. We treat it as a separate sporadic case.



(a) $L(2)$ (b) $L(3)$ (c) $T_3(2)$ (d) $T_3(3)$ (e) $T_4(3,2)$ (f) $T_4(4,3)$ (g) $T_4(4,2)$

(h) $R(2,2)$ (i) $R(2,3)$ (j) $P(2,2,2)$ (k) $P(3,2,2)$ (l) $P(2,2,3)$ (m) $H(2,2,2,2)$

(n) $H(3,2,2,3)$ (o) $H(3,2,3,3)$

Figure 5.37: The smallest members of families of convex benzenoids.

In all of the above families, except for $H(n,m,k,t)$, the benzenoid is uniquely determined by the described parameters. Note that

$$\begin{aligned} H(n,m,k,t) &\cong H(n+m-t, \min\{k,t\}, \max\{k,t\}, m) \\ &\cong H(n+k-t, \min\{m,t\}, \max\{m,t\}, k) \\ &\cong H(t, n+m-t, n+k-t, n) \\ &\cong H(m, \min\{n, n+k-t\}, \max\{n, n+k-t\}, n+m-t) \\ &\cong H(k, \min\{n, n+m-t\}, \max\{n, n+m-t\}, n+k-t). \end{aligned} \tag{5.45}$$

The above parametrisations are all valid and represent equivalent benzenoids. The set of parameters that is lexicographically the smallest of the above is called the *canonical parametrisation*. We obtain the following proposition:

Proposition 5.28. *A (finite) benzenoid $\mathcal{B}$ is convex if and only if it is equivalent to a member of one of the families: $L(n)$, $T_3(n)$, $T_4(n)$, $R(n,m)$, $P(n,m,k)$ or $H(n,m,k,t)$.* □

From Proposition 5.28 we can obtain a simple algorithm that can count (and enumerate) all convex benzenoids with $h \leq N$ hexagons for a given N. We create an array T of length N. In $T[h]$ we will keep the number of convex benzenoids with h hexagons. We iterate over all families and over all admissible parameters for which the number of hexagons remains below N. (It is easy to derive an expression for the number of hexagons in terms of the parameters for each of the families of convex benzenoids.) For each benzenoid we determine the number of hexagons, h, that it contains and increment the value of $T[h]$ by one. Some care should be taken with the hexagonal family, since we have to check whether the given parameters are indeed canonical.

By using the above procedure, we were able to extend the sequence A116513 [198] into the millions. The numbers of convex benzenoids with h hexagons for $h \leq 50$ are given in the Table 5.1. Benzene is excluded from this table.

h	L	T_3	R	T_4	P	H	Σ	h	L	T_3	R	T_4	P	H	Σ
1	0	0	0	0	0	0	0	26	1	0	1	1	2	2	7
2	1	0	0	0	0	0	1	27	1	0	1	3	2	1	8
3	1	1	0	0	0	0	2	28	1	1	2	0	0	4	8
4	1	0	1	0	0	0	2	29	1	0	0	1	4	2	8
5	1	0	0	1	0	0	2	30	1	0	3	3	1	2	10
6	1	1	1	0	0	0	3	31	1	0	0	1	1	3	6
7	1	0	0	1	0	1	3	32	1	0	2	0	3	2	8
8	1	0	1	0	1	0	3	33	1	0	1	3	2	4	11
9	1	0	1	2	0	0	4	34	1	0	1	1	3	4	10
10	1	1	1	0	0	1	4	35	1	0	1	3	3	1	9
11	1	0	0	1	1	0	3	36	1	1	4	1	1	4	12
12	1	0	2	1	0	1	5	37	1	0	0	1	2	3	7
13	1	0	0	1	1	1	4	38	1	0	1	1	2	5	10
14	1	0	1	1	1	1	5	39	1	0	1	3	5	2	12
15	1	1	1	2	1	0	6	40	1	0	3	1	0	5	10
16	1	0	2	0	0	2	5	41	1	0	0	1	4	2	8
17	1	0	0	1	2	0	4	42	1	0	3	3	2	4	13
18	1	0	2	2	0	2	7	43	1	0	0	1	3	6	11
19	1	0	0	1	2	2	6	44	1	0	2	1	4	4	12
20	1	0	2	1	1	1	6	45	1	1	2	4	2	3	13
21	1	1	1	2	1	1	7	46	1	0	1	1	2	5	10
22	1	0	1	1	1	2	6	47	1	0	0	1	4	4	10
23	1	0	0	1	2	2	6	48	1	0	4	1	3	6	15
24	1	0	3	1	2	2	9	49	1	0	1	2	4	4	12
25	1	0	1	2	1	2	7	50	1	0	2	2	3	5	13

Table 5.1: The enumeration of convex benzenoids.

5.7 Kekulé structures of hexagonal systems

In this section, we deal with counting and enumeration of Kekulé structures in benzenoids. We start with characterisation of Kekulé structures by perfect path systems. This is an important theoretical result and is widely used by theoretical chemists. For some families of benzenoids, e.g. convex benzenoids, we give formulae in closed form. Then we present an algorithm by H. Sachs [181] which can determine in linear time whether a benzenoid admits a perfect matching or not. Much research in this area was done in the past owing to the importance of benzenoids and Kekulé structures in theoretical chemistry. The monograph [53] by Cyvin and Gutman is dedicated exclusively to this subject.

5.7.1 Perfect path systems and perfect matchings

Here, we present an important theoretical result that characterises Kekulé structures in benzenoids (and also coronoids) by perfect path systems [181].

Definition 5.21. *Let $\mathcal{B}$ be a coronoid embedded in the infinite hexagonal grid $\mathcal{H}$. A vertex of $\mathcal{B}$ is called a* peak *if all of its neighbours have strictly smaller y-coordinates. A vertex of $\mathcal{B}$ is called a* valley *if all of its neighbours have strictly larger y-coordinates.*

The coronoid $\mathcal{C}$ in Figure 5.38 has 4 peaks labeled $p_1, \ldots, p_4$ and 4 valleys labeled $v_1, \ldots, v_4$. Note that every coronoid has at least one peak and at least one valley. Let $\mathcal{C}$ be a coronoid.

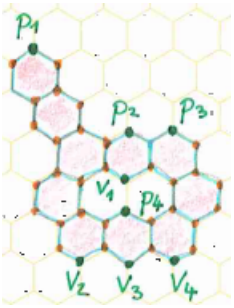


Figure 5.38: A coronoid $\mathcal{C}$ with 4 peaks and 4 valleys.

With $p(\mathcal{C})$ we denote the number of peaks and $v(\mathcal{C})$ is the number of valleys of $p(\mathcal{C})$.

Definition 5.22. *A* peak-to-valley path *(also called* monotonic path*) of a coronoid (with a fixed embedding) is a path connecting a peak to a valley such that y-coordinates of the vertices decrease at each step while traversing the path from a peak to a valley.*

Definition 5.23. *A* perfect path system *(PPS) is a collection of disjoint peak-to-valley paths, connecting all peaks and all valleys.*

Example 5.5. An example of a benzenoid with two peak-to-valley paths is shown on Figure 5.39. Because all peaks and all valleys are connected and the two paths are disjoint,

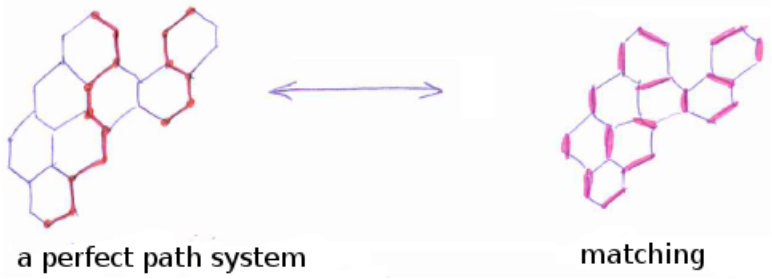


Figure 5.39: A perfect path system and the corresponding perfect matching in a benzenoid.

they form a perfect path systems. If we take all non-vertical edges that lie on peak-to-valley path and all vertical edges that do not lie on peak-to-valley paths we obtain a perfect matching. The corresponding perfect matching is shown on the right.

Theorem 5.29. *In every coronoid $\mathcal{C}$ (with fixed embedding) there is a 1-to-1 correspondence between the set of perfect matchings and the set of perfect path systems of $\mathcal{C}$.*

Proof. Let $\mathcal{C}$ be a coronoid and F one of its perfect matchings. Pick a valley b_0. Exactly one of its incident edges, say e_0, is in F. Move along edge e_0 to its other end vertex w_0. (Recall that benzenoids are bipartite graphs and their vertices can be coloured, say, black and white. The vertices that correspond to valleys are coloured black.) Vertex w_0 is white and there are two possibilities:

(i) It is a peak, in which case we just found a peak-to-valley path.

(ii) It is not a peak, which means that there is a black vertex b_1 that is connected to w_0 with a vertical edge f_0. Edge f_0 is incident to e_0, thus $f_0 \notin F$. Because of that, b_1 has a neighbour w_1, $w_1 \neq w_0$, such that $b_1w_1 \in F$.

We continue with this procedure until we eventually reach a peak w_k. This will happen, because the y-coordinates of the vertices $b_0, w_0, b_1, w_1, \ldots$ increase all the time and there are finitely many of them. The path $b_0\, w_0\, b_1\, w_1 \ldots b_k\, w_k$ is a peak-to-valley path.

It is also possible to reverse this process, i.e., start at a peak (when the coronoid is turned upside-down, peaks become valleys and valleys become peaks) and find a path to a valley. Any vertex which is an end vertex of some non-vertical edge in F lies on a peak-to-valley path. This path can be uniquely determined by using the above procedure (in both directions). Because every such vertex belongs to precisely one peak and one valley, those paths are necessarily disjoint. This also means that a PPS is uniquely determined by the perfect matching F. Note that end vertices of vertical edges from F do *not* participate in the PPS. Let $\Phi(F)$ denote the PSS obtained from F.

To show that mapping Φ is injective, consider two different perfect matchings F and F'. There exists at least one vertex v in $\mathcal{C}$ with the property that $u \neq u'$ where $vu \in F$ and $vu' \in F'$. If edges vu and vu' are both non-vertical, then the peak-to-valley path that passes through v in $\Phi(F)$ is necessarily different from the peak-to-valley path that passes through v in $\Phi(F')$. If they are not both non-vertical then we may, without loss of generality, assume that uv is a vertical edge. In that case v does not belong to any peak-to-valley path of $\Phi(F)$ and it does belong to some peak-to-valley path of $\Phi(F')$. Therefore, $\Phi(F) \neq \Phi(F')$.

To show that the mapping Φ is surjective, we have to show that every PPS can be obtained from a perfect matching. Pick any PPS. The vertices of $\mathcal{C}$ can be partitioned into two classes: those that belong to some peak-to-valley path and those that do not. Vertices that do not belong to any peak-to-valley path occur in adjacent pairs. Every such pair of vertices is connected by a vertical edge. If we take all non-vertical edges on all peak-to-valley paths and all vertical edges that are not incident with any vertex of the perfect path system then we obtain a perfect matching. □

From Theorem 5.29 we immediately obtain:

Corollary 5.30. *A coronoid $\mathcal{C}$ is Kekulean if and only if it admits a perfect path system.* □

5.7.2 Kekulé structures in fibonacenes and convex benzenoids

In the Section 5.6 we introduced the six fundamental families of convex benzenoids. The simplest case is the linear chain $L(h)$:

Proposition 5.31. *The linear chain $L(h)$ with $h \geq 1$ hexagons admits $h+1$ Kekulé structures.*

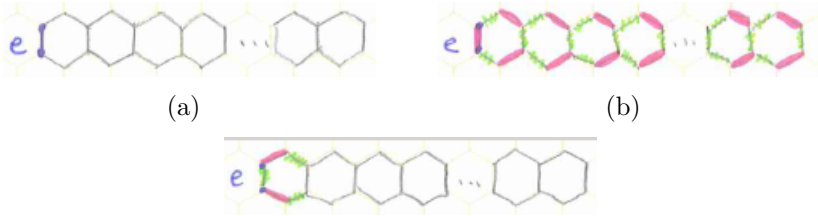
Here, we allow parameter $h = 1$ to also include the benzene. The proof is an easy exercise in mathematical induction:

Proof. Kekulé structures of benzene, $L(1)$, and naphthalene, $L(2)$, can be manually enumerated. They are shown in Figure 5.40. It is clear that $L(1) = 2$ and $L(2) = 3$.

(a) (b)

Figure 5.40: Kekulé structures of benzene (a) and naphthalene (b).

Now consider the linear chain $L(h)$ with $h \geq 3$. By the induction hypothesis, $K(L(h')) = h'+1$ for $h' < h$. The chain $L(h)$ has 2 hexagons in the L_1 mode and $h-2$ hexagons in the L_2 mode (see Figure 5.41(a)). Let e be the edge that is marked on the figure. It is clear that $K(L(h)) = K_e + K_e^*$, where K_e denotes the number of Kekulé structures that include e and K_e^* denotes the number of Kekulé structures without the edge e. It is easy to see that



(a) (b)

(c)

Figure 5.41: Kekulé structures of the linear chain $L(h)$ composed of h hexagons.

the Kekulé structure is uniquely determined if the edge e is present (see Figure 5.41(b)), thus $K_e = 1$. If the edge e is not present, then only a part of the structure is determined (see Figure 5.41(c)). The red edges are present in every such Kekulé structure and green edges are not, thus $K_e^* = K(L(h-1)) = h$. Therefore, $K(L(h)) = 1 + h$. □

Our next example is more interesting. Let us first define a family of benzenoids called *fibonacenes*:

Definition 5.24. *A* fibonacene *is an unbranched catacondensed benzenoid in which every hexagon which is adjacent to two other hexagons is in A_2 mode.*

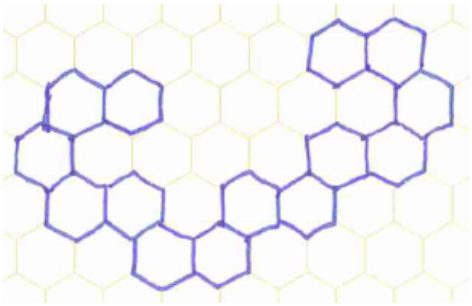


Figure 5.42: An example of a fibonacene.

Note that if $h \geq 2$, there are exactly 2 hexagons in L_1 mode in a fibonacene. Benzene is also a fibonacene. An example is shown in Figure 5.42.

The Fibonacci sequence [64] is probably the most well-known sequence amongst both professional and "recreational" mathematicians. It is defined by the recurrence relation

$$F_n = F_{n-1} + F_{n-2} \tag{5.46}$$

with the starting values $F_0 = 0$ and $F_1 = 1$. The first few terms are:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, \ldots$$

The term F_n of the Fibonacci sequence is called the n-th *Fibonacci number*. A closed-form formula for the n-th Fibonacci number is also well-known:

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n. \tag{5.47}$$

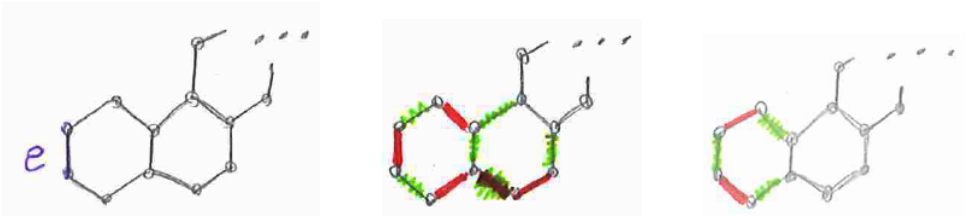
The following proposition justifies the name "fibonacene" for this family of benzenoids:

Proposition 5.32. *A fibonacene with h hexagons admits F_{h+2} Kekulé structures.*

Proof. The proof goes by induction on the number of hexagons in the fibonacene. We have already seen that benzene has $F_{1+2} = 2$ Kekulé structures and naphthalene has $F_{2+2} = 3$ Kekulé structures.

Take any fibonacene $\mathcal{F}$ with $h \geq 3$ hexagons. By the induction hypothesis, every fibonacene with $h' < h$ hexagons admits $F_{h'+2}$ Kekulé structures. Benzenoid $\mathcal{F}$ has at least one hexagon, say a, in L_1 mode. Let b denote the hexagon which is adjacent to a. Up to isometries of $\mathcal{H}$, there is only one possible arrangement of the first three hexagons (see Figure 5.43(a)). Let e be the edge of hexagon a as shown in the figure. Let K_e denote the number of Kekulé structures that include edge e and let K_e^* denote the number of Kekulé structures without the edge e. If the edge e is present, then a part of the Kekulé structure is uniquely determined (see Figure 5.43(b)). The situation is similar when the edge e is not present in the Kekulé structure, as shown in Figure 5.43(c). It is clear that $K_e = K(\mathcal{F} \setminus \{a, b\}) = F_h$ and $K_e^* = K(\mathcal{F} \setminus \{a\}) = F_{h+1}$. The total number of Kekule structures of $\mathcal{F}$ is $K(\mathcal{F}) = K_e + K_e^* = F_h + F_{h+1} = F_{h+2}$. □

Let us now consider the rhomboid family $R(h, k)$, $h \geq 1$, $k \geq 1$, of convex benzenoids:



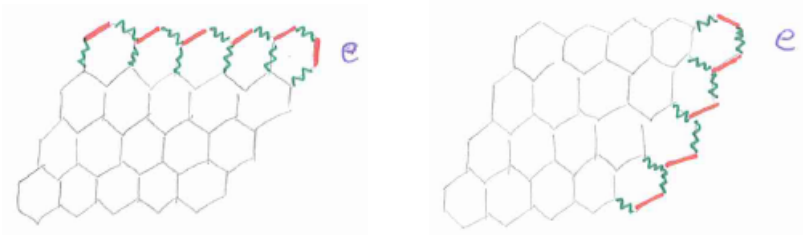
(a) (b) (c)

Figure 5.43: Arrangement of the first three hexagons in a fibonacene.

Proposition 5.33. *The rhomboid $R(h,k)$, having h hexagons on the base and k on its side, admits $\binom{h+k}{h}$ Kekulé structures.*

Proof. The proof goes by induction on $h+k$. Note that $R(h,1) \cong R(1,h) \cong L(h)$. Therefore $K(R(h,1)) = K(L(h)) = h+1 = \binom{h+1}{h} = \binom{h+1}{1}$ which establishes the base of the induction.

Now consider a rhomboid with $h,k \geq 2$. Let e be its top-most right-most vertical edge (see Figure 5.44(a)). As in the previous proof, we may write $K(R(h,k)) = K_e + K_e^*$. From



(a) (b)

Figure 5.44: Kekulé structures of the rhomboid $R(h,k)$.

Figures 5.44(a) and 5.44(b) it is clear that $K_e = K(R(h,k-1))$ and $K_e^* = K(R(h-1,k))$. It follows that $K(R(h,k)) = K(R(h,k-1)) + K(R(h-1,k)) = \binom{h+k-1}{h} + \binom{h+k-1}{h-1} = \binom{h+k}{h}$. □

Note that benzenoids $T_3(n)$, $n \geq 2$, $T_4(n,m)$, $n > m \geq 2$, and $P(n,m,k)$, $n \geq 2$, are non-Kekulean. This immediately follows from Corollary 5.30. The equilateral triangle $T_3(n)$ has $n \geq 2$ valleys, but only one peak. The equilateral trapezoid $T_4(n,m)$ has m peaks and n valleys, but $n > m$. The pentagonal benzenoid $P(n,m,k)$ has $n \geq 2$ valleys, but only one peak. It is clear that a PPS cannot exist in any of those cases.

The case of the hexagonal benzenoid $H(n,m,k,t)$ is more involved. Note that it has n valleys and t peaks. By Corollary 5.30, $n = t$ is a necessary condition for $H(n,m,k,t)$ to be Kekulean. Let us define

$$O(n,m,k) = H(n,m,k,n). \tag{5.48}$$

The equation for obtaining the Kekulé number of the hexagonal benzenoid was already considered in the early papers by Gordon and Davison [83] and Yen [201]. Gordon and Davison gave the equation

$$K(O(n,n,n)) = \prod_{i=0}^{n-1} \frac{\binom{2n+i}{n}}{\binom{n+i}{n}}, \tag{5.49}$$

for the special case where $n = k = m$. Equation (5.49) was attributed to M. R. Everett [83]. They also reported a more general equation

$$K(O(m,m,n)) = \prod_{i=0}^{m-1} \frac{\binom{m+n+i}{n}}{\binom{n+i}{n}} \tag{5.50}$$

which was deduced by M. Woodger. Unfortunately, no proofs of the above equations were given in the paper [83]. Later, Cyvin generalised the equation (5.50) to the general case:

$$K(O(k,m,n)) = \prod_{i=0}^{k-1} \frac{\binom{m+n+i}{n}}{\binom{n+i}{n}}, \tag{5.51}$$

but gave no proof. The first proof was published by Bodroža et al. [18]. We will follow their approach. Equation (5.51) is called the *Everett-Woodger-Cyvin formula* [18]. Note that the formula for the number of Kekulé structure of the rhomboid $R(n,m)$ can be obtained from (5.51) by setting $k = 0$.

The intersection graph G_{ij} of the i-th peak and the j-th valley of $\mathcal{B}$ is a subgraph that is spanned on vertices that are accessible both from peak p_i by going downwards and from valley v_j by going upwards.

Example 5.6. Consider the benzenoid in Figure 5.45. It has three peaks labeled p_1, p_2 and p_3. Its three valleys are labeled v_1, v_2 and v_3. The intersection graph $G_{3,1}$ is isomorphic

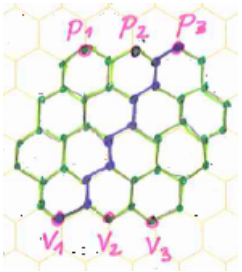


Figure 5.45: A benzenoid where $G_{3,1} \cong P_{10}$.

to the path on 10 vertices. This special case where the intersection graph is a path on an even number of vertices can be treated as "degenerated" rhomboid $R(n,0)$. □

The following equation was discovered by John and Sachs:

$$K(\mathcal{B}) = |\det W|, \tag{5.52}$$

where W is an $n \times n$ matrix defined as $W = [K(G_{ij})]_{i,j=1}^{n}$, where n is the number of peaks (or valleys). Equation (5.52) is called the *John-Sachs formula* [101]. If G_{ij} is an empty graph then $K(G_{ij}) = 0$. If peaks and valleys are labeled so that p_{i+1} is to the right of p_i and v_{i+1} is to the right of v_i then equation (5.52) can be written as

$$K(\mathcal{B}) = \det W. \tag{5.53}$$

It is easy to see that $G_{ij} \cong R(m+i-j, k-i+j)$ and consequently $K(G_{ij}) = \binom{m+k}{m+i-j}$. By using the John-Sachs formula (5.53) we obtain:

$$K(O(k,m,n)) = \left|\binom{m+k}{m+i-j}\right|_{i,j=1}^{n}. \tag{5.54}$$

In the above expression $\left|\binom{m+k}{m+i-j}\right|_{i,j=1}^{n}$ denotes the determinant of the matrix $\left[\binom{m+k}{m+i-j}\right]_{i,j=1}^{n}$.

Lemma 5.34 (Bodroža et al., Lemma 1 in [18])**.** *The determinant*

$$D_n = \left|\binom{k+n-i}{n-j}\right|_{i,j=1}^{n}$$

is equal to 1 for all $n \geq 1$.

Proof. The proof goes by induction on n. The determinant $D_1 = \binom{k}{0} = 1$ which is the base of induction.

Consider D_{n+1}. For all $i = 1, 2, \ldots, n-1$ subtract the $(i+1)$-th row of D_{n+1} from i-th row. This does not change the value of D_{n+1}. By applying the well-known combinatorial identity

$$\binom{p}{q} - \binom{p-1}{q} = \binom{p-1}{q-1}, \tag{5.55}$$

we obtain

$$D_{n+1} = \det \begin{bmatrix} D_n & 0_{n\times 1} \\ A_{1\times n} & \binom{k}{0} \end{bmatrix}, \tag{5.56}$$

where $A_{1\times n} = \begin{bmatrix}\binom{k}{n} & \binom{k}{n-1} & \ldots & \binom{k}{1}\end{bmatrix}$. From (5.56) it follows that $D_{n+1} = D_n$ which proves the lemma. □

Theorem 5.35 (Bodroža et al., Theorem 1 in [18])**.** *The following equation holds:*

$$F_n \coloneqq \left|\binom{m+k}{m+i-j}\right|_{i,j=1}^{n} = \prod_{i=0}^{n-1} \frac{\binom{m+k+i}{m}}{\binom{m+i}{m}}. \tag{5.57}$$

Proof. Add the $(i+1)$-th row of F_n to the i-th row for all $i = 1, 2, \ldots, n-1$ and apply the identity

$$\binom{p}{q} + \binom{p}{q+1} = \binom{p+1}{q+1} \tag{5.58}$$

to obtain

$$F_n^{(1)} = \left|\binom{m+k+\delta_{i\leq n-1}}{m+i-j+\delta_{i\leq n-1}}\right|_{i,j=1}^{n} \tag{5.59}$$

where $\delta_{i\leq n-1}$ is generalisation of the Kronecker δ symbol: Let P be any logical expression. Then

$$\delta_P = \begin{cases} 1, & \text{if } P; \\ 0, & \text{otherwise.} \end{cases}$$

Then for all $i = 1, 2, \ldots, n-2$ add the $(i+1)$-th row of $F_n^{(1)}$ to the i-th row which gives us

$$F_n^{(2)} = \left| \binom{m+k+\delta_{i\leq n-1}+\delta_{i\leq n-2}}{m+i-j+\delta_{i\leq n-1}+\delta_{i\leq n-2}} \right|_{i,j=1}^{n}. \tag{5.60}$$

If we continue in this manner, we obtain a series of determinants

$$F_n^{(0)}, F_n^{(1)}, F_n^{(2)}, \ldots, F_n^{(n-1)} \tag{5.61}$$

where

$$F_n^{(l)} = \left| \binom{m+k+\sum_{t=1}^{l}\delta_{i\leq n-t}}{m+i-j+\sum_{t=1}^{l}\delta_{i\leq n-t}} \right|_{i,j=1}^{n} \tag{5.62}$$

Note that $F_n = F_n^{(0)}$. It is clear that

$$F_n^{(0)} = F_n^{(1)} = F_n^{(2)} = \cdots = F_n^{(n-1)}. \tag{5.63}$$

It is easy to see that

$$F_n^{(n-1)} = \left| \binom{m+k+n-i}{m+n-j} \right|_{i,j=1}^{n}. \tag{5.64}$$

Recall that $\binom{p}{q} = \frac{p!}{q!(p-q)!}$. The right hand side of (5.64) equals

$$\begin{aligned}
&\frac{\prod_{l=1}^{n}(m+k+n-l)!}{\prod_{l=1}^{n}(m+n-l)!}\left|\frac{1}{(k-i+j)!}\right|_{i,j=1}^{n} = \\
&\quad = \frac{\prod_{l=1}^{n}(m+k+n-l)!}{\prod_{l=1}^{n}(m+n-l)!}\cdot\frac{\prod_{l=1}^{n}(n-l)!}{\prod_{l=1}^{n}(k+n-l)!}\underbrace{\left|\frac{(k+n-i)!}{(k-i+j)!(n-j)!}\right|_{i,j=1}^{n}}_{D_n=1} = \\
&\quad = \prod_{i=0}^{n-1}\frac{\frac{(m+k+i)!}{m!(k+i)!}}{\frac{(m+i)!}{m!i!}} = \prod_{i=0}^{n-1}\frac{\binom{m+k+i}{m}}{\binom{m+i}{m}}.
\end{aligned} \tag{5.65}$$

□

Note that

$$\begin{aligned}
K(O(k,m,n)) = K(O(k,n,m)) = K(O(m,n,k)) = \\
= K(O(m,k,n)) = K(O(n,m,k)) = K(O(n,k,m)),
\end{aligned} \tag{5.66}$$

since $O(k,m,n) \cong O(k,n,m) \cong \cdots \cong O(n,k,m)$, i.e., if we rotate or reflect the benzenoid its Kekulé number does not change. This means that we may permute the numbers k, m and n in (5.51) and we obtain:

Corollary 5.36. *The following equation holds:*

$$\prod_{i=0}^{k-1}\frac{\binom{m+n+i}{n}}{\binom{n+i}{n}} = \prod_{i=0}^{k-1}\frac{\binom{n+m+i}{m}}{\binom{m+i}{m}} = \cdots = \prod_{i=0}^{m-1}\frac{\binom{k+n+i}{n}}{\binom{n+i}{n}}. \tag{5.67}$$

The equation (5.51) gives us the number of Kekulé structures in the hexagonal benzenoid $H(n,m,k,t)$ for the case when $n = t$. If $n \neq t$ then from Corollary 5.30 it follows that $H(n,m,k,t)$ has no Kekulé structures.

5.7.3 A linear-time algorithm

In this section, we will need the notion of a *generalised benzenoid system*. In [181] it is defined in the following way:

Definition 5.25. *A* generalised benzenoid graph *is a finite plane subgraph of the infinite hexagonal lattice such that every bounded face is a unit hexagon of the hexagonal lattice.*

Let $H_1, \ldots, H_m$ be disjoint subgraphs of G. With $G/(H_1, \ldots, H_m)$ we will denote a new graph which is obtained from G in the following way: Start with the graph $G - (H_1 \cup \cdots \cup H_m)$. For every subgraph H_i add a new vertex h_i to that new graph. If a vertex $v \in V(G - (H_1 \cup \cdots \cup H_m))$ was adjacent to a vertex of H_i in G then connect v and h_i in the new graph. The resulting graph is $G/(H_1, \ldots, H_m)$.

Our equivalent to Definition 5.25 is:

Definition 5.26. *Let $\mathcal{B} = \mathcal{B}_1 \sqcup \cdots \sqcup \mathcal{B}_m$ be a disjoint union of (zero or more) benzenoids in the infinite hexagonal grid $\mathcal{H}$. A* generalised benzenoid system *G is a subgraph of the hexagonal lattice such that $G(\mathcal{B}) \subseteq G$ and $G/(G(\mathcal{B}_1), \ldots, G(\mathcal{B}_m))$ is a forest.*

For an example of a generalised benzenoid graph see Example 5.7. Note that a GBG (generalised benzenoid graph) is a generalisation of the class of benzenoid graphs, but is *not* a generalisation of the class of coronoid graphs. The notions of boundary- and internal vertices and edges of a benzenoid can be naturally generalised to generalised benzenoid systems. Let $b(G)$ and $w(G)$ denote the number of black and white vertices of G, respectively. Let F be a Kekulé structure of G. Here, edges F and $E(G) \setminus F$ will be called red and green edges, respectively.

Let P_1 and P_2 be two distinct points in the plane (in which the hexagonal lattice is embedded). The straight line segment P_1P_2 is called a *cut segment* if it is perpendicular bisector of two boundary edges of G and no point of P_1P_2 lies in the exterior face of G [106]. Let C denote the set of all those edges that P_1P_2 meets. Then C is called the cut corresponding to P_1P_2. We will slightly extend this notion to also include bridges of G, i.e., edges that are not adjacent to any bounded face of G. Note that $G - C$ has one more connected component than the graph G. See Example 5.7 to get familiar with the notion of cut segments and cuts.

Example 5.7. A generalised benzenoid system is in Figure 5.46. Some of its cut segments and their corresponding cuts are shown in the figure. Edges that belong to the same cut are coloured with the same colour. The singleton that contains the orange edge (it is not incident with any finite face and is therefore a bridge) is also considered to be a cut. □

Let G be a connected generalised benzenoid system and let C be a cut. Note that when the edges of C are removed, their end vertices that are of the same colour reside in the same connected component of $G - C$. Connected components of $G - C$ are called the *black bank* and the *white bank* and are denoted $B(C)$ and $W(C)$, respectively. The next lemma follows immediately:

Lemma 5.37. *Let G be a connected generalised benzenoid system that admits a Kekulé structure. Then $b(G) = w(G)$. Moreover, for every cut C the number of red edges which*

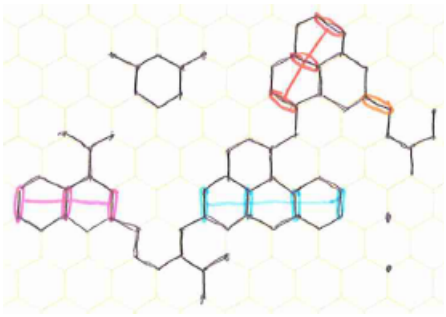


Figure 5.46: A generalised benzenoid system and some of its cuts.

are contained in C *equals* $\Delta(C) := b(B(C)) - w(B(C)) = w(W(C)) - b(W(C))$ *and* $0 \leq d(C) \leq |C|$. □

Lemma 5.38. *Let* $\mathcal{C}$ *be a coronoid with a fixed embedding in the hexagonal grid* $\mathcal{H}$*. Then*

$$w(\mathcal{C}) - b(\mathcal{C}) = p(\mathcal{C}) - v(\mathcal{C}).$$

Proof. If $\mathcal{C}$ is a coronoid with a fixed embedding, then the vertices of $\mathcal{C}$ can be partitioned into 3 classes:

(a) peaks (white vertices);

(b) valleys (black vertices);

(c) vertices that are incident to a vertical edge.

The vertices that fall under (c) can be further divided into black and white subclasses and the vertical edges define a matching between the two subclasses. Let ν denote the number of vertical edges. Then $w(\mathcal{C}) = p(\mathcal{C}) + \nu$ and $b(\mathcal{C}) = v(\mathcal{C}) + \nu$. Therefore,

$$w(\mathcal{C}) - b(\mathcal{C}) = (p(\mathcal{C}) + \nu) - (v(\mathcal{C}) + \nu) = p(\mathcal{C}) - v(\mathcal{C}).$$

□

Let H be a subgraph of G. Then $p(G, H)$ denotes the number of peaks of G that are contained in the subgraph H. Similarly, $v(G, H)$ denotes the number of valleys that are contained in H.

Lemma 5.39. *Let* C *be a horizontal cut of a benzenoid (with a fixed embedding). Then*

$$b(B(C)) - w(B(C)) = |C| + v(G, B(C)) - p(G, B(C)). \tag{5.68}$$

Proof. Vertices of $B(C)$ can be partitioned into 4 classes:

(a) end vertices of C (black vertices);

(b) vertices that are incident to a vertical edge in $B(C)$;

(c) peaks of G that lie inside $B(C)$ (white vertices);

(d) valleys of G that lie inside $B(C)$ (black vertices).

Let ν denote the number of vertical edges in $B(C)$. Then the total number of white vertices in $B(C)$ is $\nu + p(G, B(C))$ and the total number of black vertices is $|C| + \nu + v(G, B(C))$. Equation (5.68) immediately follows. □

Note that $B(C)$ "lies above" the cut C (i.e., the black end vertices of C lie above their corresponding white vertices of C).

Lemma 5.40. *Let G be a benzenoid graph with a Kekulé structure. Then $p(G) = v(G)$. For every cut C, the number of red edges (i.e., those that are contained in the Kekulé structure) equals*

$$\Delta(C) \coloneqq b(B(C)) - w(B(C)) = |C| + v(G, B(C)) - p(G, B(C)).$$

Proof. Let F be a matching (a Kekulé structure) in G. If we restrict F to graph $B(C)$, then exactly $\Delta(C)$ black vertices of $B(C)$ remain unmatched. Those vertices are matched with some white vertices from $W(C)$ and this can only be achieved by using edges of C. With other words, the number of red edges in C equals $\Delta(C)$. Everything else follows from Lemma 5.38 and Lemma 5.39. □

Corollary 5.41. *Let G be a benzenoid graph with a Kekulé structure. Then for every cut C, the number of green edges (i.e., those that are* not *contained in the Kekulé structure) equals*

$$\Delta^*(C) \coloneqq p(G, B(C)) - v(G, B(C)) = v(G, W(C)) - p(G, W(C)).$$

□

This is a straightforward consequence of $|C| = \Delta(C) + \Delta^*(C)$. Note that $0 \leq \Delta(C) \leq |C|$ and $0 \leq \Delta^*(C) \leq |C|$.

Proposition 5.42. *Let G be a benzenoid graph. Let n_2 and n_3 denote the number of degree 2 vertices on its boundary and the number of degree 3 vertices on its boundary, respectively. Then $n_2 - n_3 = 6$.* □

The proof can be found in [102] and is therefore omitted.

Corollary 5.43. *The boundary of a benzenoid has at least six $(2,2)$-edges.*

Proof. Let a, b and c be the number of $(2,2)$-edges, $(2,3)$-edges and $(3,3)$-edges on its boundary, respectively. Then $2a + b = 2n_2$ and $2c + b = 2n_3$. If we subtract the two equalities we obtain $a - c = n_2 - n_3 = 6$. Because $c \geq 0$, it follows that $a \geq 6$. □

Lemma 5.44. *Let $\mathcal{B}$ be a Kekulean benzenoid and F one of its Kekulé structures. Let C be a horizontal cut such that $p(\mathcal{B}, B(C)) - v(\mathcal{B}, B(C)) < |C|$ and the left-most edge $e \in C$ is a (2,2)-edge. Then there exists a Kekulé structure F' that contains the edge e.*

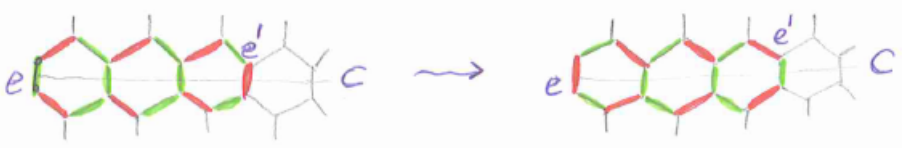


Figure 5.47: Recolouring of edges in C.

Proof. By Corollary 5.41, C contains at least one edge of F. If $e \in F$ then take $F' = F$. If $e \notin F$ then let $e' \in C$ be the left-most edge that is contained in F (see Figure 5.47). The Kekulé structure can be uniquely determined for the coloured edges in Figure 5.47. The edges on the alternating cycle can be recoloured as shown in Figure 5.47 to obtain the desired Kekulé structure F' from F. □

We will describe the algorithm by Sachs that finds a perfect matching (if it exists) in a GBG or decides that there is no perfect matching. A GBG G has a Kekulé structure if every connected component of G has a Kekulé structure, thus we can assume that G is connected. if it is not connected, run the algorithm on every connected component seperately.

At each step, the algorithm colours some edges of G (with red and green) and removes them from G. There are 2 cases to consider:

Case 1. G contains a bridge e. Consider the cut $C = \{e\}$. Determine

$$\Delta(C) = b(B(C)) - w(B(C)).$$

(i) If $\Delta(C) < 0$ or $\Delta(C) > 1$, then report that there is no perfect matching.

(ii) If $\Delta(C) = 0$, colour the edge e green.

(iii) If $\Delta(C) = 1$, colour edge e red and colour its incident edges green.

Case 2. G does not have a bridge. Then G is a benzenoid. We can find a $(2, 2)$-edge e that is on the boundary of the benzenoid and in vertical position. Let C denote the cut that contains e. Determine

$$\Delta^*(C) = p(G, B(C)) - v(G, B(C)).$$

(i) If $\Delta^*(C) < 0$ or $\Delta^*(C) > |C|$ then report that G does not admit a perfect matching.

(ii) If $\Delta^*(C) = |C|$ then colour all edges of C green.

(iii) If $\Delta^*(C) = 0$ then colour all edges of C red and their incident edges green.

(iv) If $0 < \Delta^*(C) < |C|$ then colour e red and edges that are incident to e green (this can be done due to Lemma 5.44). Moreover, if $\Delta^*(C) = |C| - 1$ then colour the remaining edges of C green.

In any case, discard the coloured edges and continue this procedure on the GBG that remains. All edges that get coloured red during this procedure belong to the matching.

Note that when implementing the above algorithm, one has to carefully choose the data structures. If one is not careful enough with the implementation then the result may be an algorithm whose time complexity is worse than linear.

Chapter 6

Patches

In this chapter we generalise benzenoids and coronoids to patches and perforated patches, respectively. Essentially, a *patch* is a (2-connected) plane graph similar to a benzenoid in which various polygons may be used instead of hexagons alone. All internal vertices are of degree 3, whilst boundary vertices are of degree 2 or 3 (see example in Figure 6.1).

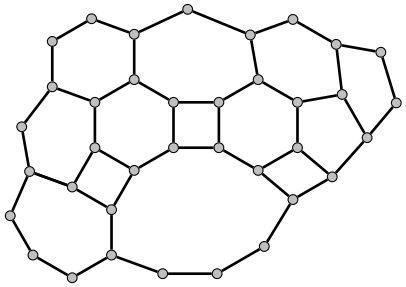


Figure 6.1: A patch.

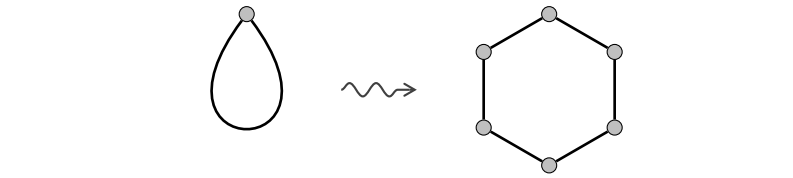
Our "traditional" definition generalises one used by Jack Graver and other authors [30, 56, 86, 87, 88, 89, 90, 91]. Each patch has a unique boundary-edges code, but in contrast to benzenoids, which are uniquely determined by their boundary-edges codes, the boundary-edges code of a patch may belong to more than one patch or even to a fullerene patch (as defined below).

Definition 6.1. *A* patch Π *is either a cycle or a 3-valent 2-connected plane (multi)graph with a distinguished outer face, where the edges of the outer face may be arbitrarily subdivided.*

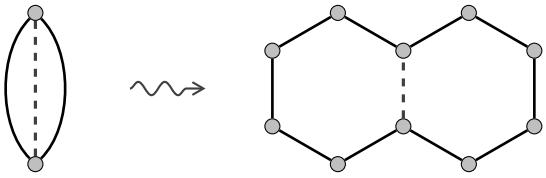
In the following example we depict some benzenoids as patches arising from plane cubic graphs.

Example 6.1. Some plane cubic 2-connected graphs may be expanded to benzenoids; see Figure 6.2. The graph on the right hand side is obtained from the graph on the left hand side by sibdividing some of the edges on the outer face. A cycle of arbitrary length can be obtained from a loop. □

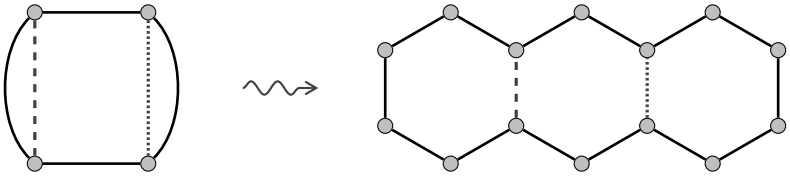
Example 6.2. The graph in Figure 6.3 does not give rise to any patch. □



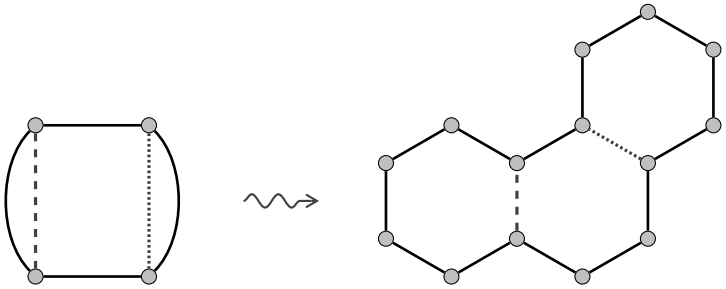
(a) benzene



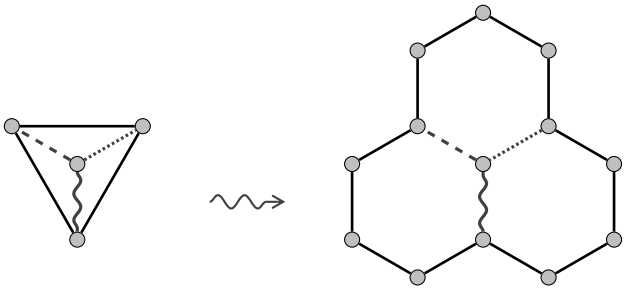
(b) naphthalene



(c) anthracene



(d) phenanthrene



(e) phenalene

Figure 6.2: Smallest benzenoids as patches. Benzene is obtained from a loop and the rest are obtained from plane cubic 2-connected graphs.

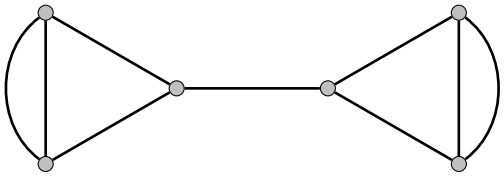


Figure 6.3: A plane cubic graph that is not 2-connected does not have a boundary cycle and is therefore not able to produce any patch.

Clearly, a patch is a proper generalization of a benzenoid.

Definition 6.2. *A patch* Φ *with interior faces pentagons and hexagons is called a* fullerene patch*.*

The above Definition 6.2 is the one that is used by Graver et al. [87, 88].

Definition 6.3. *A patch* Φ *with interior faces hexagons is called a* helicene patch*.*

Note that a helicene patch is sometimes called a helicene. Also note that the following is true:

Proposition 6.1. *Each benzenoid is a helicene patch but there are helicene patches that are not benzenoids.*

Proof. The implication follows directly from the definition. It is not an equivalence, since each proper helicene such as the one from Figure 6.4 is a counter-example to the converse. □

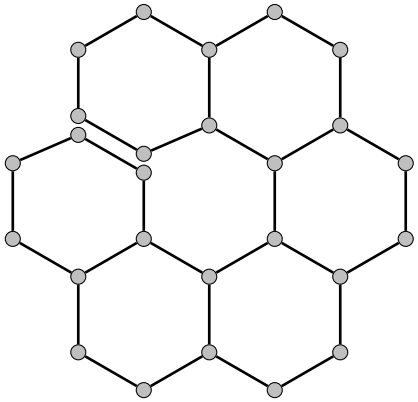


Figure 6.4: Not all patches with internal 6-cycles are benzenoids. The patch in this figure is a helicene.

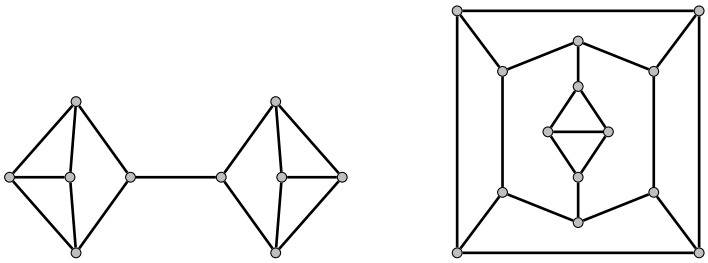
The boundary-edges code can be naturally generalised to patches. We can define:

Definition 6.4. *A patch* Π *is* convex *if its boundary-edges code contains no* 1*.*

6.1 New approach to patches and perforated patches

We will present a mathematical formalisation which is based on the treatment of coronoids and benzenoids in Section 5.4. We have seen that our definition of a *fullerene patch* is compatible with Graver's definition [87, 88]. There is also a notion of a (m, k)-patch which received a lot of attention in the past years [25, 30, 86, 95]. By our definition, faces may have a range of different degrees, but $m = 3$.

Our point of departure is a finite plane cubic (simple) graph G, which divides the Euclidean plane $\mathbb{R}^2$ into several regions called faces. The collection of all faces is denoted $\mathcal{F}_G$. One face is unbounded and the rest are bounded. Two examples of plane cubic graphs are in Figure 6.5. Note that in Section 5.4, the role of graph G was taken by an infinite



(a) (b)

Figure 6.5: Two examples of cubic plane graphs.

cubic graph with all faces hexagons which we called hexagonal lattice. Here we restrict our attention to finite graphs G. Later on, we will compare finite and infinite versions of this theory.

The hexagonal lattice has additional nice properties. First of all, no two faces of $\mathcal{H}$ share more than one edge; the graph in Figure 6.5(b) does not have that property. Also, no face of $\mathcal{H}$ is incident with itself; the graph in Figure 6.5(a) does not have that property, since one if its edges is incident to only one face (the outer face). Here, we also permit graphs which are not 2-connected, such as the one in Figure 6.5(a).

Let $\mathcal{P} \subseteq \mathcal{F}_G$ be some arbitrary subcollection of faces and let $a, b \in \mathcal{P}$. We say that a and b are *adjacent*, $a \sim b$, if they share an edge. We define relation $\equiv_{\mathcal{P}}$ by the same way as before, i.e., $a \equiv_{\mathcal{P}} b$ if there is a sequence $c_0 = a, c_1, c_2, \ldots, c_m = b$ such that $c_{i-1} \sim c_i$ and $c_i \in \mathcal{P}$. $\mathcal{P}$ is *connected* if $a \equiv_{\mathcal{P}} b$ for any $a, b \in \mathcal{P}$. The set $\mathcal{P}$ is naturally decomposed into *connected components*.

Definition 6.5. *Let G be any finite plane cubic graph. A proper subset $\mathcal{P} \subset \mathcal{F}_G$ is called a* perforated patch *if it is connected.*

Definition 6.6. *Let G be any finite plane cubic graph. A proper subset $\mathcal{P} \subset \mathcal{F}_G$ is called a* patch *if $\mathcal{P}$ is connected and $\mathcal{P}^{\complement} = \mathcal{F}_G \setminus \mathcal{P}$ is also connected.*

Observe that finiteness of set $\mathcal{K}$ in Definitions 5.5 and 5.6 implies that $\mathcal{K}$ is a proper subset of $\mathcal{H}$. In this finite version of those two definitions we had to make that explicit.

Most observations of the previous section about coronoids and benzenoids are also true for their corresponding generalisations, namely perforated patches and patches. In some

cases the proof remains essentially the same and in some cases it slightly simplifies, since we do not have to deal with infinity anymore. We will transcribe results of the previous section into this new language and omit the proofs. When appropriate, we will give some clarifications. For our considerations, only the combinatorial data are relevant (adjacency between faces and the cyclic ordering of edges incident to a common vertex). Geometric details of the drawing are unimportant. Moreover, the outer face does not play a special role. The graph G could also be embedded on the sphere S^2, where all faces would be bounded. All the combinatorial information would, of course, remain exactly the same.

Lemma 6.2. *Let G be any finite plane cubic graph and let $\emptyset \neq \mathcal{P} \subseteq \mathcal{F}_G$. Let $a \in \mathcal{P}$ be any of its faces. Then a belongs to some connected component $\mathcal{C}_i$ of $\mathcal{P}$. Let $b \in N(a)$. Then either $b \in \mathcal{C}_i$ or $b \in \mathcal{P}^{\complement}$.* □

Lemma 6.3. *Let G be any finite plane cubic graph and let $\emptyset \neq \mathcal{P} \subset \mathcal{F}_G$. Let $\mathcal{C}_a$ be any connected component of its complement $\mathcal{P}^{\complement}$. Then there exists a face $\tilde{a} \in \mathcal{C}_a$ that is adjacent to some face in $\mathcal{P}$.* □

Lemma 6.4. *Let G be any finite plane cubic graph and let $\mathcal{P}$ be a perforated patch. Let $\mathcal{C}_a$ be a connected component of its complement $\mathcal{P}^{\complement}$, $a \in \mathcal{C}_a$ and $p \in \mathcal{P}$. Then $a \equiv_{\mathcal{P} \cup \mathcal{C}_a} p$. Let $\mathcal{C}_b$ be another connected complement of $\mathcal{P}^{\complement}$. Then $a \equiv_{\mathcal{P} \cup \mathcal{C}_a \cup \mathcal{C}_b} b$.* □

The above Lemma 6.4 corresponds to both Lemma 5.5 and Corollary 5.6.

Lemma 6.5. *Let G be any finite plane cubic graph and let $\emptyset \neq \mathcal{P} \subseteq \mathcal{F}_G$. Then the complement $\mathcal{P}^{\complement}$ of $\mathcal{P}$ consists of finitely many connected components:*

$$\mathcal{P}^{\complement} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_d.$$

Each of the components $\mathcal{C}_i$, $i \geq 1$, is a perforated patch. If $\mathcal{P}$ is a perforated patch, then each $\mathcal{C}_i$ is a patch. □

The above Lemma 6.5 corresponds to both Lemma 5.4 and Theorem 5.7. Since we do not have to deal with an infinite number of faces, the proof becomes trivial. The definition of the benzenoid closure was natural. Here, one should be slightly more careful:

Definition 6.7. *Let $\mathcal{P}$ be a perforated patch. The* closure *of $\mathcal{P}$ with respect to $p \in \mathcal{P}^{\complement}$, denoted $Cl(\mathcal{P}, p)$ is the intersection of all those patches which include $\mathcal{P}$ as a subset and do not contain p among their faces, i.e.,*

$$Cl(\mathcal{P}, p) = \bigcap \{\mathcal{Q} \mid \mathcal{Q} \text{ is patch} \wedge \mathcal{P} \subseteq \mathcal{Q} \wedge p \notin \mathcal{Q}\}.$$

Let us investigate what happens if the extra condition, i.e., exclusion of a designated face p, is omitted. By Lemma 6.5, $\mathcal{P}^{\complement} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_d$ where each $\mathcal{C}_i$ is a patch. Define $\mathcal{P}_i \coloneqq \mathcal{P} \sqcup \mathcal{C}_1 \sqcup \ldots \sqcup \mathcal{C}_{i-1} \sqcup \mathcal{C}_{i+1} \sqcup \cdots \sqcup \mathcal{C}_d$. It is easy to see that each $\mathcal{P}_i$, $1 \leq i \leq d$, is a patch. Clearly, $\mathcal{P} \subseteq \mathcal{P}_i$ for each $i = 1, \ldots, d$. But $\bigcap \{\mathcal{P}_i \mid 1 \leq i \leq d\} = \mathcal{P}$. Without that extra condition the definition would not make sense. In most cases this "forbidden" face p can be chosen in advance and one can deal with only those perforated patches which do not include face p. Then we can write $Cl(\mathcal{P})$ instead of $Cl(\mathcal{P}, p)$ without introducing any ambiguity. The most natural candidate for the forbidden face is, of course, the outer face. Let us denote $Pat(\mathcal{P}, p) = \{\mathcal{Q} \mid \mathcal{Q} \text{ is patch} \wedge \mathcal{P} \subseteq \mathcal{Q} \wedge p \notin \mathcal{Q}\}$ for convenience. Lemma 5.8 and Proposition 5.9 give rise to the following analogue in the theory of perforated patches:

Lemma 6.6. *Let $\mathcal{P}$ be a perforated patch and $p \in \mathcal{P}^{\complement}$ a face in its complement. Let $\mathcal{P}^{\complement} = \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \cdots \sqcup \mathcal{C}_d$ as in Lemma 6.5. Then the closure of $\mathcal{P}$ with respect to p is*

$$Cl(\mathcal{P}, p) = \mathcal{P} \sqcup \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_{j-1} \sqcup \mathcal{C}_{j+1} \sqcup \cdots \sqcup \mathcal{C}_d,$$

where $\mathcal{P}_j$ is the connected component of $\mathcal{P}^{\complement}$ which includes p among its faces, i.e., $p \in \mathcal{C}_j$. Moreover, the closure with respect to p is an operation on the set of perforated patches without p that satisfies the following conditions:

(a) $\mathcal{P} \subseteq Cl(\mathcal{P}, p)$,

(b) $\mathcal{P} \subseteq \mathcal{Q} \implies Cl(\mathcal{P}, p) \subseteq Cl(\mathcal{Q}, p)$, *and*

(c) $Cl(Cl(\mathcal{P}, p), p) = Cl(\mathcal{P}, p)$.

□

The following lemma is an analogue to Definition 5.8 and Lemma 5.10 from the theory of coronoid hydrocarbons:

Lemma 6.7. *Let $\mathcal{P}$ be a perforated patch such that $p \notin \mathcal{P}$. Let $\widetilde{\mathcal{P}} \subseteq \mathcal{F}_G$ satisfy the following conditions:*

(i) $\widetilde{\mathcal{P}}$ *is a patch without* p,

(ii) $\mathcal{P} \subseteq \widetilde{\mathcal{P}}$, *and*

(iii) $\mathcal{P} \subseteq \mathcal{Q} \implies \widetilde{\mathcal{P}} \subseteq \mathcal{Q}$ *for every patch* $\mathcal{Q}$ *without* p.

This $\widetilde{\mathcal{P}}$ is unique and $\widetilde{\mathcal{P}} = Cl(\mathcal{P}, p)$. □

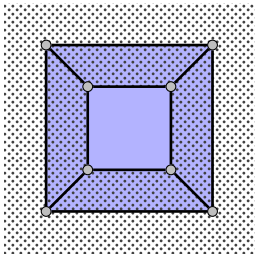


Figure 6.6: The intersection of two patches may be a perforated patch.

Some caution should be used in making the analogy with Lemma 5.11: as seen in Figure 6.6, the intersection of two patches may be a perforated patch. However, the following is true:

Lemma 6.8. *Let $\mathcal{P}$ and $\mathcal{Q}$ be two patches such that their complements share a face, i.e., $\mathcal{P}^{\complement} \cap \mathcal{Q}^{\complement} \neq \emptyset$. Then the intersection of patches $\mathcal{P}$ and $\mathcal{Q}$ is a union of disjoint patches.* □

To obtain a perforated patch we choose a set of faces of the plane cubic graph G that constitute a connected region. Sometimes we talk about "removing faces". This means that we select members of the complement, i.e., faces that will not be present in the perforated patch.
Given a perforated patch as a plane graph it is not possible to detect (in the general case) which faces are corona holes and which not. An example is given in Figure 6.7. Therefore, plane graphs do not give a sufficient model for perforated patches. One has to indicate which faces are actually present, and which are not. This means that Lemma 5.16 has no equivalent in this theory.

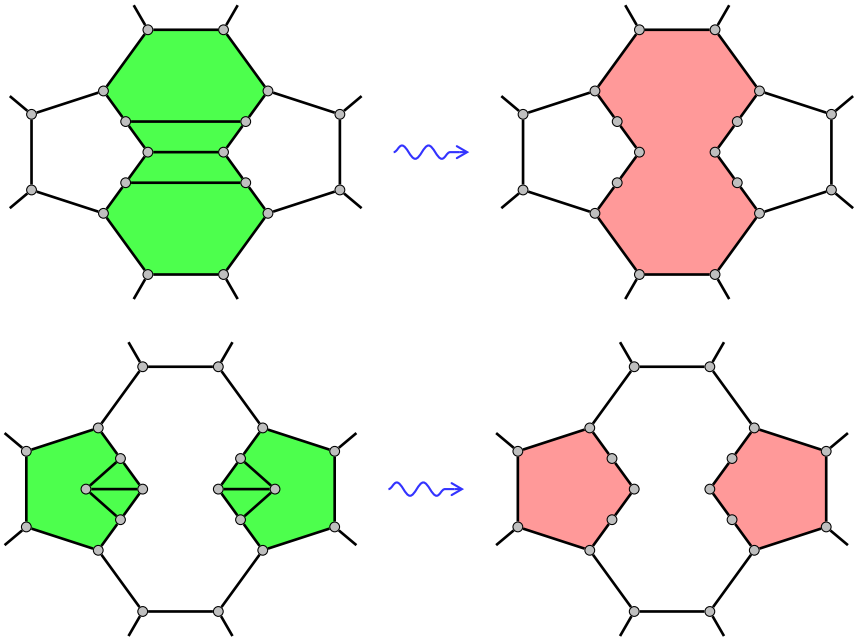


Figure 6.7: Two distinct perforated patches with the same skeleton. The faces that we removed, i.e., the faces in the complement, are indicated by shading and the corresponding hole is also shaded.

A patch is called *k-connected* when its skeleton is k-connected. A face which is incident to an edge from both sides is called an *ill-behaved* face. The following proposition precisely characterizes 2-connected patches:

Proposition 6.9. *A patch $\mathcal{P}$ is 2-connected if and only if it contains no ill-behaved faces.*

Proof. An ill-behaved face implies there is a bridge (cut-edge) so it is not 2-connected. If there are no ill-behaved faces: start with a single face. Its skeleton is a cycle which is 2-connected. Then iteratively add adjacent faces. The newly added edges form one or more paths that are connected to the existing graph. In this way we construct an ear-decomposition. □

It is possible to generalise this theory for the case when the cubic planar graph G is infinite. The standard embedding of the hexagonal lattice is such an example. If one tries to use an arbitrary infinite cubic graph, problems of topological nature may arise. Some remarks

must therefore be made. The hexagonal grid, as we will see, has certain nice properties. In the proof of Lemma 5.3 we use the fact that $\cup_{n=0}^{\infty}\mathcal{P}_n = \mathcal{H}$, with $\mathcal{P}_n$ defined as in the proof of Lemma 5.3, without proving it. This holds when there exists a finite pathway of faces between any two faces of the plane graph. With other words, the distance between every two vertices of the dual graph is finite. The hexagonal grid is an example. If this was not the case, the proof of Lemma 5.3 would fail. Another important building block of this theory is Lemma 5.4. In its proof, when we claim that $\mathcal{P}^{\complement}$ is connected, we implicitly use the Jordan curve theorem. This renowned theorem seems obvious at the first sight, but its proof happens to be involved (if one attempts to prove it in its full generality). Another sensitive part of the proof is when we claim that $\mathcal{P}_n$ contains finitely many faces. A nifty topologist could construct such an example where this would fail. Luckily, every bounded region of the infinite hexagonal grid contains finitely many faces. This is another condition on the infinite plane graph G that has to be met.

Chapter 7

Altans

Recently a class of molecular graphs, called *altans*, became a focus of attention of several theoretical chemists and mathematicians. In this section we study primary *iterated altans* and show, among other things, their connections with *nanotubes* and *nanocaps*. The question of classification of bipartite altans is also addressed. Using the results of Gutman we are able to enumerate Kekulé structures of several nanocaps of arbitrary length.

7.1 Introduction to altans

Altans were first introduced as special planar systems, obtained from benzenoids by attachment of a ring to all outer vertices of valence two [100, 147], in particular in connection with concentric decoupled nature of the ring currents (see papers by Zanasi et al. [146, 147, 149] and also Mallion and Dickens [59, 60]). In Section 3.4.3 we have presented conjugated circuits, i.e., the graph-theoretical approach to ring current which was initiated by Randić in 1976 [173]. It was also studied by Gomes and Mallion [82]. Full description is provided for instance in references [175, 176]. See also paper [72] by Fowler and Myrvold on 'The Anthracene Problem'.

Later altans were generalized by Gutman [99] to arbitrary graphs. We essentially follow Gutman's approach. Our point of departure is a *peripherally rooted graph*, i.e., an ordered pair (G, S) which consists of an arbitrary graph G and a cyclically ordered subset S of its vertices, called the *peripheral root*.

Let n denote the order of G and let k denote the cardinality of S. Assume that $V(G) = \{0, 1, \ldots, n-1\}$. The operation $A(G, S)$ maps the pair (G, S) to a new pair (G_1, S_1) as follows: Let

$$S_0 = \{n, n+1, \ldots, n+k-1\} \text{ and } S_1 = \{n+k, n+k+1, \ldots, n+2k-1\}. \tag{7.1}$$

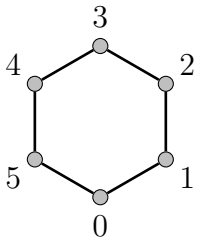
Let the vertex set of G be augmented by $S_0 \cup S_1$. Through the vertices $S_0 \cup S_1$, we construct a *peripheral* cycle graph C of length $2k$ in the cyclic order

$$(n, n+k, n+1, n+k+1, n+2, \ldots, n+k-1, n+2k-1, n). \tag{7.2}$$

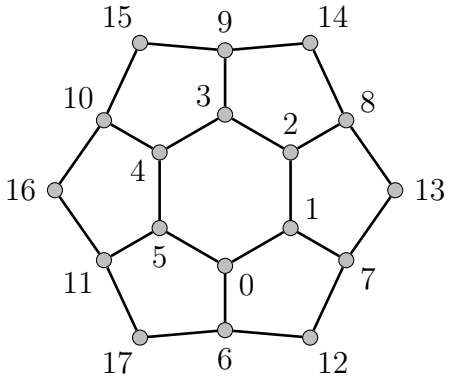
Finally, we attach C to G by k edges between S and S_0 of the form $(s_i, n+i)$, $0 \leq i < k$, where s_i is the i-th vertex of S. The vertices of C that have valence 2 in the final construction

are exactly the ones originating from S_1 and are the new peripheral root of the altan. The new peripheral root, S_1, is ordered in the natural way.

Example 7.1. A bipartite graph may give rise to non-bipartite or bipartite altans. Let



(a) $G = C_6$, $S = (0, 1, 2, 3, 4, 5)$

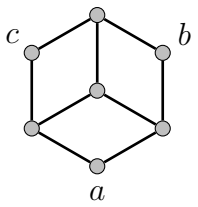


(b) $A(G, S) = (G_1, S_1)$. $S_1 = (12, 13, 14, 15, 16, 17)$ consists of vertices of valence 2 in the natural cyclic order.

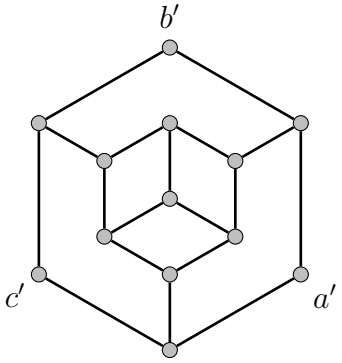
Figure 7.1: The altan of benzene is not bipartite since $A(G, S)$ contains pentagons.

$G = C_6$ and $S = (0, 1, 2, 3, 4, 5)$. Graph G and the non-bipartite $A(G, S)$ are depicted in Figure 7.1. □

Example 7.2. The altan of the graph G in Figure 7.2 is bipartite. □



(a) G, $S = (a, b, c)$



(b) G, $S_1 = (a', b', c')$

Figure 7.2: Since G is bipartite and all vertices of S belong to the same set of bipartition, $A(G, S)$ is also bipartite.

Note that the altan in Example 7.1 is non-bipartite, whilst the one in Example 7.2 is bipartite. We can classify bipartite altans:

Theorem 7.1. *Let (G, S) be a graph G with a peripheral root S. The altan $A(G, S)$ is bipartite if and only if:*

(a) G is bipartite and

(b) members of S belong to the same bipartition set.

Proof. First, we will show that conditions (a) and (b) imply that $A(G,S) = (G_1, S_1)$ is bipartite. Let G be bipartite and let us colour its vertices black and white. We may assume that S has all of its vertices coloured black. Their adjacent vertices, S_0, can be coloured white. This can be further extended, by colouring S_1 black, to a proper black and white vertex colouring of G_1. Hence G_1 is bipartite. Furthermore, (G_1, S_1) also satisfies the conditions (a) and (b) of the Theorem.

Now we prove the other direction. Let $A(G,S)$ be bipartite. Graph G is a subgraph of G_1, so it is also bipartite. If G is bipartite but not all vertices of S are coloured with the same colour, then two consecutive vertices of S, say u and v, would be coloured differently. Recall that vertices of S are cyclically ordered. Hence there is a u,v-path in G of odd length. By attaching the cycle C to G to form $A(G,S)$ it is possible to connect u to v by a path of length 4. This means there is a cycle of odd length in graph G_1, a contradiction. □

From the definition of the altan operation it follows that we may repeat it several times. Let $A^n(G,S)$ denote the n-th altan of (G,S), i.e.,

$$\underbrace{A(A(\cdots A}_{n}(G,S)\cdots)). \tag{7.3}$$

We obtain the following consequence of Theorem 7.1:

Corollary 7.2. *Let (G,S) be a graph with peripheral root S and let $n \geq 1$ be an arbitrary integer. $A^n(G,S)$ is bipartite if and only if (G,S) satisfies the conditions of Theorem 7.1.*

Proof. It follows by induction. The basis of induction is given by Theorem 7.1. □

7.2 Altans of benzenoid systems

Let $\mathcal{B}$ be a (finite) benzenoid. The altan of $\mathcal{B}$ is assumed to have the cyclically ordered peripheral vertices, S, of valence 2. Their order is obtained by traversing $\mathcal{B}$ along its perimeter.

Theorem 7.3. *For any finite benzenoid $\mathcal{B}$, the altan $A(\mathcal{B})$ is non-bipartite.*

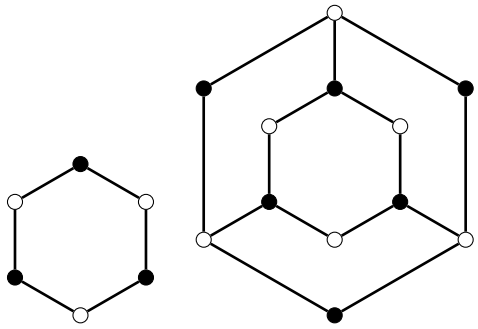
This theorem is evident from the considerations in [100]. Nevertheless, we will give a short formal proof here:

Proof. By [102] each finite benzenoid has two consecutive peripheral vertices of valence 2, say u and v. Both vertices u and v are attached to the outer cycle C of $A(\mathcal{B})$ to vertices u' and v' respectively that are two-apart in C. Let w' be the vertex on C adjacent to both u' and v'. Vertices $uvv'w'u'$ form a cycle of length 5. Altan $A(\mathcal{B})$ is not bipartite.

Note that the result follows also from Theorem 7.1. Although $\mathcal{B}$ is bipartite, the corresponding peripheral root S is not all coloured with the same color. □

For a bipartite graph G and a peripheral root S we may define partition S_b and S_w with black and white coloured sets. For a connected G the partition is unique. We can consider two bipartite altans $A(G, S_b)$ and $A(G, S_w)$. In case of benzenoids the definition is natural and the bipartite altans are determined by the benzenoid itself.

Example 7.3. Black and white altans may be isomorphic (see Figures 7.3 and 7.4) or not (see Figure 7.5.) □



(a) $\mathcal{B}_1$ (b) $A_w(\mathcal{B}_1) \cong A_b(\mathcal{B}_1)$

Figure 7.3: Black and white altans are isomorphic: $A_w(\mathcal{B}_1) \cong A_b(\mathcal{B}_1)$.

Proposition 7.4. *If G is bipartite, then both $A(G, S_b)$ and $A(G, S_w)$ satisfy conditions of Theorem 7.1.*

Proof. Since all vertices of S_b are coloured black and all vertices of S_w are coloured white, Theorem 7.1 applies to both black and white altans and the conclusion follows in each case. □

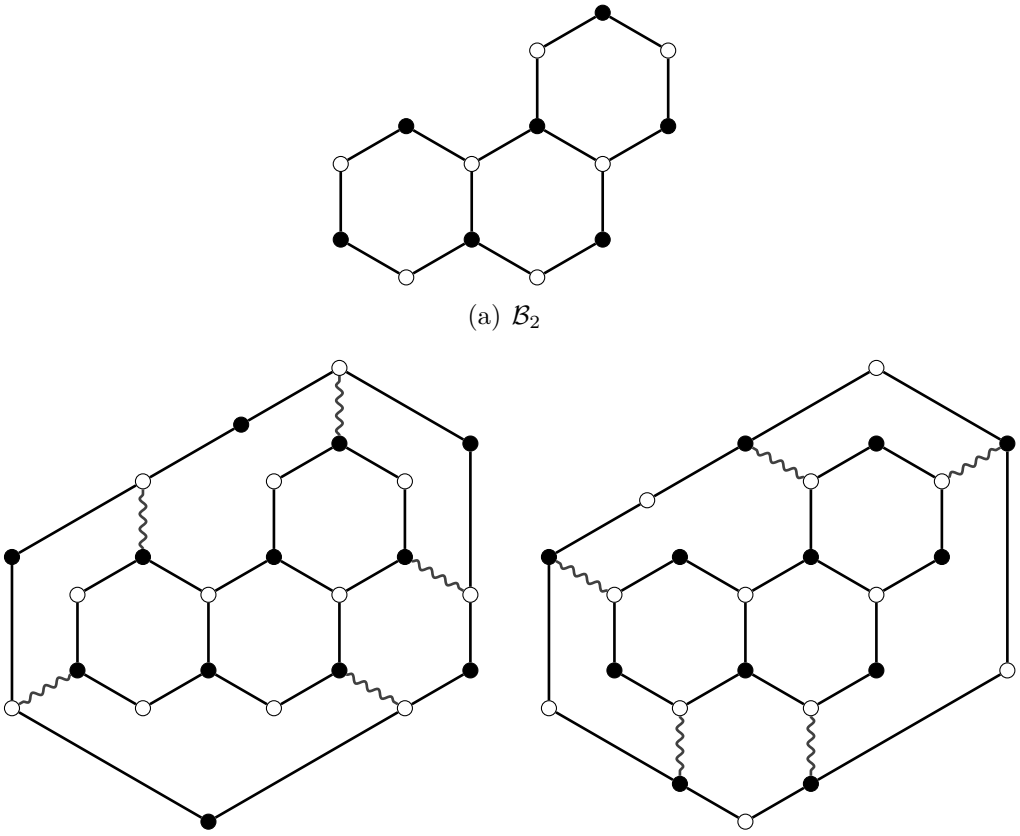
In [99, 100] it was shown that $A(G, S)$ has twice as many Kekulé structures as G. We note that this number is independent of the order in which we choose the vertices of S. It is even independent of the choice of S itself. It is not hard to see that $A^n(G)$ at some point becomes similar to a hexagonal nanotube with the original graph G as part of the cap. The structure of the periphery of $A^n(G, S)$ is composed of a ring of k hexagons where $k = |S|$. Here we give the answers to the question of which benzenoids $\mathcal{B}$ or more general fullerene patches will give rise to a capped nanotube $A^n(\mathcal{B})$.

7.3 Altans of fullerene- and other patches

For a patch Π from Definition 6.1, the altan of the patch, $A(\Pi)$, is clearly defined, i.e., its peripheral root contains cyclically ordered vertices of valence 2 on the perimeter.

Proposition 7.5. *If Π is a patch with k vertices of valence 2, then $A(\Pi)$ is a patch with k vertices of valence 2 and boundary-edges code $\mathbf{2}^k$. Furthermore, the faces of Π are augmented by a ring of k faces to form the internal faces of $A(\Pi)$.*

Proof. This follows directly from the definition of altans. □



(a) $\mathcal{B}_2$

(b) $A_b(\mathcal{B}_2)$ (c) $A_w(\mathcal{B}_2)$

Figure 7.4: Black and white altans are isomorphic, $A_b(\mathcal{B}_2) \cong A_w(\mathcal{B}_2)$, but are oppositely oriented.

A $(k, 1)$-nanotube is a ring of k hexagons. We may view it as the black altan of a cycle on $2k$ vertices: $A_b(C_{2k})$. If we glue s such rings one on top of the other we obtain a (k, s)-nanotube. Be warned that this notation is used by various authors to denote something else. From Proposition 7.5 we obtain:

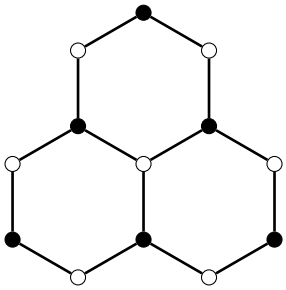
Corollary 7.6. *If* Π *is a patch with boundary code* $\mathcal{2}^k$ *then the ring of faces attached to* Π *when forming* $A(\Pi)$ *is a* $(k, 1)$*-nanotube.* □

Corollary 7.7. *Let* Π *be an arbitrary patch with* k *vertices of valence 2. The iterated altan* $A^n(\Pi)$ *is composed of* $A(\Pi)$ *to which a* $(k, n-1)$*-nanotube is attached.*

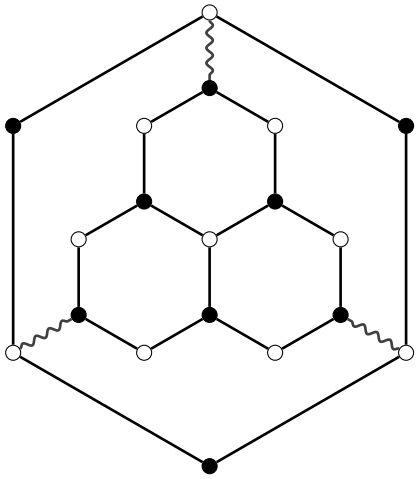
Proof. It follows easily by induction. □

Proposition 7.8. *Let* Π *be a patch. The ring of new faces of* $A(\Pi)$ *contains only faces of length* ≥ 5 *and* $A(\Pi)$ *is convex.*

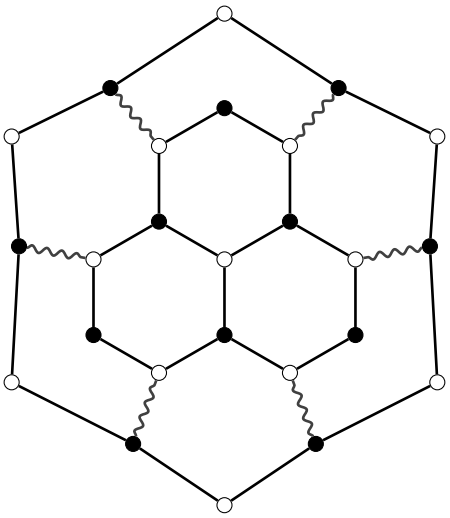
Proof. Each vertex of valence 2 on the peripheral cycle is adjacent to two vertices of valence 3. Therefore each new face of $A(\Pi)$ contains at least 3 peripheral vertices. It must contain at least two old vertices. Actually it will contain two consecutive old vertices if and only if



(a) $\mathcal{B}_3$



(b) $A_b(\mathcal{B}_3)$



(c) $A_w(\mathcal{B}_3)$

Figure 7.5: Black and white non-isomorphic altans $A_b(\mathcal{B}_3) \ncong A_w(\mathcal{B}_3)$.

they were adjacent boundary vertices of valence 2 of Π. By definition $A(\Pi)$ is convex by Proposition 7.5. □

Proposition 7.9. *Let Π be a patch. The ring of new faces of $A(\Pi)$ contains only faces of length 5 or 6 if and only if Π is convex.*

Proof. The pentagons are covered already in the proof of Proposition 7.8. A face of length > 6 appears if and only if the boundary edges code contains one or more consecutive 1s. □

Theorem 7.10. *A benzenoid $\mathcal{B}$ will give rise (by applications of the altan operation) to a fullerene nanotube (capped on one side) if and only if $\mathcal{B}$ is convex.*

Proof. It follows by induction from Proposition 7.9. □

Theorem 7.11. *If Π is a convex fullerene patch with p pentagons and boundary-edges code $BEC(\Pi) = (2+a_1)(2+a_2)\ldots(2+a_k)$, $a_i \geq 0$, $d = a_1 + a_2 + \ldots + a_k$. Then $d + p = 6$ and $A(\Pi)$ is a fullerene patch with boundary-edges code 2^{k+d} and 6 pentagons.*

The patch with a single hexagon (its boundary-edges code is 6) is an exception. Its altan is a fullerene patch with 6 pentagons and boundary-edges code 2^6.

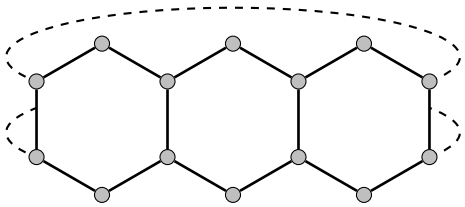


Figure 7.6: $(k, 1)$-nanotube

This result may be viewed as a consequence of the fact that a nanotube may be capped only by a fullerene patch having 6 pentagons. All action takes place at the initial step passing from Π to $A(\Pi)$ since the width of the tube does not change after eventual iterated applications of altan operation.

7.4 Kekulé structures of iterated altans

We have seen that $A^n(G)$ behaves essentially like a capped nanotube. In the first step from G to $A(G)$ all irregularities happen. After that each $A^{n+1}(G)$ is obtained from $A^n(G)$ by attaching a ring of hexagons to its periphery.

Gutman [99] proved the following:

Lemma 7.12 (Gutman, 2014)**.** *Let G be a graph having $K = K(G)$ Kekulé structures. Then any of its altans $A(G)$ has $2K$ Kekulé structure, i.e., twice the number of Kekulé structures.* □

We may apply it to the iterated altans:

Theorem 7.13. *The number of Kekulé structures in the n-th iterated altan of G is $2^n K(G)$.*

Proof. From Gutman's Lemma by induction. □

This result has many interesting consequences. The first one is confirmation of the result of Sachs et al. [182] that the number of Kekulé structures of a (k, s)-nanotube is independent of k and is equal to 2^{s+1}. More generally we may compute the number of Kekulé structures of any patch extended by a nanotube.

Corollary 7.14. *Let Π be a patch then $K(A^n(\Pi)) = 2^n K(\Pi)$.*

In particular this means that a nanotube capped by six pentagons (half of a dodecahedron) has no Kekulé structures, while the nanotube capped by half of the buckyball and its 11 Kekulé structures gives rise to the total of 11×2^n Kekulé structures. We should mention that our nanotubes correspond to a very special *untwisted* case of much more general nanotubes, alias *tubules* considered in the 73-page paper [182].

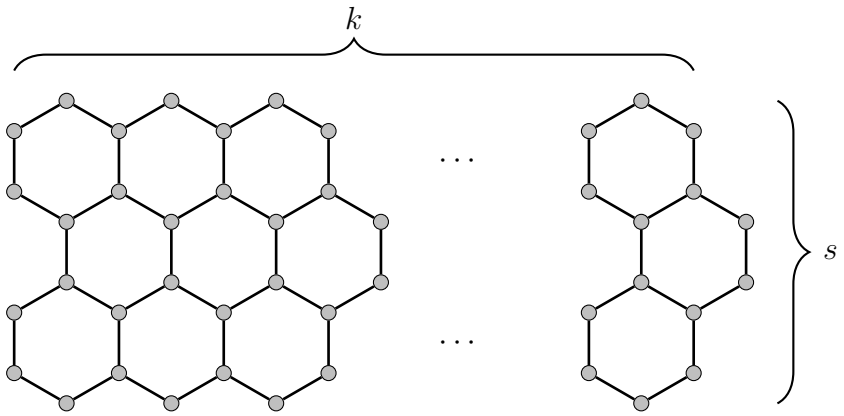


Figure 7.7: (k, s)-nanotube

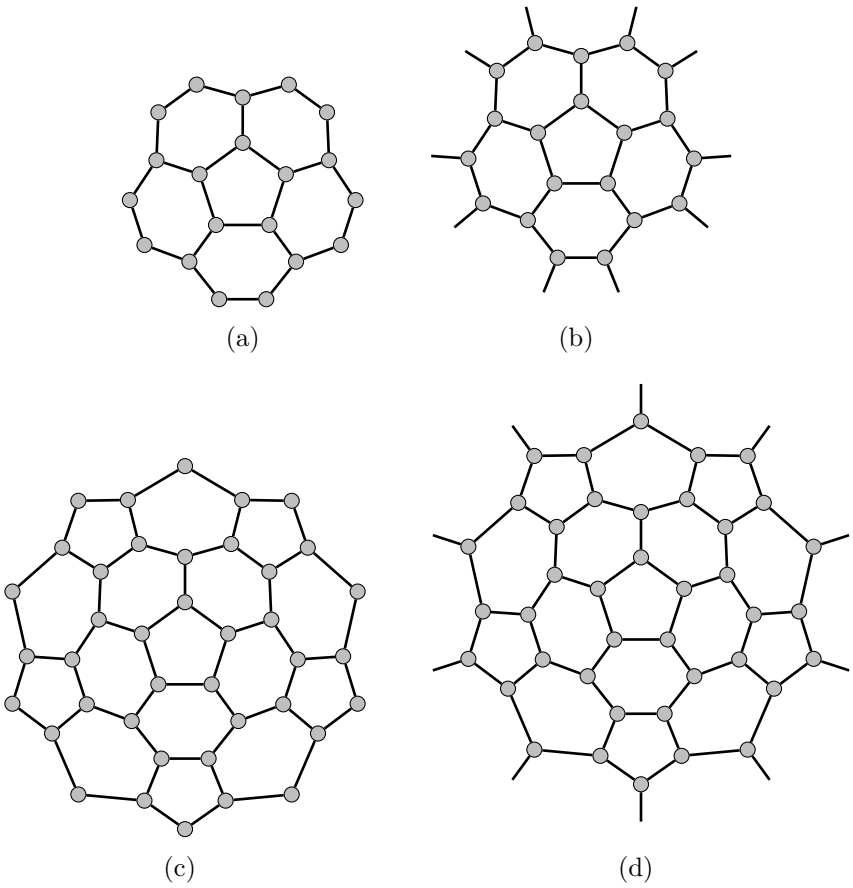
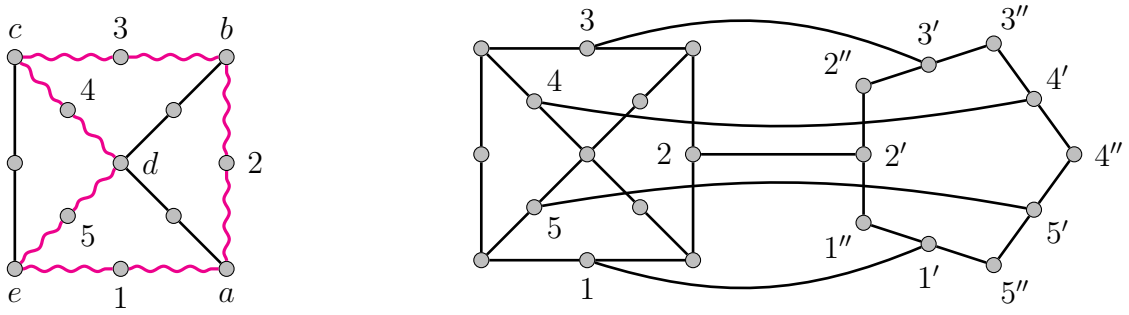


Figure 7.8: A cap Π with a single pentagon (see (a), (b)) that turns into a bucky-ball dome $A(\Pi)$ with six pentagons (see (c), (d)) after the first altan and to a longer capped nanotube after any additional altan operation.

7.5 Generalised altans and iterated altans

We will start by making a small extension to the definition of an altan as presented previously. Let G be a graph and let C, *the perimeter*, be a cycle in G having $k \geq 2$ vertices of degree 2. Then $A(G, C)$ will be the altan as defined previously with respect to the degree-2 vertices of C. Those edges which connect degree-3 vertices on the new cycle with C will be called *spokes*.

Example 7.4. See Figure 7.9. □



(a) G, $C = (1, a, 2, b, 3, c, 4, d, 5, e)$ (b) $G' = A(G, C), C' = (1', 1'', 2', 2'', 3', 3'', 4', 4'', 5', 5'')$

Figure 7.9: A graph G with a designated perimeter C (on the left), and its altan $A(G, C)$ (on the right).

A generalised altan is obtained by selecting a collection of cycles $C_1, C_2, \ldots, C_k$ in G with the property that any degree-2 vertex of G appears on at most one of the cycles and that each of the cycles C_i contains at least 2 vertices of degree 2. We call $(G; C_1, \ldots, C_k)$ an *admissible structure*. In addition we select a non-empty subset of indices $J \subseteq \{1, 2, \ldots, k\}$ and perform the altan operation on all cycles C_j, $j \in J$. We define

$$A(G; C_1, \ldots, C_k; \emptyset) = (G; C_1, C_2, \ldots, C_k),$$

and

$$A(G; C_1, \ldots, C_k; J) = A(G'; C_1, \ldots, C_{j-1}, C'_j, C_{j+1}, \ldots, C_k; J \setminus j),$$

for $J \neq \emptyset$, where $j = \min J$. Graph G' is obtained from G by adding a new copy, denoted C'_j, of a cycle on $2d$ vertices, where d is the number of degree-2 vertices on cycle C_j. Every second vertex on cycle C'_j is attached to a degree-2 vertex on C_j. Usually we take either $|J| = 1$ or $|J| = k$. In the former case we are dealing with the ordinary altan operation. In the latter case all perimeters are used.

Example 7.5. Let G be the subdivided cube in Figure 7.10. Let $C_1 = (1, 2, 3, 4, 7, 6, 5)$, $C_2 = (1, 5, 9, 10, 14, 8)$ and $C_3 = (10, 11, 12, 13, 16, 15, 14)$. Then $(G; C_1, C_2, C_3)$ is an admissible structure. Note that cycles $C_1, \ldots, C_k$ of an admissible structure need not be disjoint as long as no degree-2 vertex lies on a shared part. The generalised altan $A(G; C_1, C_2, C_3; \{1, 2, 3\})$ is on the right in Figure 7.10.

We may apply the generalised altan operation iteratively. The order in which we apply individual "local" altan operations is irrelevant.

7.5.1 Iterated generalised altans

Let $(G; C_1, C_2, \ldots, C_k)$ be an admissible structure. Let $\mathbf{n} = (n_1, n_2, \ldots, n_k)$ be an integer vector with $n_i \geq 0$. Let $\mathbf{n}^\dagger = (n_1^\dagger, n_2^\dagger, \ldots, n_k^\dagger)$ where

$$n_i^\dagger = \begin{cases} n_i - 1, & \text{if } n_i > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, let $J_{\mathbf{n}}^\dagger = \{i \mid n_i > 0\}$. Then

$$A^{\mathbf{n}}(G; C_1, \ldots, C_k)$$

denotes the iterated generalised altan which is defined in terms of the generalised altan in the following way:

$$A^{\mathbf{n}}(G; C_1, \ldots, C_k) = A^{\mathbf{n}^\dagger}(A(G; C_1, \ldots, C_k; J_{\mathbf{n}}^\dagger)).$$

Naturally, $A^{\mathbf{0}}(G; C_1, \ldots, C_k) = (G; C_1, \ldots, C_k)$ where $\mathbf{0} = (0, 0, \ldots, 0)$.

Example 7.6. Consider the admissible structure $(G; C_1, C_2, C_3)$ from Example 7.5. The iterated generalised altan $A^{(1,0,2)}(G; C_1, C_2, C_3)$ is shown in Figure 7.11.

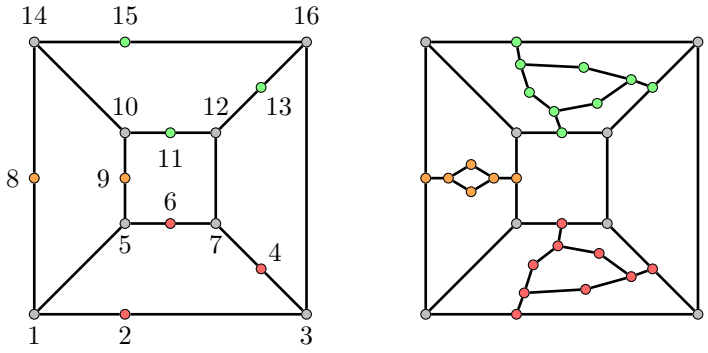


Figure 7.10: A subdivided cube (on left) and its generalised altan (on right).

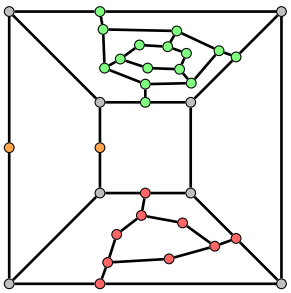


Figure 7.11: The iterated generalised altan $A^{(1,0,2)}(G; C_1, C_2, C_3)$.

7.5.2 Altans of coronoids and perforated patches

From now on, by a coronoid we mean a non-degenerate coronoid. Each coronoid $\mathcal{K}$ with its perimeters is an admissible generalised altan structure. Hence, $A^{\mathbf{n}}(\mathcal{K})$ is well-defined as soon as we label its perimeters. Note that the cycles are exactly perimeters and they are disjoint. When $\mathbf{n} \neq \mathbf{0}$, we call $A^{\mathbf{n}}(\mathcal{K})$ a *proper generalised altan*.

Example 7.7. Let $\mathcal{K}$ be the coronoid in Figure 7.11 (the part consisting of dotted faces). Let C_1 denote the left inner perimeter, C_2 the right inner perimeter and C_3 the outer perimeter. Generalised altan $A^{(2,3,0)}(\mathcal{K})$ consists of shaded and dotted faces in Figure 7.12.

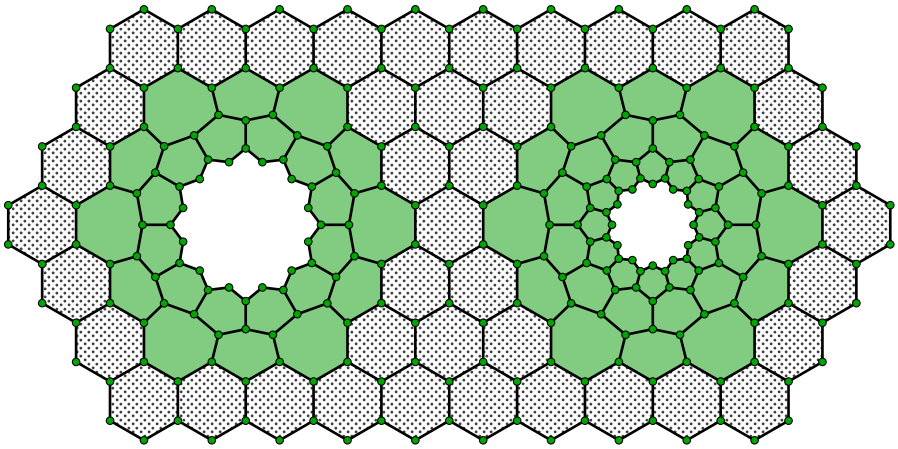
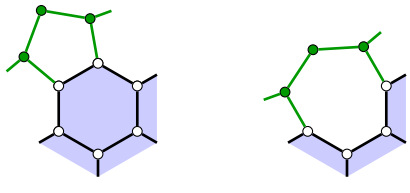


Figure 7.12: $A^{(2,3,0)}(\mathcal{K})$.

Proposition 7.15. *A proper generalised altan of a coronoid is not a coronoid.*

Proof. In the case of a benzenoid, we restate Gutman's observation [100] on the structural features of altan-benzenoids. Every benzenoids contains a $(2, 2)$-edge, i.e., an edge connecting two degree-2 vertices. Figure 7.13(a) shows a fragment of a benzenoid with a $(2, 2)$-edge. This gives rise to a pentagon in its altan. (The new vertices that are obtained by the altan operation are distinguished by shading.)



(a) (b)

Figure 7.13: The altan of a coronoid is no longer a coronoid.

In the case of a coronoid, the same proof works for the outer perimeter. For an inner perimeter, one can observe that it must contain at least one $(3, 3)$-edge which corresponds to a $(2, 2)$-edge in its corona hole. This means that there are at least two vertices of degree

3 between some pair of degree-2 vertices. This give rise to a heptagon or an even larger face (see Figure 7.13(b)). □

However, Gutman has shown [100] that when the altan operation is performed on a convex benzenoid [47], degrees of the newly obtained faces are limited to 5 and 6. Traditionally, a patch is defined as a subcubic plane graph that has all its degree-2 vertices on its outer perimeter. Clearly, the skeleton $G(\mathcal{P})$ of a patch $\mathcal{P}$ is such a graph. The following result shows that our definition actually coincides with the traditional one:

Proposition 7.16. *Let G be a plane 2-connected subcubic graph with all the degree-2 vertices on its outer perimeter. Then there exists a plane cubic graph $\widetilde{G}$, such that $G \subseteq \widetilde{G}$, all inner faces of G are also faces of $\widetilde{G}$ and there exists a patch $\mathcal{P} \subseteq \mathcal{F}_{\widetilde{G}}$ such that $G = G(\mathcal{P})$. Moreover, if G has at least two degree-2 vertices then there exists a 2-connected graph $\widetilde{G}$.*

Proof. If G contains three or more degree-2 vertices, choose C to be the outer perimeter of G and make altan $A(G, C)$. Then remove all the newly obtained degree-2 vertices and reconnect its neighbours (reverse operation to subdivision) as shown in Figure 7.14. It is

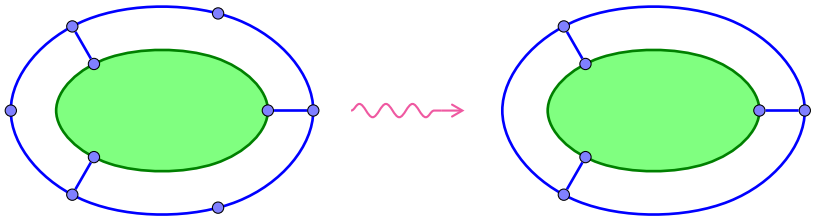


Figure 7.14: Obtaining $\widetilde{G}$ from G with three or more degree-2 vertices.

trivial to verify that this graph is indeed the desired $\widetilde{G}$. If G was 2-connected then its ear-decomposition can easily be extended to include the newly obtained edges. This shows that $\widetilde{G}$ is also 2-connected.

If G had only two degree-2 vertices, then the above procedure would yield a multigraph. It can be fixed by subdiving its edges on the new outer perimeter to obtain 3 or more degree-2 vertices and repeating this operation as shown in Figure 7.15. Again, it is not

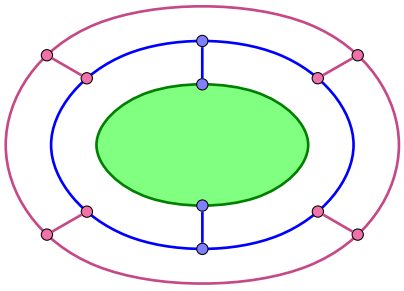


Figure 7.15: Obtaining $\widetilde{G}$ from G with two degree-2 vertices.

hard to see that this yields the desired $\widetilde{G}$ and that the ear-decomposition of G can be extended to $\widetilde{G}$.

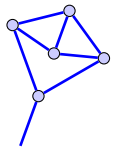


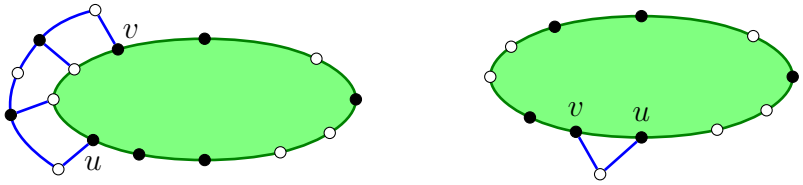
Figure 7.16: A "cherry".

If G has only one degree-2 vertex, there is no hope of obtaining 2-connected $\widetilde{G}$. Graph $\widetilde{G}$ can be obtain from G by attaching a "cherry" (see Figure 7.16) to its degree 2 vertex. □

We can say more:

Proposition 7.17. *Let G be a 2-connected plane subcubic* bipartite *graph with all degree-2 vertices on its outer perimeter. Then there exists a plane cubic* bipartite *graph $\widetilde{G}$, such that $G \subseteq \widetilde{G}$, all faces of G are also faces of $\widetilde{G}$ and there exists a patch $\mathcal{P} \subseteq \mathcal{F}_{\widetilde{G}}$ such that $G = G(\mathcal{P})$. Moreover, there exists a 2-connected graph $\widetilde{G}$ with desired properties.*

Proof. Suppose there are at least two degree-2 vertices of the same colour, i.e., belong to the same set of the bipartition, in graph G and denote those two vertices by u and v. Without loss of generality, we can assume they are coloured black. Those two vertices can be choosen in such way that there are no other black vertices between them when we traverse the perimeter from u to v in the clockwise direction (see Figure 7.17(a)). However,



(a) (b)

Figure 7.17: A step in obtaining bipartite $\widetilde{G}$ from bipartite G.

there may be 0 or more white vertices on that path. Say there are one or more white vertices. Label those vertices $w_1, w_2, \ldots, w_l$, where $l \geq 1$. Make an altan of G with vertices $(u, w_1, \ldots, w_l, v)$ as its *peripheral root*. Label the newly obtained neighbours of vertices u and v with u' and v', respectively. Then make a reverse subdivision operation which removes all degree-2 vertices in the neighbourhood of u' and v'. Also, remove the edge $u'v'$ to obtain the graph in Figure 7.17(a). If there are no white vertices between u and v, add a new vertex to the graph and connect it to u and v as shown in Figure 7.17(b). In both cases, this graph is clearly bipartite and 2-connected if the graph G was 2-connected. Also, two black degree-2 vertices and l white degree-2 vertices have disappered and $l + 1$ new white degree-2 vertices have emerged. The total number of degree-2 vertices is therefore decreased by one.

This procedure terminates when there are only two vertices left, which have to be of different colours. (The situation with only one degree-2 vertex, say a white vertex, cannot occur. The number of edges should be divisible by 3, because every black vertex has degree

3. On the other hand, there is one white vertex of degree 2 and the rest have degree 3, which implies that the number of edges is congruent 2 modulo 3, a contradiction.) If those

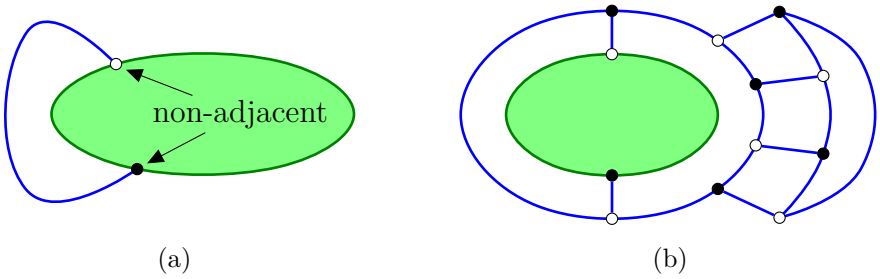


Figure 7.18: The final step in obtaining bipartite $\widetilde{G}$ from bipartite G.

final two vertices (which have to be of different colour) are non-adjacent, connect them by an edge as shown in Figure 7.18(a). If they are adjacent, we would create a multigraph. In that case, use the construction shown in Figure 7.18(b) to avoid the multigraph. □

A perforated patch with pentagonal and hexagonal faces is called a *perforated fullerene patch*. Similary, a patch with pentagonal and hexagonal faces is called a *fullerene patch*. One should be aware that the restriction to hexagonal and pentagonal faces applies within the patch $\mathcal{P}$; other faces of the cubic graph G from which the patch was derived may be of other sizes. If such a graph G with exclusively pentagonal and hexagonal faces exists, then the patch (or perforated patch) can be extended to a fullerene. It is not easy to verify the existence of such G.

Example 7.8. Figure 7.19 shows that various possibilities can occur when we apply the altan operation to a fullerene patch. In the first case (left-hand side of Figure 7.19), the altan contains a 7-gon. In the second case (right-hand side), the altan is again a fullerene patch (which may or may not extend to a fullerene).

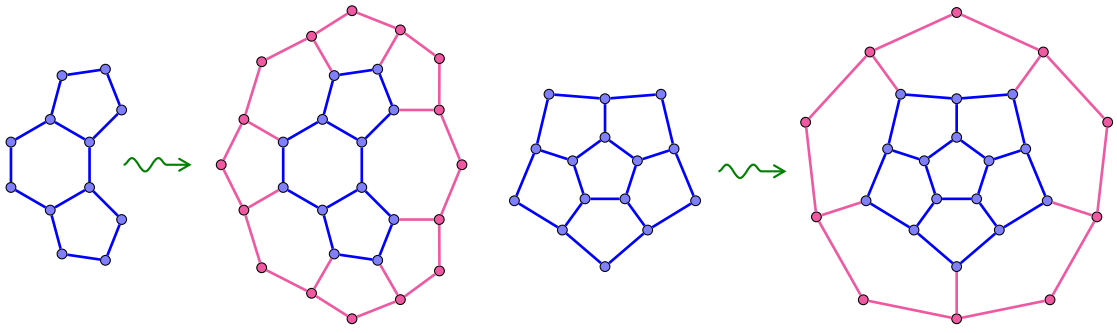


Figure 7.19: Altans of fullerene patches.

There is another viewpoint we can take when dealing with (perforated) patches. In addition to the skeleton $G(\mathcal{P})$, one can also obtain a planar *pre-graph*, denoted $P(\mathcal{P})$. It can be

obtained from the plane graph G by removing all vertices that are not incident to any face of $\mathcal{P}$, together with all semiedges that are incident to removed vertices. In addition, edges incident to two faces from $\mathcal{P}^{\complement}$ are also removed and replaced with two half edges (as if the edge was cut in the middle). An example is given in Figure 7.20.

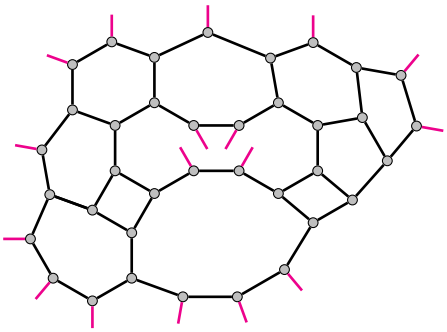


Figure 7.20: A pre-graph of a patch. Half edges are "without vertices" on one end, i.e., they are "dangling".

Theorem 7.18. *The generalised altan of a perforated patch $\mathcal{P}$ is a perforated patch. Moreover, if $G(\mathcal{P})$ is 2-connected then $G(A^{\mathbf{n}}(\mathcal{P}))$ is also 2-connected.*

Proof. The pre-graph of a perforated patch $\mathcal{P}$ is schematically illustrated in Figure 7.21. Each hole corresponds to a void space that can be filled with an open disc. There are also half edges (attached to degree-2 vertices of $G(\mathcal{P})$) which are drawn inside those holes. When we perform an altan operation on that perforated patch, a cycle is drawn inside every

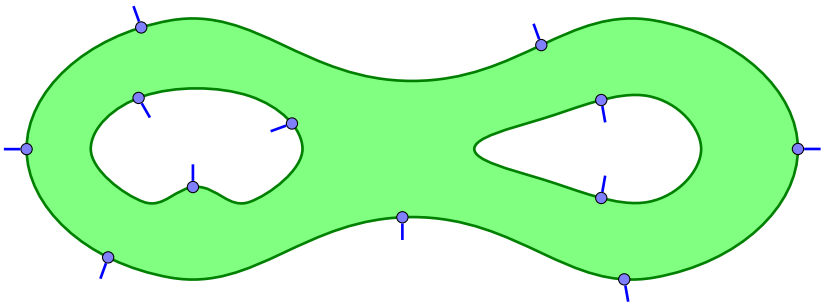


Figure 7.21: Pre-graph of a perforated patch.

hole on which the altan operation is performed and an annulus of new faces (bounded by the new and the old perimeter) is added to the patch. See Figure 7.22 for an illustration. New holes have the same number of half edges as they had before the operation. The parts that were removed from G can be reattached to form the plane cubic graph.

It is clear that $G(A^{\mathbf{n}}(\mathcal{P}))$ is 2-connected if $G(\mathcal{P})$ is 2-connected. The ear decomposition of $G(\mathcal{P})$ can easily be extended to include the newly obtained edges. □

The following corollary obviously follows from Theorem 7.18:

Corollary 7.19. *The generalised altan of a patch $\mathcal{P}$ is a patch. Moreover, if $G(\mathcal{P})$ is 2-connected then $G(A^{\mathbf{n}}(\mathcal{P}))$ is also 2-connected.*

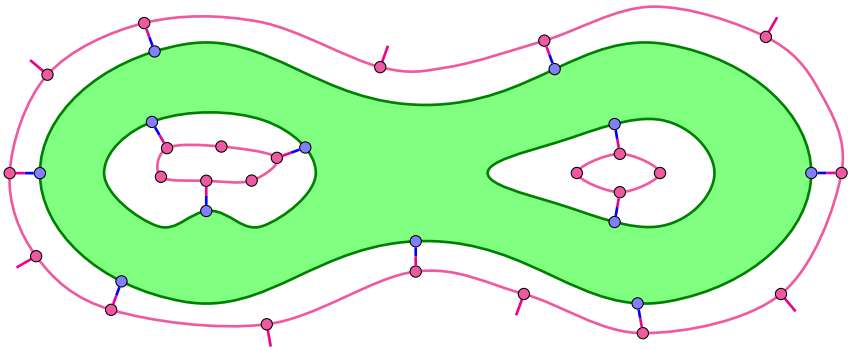


Figure 7.22: Pre-graph of an altan of a perforated patch.

7.6 Kekulé structures and Pauling Bond Orders

It is not hard to see that iteration of the generalised altan operation on coronoids and perforated patches grows tubes on each perimeter, i.e., we can visualize an embedded version in a way that is reminiscent of the classic ruled surface of the graph in which some central planar perforated patch has tubular towers growing out of it (in either up or down directions). See Figure 7.24.

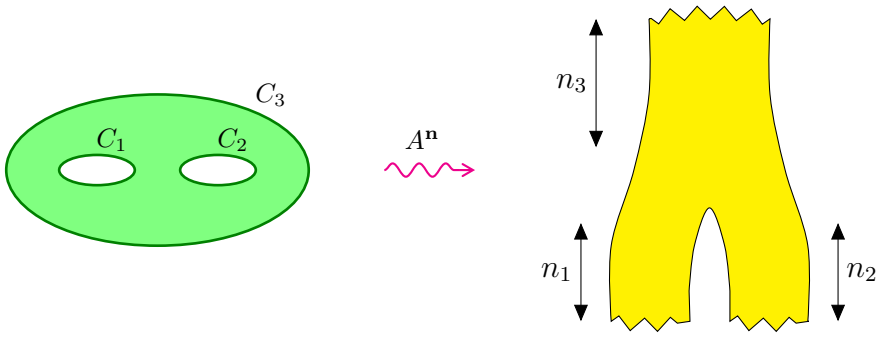


Figure 7.23: Pants resulting from a disk with two holes by applying iterated altan operation. If the disk has K Kekulé structures then the pants have $K' = 2^{n_1+n_2+n_3}K$ Kekulé structures.

The *binary boundary code* for benzenoids is described in [170]. (This code is also known as PC-1 [110].) The binary boundary code of a benzenoid is a sequence of degrees of consecutive vertices alongs its perimeter. Cyclic shifts and reversal of the sequence are considered as equivalent codes. Traditionally, ones and zeroes were used, but we will use 3s and 2s instead. To each perimeter C_i of an admissible structure $(G; C_1, \ldots, C_k)$ we will assign a binary boundary code, denoted $BBC(C_i)$. Boundary-edges codes for benzenoids, introduced in [107], are useful on many occasions [125], but in this case binary boundary codes are more natural.

Example 7.9. The first coronoid in Figure 5.12 has three perimeters. Let C_∞ denote the outer perimeter and C_1 and C_2 the inner perimeters. Then

$$BBC(C_\infty) = 32322233232233222323223232322322 32$$

and

$$BBC(C_1) = BBC(C_2) = 3333233332.$$

We will use 3^k as a short form of $\underbrace{33\ldots3}_{k}$. In this convention,

$$BBC(C_1) = 3^4234^2 = (3^42)^2.$$

Theorem 7.20. *Let G be a coronoid and let $BBC(C) = 23^{\ell_1}23^{\ell_2}2\ldots23^{\ell_d}$, where $l_i \geq 0$ for $i = 1, \ldots, d$, be the binary boundary code for one of its perimeters C. The degree of the i-th newly obtained face, $1 \leq i \leq d$, is $\ell_i + 5$. Moreover, the binary boundary code of the new boundary is $(32)^d$.*

Theorem 7.21. *Let G be a perforated patch with K Kekulé structures and let $G' = A^{\mathbf{n}}(G)$ be any of its generalised altans. Then the number of Kekulé structures in G' is given by $K' = 2^{|\mathbf{n}|}K$, where $|\mathbf{n}| = n_1 + n_2 + \cdots + n_k$. Furthermore:*

(a) No spoke belongs to a Kekulé structure.

(b) If $n_i > 0$, all edges on the new perimeter belong to the same number, $\frac{K'}{2}$, of Kekulé structures.

The following corollary follows straightforwardly from the above theorem:

Corollary 7.22. *A generalised altan $A^{\mathbf{n}}(G)$ is Kekulean if and only if G is Kekulean.* □

Corollary 7.23. *Let G be a perforated patch and let $G' = A^{\mathbf{n}}(G)$ be any of its generalised altans. The Pauling Bond Order of the newly obtained edge e is:*

(a) 0 if e is a spoke,

(b) $\frac{1}{2}$ if e is not a spoke.

Pauling Bond Orders of the edges that belong to the original graph G remain the same. □

Note that graph $A^{\mathbf{n}}(G)$ was obtained from G by adding new vertices and edges. Therefore, G is a subgraph of $A^{\mathbf{n}}(G)$ in a natural way.

Generalised altans are models of carbon nanostructures that are constructed by attachment of carbon towers [131, 182] to the holes in coronoid patches. Kekulé structures and Pauling Bond Orders (and by implication ring currents [72, 82, 176]) of the nanostructure can be derived in terms of those of the undecorated structure.

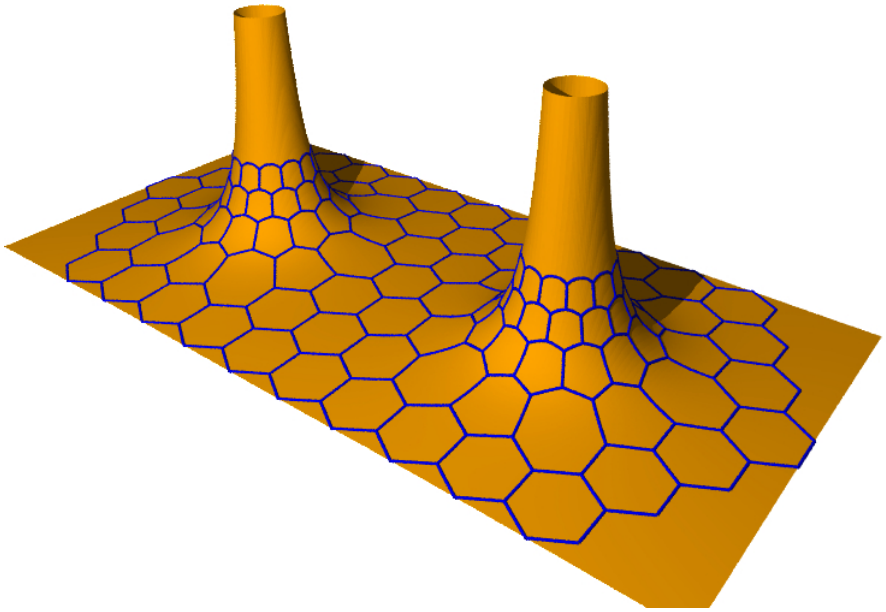


Figure 7.24: Iteration of the altan construction leads to a carbon nanostructure in which nanotubes grow out of the original holes of the coronoid. A given tube may grow up or down, as a 'chimney-stack' or a 'mine-shaft' on the graphene-like landscape, leading to isomeric structures that share a common molecular graph.

Chapter 8

Conclusions

Chemical graph theory is one amongst the most applied areas within graph theory. It began in the 19th century with enumeration of *alkanes*, i.e., enumeration of non-isomorphic trees with bounded degree. After the rise of *quantum mechanics* in the first half of the 20th century, the *Hückel Molecular-Orbital theory* played an important role. In the second half of the 20th century, the study of *topological indices* gained on popularity. Topological indices (in graph theory they are often referred to as *graph invariants*) are molecular descriptors based on the molecular graph of a compound under investigation. They made it possible to obtain good estimates of molecules' properties "in silico", i.e., by computer searches within certain families of graphs. Experiments "in vitro" and "in vivo", on the other hand, are much more expensive. While we are happy to observe that our work has been noted in the mathematical and chemical community [16, 204], there is still a lot of work that needs to be done.

In this work, we have revisited the *molecular symmetry* from both mathematical and chemical viewpoint. Chapter 3 which introduces the Hückel theory (and several other important chemical concepts, e.g. Kekulé structures) from a mathematician's perspective, serves as a between mathematics and chemistry.

In Section 3.5 we have touched the field of *synthetic biochemistry* by introducing mathematical foundations behind the polyhedral self-assembly. We described a *dynamic programming* algorithm for enumeration of *map traces*. We will work further in this area to obtain an algorithm that will be able to enumerate exclusively *parallel* or exclusively *antiparallel* stable traces which are of great interest to biochemists. We also plan to determine the strong traces of *fullerenes*. Some day, chemistry might be able to provide large enough sets of *orthogonal peptide pairs* to enable synthesis of *fullerene-shaped nanostructures* out of proteins.

The *Coulson conjecture on maximal bond number* still remains unproved in its full generality. In Chapter 4, we gave a few partial results which may lead to the ultimate solution. We also posed some conjectures which are also interesting from a purely mathematical viewpoint. There are several other approaches to this that can be considered. We plan to study how operations on graphs affect the Coulson bond number. In ideal scenario, we hope to be able to find operations that strictly increase the maximal bond number and a set of base graphs from which all maximal elements are obtained by iteratively using this

operation.

Another thing that can be done is the classification and enumeration of chemical graphs with a given number of positive eigenvalues n_+. So far, we only provide the proof for the case when $n_+ = 1$.

In the paper [43], Coulson provides alternative formulae for the Coulson π bond order. Let $\Delta(\varepsilon)$ denote the secular determinant, i.e., $\Delta(\varepsilon) = \det(\mathbf{H} - \varepsilon I_{n,n})$. Let m be the index of the highest bonding molecular orbital and assume that all bonding molecular orbitals are doubly occupied. Then

$$P^{\pi}_{rs} = 2(-1)^{r+s+1} \sum_{j=1}^{m} \frac{\Delta_{r,s}(\varepsilon_j)}{\Delta'(\varepsilon_j)}, \tag{8.1}$$

where $\Delta_{r,s}(\varepsilon)$ is the determinant obtained from $\Delta(\varepsilon)$ by striking out the r-th row and s-th column; $\Delta'(\varepsilon_j)$ denotes $\frac{\partial\Delta(\varepsilon)}{\partial\varepsilon}$ evaluated at ε_j. Another formula provided by Coulson is

$$P^{\pi}_{rs} = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial}{\partial\beta_{rs}} \log\Delta(iy)\,\mathrm{d}y = (-1)^{r+s+1} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\Delta_{r,s}(iy)}{\Delta(iy)}\,\mathrm{d}y. \tag{8.2}$$

We will attempt to derive an analytical proof of this conjecture. Moreover, our current algorithms that calculate the Coulson bond number of families of graphs can easily be parallelized. We could extend the lists even further by the use of grid computers.

The most important contributions reside in Chapters 5, 6 and 7. New foundations have been laid in the theory of benzenoids and their generalisations (i.e., coronoids, patches and perforated patches). The family of *convex benzenoids* was studied in great detail. We plan to study the properties of *pseudo-convex* benzenoids to the same extent in the near future.

The *altan* operation, which comes from chemistry, was given graph-theoretical foundations and was further generalised and then applied back to chemistry. Note that in a recent paper by Fowler and Myrvold [73] it is shown that the ring currents in boundary hexagonal rings of Kekulean benzenoids, as estimated within the conjugated-circuit model, can be calculated directly from the Pauling bond orders of boundary edges and the Kekulé number of the benzenoid. In this work, the ground work has been layed for determining the π electron currents on chimney-stack graphene-like structures that we obtained by the altan operation.

A small original contribution is the introduction of the *Pentagonal Incidence Partition* (abbreviated as PIP) of *fullerenes*. The notion of a PIP fullerene can be further refined. Assume that the $PIP(\mathcal{F})$ of a fullerene $\mathcal{F}$ has a part 4. This can be realised with different pentagonal patches. We plan to refine the notion yet again to distinguish between differently shaped patches. We plan to extend the search beyond our current limit $n = 100$ vertices and classify those partitions into *admissible* and *non-admissible*. We also investigate whether for a given PIP there exist infinite families of fullerenes that realise them or only a few small sporadic cases.

During our study, a lot of computer code emerged. We plan to refactor and fully document the code and, eventually, make it publicly available in the form of a library that will be called **ChemVEGA**.

Bibliography

[1] OEIS Foundation Inc. (2011), *The On-Line Encyclopedia of Integer Sequences*, `http://oeis.org`, [Online; accessed: 1-August-2015].

[2] E. Albertazzi, C. Domene, P. W. Fowler, T. Heine, G. Seifert, C. Van Alsenoy, and F. Zerbetto, *Pentagon adjacency as a determinant of fullerene stability*, Phys. Chem. Chem. Phys. **1** (1999), no. 12, 2913–2918.

[3] S. L. Altmann and E. J. Bowen, *Charles Alfred Coulson. 1910-1974*, Biographical Memoirs of Fellows of the Royal Society **20** (1974), 74–134.

[4] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3 ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.

[5] E. D. Andrulis, *Theory of the origin, evolution, and nature of life*, Life **2** (2012), no. 1, 1–105.

[6] P. W. Atkins, *Physical Chemistry*, W. H. Freeman and Company, 1994.

[7] D. Babić, A. Graovac, B. Mohar, and T. Pisanski, *The matching polynomial of a polygraph*, Discrete Appl. Math. **15** (1986), no. 1, 11–24.

[8] ———, *Recursion for the matching polynomials of fasciagraphs and rotagraphs*, Proc. Sixth Yugosl. Sem. on Graph Theory (Novi Sad), University of Novi Sad, 1986, pp. 27–41.

[9] A. T. Balaban, *Chemical graphs VII. Proposed nomenclature of branched catacondensed benzenoid polycyclic hydrocarbons*, Tetrahedron **25** (1969), no. 15, 2949–2956.

[10] H.-J. Bandelt and V. Chepoi, *Metric graph theory and geometry: a survey*, Surveys on Discrete and Computational Geometry, Contemp. Math., vol. 453, Amer. Math. Soc., Providence, RI, 2008, pp. 49–86.

[11] N. Bašić and T. Pisanski, *Iterated altans and their properties*, MATCH Commun. Math. Comput. Chem. **74** (2015), no. 3, 653–666.

[12] N. Bašić, D. Bokal, T. Boothby, and J. Rus, *An algebraic approach to enumerating non-equivalent double traces*, submitted, 2015.

[13] N. Bašić, D. Bokal, T. Pisanski, and J. Rus, *Graph embeddings yield natural strong trace realizations*, in preparation.

[14] N. Bašić, P. W. Fowler, and T. Pisanski, *Coronoids, patches and generalised altans*, J. Math. Chem. **54** (2016), no. 4, 977–1009.

[15] N. Bašić, R. Jerala, and T. Pisanski, *Stable polyhedral self-assembly from dimers*, in preparation.

[16] A. Behmaram, T. Došlić, and S. Friedland, *Matchings in m-generalized fullerene graphs*, Ars Math. Contemp. **11** (2016), no. 2, 301–313.

[17] D. M. Bishop, *Group Theory and Chemistry*, Dover Books on Chemistry, Dover Publications, 2012.

[18] O. Bodroža, I. Gutman, S. J. Cyvin, and R. Tošić, *Number of Kekulé structures of hexagon-shaped benzenoids*, J. Math. Chem. **2** (1988), no. 3, 287–298.

[19] A. Bondy and U. S. R. Murty, *Graph Theory*, Graduate Texts in Mathematics, Springer, London, 2011.

[20] K. S. Booth, *Lexicographically least circular substrings*, Inform. Process. Lett. **10** (1980), no. 4-5, 240–242.

[21] S. Brin and L. Page, *The anatomy of a large-scale hypertextual web search engine*, Comput. Networks ISDN **30** (1998), no. 1-7, 107–117, 7th International World Wide Web Conference, Brisbane, Australia, Apr 14-18, 1998.

[22] G. Brinkmann, G. Caporossi, and P. Hansen, *A constructive enumeration of fusenes and benzenoids*, J. Algorithms **45** (2002), 155–166.

[23] ———, *A survey and new results on computer enumeration of polyhex and fusene hydrocarbons*, J. Chem. Inf. Comput. Sci. **43** (2003), no. 3, 842–851.

[24] G. Brinkmann, O. Delgado-Friedrichs, S. Lisken, A. Peeters, and N. Van Cleemput, *CaGe – a virtual environment for studying some special classes of plane graphs – an update*, MATCH Commun. Math. Comput. Chem. **63** (2010), no. 3, 533–552.

[25] G. Brinkmann, O. Delgado-Friedrichs, and U. von Nathusius, *Numbers of faces and boundary encodings of patches*, Graphs and Discovery, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 69, Amer. Math. Soc., 2005, pp. 27–38.

[26] G. Brinkmann and A. W. M. Dress, *A constructive enumeration of fullerenes*, J. Algorithms **23** (1997), no. 2, 345–358.

[27] G. Brinkmann, P. W. Fowler, and M. Yoshida, *New non-spiral fullerenes from old: Generalised truncations of isolated-pentagon-triple carbon cages*, MATCH Commun. Math. Comput. Chem. (1998), no. 38, 7–17.

[28] G. Brinkmann, J. Goedgebeur, and B. D. McKay, *The generation of fullerenes*, J. Chem. Inf. Model. **52** (2012), no. 11, 2910–2918.

[29] ———, *The smallest fullerene without a spiral*, Chem. Phys. Lett. **522** (2012), 54–55.

[30] G. Brinkmann, J. E. Graver, and C. Justus, *Numbers of faces in disordered patches*, J. Math. Chem. **45** (2009), no. 2, 263–278.

[31] G. Brinkmann, C. Grothaus, and I. Gutman, *Fusenes and benzenoids with perfect matchings*, J. Math. Chem. **42** (2007), no. 4, 909–924.

[32] A. E. Brouwer and W. H. Haemers, *Spectra of Graphs*, Universitext, Springer, New York, 2011.

[33] F. H. Burkitt, C. A. Coulson, and H. C. Longuet-Higgins, *Free valence in unsaturated hydrocarbons*, Trans. Faraday Soc. **47** (1951), no. 6, 553–564.

[34] P. J. Cameron, *Permutation Groups*, London Mathematical Society Student Texts, Cambridge University Press, 1999.

[35] A. Ceulemans, E. Lijnen, P. W. Fowler, R. B. Mallion, and T. Pisanski, *Graph theory and the Jahn-Teller theorem*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **468** (2012), no. 2140, 971–989.

[36] ———, *S_5 graphs as model systems for icosahedral Jahn–Teller problems*, Theor. Chem. Acc. **131** (2012), 1–10.

[37] D. Chen, H. Deng, and Q. Guo, *Zhang-Zhang polynomials of a class of pericondensed benzenoid graphs*, MATCH Commun. Math. Comput. Chem. **63** (2010), no. 2, 401–410.

[38] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 3 ed., MIT Press, 2009.

[39] F. A. Cotton, *Chemical Applications of Group Theory*, 3 ed., Wiley-Interscience, 1990.

[40] C. A. Coulson, *The theory of the structure of free radicals*, Discuss. Faraday Soc. **2** (1947), 9–18.

[41] ———, *Free valence in organic reactions*, J. Chim. Phys. Phys.-Chim. Biol. **45** (1948), no. 9-10, 247–248.

[42] ———, *Prof. William Moffitt (Obituary)*, The Times (1958), 8, Tuesday, Dec 30, 1958.

[43] C. A. Coulson and H. C. Longuet-Higgins, *The electronic structure of conjugated systems. I. General theory*, Proc. R. Soc. A **191** (1947), no. 1024, 39–60.

[44] C. A. Coulson, B. O'Leary, and R. B. Mallion, *Hückel Theory for Organic Chemists*, Academic Press, 1978.

[45] H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer-Verlag, 1980.

[46] R. Cruz, H. Giraldo, and J. Rada, *Extremal values of vertex-degree topological indices over hexagonal systems*, MATCH Commun. Math. Comput. Chem. **70** (2013), no. 2, 501–512.

[47] R. Cruz, I. Gutman, and J. Rada, *Convex hexagonal systems and their topological indices*, MATCH Commun. Math. Comput. Chem. **68** (2012), no. 1, 97–108.

[48] ———, *On benzenoid systems with a minimal number of inlets*, Serb. Chem. Soc. **78** (2013), no. 9, 1351–1357.

[49] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Applications*, 3 ed., Johann Ambrosius Barth, 1995.

[50] D. M. Cvetković, P. W. Fowler, P. Rowlinson, and D. Stevanović, *Constructing fullerene graphs from their eigenvalues and angles*, Linear Algebra Appl. **356** (2002), 37–56.

[51] S. J. Cyvin, J. Brunvoll, R. S. Chen, B. N. Cyvin, and F. J. Zhang, *Theory of Coronoid Hydrocarbons II*, Lecture Notes in Chemistry, Springer, Berlin, Heidelberg, 2012.

[52] S. J. Cyvin, J. Brunvoll, and B. N. Cyvin, *Theory of Coronoid Hydrocarbons*, Lecture Notes in Chemistry, vol. 54, Springer-Verlag, 1991.

[53] S. J. Cyvin and I. Gutman, *Kekulé Structures in Benzenoid Hydrocarbons*, Lecture Notes in Chemistry, vol. 48, Springer-Verlag, 1988.

[54] M. Daniel, S. Gravier, and J. Moncel, *Identifying codes in some subgraphs of the square lattice*, Theoret. Comput. Sci. **319** (2004), no. 1-3, 411–421.

[55] J. W. Demmel, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, 1997.

[56] M. Deza, P. W. Fowler, and V. Grishukhin, *Allowed boundary sequences for fused polycyclic patches and related algorithmic problems*, J. Chem. Inf. Comput. Sci. **41** (2001), no. 2, 300–308.

[57] J. R. Dias, *Handbook of Polycyclic Hydrocarbons: Part A, Benzenoid Hydrocarbons*, Physical Sciences Data, Elsevier, 1987.

[58] ———, *Handbook of Polycyclic Hydrocarbons: Part B, Polycyclic Isomers and Heteroatom Analogs of Benzenoid Hydrocarbons*, Physical Sciences Data, Elsevier, 1988.

[59] T. K. Dickens and R. B. Mallion, *π-Electron ring-currents and bond-currents in some conjugated altan-structures*, J. Phys. Chem. A **118** (2014), no. 20, 3688–3697.

[60] ———, *Topological Huckel-London-Pople-McWeeny ring currents and bond currents in altan-corannulene and altan-coronene*, J. Phys. Chem. A **118** (2014), no. 5, 933–939.

[61] R. Diestel, *Graph Theory*, The Electronic Library of Mathematics, Springer, 2006.

[62] P. Dowd, *Trimethylenemethane*, J. Am. Chem. Soc. **88** (1966), no. 11, 2587–2589.

[63] A. Dress and G. Brinkmann, *Phantasmagorical fulleroids*, MATCH Commun. Math. Comput. Chem. (1996), no. 33, 87–100.

[64] F. Dubeau, *The rabbit problem revisited*, Fibonacci Quart. **31** (1993), no. 3, 268–274.

[65] E. H. Farmer, F. S. Dainton, A. G. Mazurkiewicz, A. E. Rout, Prettre, W. A. Waters, and Walsh, *General discussion*, Trans. Faraday Soc. **42** (1946), 279–283.

[66] D. Feller, W. T. Borden, and E. R. Davidson, *Calculation of zero-field splitting parameters for trimethylenemethane*, J. Chem. Phys. **74** (1981), no. 4, 2256–2259.

[67] G. Fijavž, T. Pisanski, and J. Rus, *Strong traces model of self-assembly polypeptide structures*, MATCH Commun. Math. Comput. Chem. **71** (2014), no. 1, 199–212.

[68] P. W. Fowler, J. E. Cremona, and J. I. Steer, *Systematics of bonding in non-icosahedral carbon clusters*, Theor. Chim. Acta **73** (1988), no. 1, 1–26.

[69] P. W. Fowler and D. E. Manolopoulos, *Molecular graphs, point groups, and fullerenes*, J. Chem. Phys. **96** (1992), no. 10, 7603–7614.

[70] ———, *An Atlas of Fullerenes*, Dover Books on Chemistry, Dover Publications, 2006.

[71] P. W. Fowler, D. E. Manolopoulos, D. B. Redmond, and R. P. Ryan, *Possible symmetries of fullerene structures*, Chem. Phys. Lett. **202** (1993), no. 5, 371–378.

[72] P. W. Fowler and W. Myrvold, *The "anthracene problem": Closed-form conjugated-circuit models of ring currents in linear polyacenes*, J. Phys. Chem. A **115** (2011), no. 45, 13191–13200.

[73] P. W. Fowler, W. Myrvold, D. Jenkinson, and W. H. Bird, *Perimeter ring currents in benzenoids from Pauling bond orders*, Phys. Chem. Chem. Phys. **18** (2016), 11756–11764.

[74] P. W. Fowler and T. Pisanski, *HOMO-LUMO maps for chemical graphs*, MATCH Commun. Math. Comput. Chem. **64** (2010), no. 2, 373–390.

[75] ———, *HOMO-LUMO maps for fullerenes*, Acta Chim. Slov. **57** (2010), no. 3, 513–517.

[76] P. W. Fowler, T. Pisanski, and J. Shawe-Taylor, *Molecular graph eigenvectors for molecular coordinates*, Graph Drawing (R. Tamassia and I. G. Tollis, eds.), Lecture Notes in Computer Science, Springer, 1995, pp. 367–377.

[77] P. W. Fowler and C. M. Quinn, *σ, π and δ representations of the molecular point groups*, Theor. Chim. Acta **70** (1986), no. 5, 333–350.

[78] M. E. Franks, G. R. Macpherson, and W. D. Figg, *Thalidomide*, Lancet **363** (2004), no. 9423, 1802–1811.

[79] A. A. Frost and B. Musulin, *A mnemonic device for molecular orbital energies*, J. Chem. Phys. **21** (1953), no. 3, 572–573.

[80] K. Gavroglu and A. Simões, *Neither Physics nor Chemistry: A History of Quantum Chemistry*, Transformations-Studies in the History of Science and Technology, The MIT Press, 2012.

[81] J. Goedgebeur and B. D. McKay, *Recursive generation of IPR fullerenes*, J. Math. Chem. **53** (2015), no. 8, 1702–1724.

[82] J. A. N. F. Gomes and R. B. Mallion, *A quasi-topological method for the calculation of relative 'ring current' intensities in polycyclic, conjugated hydrocarbons*, Rev. Port. Quim. **21** (1979), 82–89.

[83] M. Gordon and W. H. T. Davison, *Theory of resonance topology of fully aromatic hydrocarbons. I*, J. Chem. Phys. **20** (1952), no. 3, 428–435.

[84] H. Gradišar, S. Božič, T. Doles, D. Vengust, I. Hafner Bratkovič, A. Mertelj, B. Webb, A. Šali, S. Klavžar, and R. Jerala, *Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments*, Nat. Chem. Biol. **9** (2013), no. 6, 362–366.

[85] H. Gradišar and R. Jerala, *De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers*, J. Pept. Sci. **17** (2011), no. 2, 100–106.

[86] J. E. Graver, *The (m, k)-patch boundary code problem*, MATCH Commun. Math. Comput. Chem. **48** (2003), 189–196.

[87] J. E. Graver and C. M. Graves, *Fullerene patches I*, Ars Math. Contemp. **3** (2010), no. 1, 109–120.

[88] J. E. Graver, C. M. Graves, and S. J. Graves, *Fullerene patches II*, Ars Math. Contemp. **7** (2014), no. 2, 405–421.

[89] C. M. Graves and S. J. Graves, *Counting symmetric and near-symmetric fullerene patches*, J. Math. Chem. **52** (2014), no. 9, 2423–2441.

[90] C. M. Graves and J. McLoud-Mann, *Side lengths of pseudoconvex fullerene patches*, Ars Math. Contemp. **5** (2012), no. 2, 291–302.

[91] C. M. Graves, J. McLoud-Mann, and K. Stagg Rovira, *Extending patches to fullerenes*, Ars Math. Contemp. **9** (2015), no. 2, 219–232.

[92] D. J. Griffiths, *Introduction to Quantum Mechanics*, 2 ed., Benjamin Cummings, 2004.

[93] J. L. Gross and T. W. Tucker, *Topological Graph Theory*, John Wiley & Sons, 1987.

[94] B. Grünbaum and G. C. Shephard, *Tilings and Patterns*, Dover Books on Mathematics, Dover Publications, 2013.

[95] X. Guo, P. Hansen, and M. Zheng, *Boundary uniqueness of fusenes*, Disc. Appl. Math. **118** (2002), no. 3, 209–222.

[96] I. Gutman, *Topological properties of benzenoid systems. An identity for the sextet polynomial*, Theor. Chim. Acta **45** (1977), no. 4, 309–315.

[97] ———, *The energy of a graph*, Ber. Math.-Statist. Sekt. Forsch. Graz (1978), no. 100-105, Ber. No. 103, 22, 10. Steiermärkisches Mathematisches Symposium (Stift Rein, Graz, 1978).

[98] ———, *Topological properties of benzenoid systems. IX. On the sextet polynomial*, Z. Naturforsch. A **37** (1982), no. 1, 69–73.

[99] I. Gutman, *Altan derivatives of a graph*, Iranian J. Math. Chem. **5** (2014), 85–90.

[100] I. Gutman, *Topological properties of altan-benzenoid hydrocarbons*, Serb. Chem. Soc. **79** (2014), no. 12, 1515–1521.

[101] I. Gutman and S. J. Cyvin, *A new method for the enumeration of Kekulé structures*, Chem. Phys. Lett. **136** (1987), no. 2, 137–140.

[102] ———, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer-Verlag, 1989.

[103] I. Gutman, M. Milun, and N. Trinajstić, *Topological definition of delocalisation energy*, MATCH Commun. Math. Comput. Chem. **1** (1975), 171–175.

[104] I. Gutman and B. Mohar, *More difficulties with topological resonance energy*, Chem. Phys. Lett. **77** (1981), no. 3, 567–570.

[105] I. Gutman and M. Randić, *A correlation between Kekulé valence structures and conjugated circuits*, Chem. Phys. **41** (1979), 265–270.

[106] R. Hammack, W. Imrich, and S. Klavžar, *Handbook of Product Graphs*, CRC Press, 2011.

[107] P. Hansen, C. Lebatteux, and M. Zheng, *The boundary-edges code for polyhexes*, J. Mol. Struct. (Theochem) **363** (1996), no. 2, 237–247.

[108] F. Harary and H. Harborth, *Extremal animals*, J. Combinatorics Information Syst. Sci. **1** (1976), no. 1, 1–8.

[109] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[110] W. C. Herndon and A. J. Bruce, *Perimeter code for benzenoid aromatic hydrocarbons*, Graph Theory and Topology in Chemistry, Studies in physical and theoretical chemistry, vol. 51, 1987, pp. 491–513.

[111] E. Hückel, *Quantentheoretische Beiträge zum Benzolproblem. I. Die Elektronenkonfiguration des Benzols und verwandter Verbindungen*, Z. Phys. **70** (1931), no. 3, 204–286.

[112] ———, *Quantentheoretische Beiträge zum Benzolproblem. II. Quantentheorie der induzierten Polaritäten*, Z. Phys. **72** (1931), no. 5, 310–337.

[113] ———, *Quantentheoretische Beiträge zum Problem der aromatischen und ungesättigten Verbindungen. III*, Z. Phys. **76** (1932), no. 9, 628–648.

[114] ———, *Die freien Radikale der organischen Chemie. Quantentheoretische Beiträge zum Problem der aromatischen und ungesättigten Verbindungen. IV*, Z. Phys. **83** (1933), no. 9, 632–668.

[115] G. Jaklič, P. W. Fowler, and T. Pisanski, *HL-index of a graph*, Ars Math. Contemp. **5** (2012), no. 1, 99–105.

[116] P. K. Jha, S. Klavžar, and A. Vesel, *$L(2,1)$-labeling of direct product of paths and cycles*, Discrete Appl. Math. **145** (2005), no. 2, 317–325.

[117] F. Johansson et al., *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)*, December 2013, `http://mpmath.org/`.

[118] E. Jones, T. Oliphant, P. Peterson, et al., *SciPy: Open source scientific tools for Python*, 2001–, [Online; accessed 2016-04-05].

[119] Kenner, H. Theorell, H. P. Koch, B. Dudley Sully, M. Magat, P. George, A. Robertson, E. W. J. Mardles, Z. Karpinski, A. D. Walsh, C. A. Coulson, A. S. Carpenter, L. Bateman, G. Gee, and Horrobin, *General discussion*, Trans. Faraday Soc. **42** (1946), 258–269.

[120] S. Klavžar and J. Rus, *Stable traces as a model for self-assembly of polypeptide nanoscale polyhedrons*, MATCH Commun. Math. Comput. Chem. **70** (2013), no. 1, 317–330.

[121] S. Klavžar and A. Vesel, *Computing graph invariants on rotagraphs using dynamic algorithm approach: the case of $(2,1)$-colorings and independence numbers*, Discrete Appl. Math. **129** (2003), no. 2-3, 449–460.

[122] S. Klavžar and J. Žerovnik, *Algebraic approach to fasciagraphs and rotagraphs*, Discrete Appl. Math. **68** (1996), no. 1-2, 93–100.

[123] V. Kočar, S. Božič Abram, T. Doles, N. Bašić, H. Gradišar, T. Pisanski, and R. Jerala, *TOPOFOLD, the designed modular biomolecular folds: polypeptide-based molecular origami nanostructures following the footsteps of DNA*, WIREs Nanomed. Nanobiotechnol. **7** (2015), no. 2, 218–237.

[124] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Algorithms and Combinatorics, Springer, Berlin, Heidelberg, 2014.

[125] J. Kovič, T. Pisanski, A. T. Balaban, and P. W. Fowler, *On symmetries of benzenoid systems*, MATCH Commun. Math. Comput. Chem. **72** (2014), no. 1, 3–26.

[126] V. Kočar, J. S. Schreck, S. Čeru, H. Gradišar, N. Bašić, T. Pisanski, J. P. K. Doye, and R. Jerala, *Design principles for rapid folding of knotted dna nanostructures*, Nat. Commun. **7** (2016).

[127] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, C_{60}*: Buckminsterfullerene*, Nature **318** (1985), no. 6042, 162–163.

[128] I. László, A. Graovac, T. Pisanski, and D. Plavšić, *Graph drawing with eigenvectors*, Carbon Bonding and Structures: Advances in Physics and Chemistry (M. V. Putz, ed.), Carbon Materials-Chemistry and Physics, vol. 5, 2011, pp. 95–115.

[129] H. C. Longuet-Higgins and G. W. Wheland, *Theories of valence*, Annu. Rev. Phys. Chem. **1** (1950), no. 1, 133–150.

[130] L. Lovász, *Semidefinite programs and combinatorial optimization*, Recent Advances in Algorithms and Combinatorics, CMS Books Math. / Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 137–194.

[131] I. Lukovits, A. Graovac, E. Kalman, G. Kaptay, P. Nagy, S. Nikolić, J. Sytchev, and N. Trinajstić, *Nanotubes: Number of Kekulé structures and aromaticity*, J. Chem. Inf. Comput. Sci. **43** (2003), no. 2, 609–614.

[132] R. B. Mallion and D. H. Rouvray, *The golden jubilee of the Coulson-Rushbrooke pairing theorem*, J. Math. Chem. **5** (1990), no. 1, 1–21.

[133] P. Mani, *Automorphismen von polyedrischen graphen*, Math. Ann. **192** (1971), no. 4, 279–303.

[134] D. E. Manolopoulos and P. W. Fowler, *A fullerene without a spiral*, Chem. Phys. Lett. **204** (1993), no. 1-2, 1–7.

[135] D. E. Manolopoulos, J. C. May, and S. E. Down, *Theoretical studies of the fullerenes:* C_{34} *to* C_{70}, Chem. Phys. Lett. **181** (1991), no. 2-3, 105–111.

[136] B. D. McKay, *Practical graph isomorphism*, Proceedings of the Tenth Manitoba Conference on Numerical Mathematics and Computing, Vol. I (Winnipeg, Man., 1980), vol. 30, 1981, pp. 45–87.

[137] B. D. McKay and A. Piperno, *Practical graph isomorphism, II*, J. Symb. Comput. **60** (2014), no. 0, 94–112.

[138] A. D. McNaught and A. Wilkinson, *IUPAC Compendium of Chemical Terminology (The "Gold Book")*, 2 ed., Wiley-Blackwell, Oxford, 1997.

[139] M. Miller and J. Širáň, *Moore graphs and beyond: A survey of the degree/diameter problem*, Electron. J. Combin. (2005).

[140] W. E. Moffitt, *On the Electronic Structure of Molecules*, D.Phil. Thesis, University of Oxford, Oxford, 1948.

[141] W. E. Moffitt, *The residual affinity of conjugated and resonating hydrocarbons*, Trans. Faraday Soc. **45** (1949), no. 4, 373–385.

[142] B. Mohar, *Median eigenvalues of bipartite planar graphs*, MATCH Commun. Math. Comput. Chem. **70** (2013), no. 1, 79–84.

[143] ———, *Median eigenvalues and the HOMO-LUMO index of graphs*, J. Combin. Theory Ser. B **112** (2015), 78–92.

[144] B. Mohar and B. Tayfeh-Rezaie, *Median eigenvalues of bipartite graphs*, J. Algebraic Combin. **41** (2015), no. 3, 899–909.

[145] B. Mohar and C. Thomassen, *Graphs on Surfaces*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 2001.

[146] G. Monaco, M. Memoli, and R. Zanasi, *Additivity of current density patterns in altan-molecules*, J. Phys. Org. Chem. **26** (2013), no. 2, 109–114.

[147] G. Monaco and R. Zanasi, *On the additivity of current density in polycyclic aromatic hydrocarbons*, J. Chem. Phys. **131** (2009), no. 4, 044126.

[148] ———, *Three contra-rotating currents from a rational design of polycyclic aromatic hydrocarbons: altan-corannulene and altan-coronene*, J. Phys. Chem. A **116** (2012), no. 36, 9020–9026.

[149] ———, *Anionic derivatives of altan-corannulene*, J. Phys. Org. Chem. **26** (2013), 730–736.

[150] L. Moura and I. Stojmenović, *Backtracking and Isomorph-Free Generation of Polyhexes*, pp. 39–83, John Wiley & Sons, 2007.

[151] A. Mowshowitz, *The characteristic polynomial of a graph*, **12** (1972), 177–193.

[152] T. Muir and W. H. Metzler, *A Treatise on the Theory of Determinants*, Dover Phoenix Editions, Dover Publications, 2003.

[153] W. Myrvold, B. Bultena, S. Daugherty, B. Debroni, S. Girn, M. Minchenko, J. Woodcock, and P. W. Fowler, *FuiGui: A graphical user interface for investigating conjectures about fullerenes*, MATCH Commun. Math. Comput. Chem. **58** (2007), no. 2, 403–422.

[154] L. Pauling, *Bond numbers and bond lengths in tetrabenzo*$[de, no, st, c_1d_1]$*heptacene and other condensed aromatic hydrocarbons: A valence-bond treatment*, Acta Crystallogr. Sect. B-Struct. Sci. **36** (1980), 1898–1901.

[155] ———, *General Chemistry*, Dover Books on Chemistry, Dover Publications, 2014.

[156] L. Pauling, L. O. Brockway, and J. Y. Beach, *The dependence of interatomic distance on single bond-double bond resonance*, J. Am. Chem. Soc. **57** (1935), 2705–2709.

[157] L. Pauling and E. B. Wilson, *Introduction to Quantum Mechanics with Applications to Chemistry*, Dover Books on Physics, Dover Publications, 2012.

[158] A. Piperno, *Search space contraction in canonical labeling of graphs*, ArXiv e-prints (2008).

[159] T. Pisanski and A. T. Balaban, *Flag graphs and their applications in mathematical chemistry for benzenoids*, J. Math. Chem. **50** (2012), no. 4, 893–903.

[160] T. Pisanski and P. W. Fowler, *Golden graphs*, Int. J. Chem. Mod. **3** (2011), no. 1-2, 7–13.

[161] T. Pisanski and D. Marušič, *Symmetries of hexagonal molecular graphs on the torus*, Croat. Chem. Acta **73** (2000), no. 4, 969–981.

[162] T. Pisanski and B. Servatius, *Configurations from a Graphical Viewpoint*, Basler Lehrbücher, Birkhäuser Boston, 2012.

[163] T. Pisanski and J. Shawe-Taylor, *Characterizing graph drawing with eigenvectors*, J. Chem. Inf. Comput. Sci. **40** (2000), no. 3, 567–571.

[164] T. Pisanski and A. Žitnik, *Representing Graphs and Maps*, Topics in Topological Graph Theory, Encyclopedia Math. Appl., vol. 128, Cambridge Univ. Press, Cambridge, 2009, pp. 151–180.

[165] T. Pisanski, A. Žitnik, A. Graovac, and A. Baumgartner, *Rotagraphs and their generalizations*, J. Chem. Inf. Comput. Sci. **34** (1994), no. 5, 1090–1093.

[166] M. D. Plummer and L. Lovász, *Matching Theory*, North-Holland Mathematics Studies, Elsevier Science, 1986.

[167] O. E. Polansky and N. N. Tyutyulkov, *Structural graphs of regular polymers and their properties*, MATCH Commun. Math. Comput. Chem. **3** (1977), 149–223.

[168] A. Pullman, *The electronic structure of some free radicals*, Disc. Faraday Soc. **2** (1947), 26–35.

[169] B. Pullman and A. Pullman, *Quantum Biochemistry*, Interscience Publishers, 1963.

[170] M. V. Putz, *Carbon Bonding and Structures: Advances in Physics and Chemistry*, Carbon Materials: Chemistry and Physics, Springer, Netherlands, 2011.

[171] J. Rada, R. Cruz, and I. Gutman, *Vertex-degree-based topological indices of catacondensed hexagonal systems*, Chem. Phys. Lett. **572** (2013), 154–157.

[172] ———, *Benzenoid systems with extremal vertex-degree-based topological indices*, MATCH Commun. Math. Comput. Chem. **72** (2014), no. 1, 125–136.

[173] M. Randić, *Conjugated circuits and resonance energies of benzenoid hydrocarbons*, Chem. Phys. Lett. **38** (1976), no. 1, 68–70.

[174] ———, *Enumeration of Kekulé structures in conjugated hydrocarbons*, J. Chem. Soc. Faraday Trans. **72** (1976), 232–243.

[175] M. Randić, *Aromaticity of polycyclic conjugated hydrocarbons*, Chem. Rev. **103** (2003), no. 9, 3449–3605.

[176] ———, *Graph theoretical approach to π-electron currents in polycyclic conjugated hydrocarbons*, Chem. Phys. Lett. **500** (2010), no. 1–3, 123–127.

[177] H. Ren and F. Zhang, *Extremal double hexagonal chains with respect to k-matchings and k-independent sets*, Discrete Appl. Math. **155** (2007), no. 17, 2269–2281.

[178] S. Reynolds, *Convex Block — Polyform Wiki*, `http://polyform.wikia.com/wiki/Convex_Block#Polyhexes`, [Online; accessed: 1-August-2015].

[179] J. D. Roberts, A. Streitwieser, and C. M. Regan, *Small-ring compounds. X. Molecular orbital calculations of properties of some small-ring hydrocarbons and free radicals*, J. Am. Chem. Soc. **74** (1952), no. 18, 4579–4582.

[180] E. A. Rohlfing, D. M. Cox, and A. Kaldor, *Production and characterization of supersonic carbon cluster beams*, J. Chem. Phys. **81** (1984), no. 7, 3322–3330.

[181] H. Sachs, *Perfect matchings in hexagonal systems*, Combinatorica **4** (1984), no. 1, 89–99.

[182] H. Sachs, P. Hansen, and M. L. Zheng, *Kekulé count in tubular hydrocarbons*, MATCH Commun. Math. Comput. Chem. **33** (1996), 169–241.

[183] C. H. Sah, *A generalized leapfrog for fullerene structures*, Fullerene Sci. Techn. **2** (1994), no. 4, 445–458.

[184] A. M. Schönflies, *Krystallsysteme und Krystallstructur*, Teubner, 1891.

[185] P. Schwerdtfeger, L. Wirz, and J. Avery, *Program fullerene – a software package for constructing and analyzing structures of regular fullerenes*, J. Comput. Chem. **34** (2013), no. 17, 1508–1526.

[186] P. Schwerdtfeger, L. N. Wirz, and J. Avery, *The topology of fullerenes*, WIREs Comput. Mol. Sci. **5** (2015), no. 1, 96–145.

[187] I. Sciriha, *On the construction of graphs of nullity one*, Discrete Math. **181** (1998), no. 1-3, 193–211.

[188] I. Sciriha and I. Gutman, *Nut graphs: maximally extending cores*, Util. Math. **54** (1998), 257–272.

[189] J. H. Smith, *Some properties of the spectrum of a graph*, Combinatorial Structures and their Applications (Proc. Calgary Internat. Conf., Calgary, Alta., 1969), Gordon and Breach, New York, 1970, pp. 403–406.

[190] G. Strang, *Linear Algebra and Its Applications*, 2 ed., Academic Press, Orlando, Florida, 1980.

[191] A. Streitwieser, *Molecular Orbital Theory for Organic Chemists*, Wiley, 1961.

[192] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry, Dover Publications, 2012.

[193] R. Tamassia, *Handbook of Graph Drawing and Visualization*, Discrete Mathematics and Its Applications, Taylor & Francis, 2013.

[194] C. Thomassen, *The Jordan-Schönflies theorem and the classification of surfaces*, Amer. Math. Monthly **99** (1992), no. 2, 116–130.

[195] M. Tinkham, *Group Theory and Quantum Mechanics*, Dover Books on Chemistry, Dover Publications, 2012.

[196] N. Trinajstić, *Chemical Graph Theory*, 2 ed., CRC Press, Boca Raton, Florida, 1992.

[197] J. F. Webb, *Canadian thalidomide experience*, Can. Med. Assoc. J. **89** (1963), no. 19, 987–992.

[198] A. C. Wechsler, *Sequence A116513 in The On-Line Encyclopedia of Integer Sequences (2011)*, `http://oeis.org`, [Online; accessed: 1-August-2013].

[199] P. G Wenthold, J. Hu, R. R. Squires, and W. C. Lineberger, *Photoelectron spectroscopy of the trimethylenemethane negative ion. The singlet-triplet splitting of trimethylenemethane*, J. Am. Chem. Soc. **118** (1996), no. 2, 475–476.

[200] Z. C. Wu, D. A. Jelski, and T. F. George, *Vibrational motions of buckminsterfullerene*, Chem. Phys. Lett. **137** (1987), no. 3, 291–294.

[201] T. F. Yen, *Resonance topology of polynuclear aromatic hydrocarbons*, Theor. Chim. Acta. **20** (1971), no. 4, 399–404.

[202] M. Yoshida and P. W. Fowler, *Dihedral fullerenes of threefold symmetry with and without face spirals*, J. Chem. Soc. Faraday Trans. **93** (1997), no. 18, 3289–3294.

[203] ———, *Systematic relationships between fullerenes without spirals*, Chem. Phys. Lett. **278** (1997), no. 4-6, 256–261.

[204] R. Zanasi, P. Della Porta, and G. Monaco, *The intriguing class of altan-molecules*, J. Phys. Org. Chem. (2016), 6 pages.

Appendices

Appendix A

Fullerenes and the Pentagonal Incidence Partition

This appendix provides detailed lists and pictures of fullerenes with a certain Pentagonal Incidence Partition.

Fullerenes $\mathcal{F}$ where one part of $PIP(\mathcal{F})$ is strictly greater than 6

The table below lists all fullerenes $\mathcal{F}$ where at least one part of the partition $PIP(\mathcal{F})$ is strictly greater than 6. They are listed by $PIP(\mathcal{F})$. Within each group, they are further divided with respect to their symmetry group. The third column in the table shows the total number of fullerenes with a prescribed PIP and symmetry group (the bold number is the cumulative over all groups). They are listed in the fourth column where the notation n:m means that it is the m-th fullerene which is output by the **fullgen** program when all fullerenes on n vertices are generated. (Note that the notation n:m used here is note the same as the spiral notation.)

PIP	Group	Count		List of fullerenes
12	$\mathbf{C}_1$	**41**	6	36:7, 38:3, 38:4, 38:5, 40:16, 42:18
	$\mathbf{C}_2$		18	32:1, 32:5, 34:3, 34:4, 34:6, 36:6, 36:9, 36:13, 38:7, 40:15, 40:17, 40:35, 40:38, 42:16, 42:19, 44:34, 44:63, 46:93
	$\mathbf{C}_s$		1	34:2
	$\mathbf{D}_2$		3	28:2, 36:5, 44:36
	$\mathbf{C}_{2v}$		3	30:2, 30:3, 38:12
	$\mathbf{D}_3$		1	32:3
	$\mathbf{C}_{3v}$		1	34:5

Continues on the next page ...

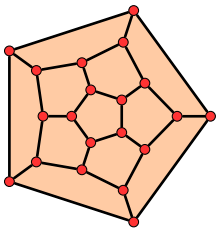
Cont'd ...				
	$\mathbf{D}_{2d}$		1	36:11
	$\mathbf{D}_{3h}$		2	26:1, 32:2
	$\mathbf{D}_{3d}$		1	44:37
	$\mathbf{D}_{6d}$		2	24:1, 48:146
	$\mathbf{T}_{d}$		1	28:1
	$\mathbf{I}_{h}$		1	20:1
$11+1$	$\mathbf{C}_{s}$	**2**	2	40:21, 42:17
$10+2$	$\mathbf{C}_{2v}$	**1**	1	40:14
$10+1+1$	$\mathbf{D}_{5d}$	**1**	1	40:19
$9+3$	$\mathbf{C}_{s}$	**2**	1	44:76
	$\mathbf{C}_{3v}$		1	38:6
$8+4$	$\mathbf{C}_{2v}$	**16**	1	36:8
	$\mathbf{C}_{1}$		6	38:9, 42:20, 42:35, 46:38, 48:34, 48:153
	$\mathbf{C}_{2}$		7	40:23, 40:32, 44:74, 48:33, 48:114, 48:154, 52:370
	$\mathbf{C}_{s}$		2	46:35, 46:39
$7+5$	$\mathbf{C}_{1}$	**69**	52	36:2, 38:2, 38:13, 38:14, 40:6, 40:20, 40:28, 40:29, 42:15, 42:23, 42:25, 42:26, 42:32, 42:34, 42:37, 44:46, 44:47, 44:48, 44:82, 44:84, 44:86, 46:75, 46:85, 46:98, 46:99, 46:109, 46:112, 48:37, 48:38, 48:126, 48:180, 50:31, 50:32, 50:34, 50:35, 50:142, 50:146, 50:265, 50:269, 52:31, 52:32, 52:313, 52:314, 52:427, 54:32, 54:33, 54:406, 56:660, 56:856, 58:35, 58:36, 60:1260
	$\mathbf{C}_{s}$		17	34:1, 36:3, 36:10, 40:31, 40:33, 40:39, 42:24, 44:65, 44:68, 46:96, 48:137, 50:165, 54:402, 54:536, 58:1089, 60:1641, 64:3068
$7+4+1$	$\mathbf{C}_{1}$	**12**	8	44:31, 46:19, 46:33, 46:34, 46:36, 48:108, 50:206, 52:302
	$\mathbf{C}_{s}$		4	44:25, 44:30, 46:20, 54:537
$7+3+2$	$\mathbf{C}_{s}$	**1**	1	48:120

Pictures of all fullerenes from the above table are shown below. They are ordered by the PIP. Fullerenes $\mathcal{F}$ for each given PIP are listed by increasing number of vertices. Below every picture the following information is listed:

(a) n – the number of vertices;
(b) *id* – the place in which this fullerene appears when all fullerenes on n vertices are generated by the **fullgen** program;
(c) point group.

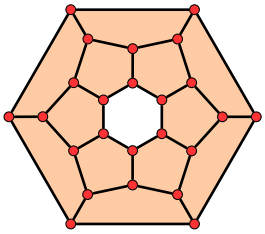
The 41 fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 12$

(1)



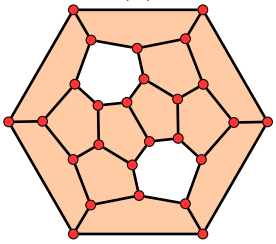
$20, 1, \mathbf{I}_h$

(2)



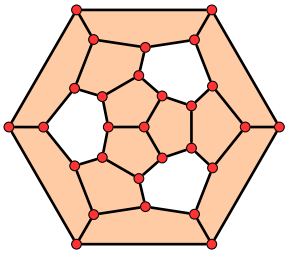
$24, 1, \mathbf{D}_{6d}$

(3)



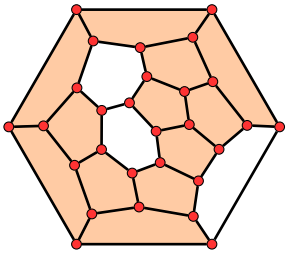
$26, 1, \mathbf{D}_{3h}$

(4)



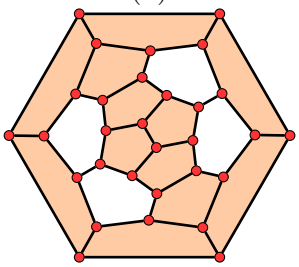
$28, 1, \mathbf{T}_d$

(5)



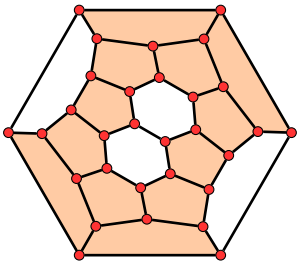
$28, 2, \mathbf{D}_2$

(6)



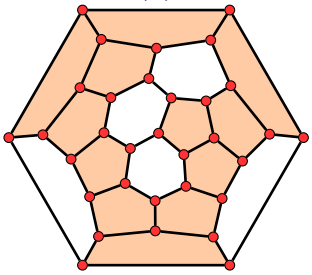
$30, 2, \mathbf{C}_{2v}$

(7)



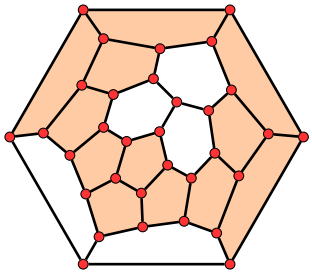
$30, 3, \mathbf{C}_{2v}$

(8)



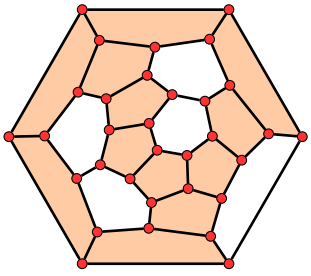
$32, 1, \mathbf{C}_2$

(9)



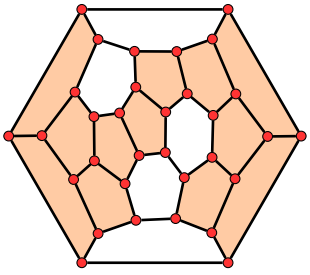
$32, 2, \mathbf{D}_{3h}$

(10)



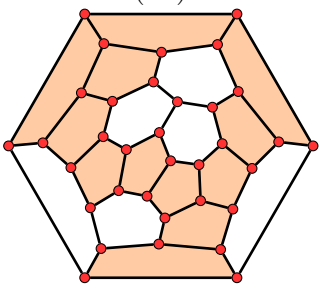
$32, 3, \mathbf{D}_3$

(11)



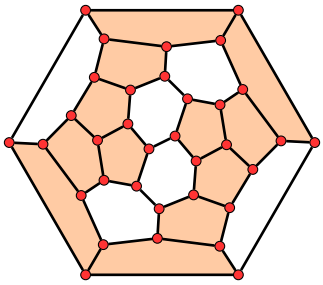
$32, 5, \mathbf{C}_2$

(12)



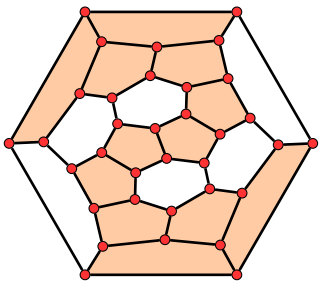
$34, 2, \mathbf{C}_s$

(13)



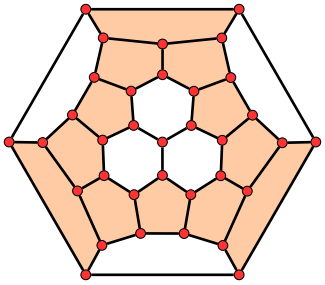
$34, 3, \mathbf{C}_2$

(14)



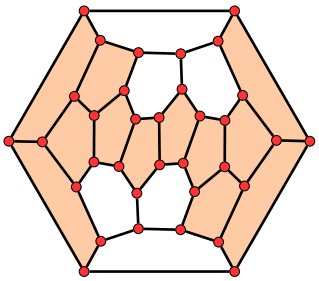
$34, 4, \mathbf{C}_2$

(15)



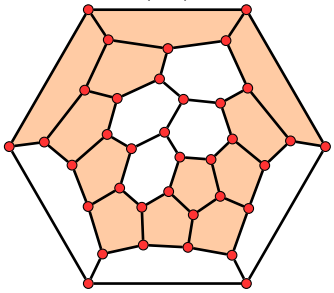
$34, 5, \mathbf{C}_{3v}$

(16)



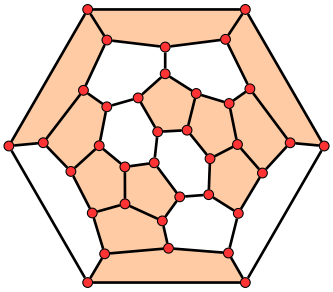
$34, 6, \mathbf{C}_2$

(17)



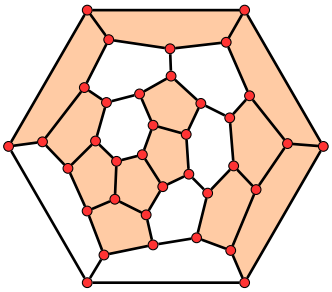
$36, 5, \mathbf{D}_2$

(18)



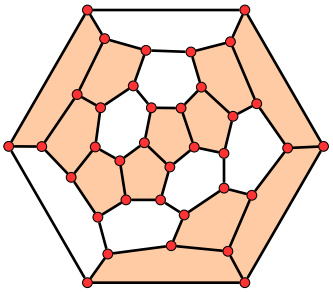
$36, 6, \mathbf{C}_2$

(19)



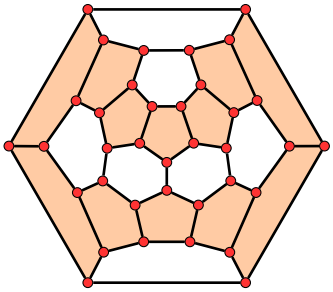
$36, 7, \mathbf{C}_1$

(20)



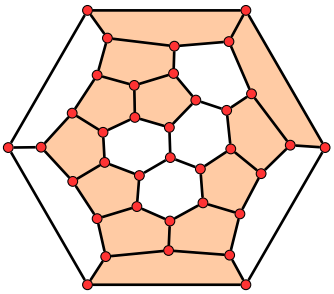
$36, 9, \mathbf{C}_2$

(21)



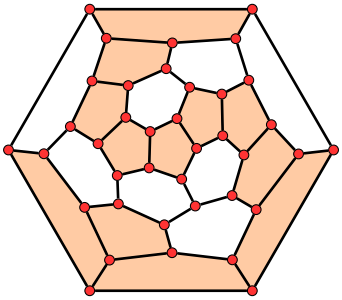
$36, 11, \mathbf{D}_{2d}$

(22)



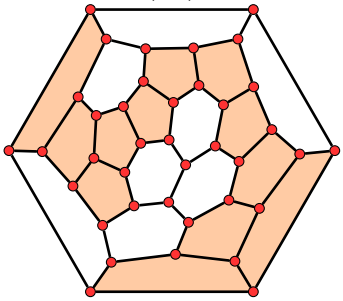
$36, 13, \mathbf{C}_2$

(23)



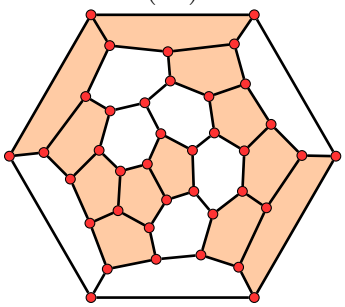
$38, 3, \mathbf{C}_1$

(24)



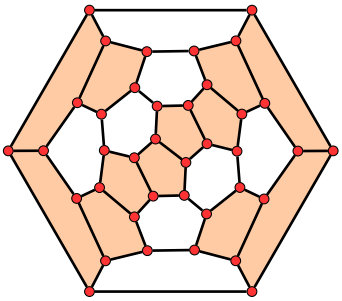
$38, 4, \mathbf{C}_1$

(25)



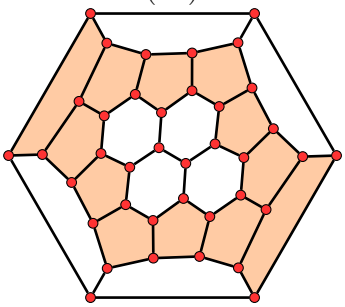
$38, 5, \mathbf{C}_1$

(26)



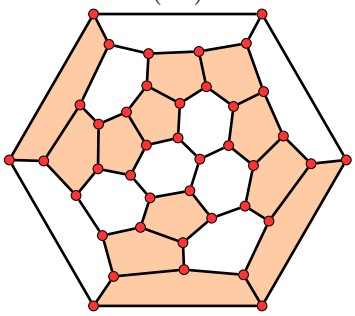
$38, 7, \mathbf{C}_2$

(27)



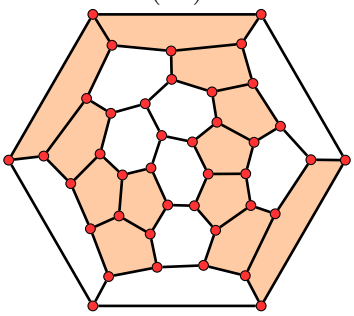
$38, 12, \mathbf{C}_{2v}$

(28)



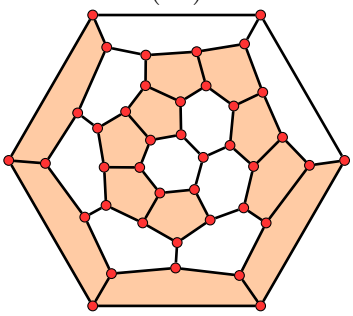
$40, 15, \mathbf{C}_2$

(29)



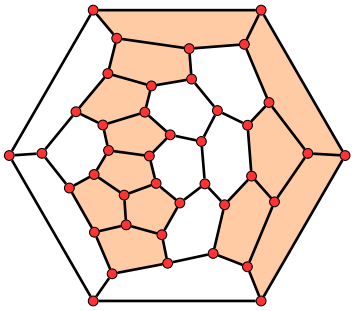
$40, 16, \mathbf{C}_1$

(30)



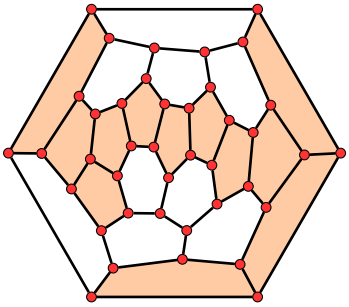
$40, 17, \mathbf{C}_2$

(31)



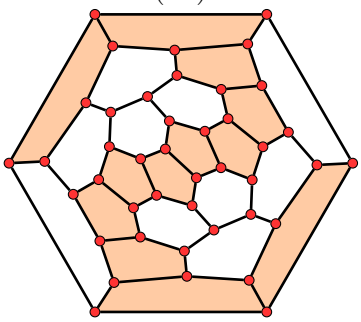
$40, 35, \mathbf{C}_2$

(32)



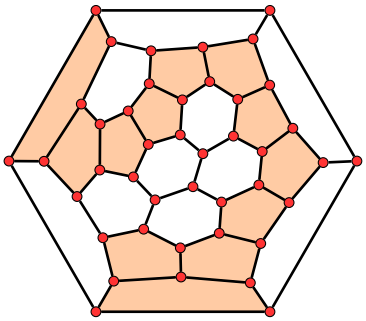
$40, 38, \mathbf{C}_2$

(33)



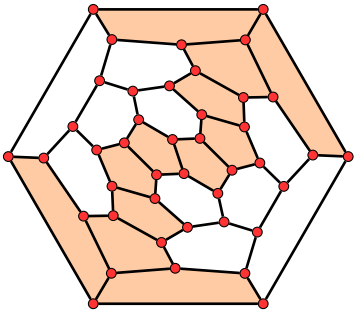
$42, 16, \mathbf{C}_2$

(34)



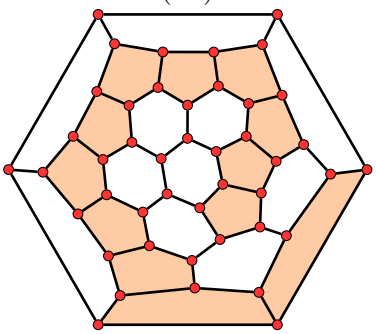
$42, 18, \mathbf{C}_1$

(35)



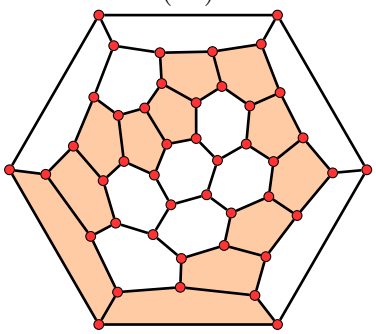
$42, 19, \mathbf{C}_2$

(36)



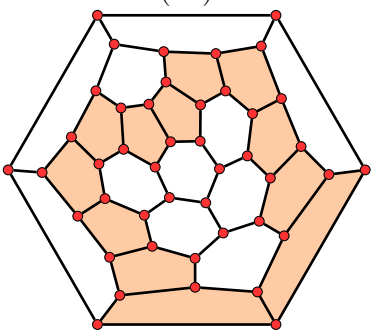
$44, 34, \mathbf{C}_2$

(37)



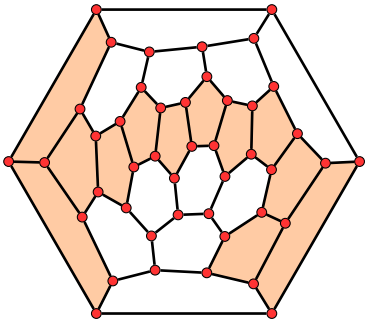
$44, 36, \mathbf{D}_2$

(38)



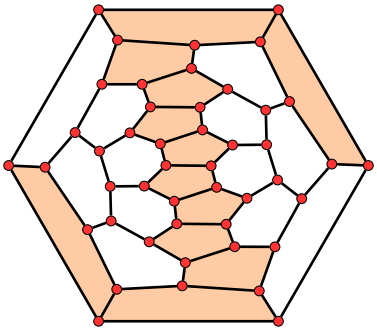
$44, 37, \mathbf{D}_{3d}$

(39)



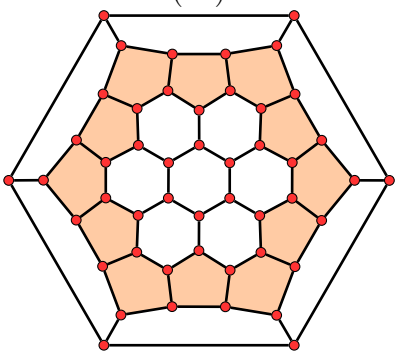
$44, 63, \mathbf{C}_2$

(40)



$46, 93, \mathbf{C}_2$

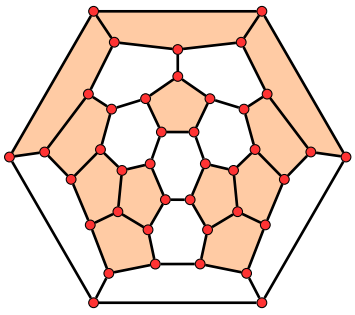
(41)



$48, 146, \mathbf{D}_{6d}$

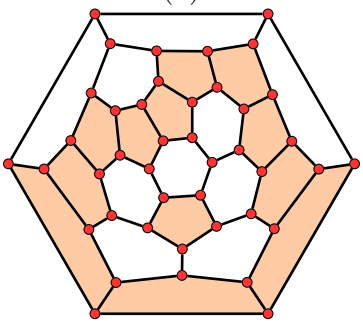
The 2 fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 11 + 1$

(1)



$40, 21, \mathbf{C}_s$

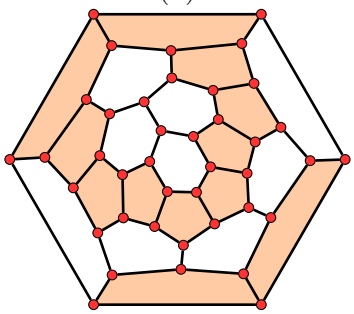
(2)



$42, 17, \mathbf{C}_s$

The fullerene $\mathcal{F}$ with $PIP(\mathcal{F}) = 10 + 2$

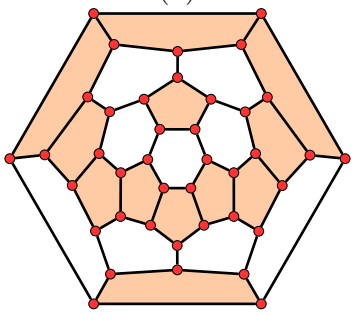
(1)



$40, 14, \mathbf{C}_{2v}$

The fullerene $\mathcal{F}$ with $PIP(\mathcal{F}) = 10 + 1 + 1$

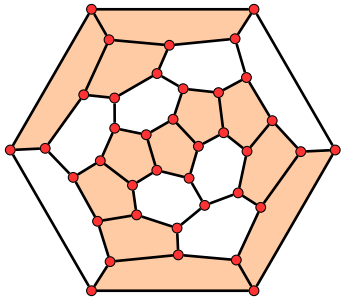
(1)



$40, 19, \mathbf{D}_{5d}$

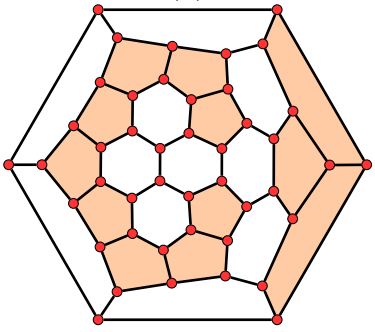
The 2 fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 9 + 3$

(1)



$38, 6, \mathbf{C}_{3v}$

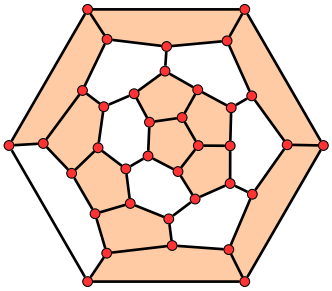
(2)



$44, 76, \mathbf{C}_{s}$

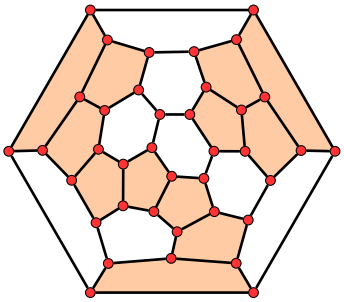
The 16 fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 8 + 4$

(1)



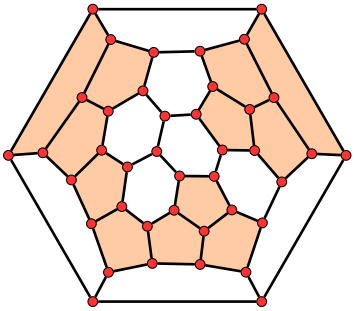
$36, 8, \mathbf{C}_{2v}$

(2)



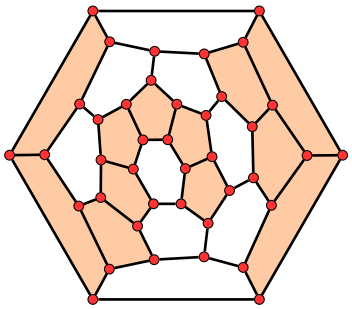
$38, 9, \mathbf{C}_1$

(3)



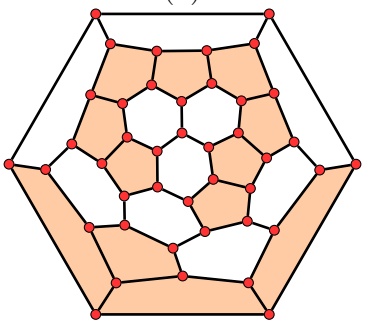
$40, 23, \mathbf{C}_2$

(4)



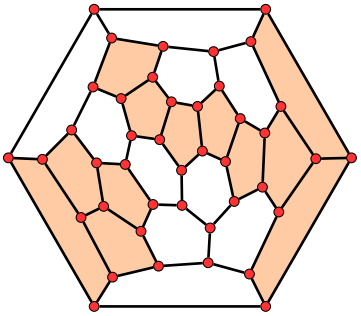
$40, 32, \mathbf{C}_2$

(5)



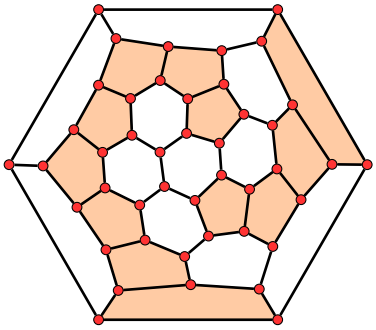
$42, 20, \mathbf{C}_1$

(6)



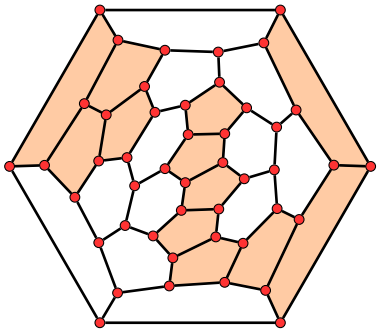
$42, 35, \mathbf{C}_1$

(7)



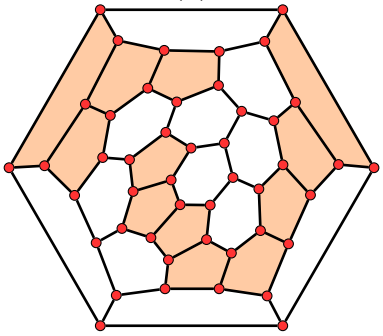
$44, 74, \mathbf{C}_2$

(8)



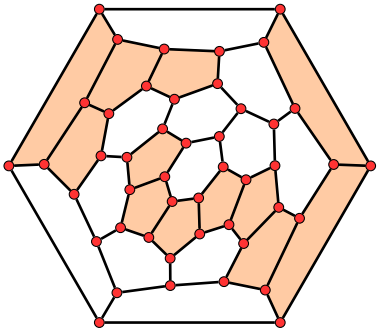
$46, 35, \mathbf{C}_s$

(9)



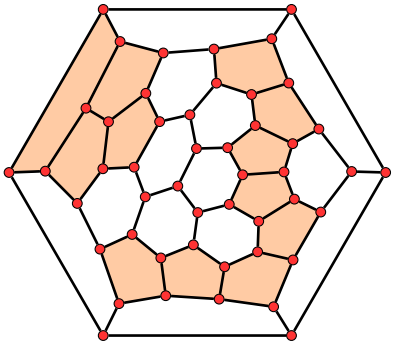
$46, 38, \mathbf{C}_1$

(10)



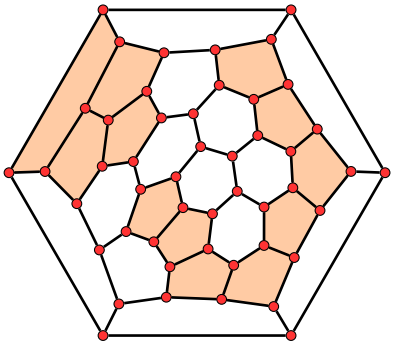
$46, 39, \mathbf{C}_s$

(11)



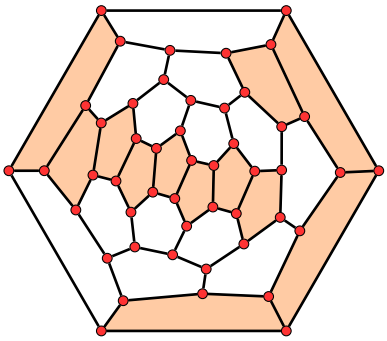
$48, 33, \mathbf{C}_2$

(12)



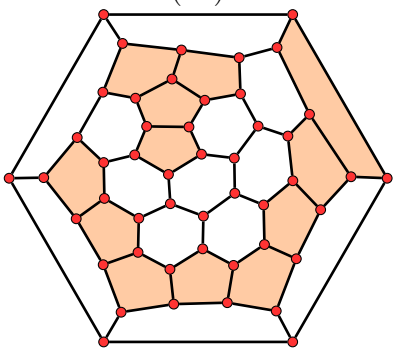
$48, 34, \mathbf{C}_1$

(13)



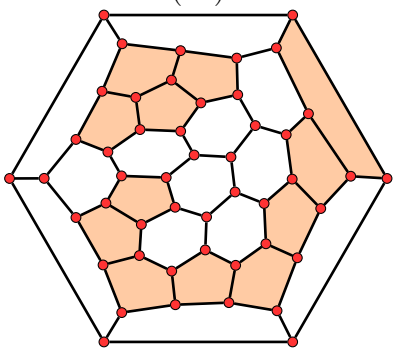
$48, 114, \mathbf{C}_2$

(14)



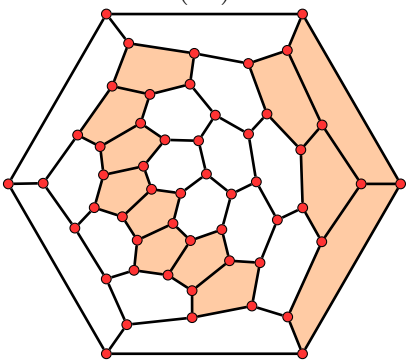
$48, 153, \mathbf{C}_1$

(15)



$48, 154, \mathbf{C}_2$

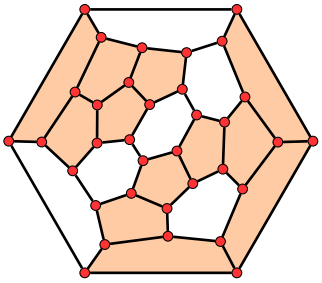
(16)



$52, 370, \mathbf{C}_2$

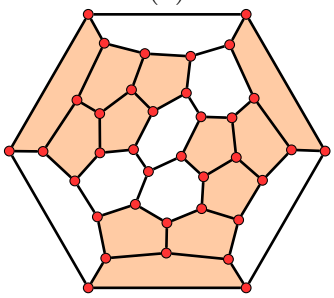
The 69 fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 7 + 5$

(1)



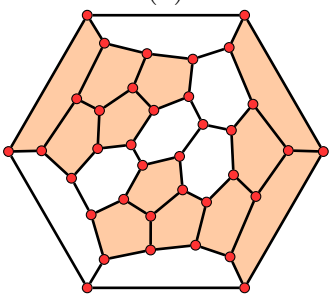
$34, 1, \mathbf{C}_s$

(2)



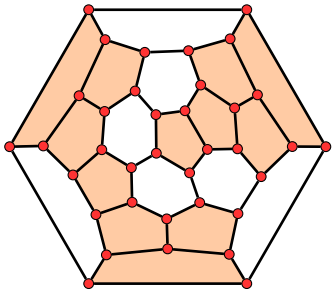
$36, 2, \mathbf{C}_1$

(3)



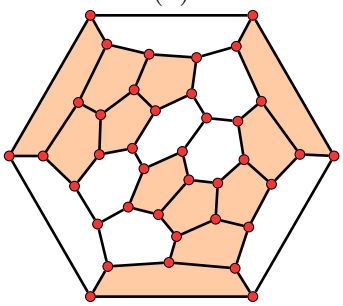
$36, 3, \mathbf{C}_s$

(4)



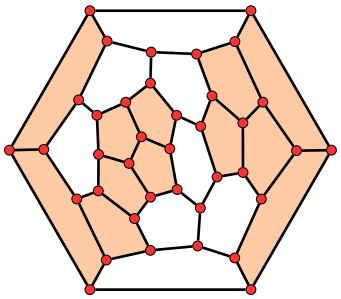
$36, 10, \mathbf{C}_s$

(5)



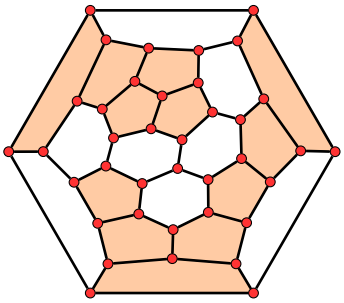
$38, 2, \mathbf{C}_1$

(6)



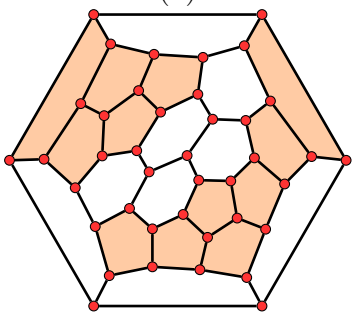
$38, 13, \mathbf{C}_1$

(7)



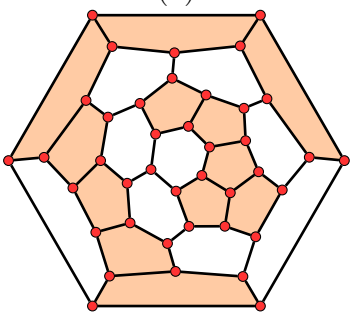
$38, 14, \mathbf{C}_1$

(8)



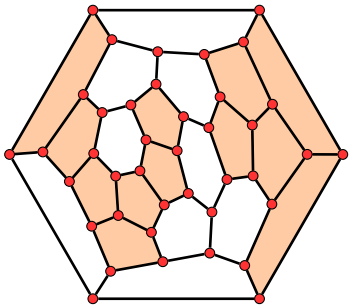
$40, 6, \mathbf{C}_1$

(9)



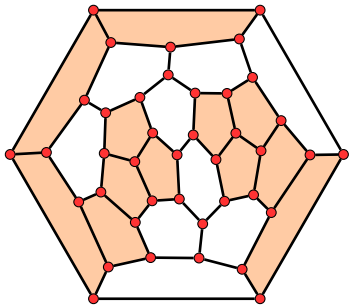
$40, 20, \mathbf{C}_1$

(10)



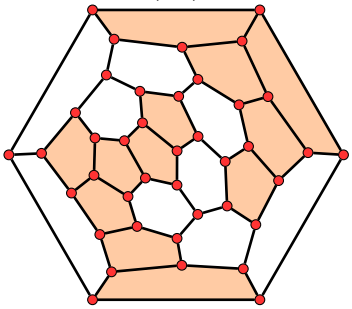
$40, 28, \mathbf{C}_1$

(11)



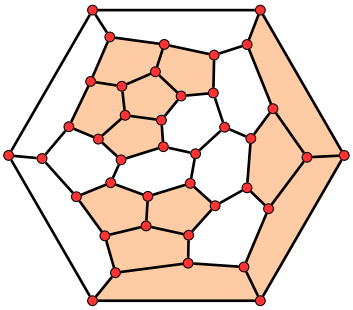
$40, 29, \mathbf{C}_1$

(12)



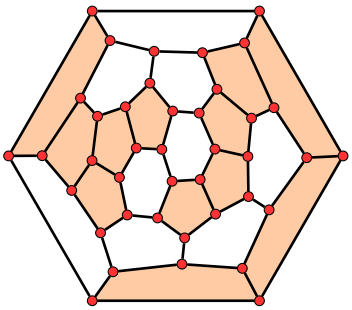
$40, 31, \mathbf{C}_s$

(13)



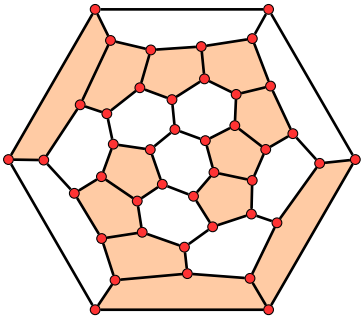
$40, 33, \mathbf{C}_s$

(14)



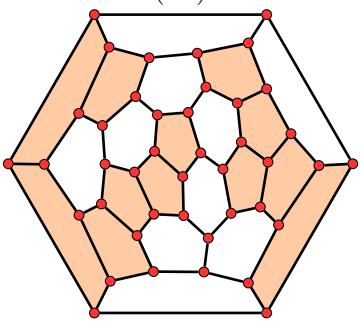
$40, 39, \mathbf{C}_s$

(15)



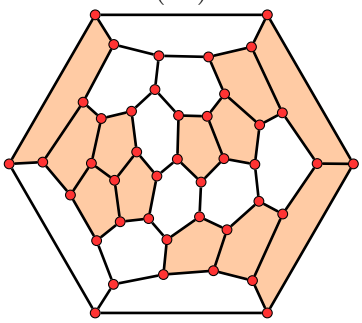
$42, 15, \mathbf{C}_1$

(16)



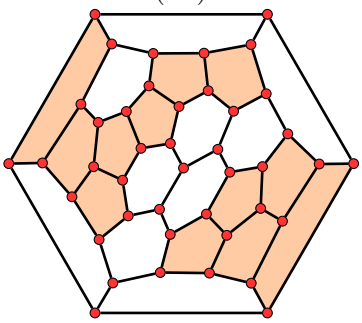
$42, 23, \mathbf{C}_1$

(17)



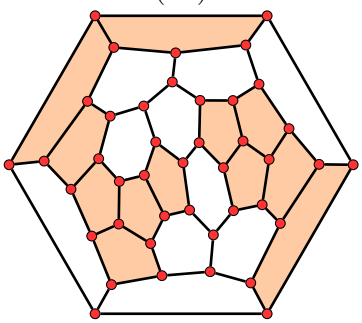
$42, 24, \mathbf{C}_s$

(18)



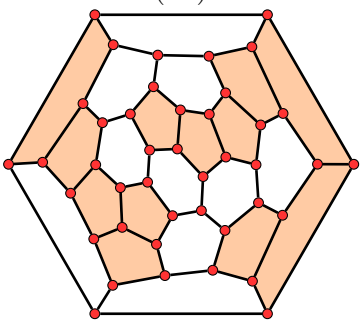
$42, 25, \mathbf{C}_1$

(19)



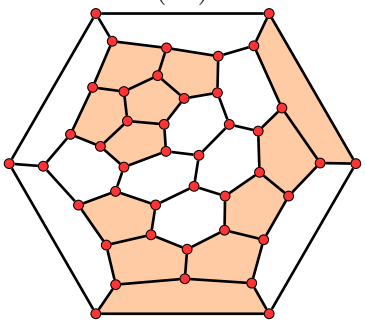
$42, 26, \mathbf{C}_1$

(20)



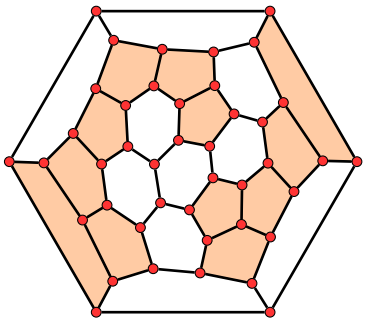
$42, 32, \mathbf{C}_1$

(21)



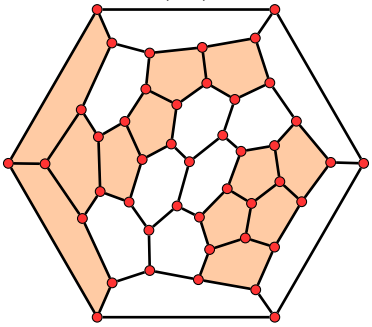
$42, 34, \mathbf{C}_1$

(22)



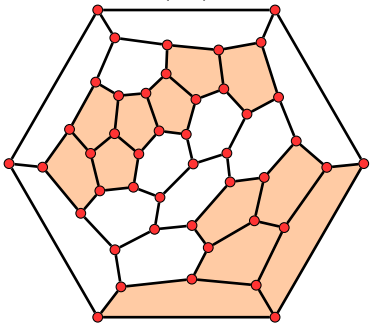
$42, 37, \mathbf{C}_1$

(23)



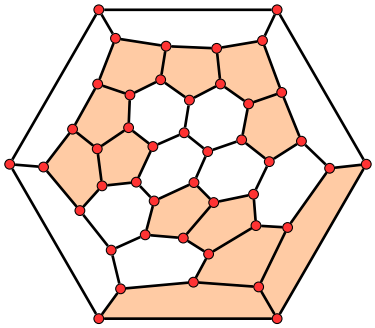
$44, 46, \mathbf{C}_1$

(24)



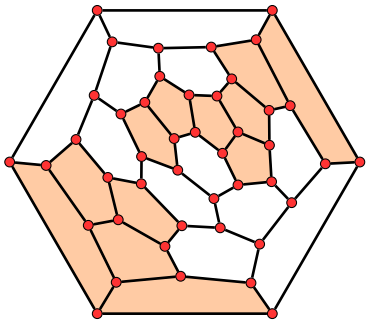
$44, 47, \mathbf{C}_1$

(25)



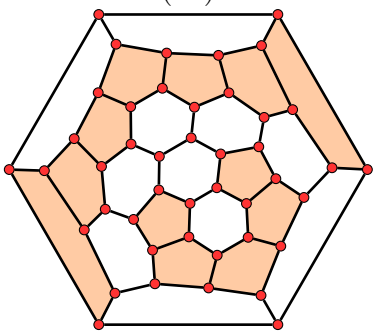
$44, 48, \mathbf{C}_1$

(26)



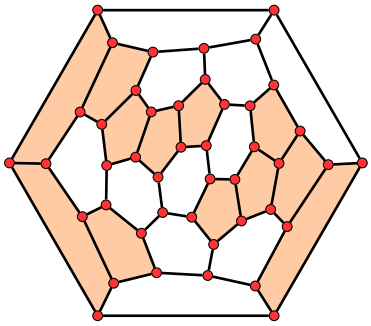
$44, 65, \mathbf{C}_s$

(27)



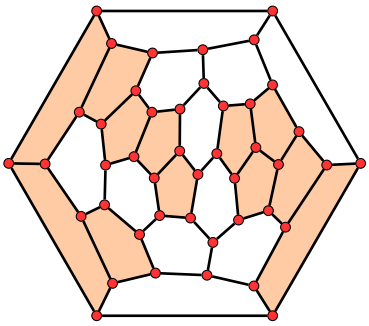
$44, 68, \mathbf{C}_s$

(28)



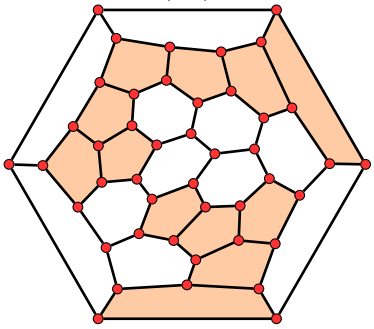
$44, 82, \mathbf{C}_1$

(29)



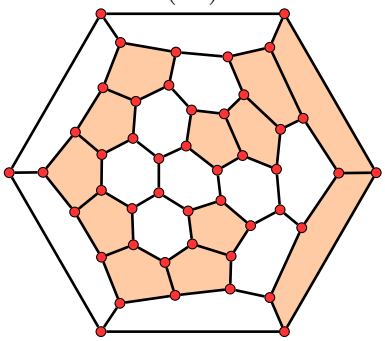
$44, 84, \mathbf{C}_1$

(30)



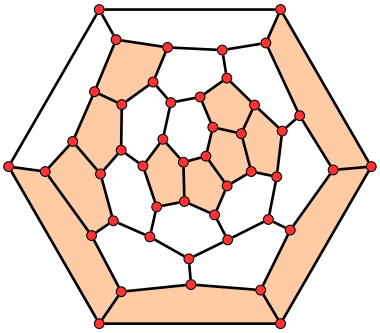
$44, 86, \mathbf{C}_1$

(31)



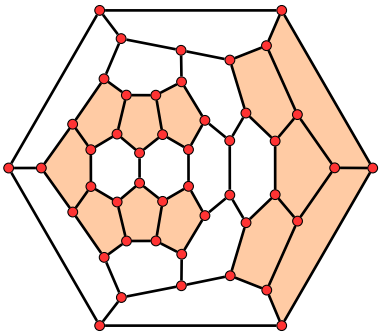
$46, 75, \mathbf{C}_1$

(32)



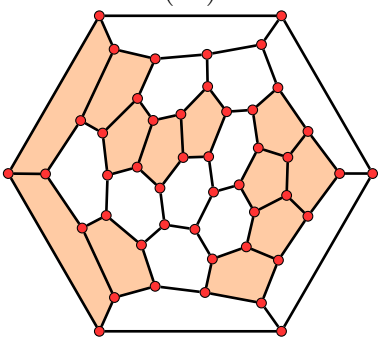
$46, 85, \mathbf{C}_1$

(33)



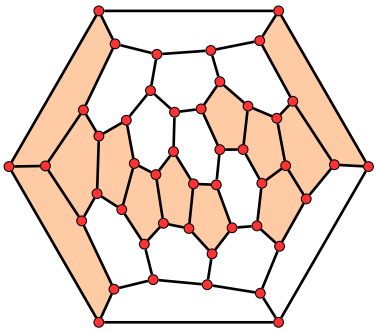
$46, 96, \mathbf{C}_s$

(34)



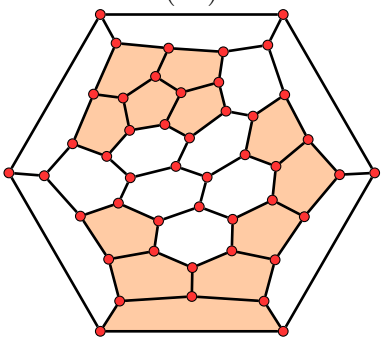
$46, 98, \mathbf{C}_1$

(35)



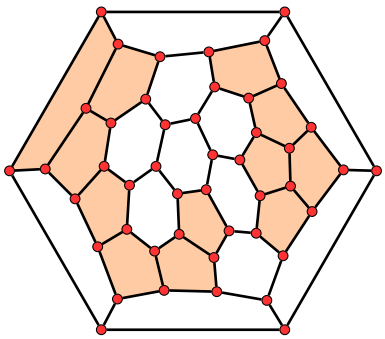
$46, 99, \mathbf{C}_1$

(36)



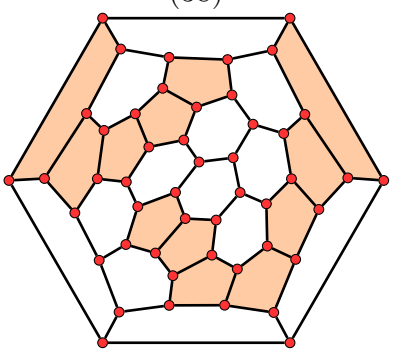
$46, 109, \mathbf{C}_1$

(37)



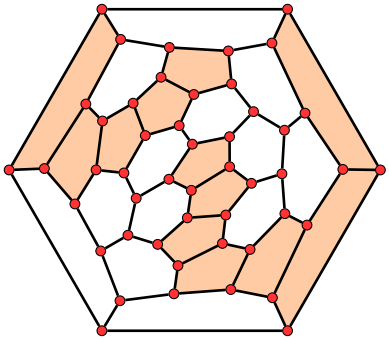
$46, 112, \mathbf{C}_1$

(38)



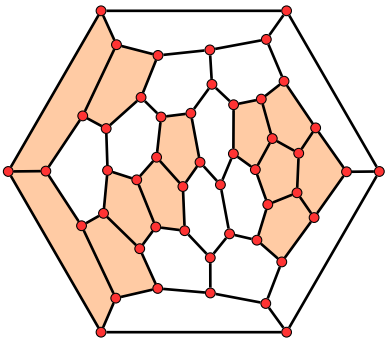
$48, 37, \mathbf{C}_1$

(39)



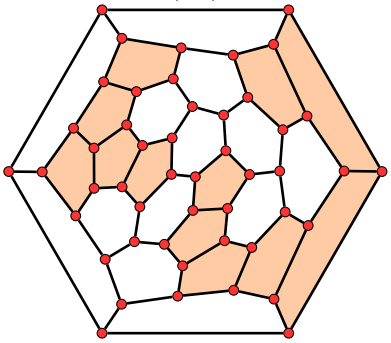
$48, 38, \mathbf{C}_1$

(40)



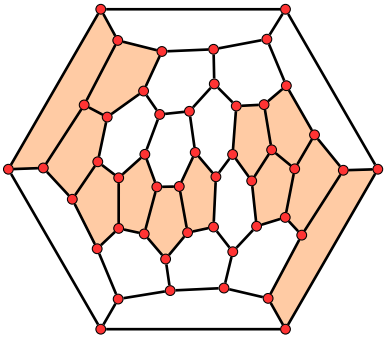
$48, 126, \mathbf{C}_1$

(41)



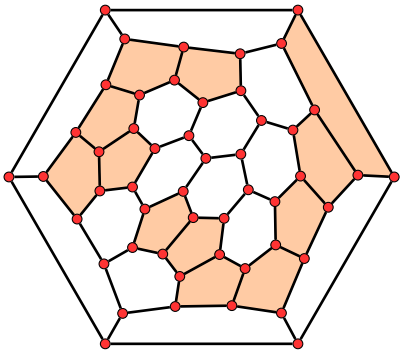
$48, 137, \mathbf{C}_s$

(42)



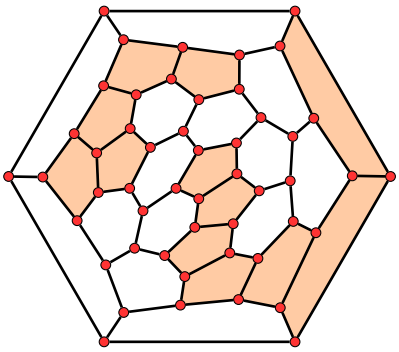
$48, 180, \mathbf{C}_1$

(43)



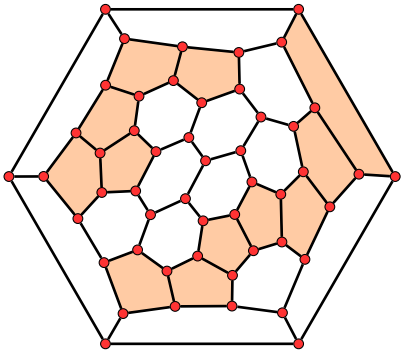
$50, 31, \mathbf{C}_1$

(44)



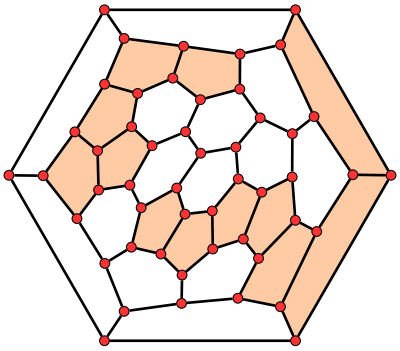
$50, 32, \mathbf{C}_1$

(45)



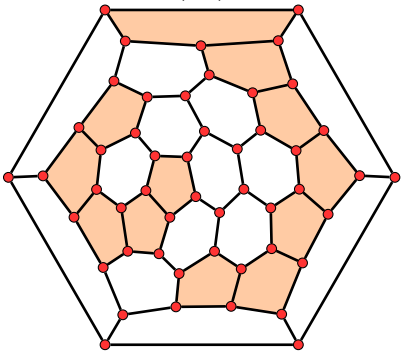
$50, 34, \mathbf{C}_1$

(46)



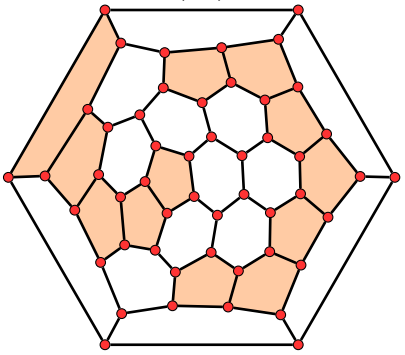
$50, 35, \mathbf{C}_1$

(47)



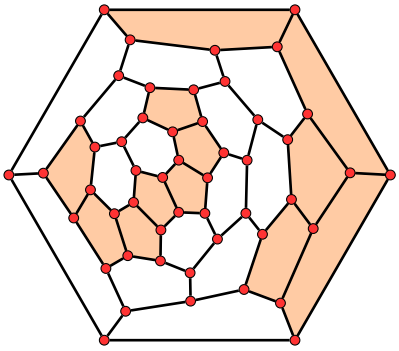
$50, 142, \mathbf{C}_1$

(48)



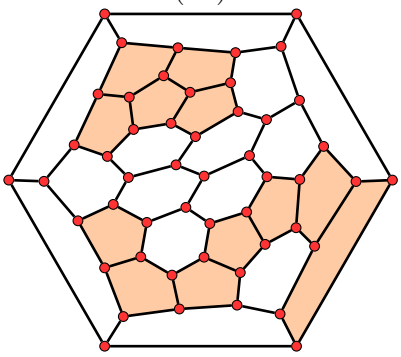
$50, 146, \mathbf{C}_1$

(49)



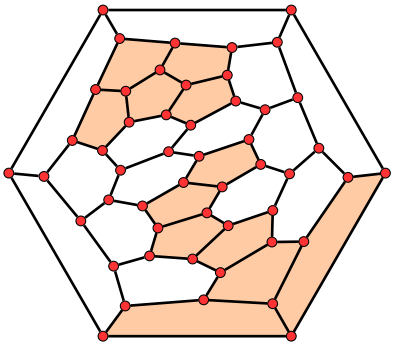
$50, 165, \mathbf{C}_s$

(50)



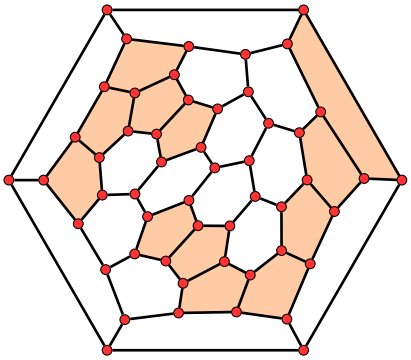
$50, 265, \mathbf{C}_1$

(51)



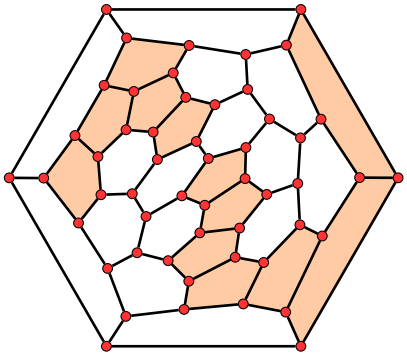
$50, 269, \mathbf{C}_1$

(52)



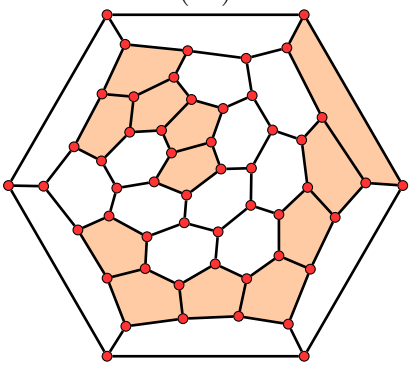
$52, 31, \mathbf{C}_1$

(53)



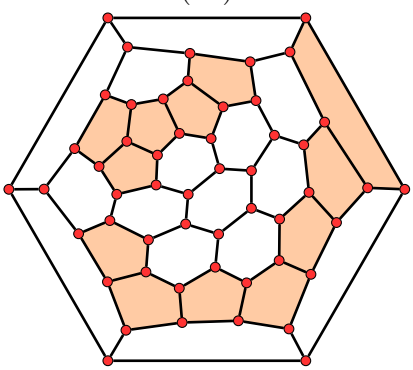
$52, 32, \mathbf{C}_1$

(54)



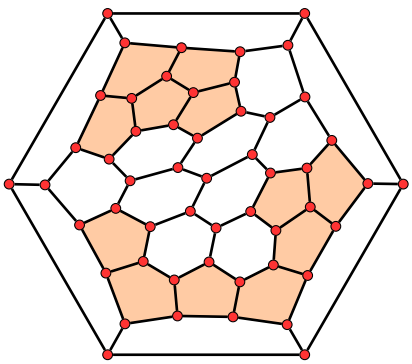
$52, 313, \mathbf{C}_1$

(55)



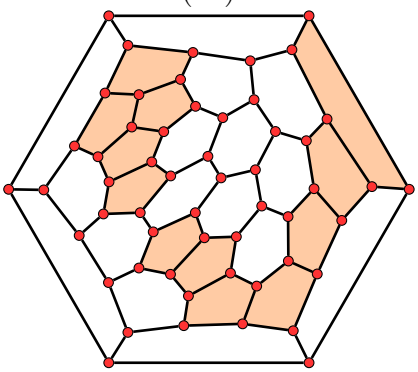
$52, 314, \mathbf{C}_1$

(56)



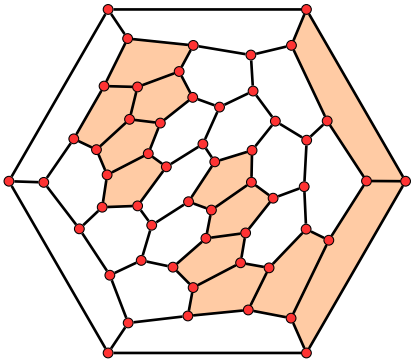
$52, 427, \mathbf{C}_1$

(57)



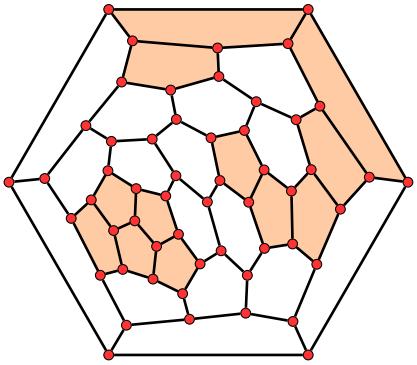
$54, 32, \mathbf{C}_1$

(58)



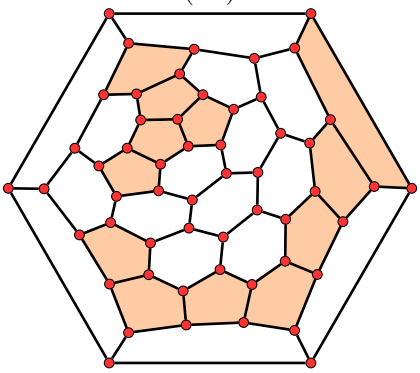
$54, 33, \mathbf{C}_1$

(59)



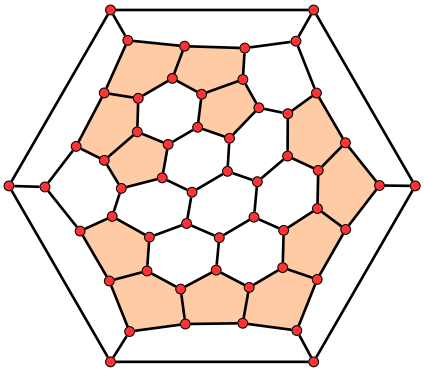
$54, 402, \mathbf{C}_s$

(60)



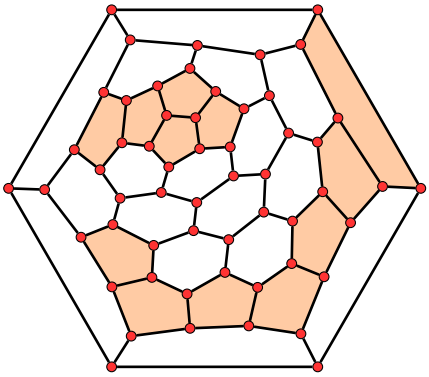
$54, 406, \mathbf{C}_1$

(61)



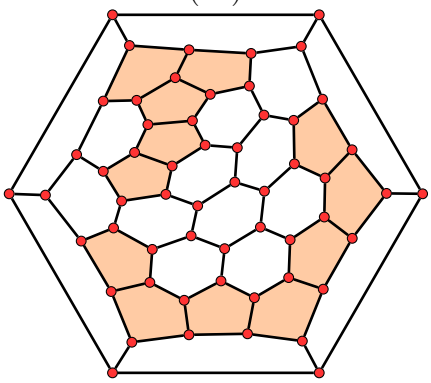
$54, 536, \mathbf{C}_s$

(62)



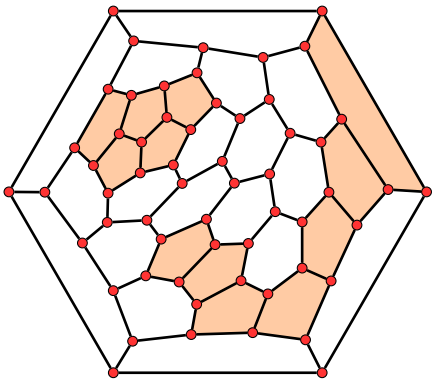
$56, 660, \mathbf{C}_1$

(63)



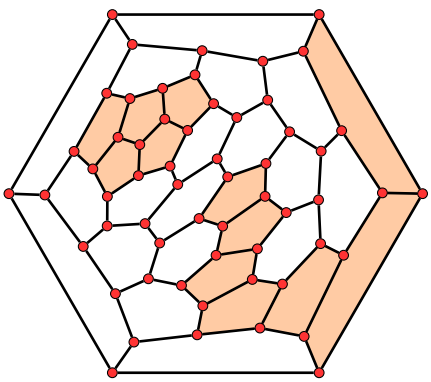
$56, 856, \mathbf{C}_1$

(64)



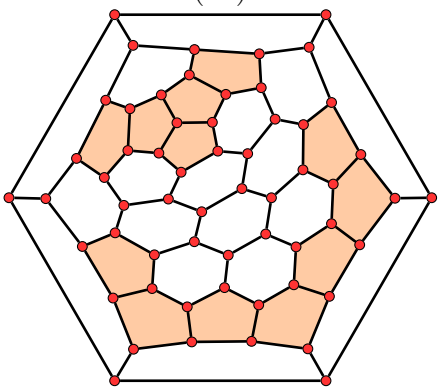
$58, 35, \mathbf{C}_1$

(65)



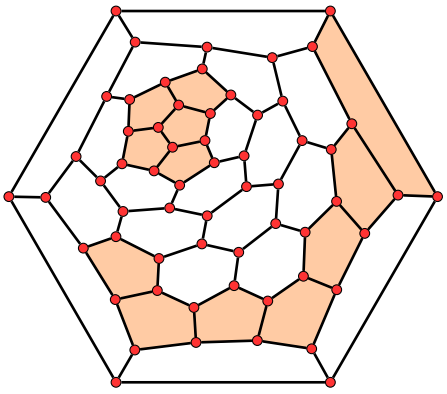
$58, 36, \mathbf{C}_1$

(66)



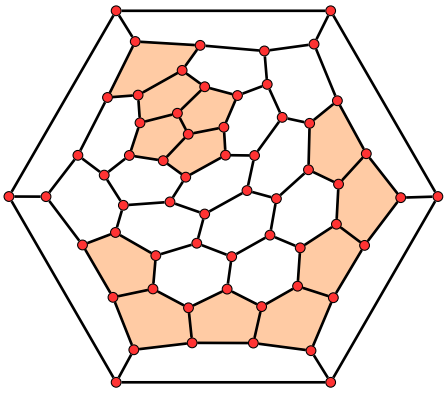
$58, 1089, \mathbf{C}_s$

(67)



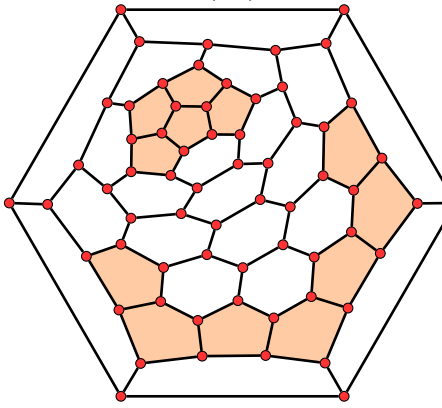
$60, 1260, \mathbf{C}_1$

(68)



$60, 1641, \mathbf{C}_s$

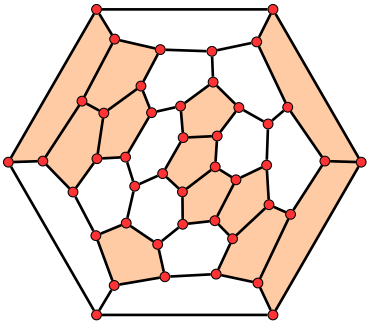
(69)



$64, 3068, \mathbf{C}_s$

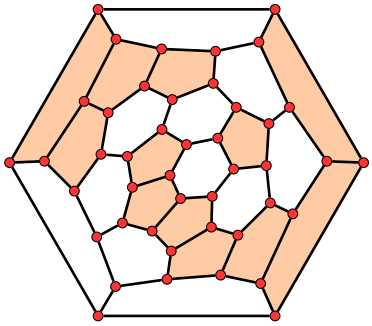
The 12 fullerenes $\mathcal{F}$ with $PIP(\mathcal{F}) = 7 + 4 + 1$

(1)



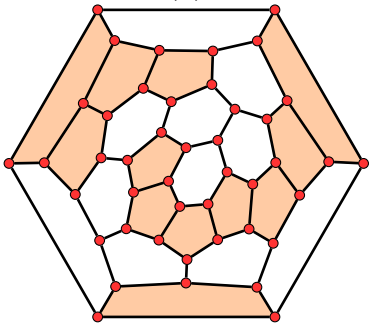
$44, 25, \mathbf{C}_s$

(2)



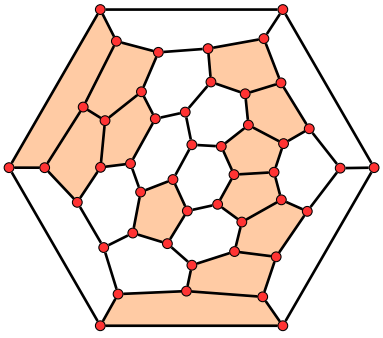
$44, 30, \mathbf{C}_s$

(3)



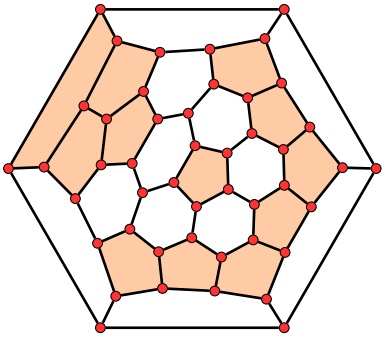
$44, 31, \mathbf{C}_1$

(4)



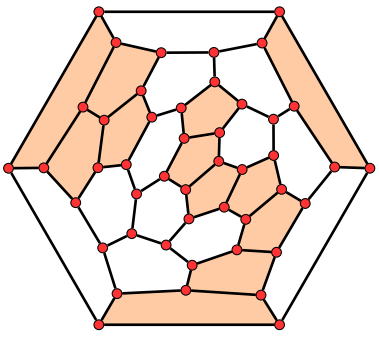
$46, 19, \mathbf{C}_1$

(5)



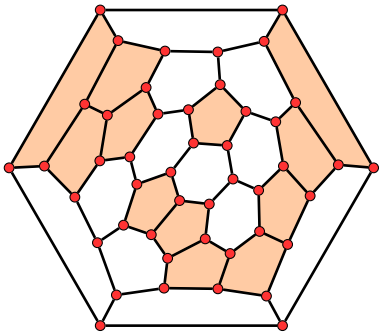
$46, 20, \mathbf{C}_s$

(6)



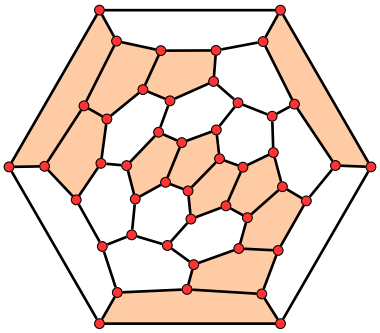
$46, 33, \mathbf{C}_1$

(7)



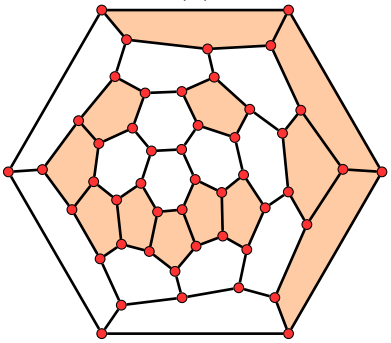
$46, 34, \mathbf{C}_1$

(8)



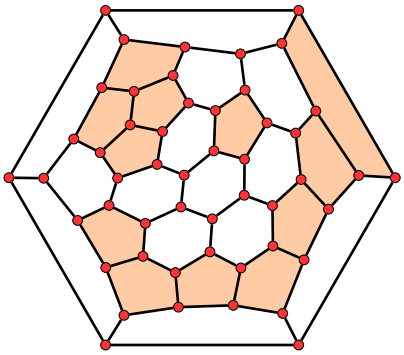
$46, 36, \mathbf{C}_1$

(9)



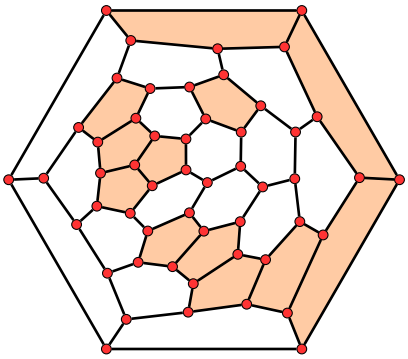
$48, 108, \mathbf{C}_1$

(10)



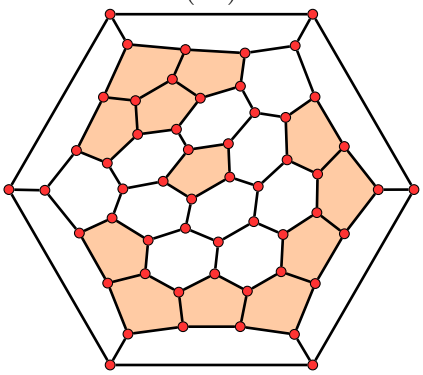
$50, 206, \mathbf{C}_1$

(11)



$52, 302, \mathbf{C}_1$

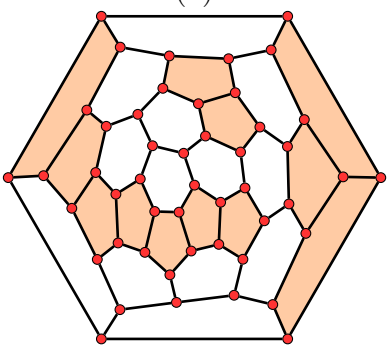
(12)



$54, 537, \mathbf{C}_s$

The fullerene $\mathcal{F}$ with $PIP(\mathcal{F}) = 7 + 3 + 2$

(1)



$48, 120, \mathbf{C}_s$

Appendix B

The Coulson Conjecture

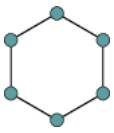
B.1 Benzenoids

Pictures of all benzenoids from Table 4.7 are shown here. Each figure contains four items:

(a) the number in paranthesis is the number of vertices;
(b) the next number is the boundary-edges code;
(c) below the boundary-edges code is the picture of the benzenoid (centres that attain the maximum bond number are coloured blue);
(d) the last number is the $\max_r N_r$ for the benzenoid.

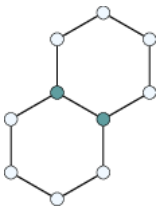
Benzenoids that attain the maximal bond number

(6)
6



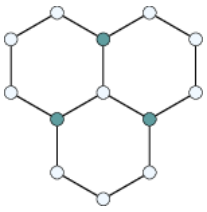
1.33333333

(10)
55



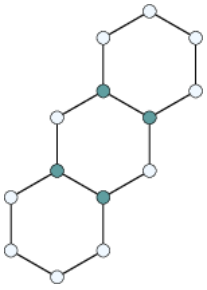
1.62763379

(13)
444



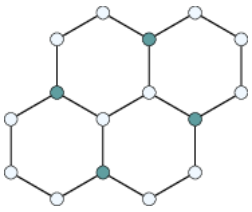
1.63054753

(14)
5252



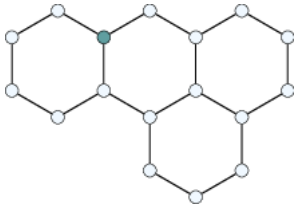
1.62631158

(16)
4343



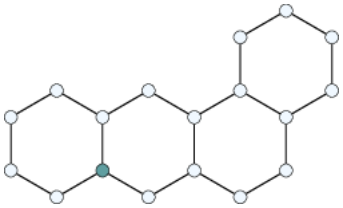
1.62132013

(17)
52441



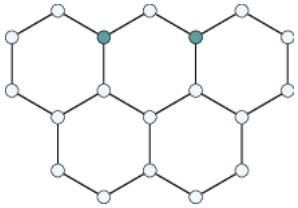
1.63220884

(18)
532521



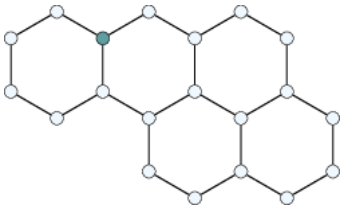
1.62715855

(19)
43342



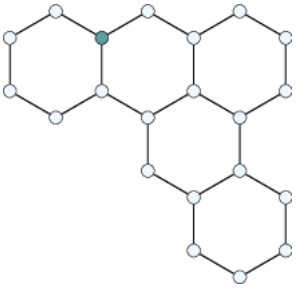
1.63164451

(20)
523431



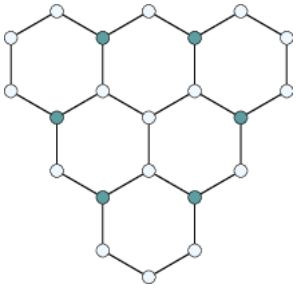
1.63130593

(21)
5241521



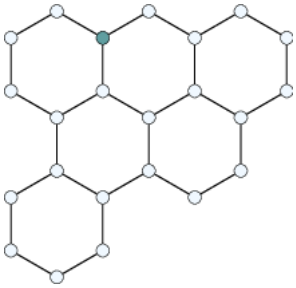
1.63182435

(22)
424242



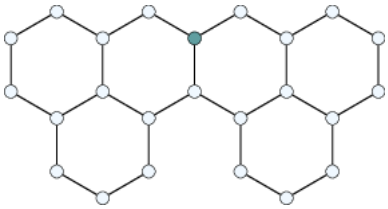
1.63238336

(23)
5142431



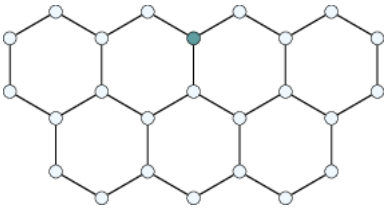
1.63206671

(24)
44224411



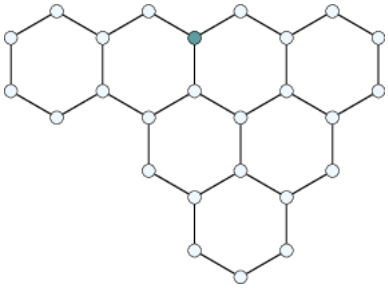
1.63672514

(25)
4323422



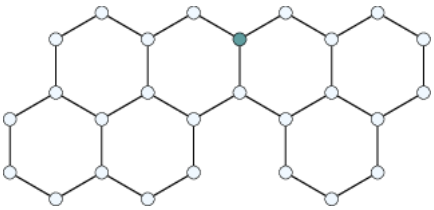
1.63252587

(26)
52242421



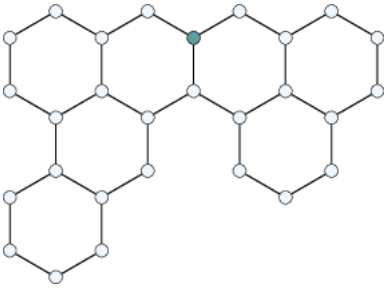
1.63273787

(27)
442234311



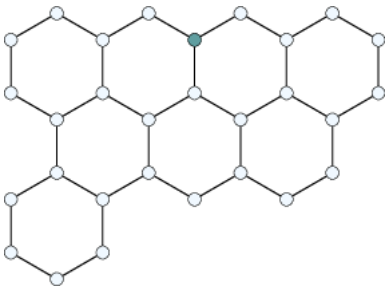
1.63584712

(28)
5211442241



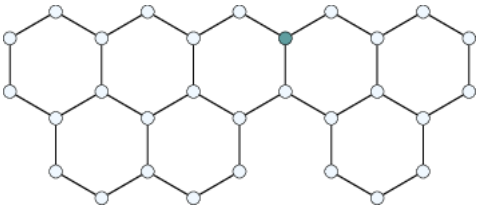
1.63635309

(29)
514224321



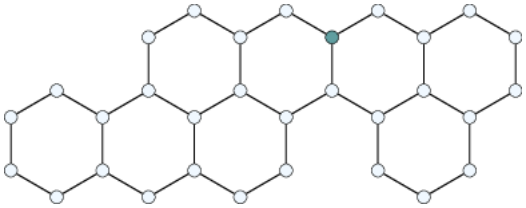
1.63223591

(30)
4422243311



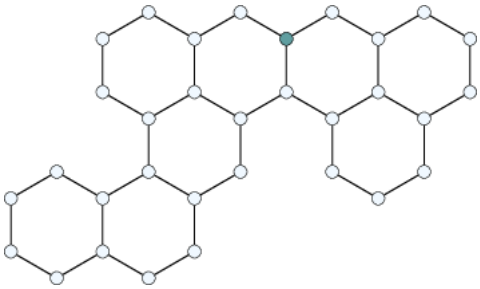
1.63620345

(31)
52311442231



1.63575386

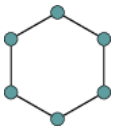
(32)
532114422411



1.63644271

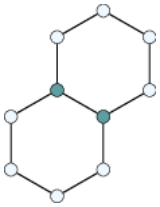
Benzenoids that minimize the maximum bond number

(6)
6



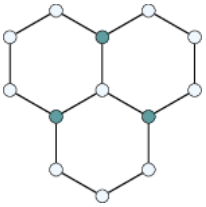
1.33333333

(10)
55



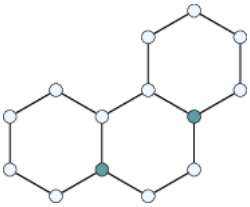
1.62763379

(13)
444



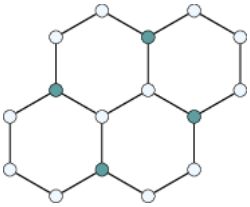
1.63054753

(14)
5351



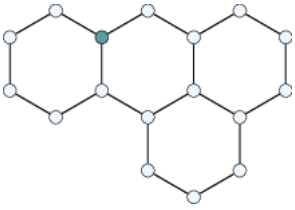
1.62334845

(16)
4343



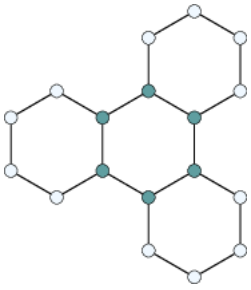
1.62132013

(17)
52441



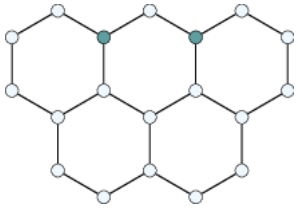
1.63220884

(18)
515151



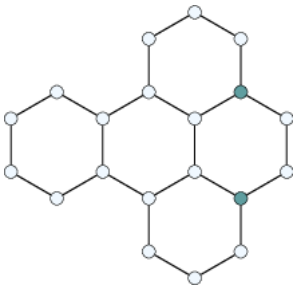
1.59242502

(19)
43342



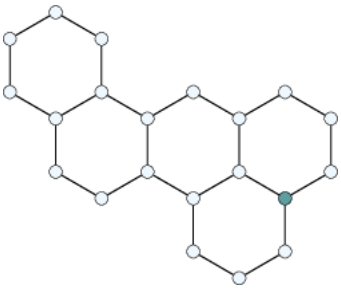
1.63164451

(20)
514341



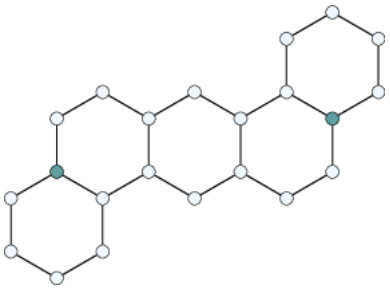
1.62228423

(21)
5314421



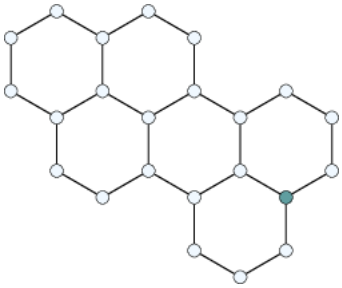
1.63060923

(22)
53215321



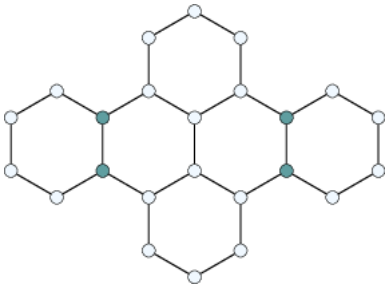
1.62191039

(23)
4413431



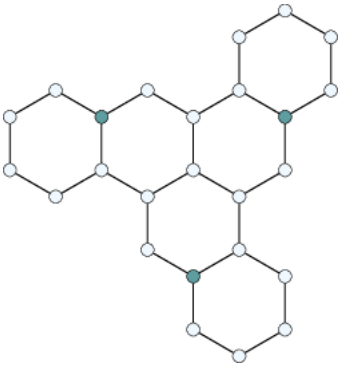
1.63032477

(24)
51415141



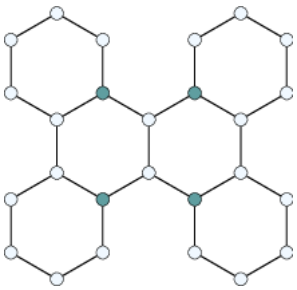
1.59253929

(25)
521521521



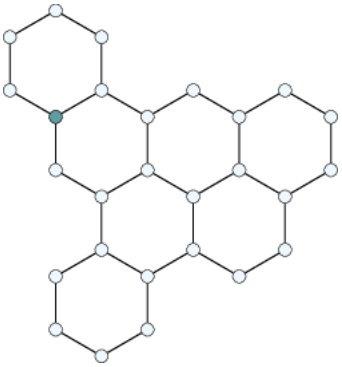
1.62993398

(26)
5151151511



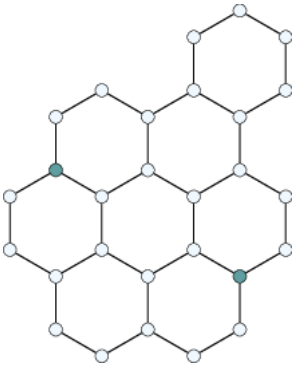
1.59980247

(27)
521513421



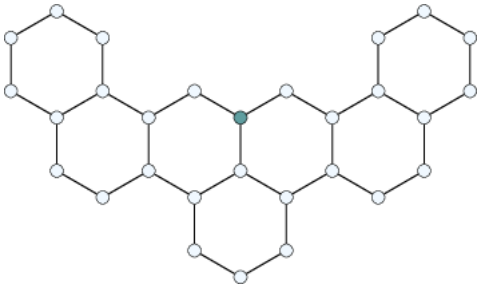
1.62969256

(28)
51333331



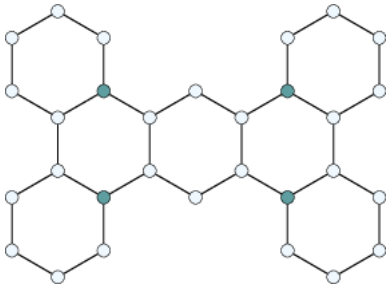
1.61604295

(29)
53141351221



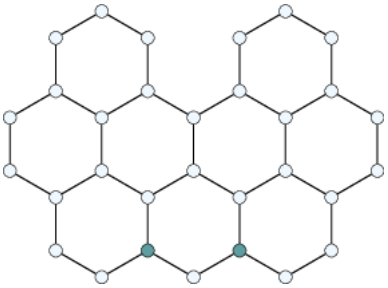
1.62602800

(30)
515121515121



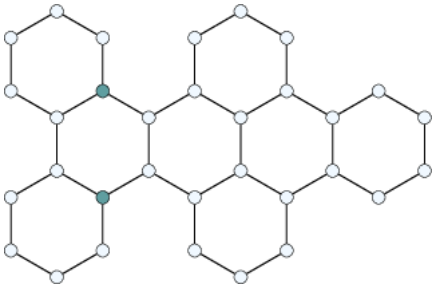
1.59179608

(31)
433233411



1.62362874

(32)
515114151411



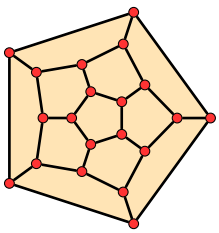
1.59969316

B.2 Fullerenes

Pictures of all fullerenes from Tables 4.5 and 4.6 are shown here. The notation used here is the same as in Appendix A.

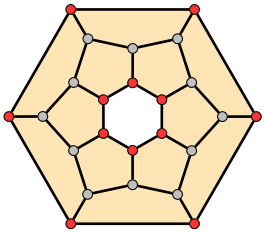
Fullerenes that minimize the maximum bond order

(1)



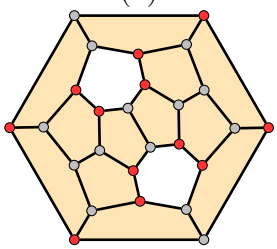
$20, 1, \mathbf{I}_h$

(2)



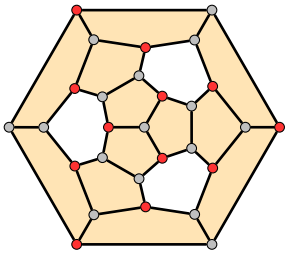
$24, 1, \mathbf{D}_{6d}$

(3)



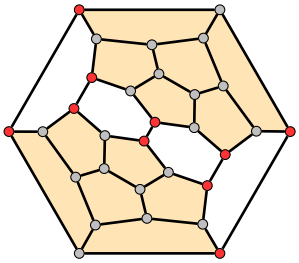
$26, 1, \mathbf{D}_{3h}$

(4)



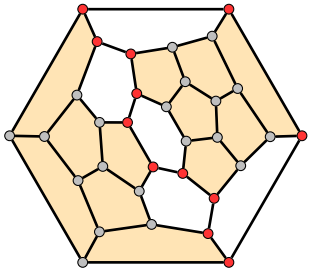
$28, 1, \mathbf{T}_d$

(5)



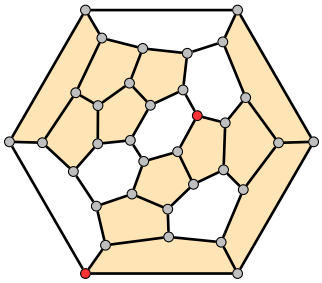
$30, 1, \mathbf{D}_{5h}$

(6)



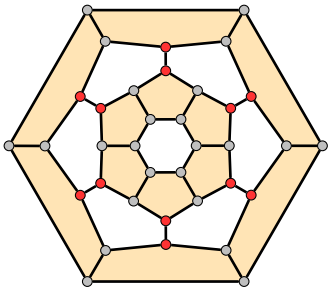
$32, 4, \mathbf{D}_{3d}$

(7)



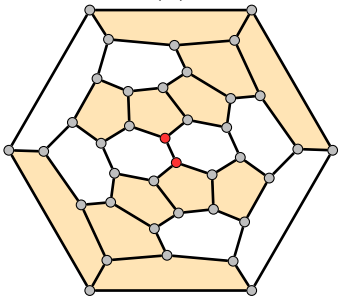
$34, 1, \mathbf{C}_s$

(8)



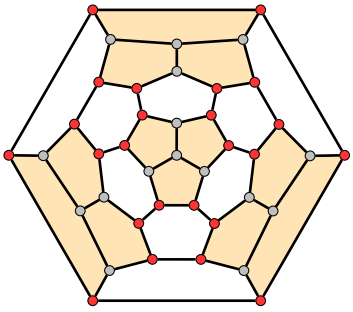
$36, 1, \mathbf{D}_{6h}$

(9)



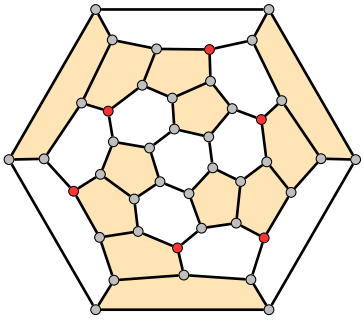
$38, 8, \mathbf{C}_2$

(10)



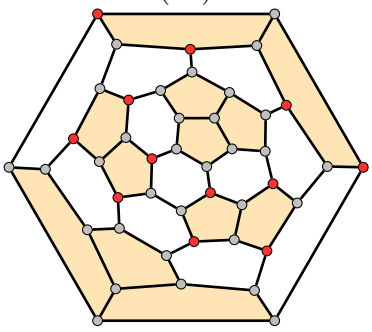
$40, 34, \mathbf{T}_d$

(11)



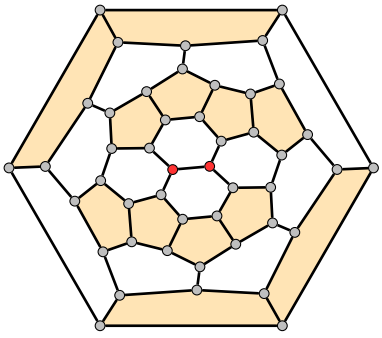
$42, 39, \mathbf{D}_3$

(12)



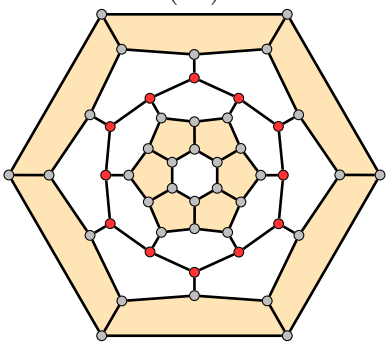
$44, 52, \mathbf{D}_{3h}$

(13)



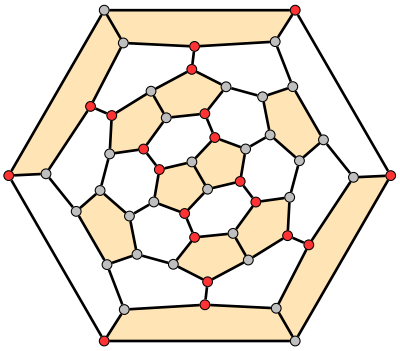
$46, 44, \mathbf{C}_2$

(14)



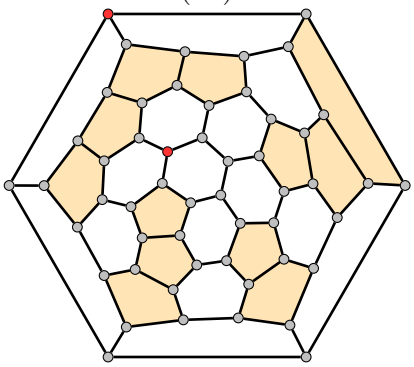
$48, 1, \mathbf{D}_{6d}$

(15)



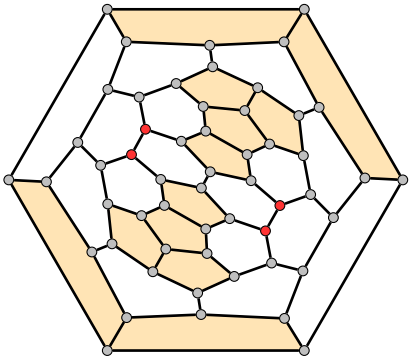
$50, 105, \mathbf{D}_{5h}$

(16)



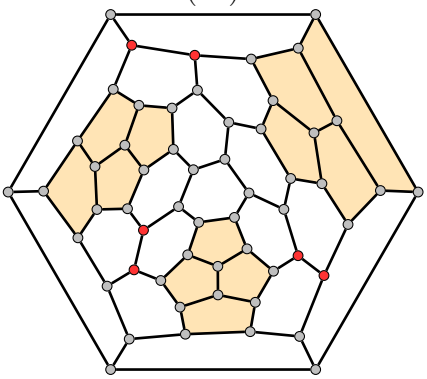
$52, 140, \mathbf{C}_2$

(17)



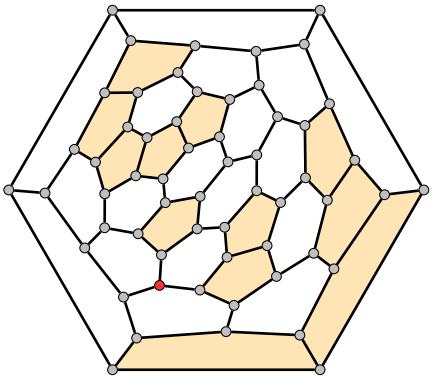
$54, 30, \mathbf{C}_{2v}$

(18)



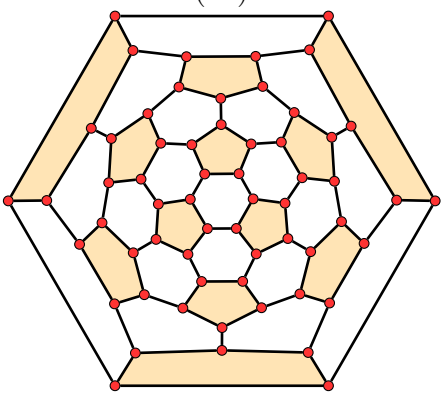
$56, 311, \mathbf{D}_3$

(19)



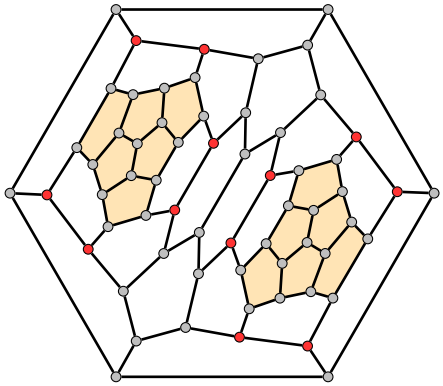
$58, 794, \mathbf{C}_1$

(20)



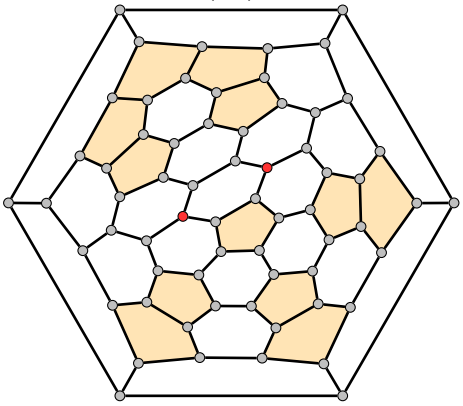
$60, 936, \mathbf{I}_h$

(21)



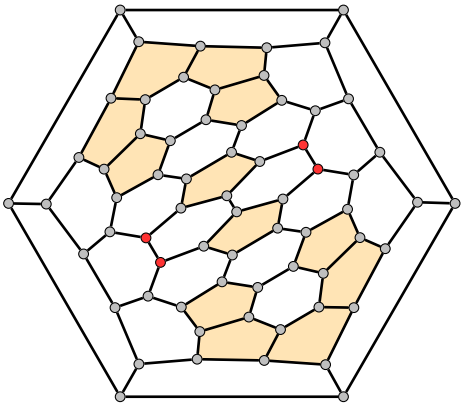
$62, 1612, \mathbf{D}_{3h}$

(22)



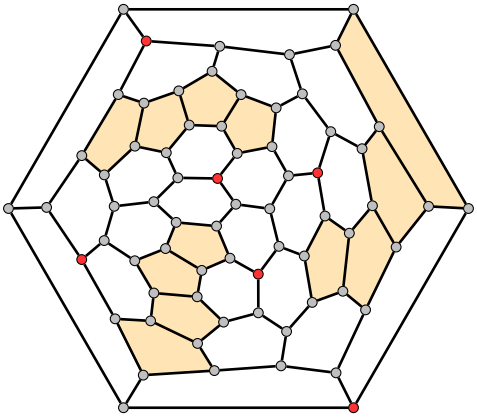
$64, 150, \mathbf{C}_s$

(23)



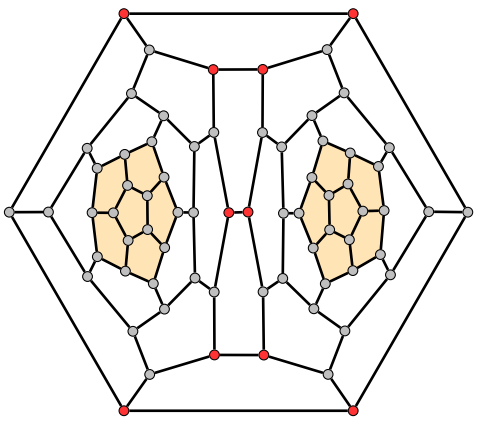
$66, 56, \mathbf{C}_{2v}$

(24)



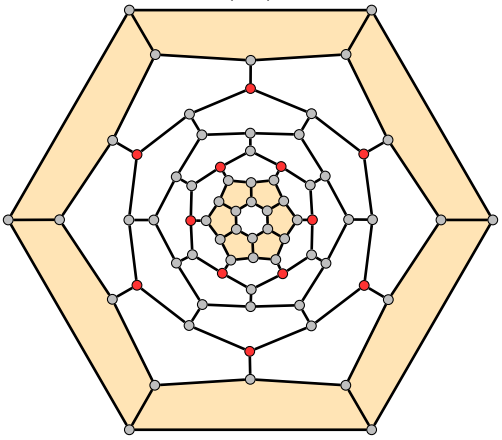
$68, 4656, \mathbf{D}_3$

(25)



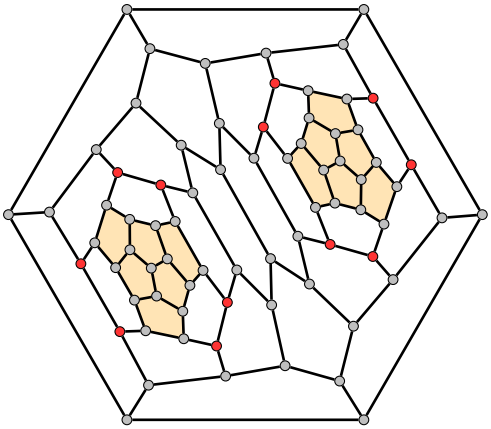
$70, 1, \mathbf{D}_{5h}$

(26)



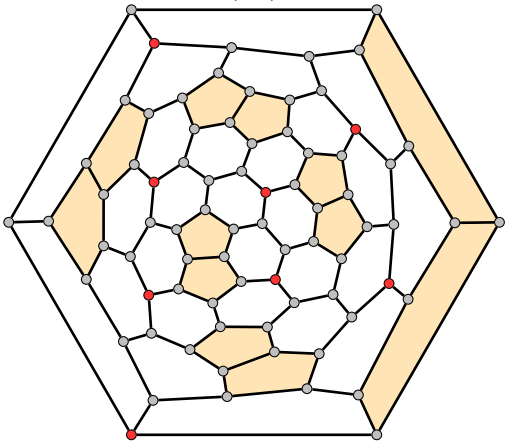
$72, 1, \mathbf{D}_{6d}$

(27)



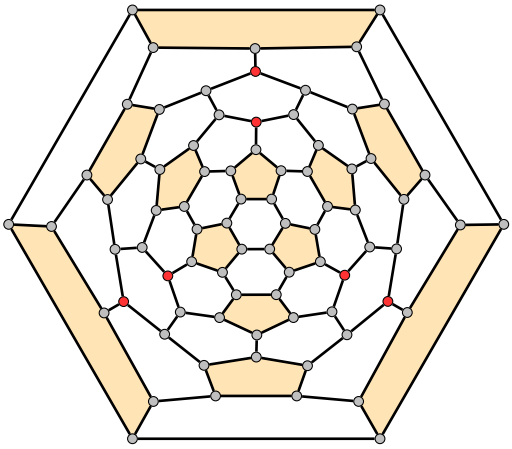
$74, 9249, \mathbf{D}_{3h}$

(28)



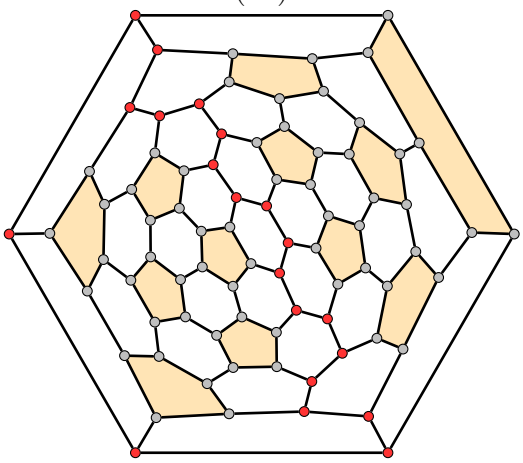
$76, 7956, \mathbf{D}_{2d}$

(29)



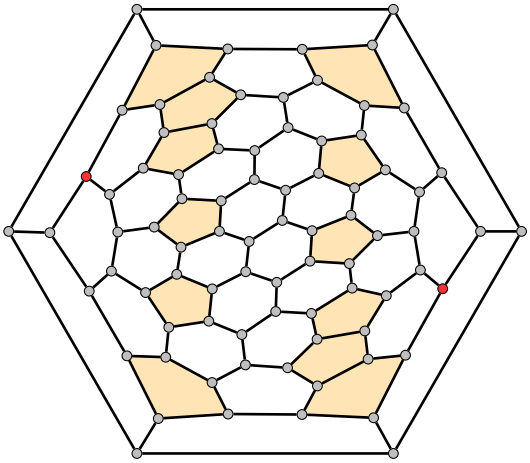
$78, 1992, \mathbf{D}_{3h}$

(30)



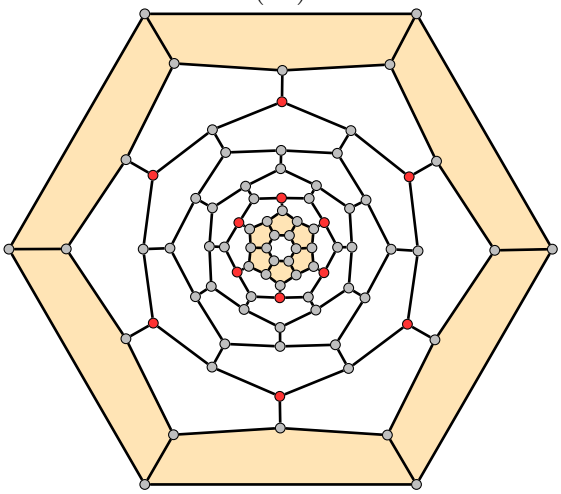
$80, 18087, \mathbf{D}_{5d}$

(31)



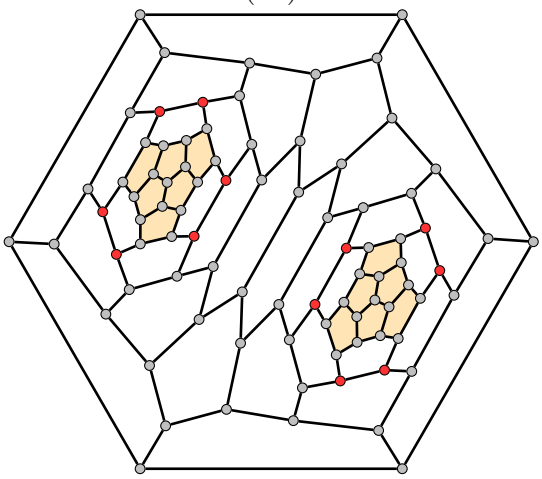
$82, 36884, \mathbf{C}_2$

(32)



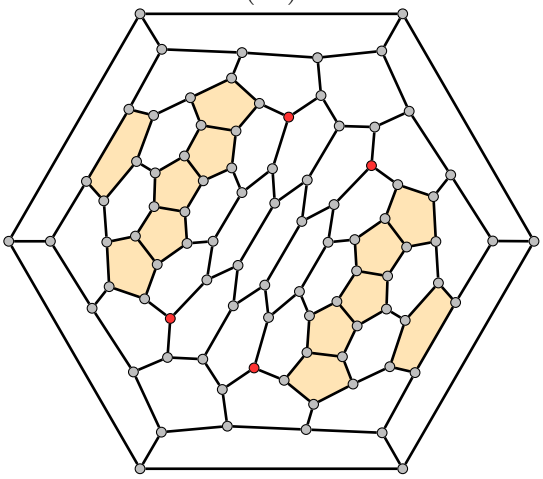
$84, 1, \mathbf{D}_{6h}$

(33)



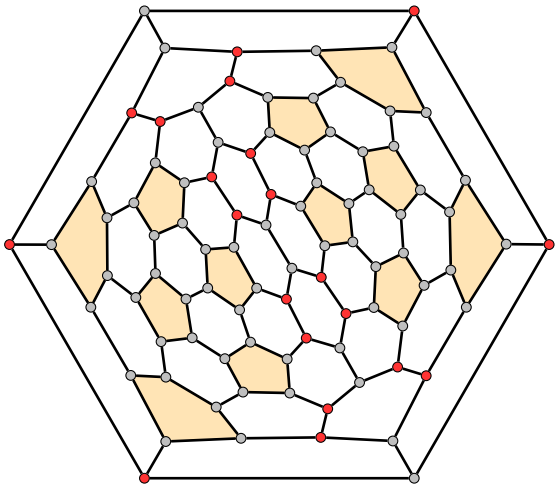
$86, 40357, \mathbf{D}_{3h}$

(34)



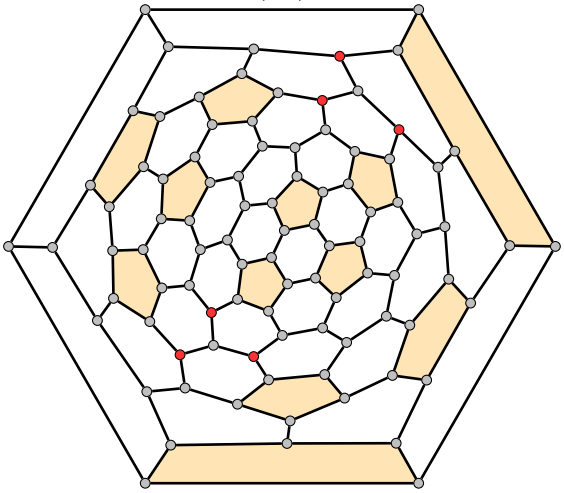
$88, 50148, \mathbf{C}_{2v}$

(35)



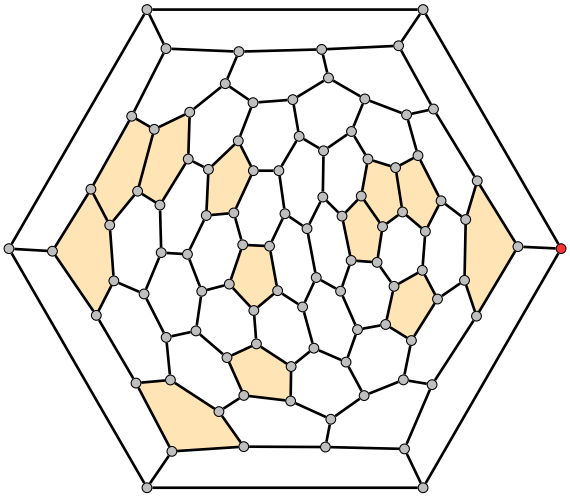
$90, 54468, \mathbf{D}_{5h}$

(36)



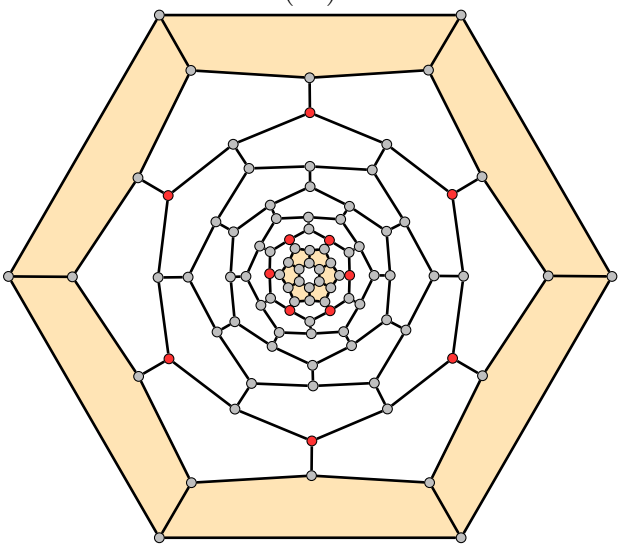
$92, 43314, \mathbf{D}_3$

(37)



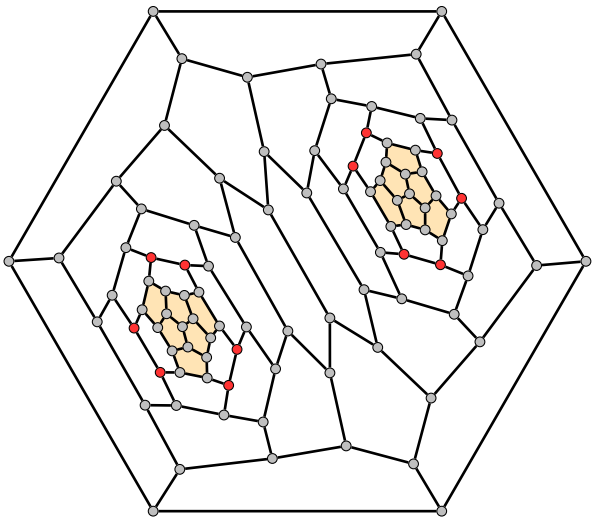
$94, 75372, \mathbf{C}_1$

(38)



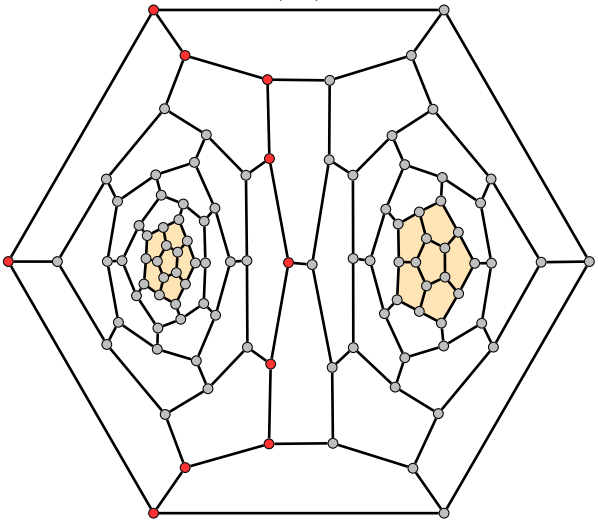
$96, 1, \mathbf{D}_{6d}$

(39)



$98, 142141, \mathbf{D}_{3h}$

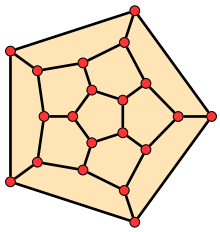
(40)



$100, 1, \mathbf{D}_{5d}$

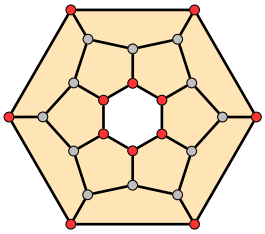
Fullerenes that obtain the maximum bond order

(1)



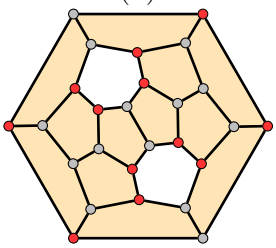
$20, 1, \mathbf{I}_h$

(2)



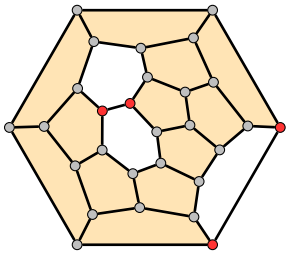
$24, 1, \mathbf{D}_{6d}$

(3)



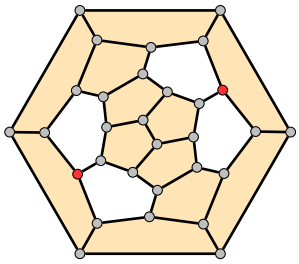
$26, 1, \mathbf{D}_{3h}$

(4)



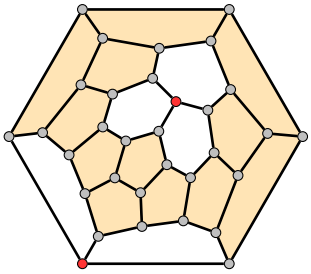
$28, 2, \mathbf{D}_2$

(5)



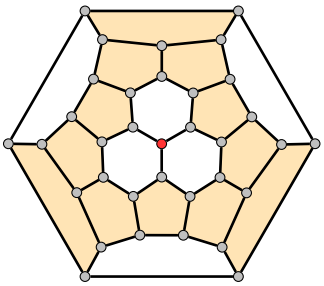
$30, 2, \mathbf{C}_{2v}$

(6)



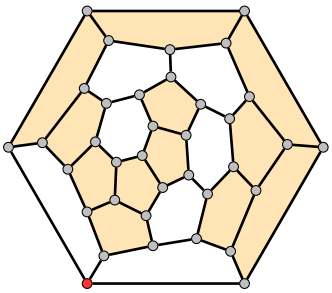
$32, 2, \mathbf{D}_{3h}$

(7)



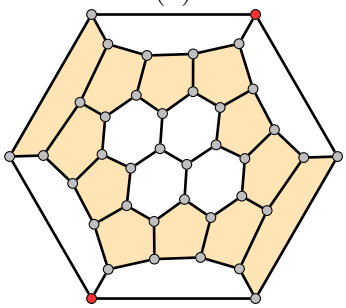
$34, 5, \mathbf{C}_{3v}$

(8)



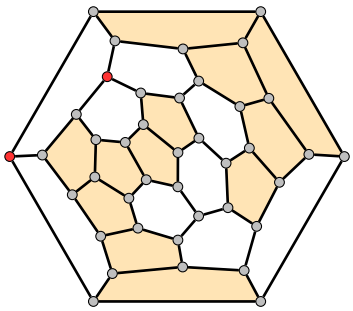
$36, 7, \mathbf{C}_1$

(9)



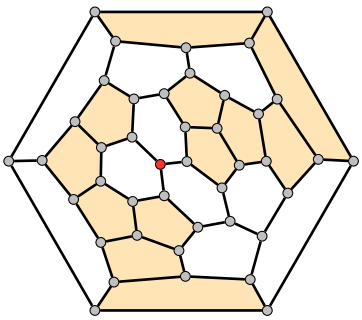
$38, 12, \mathbf{C}_{2v}$

(10)



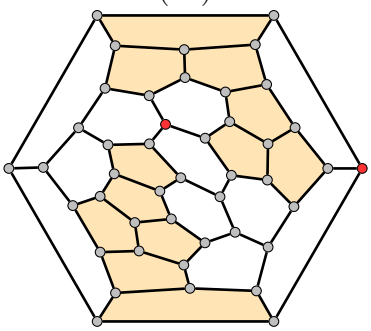
$40, 31, \mathbf{C}_s$

(11)



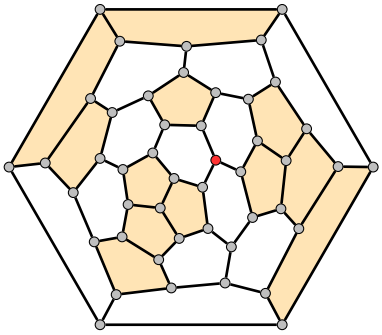
$42, 21, \mathbf{C}_1$

(12)



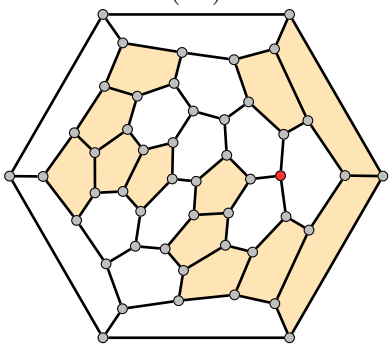
$44, 1, \mathbf{C}_{2v}$

(13)



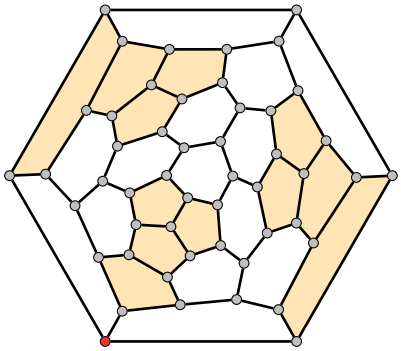
$46, 50, \mathbf{C}_s$

(14)



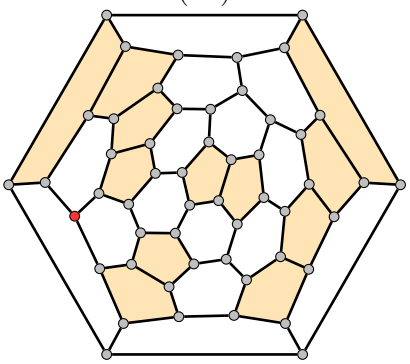
$48, 137, \mathbf{C}_s$

(15)



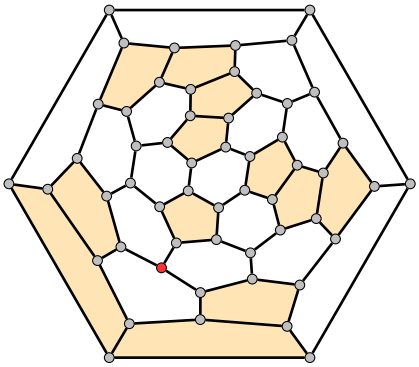
$50, 55, \mathbf{C}_{3v}$

(16)



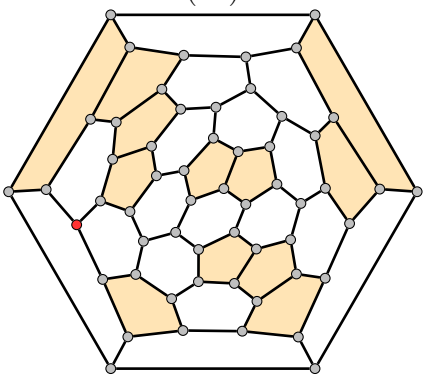
$52, 128, \mathbf{C}_1$

(17)



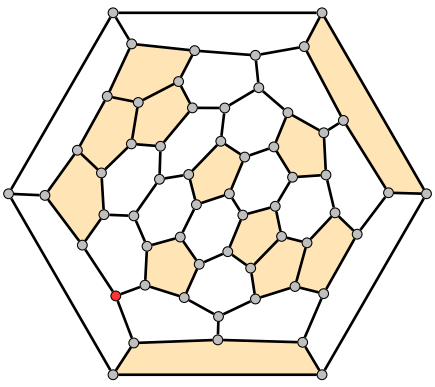
$54, 194, \mathbf{C}_1$

(18)



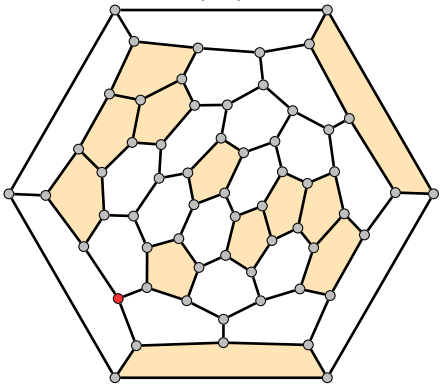
$56, 293, \mathbf{C}_s$

(19)



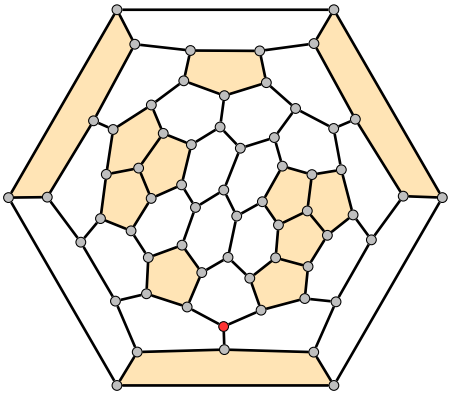
$58, 255, \mathbf{C}_1$

(20)



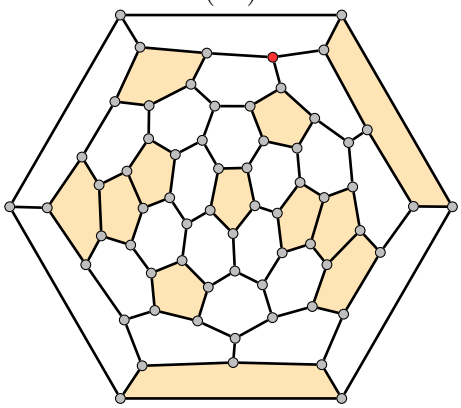
$60, 359, \mathbf{C}_s$

(21)



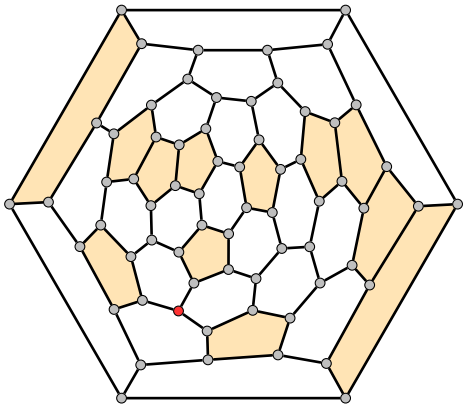
$62, 493, \mathbf{C}_1$

(22)



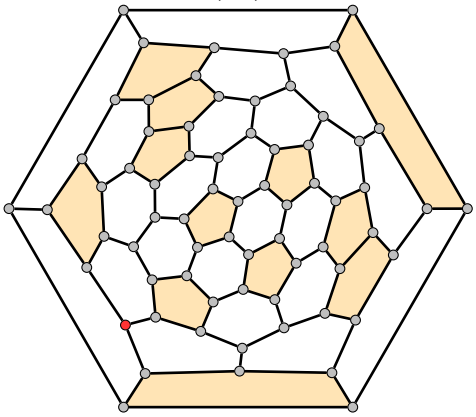
$64, 1481, \mathbf{C}_s$

(23)



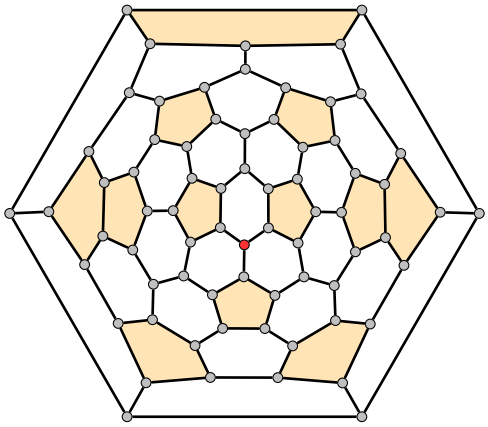
$66, 3652, \mathbf{C}_1$

(24)



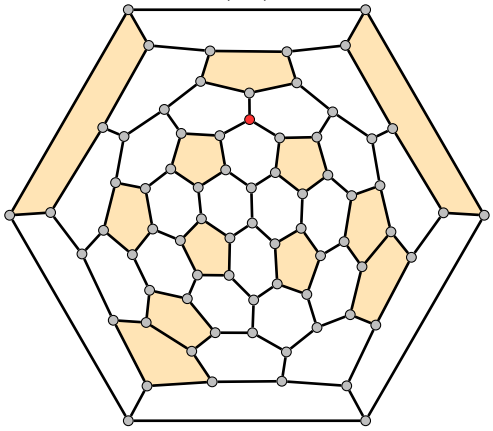
$68, 1959, \mathbf{C}_1$

(25)



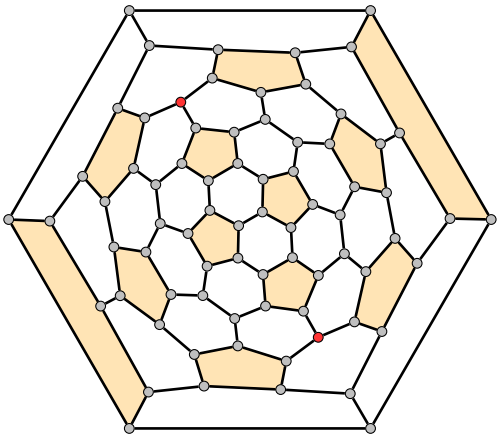
$70, 4087, \mathbf{C}_s$

(26)



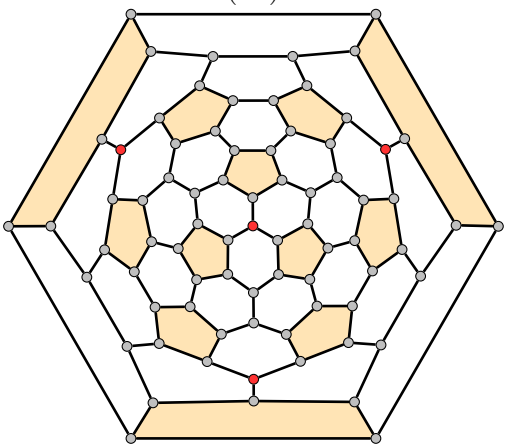
$72, 5633, \mathbf{C}_1$

(27)



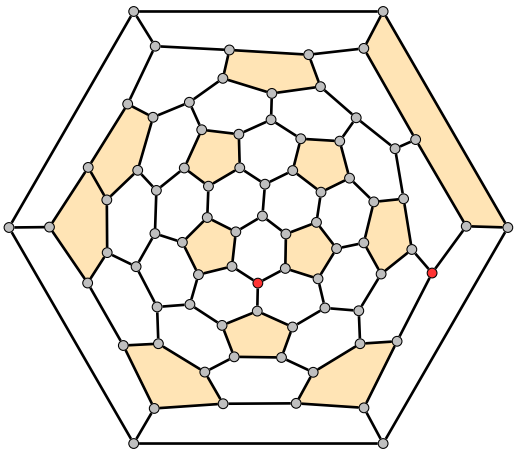
$74, 5814, \mathbf{D}_{3h}$

(28)



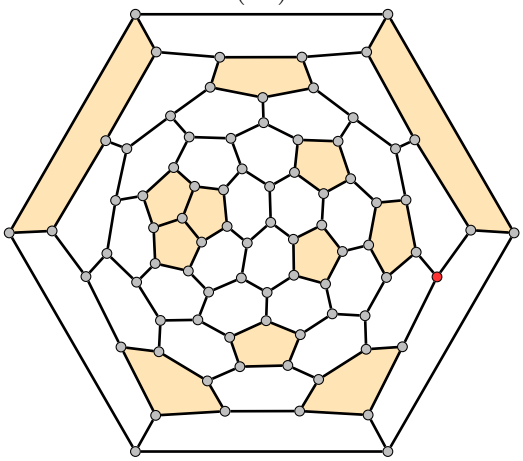
$76, 9132, \mathbf{T}_d$

(29)



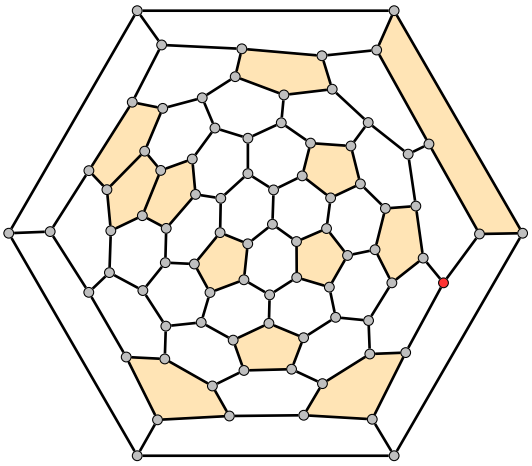
$78, 10288, \mathbf{C}_s$

(30)



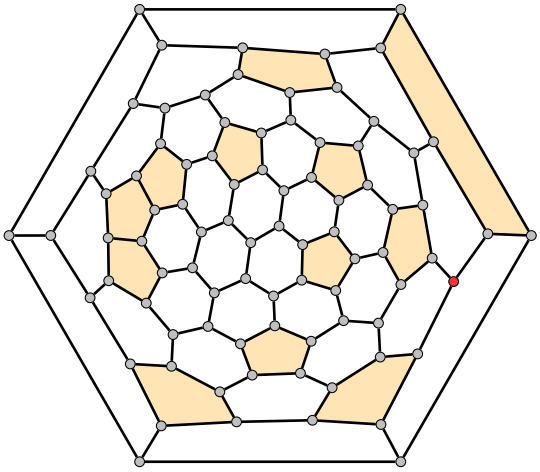
$80, 11391, \mathbf{C}_{3v}$

(31)



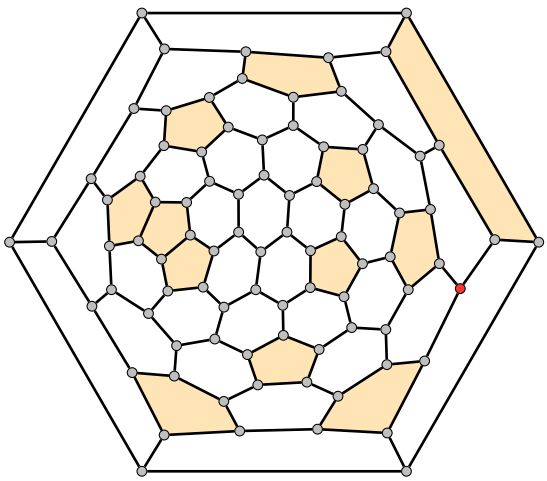
$82, 13243, \mathbf{C}_1$

(32)



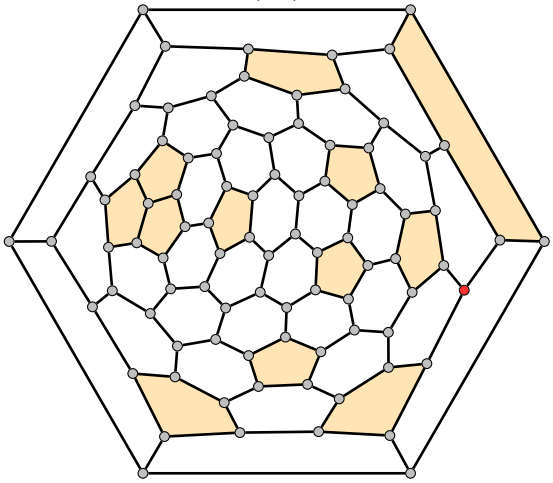
$84, 14510, \mathbf{C}_1$

(33)



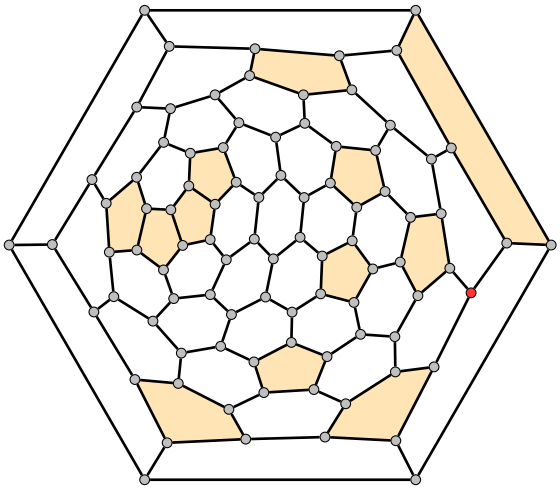
$86, 15156, \mathbf{C}_1$

(34)



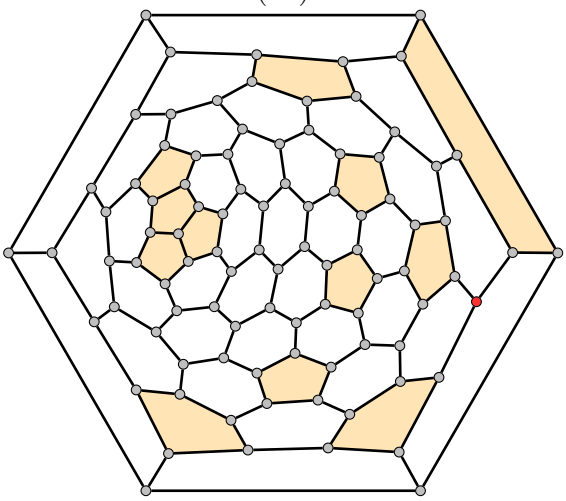
$88, 14729, \mathbf{C}_1$

(35)



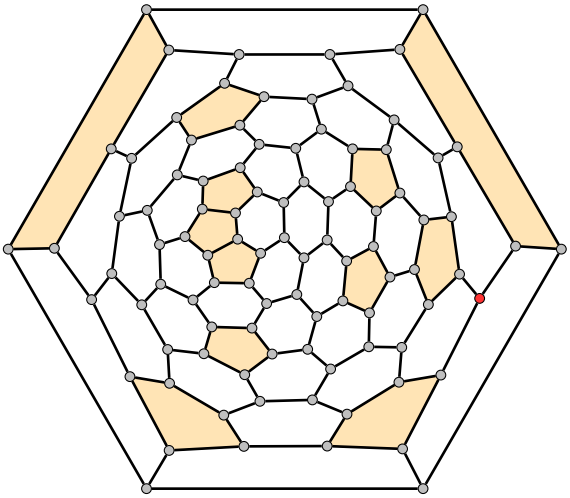
$90, 16030, \mathbf{C}_1$

(36)



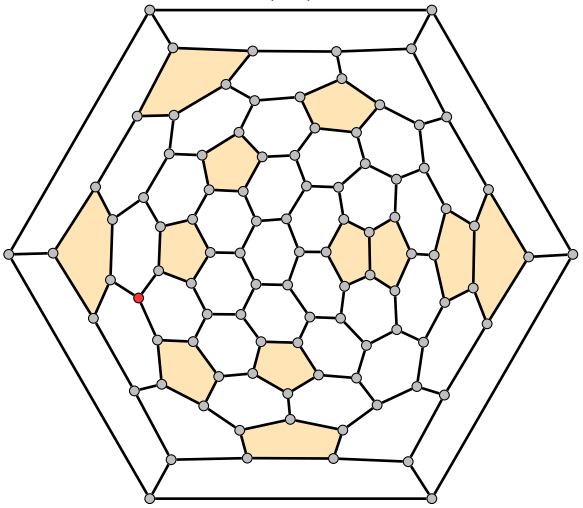
$92, 16163, \mathbf{C}_1$

(37)



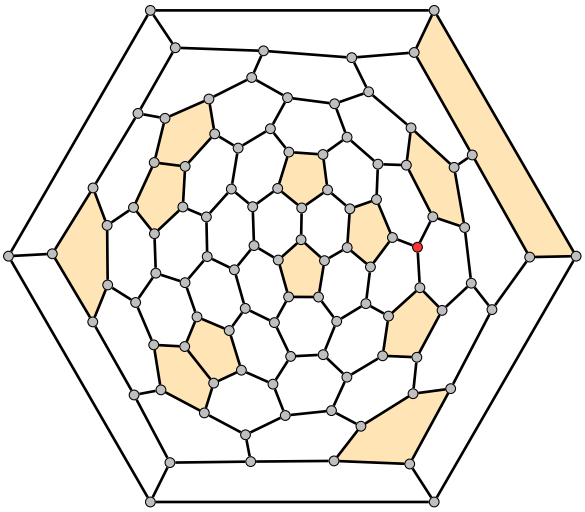
$94, 47431, \mathbf{C}_s$

(38)



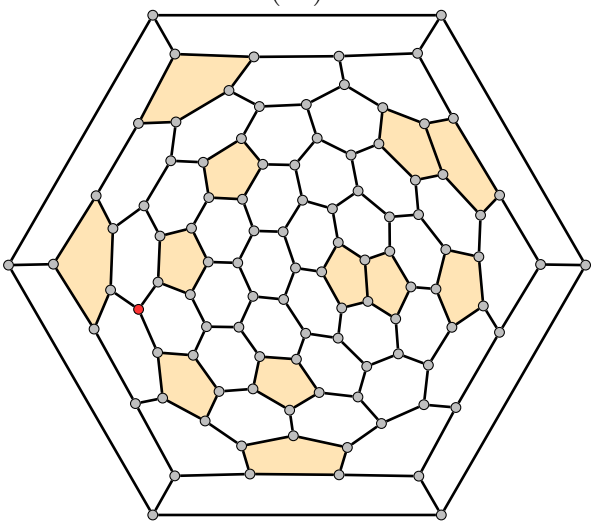
$96, 126731, \mathbf{C}_s$

(39)



$98, 101379, \mathbf{C}_1$

(40)



$100, 186008, \mathbf{C}_1$

Index

Razširjeni slovenski povzetek

V zadnjem času so na področju matematične kemije nastala številna dela, ki opisujejo uporabo matematičnih orodij pri analizi fizikalnih lastnosti in pojavov, ki so pomembni v fizikalni kemiji in molekularni fiziki. To je vzbudilo pozornost nekaterih matematikov, še posebej tistih, ki se ukvarjajo s teorijo grafov in uporabo linearne algebre v teoriji grafov. Po drugi strani se je izkazalo, da so številni koncepti iz algebrajske in topološke teorije grafov primerni za opis in analizo ogljikovodikov, kot so npr. heksagonalni sistemi in fulereni. V tem delu obravnavamo več problemov iz matematične kemije s povsem matematičnega vidika. Poleg tega predstavimo nekaj kemijskih konceptov preko teorije grafov. Namen tega je vzpostavitev tesnejše povezave med matematiko in kemijo. Med molekularnimi grafi, ki jih obravnavamo, namenimo posebno pozornost heksagonalnim sistemom.

Drugo poglavje je preliminarno. V njem vpeljemo matematične objekte, ki jih uporabljamo v tem delu. Najprej spoznamo grafe. Posebno pozornost namenimo dvodelnim in ravninskim grafom. Ogledamo si tudi mnogokotniške ploskve in zemljevide. Zemljevidi so kombinatorični opisi celičnih vložitev grafov v ploskve. Kemijski graf je lahko načeloma vsak graf, ki ima določeno uporabo v matematični kemiji. V ožjem pomenu besedne zveze "kemijski graf" imamo običajno v mislih subkubične grafe. Nato pokažemo, kako lahko molekule opišemo z grafi. Posebej pomembni so Hücklovi grafi, tj. grafi ogljikovodikov, kjer vodikove atome zanemarimo. Razdelek 2.4 je posvečen točkovnim grupam. Z njimi opišemo simetrijo molekul, pri čemer upoštevamo tudi njihovo geometrijo, tj. molekulo podamo s položaji njenih atomov v trirazsežnem prostoru. Pri tem uporabljamo Schönfliesovo notacijo, ki je razširjena v kemiji in kristalografiji. Zanimajo nas le tiste podgrupe grupe $O(3)$, ki pridejo v poštev kot grupe simetrij molekul. To so grupe $\mathbf{D}_{\infty h}$, $\mathbf{C}_{\infty v}$, $\mathbf{C}_1$, $\mathbf{C}_s$, $\mathbf{C}_i$, $\mathbf{T}_d$, $\mathbf{T}$, $\mathbf{T}_h$, $\mathbf{O}$, $\mathbf{O}_h$, $\mathbf{I}$, $\mathbf{I}_h$ in naslednje družine:

$$\mathbf{C}_n,\ \mathbf{S}_n,\ \mathbf{D}_n\,\mathbf{C}_{nh},\ \mathbf{C}_{nv},\ \mathbf{D}_{nh} \text{ ter } \mathbf{D}_{nd}. \tag{8.1}$$

Parameter $n \geq 2$ je naravno število. Pri grupi $\mathbf{S}_n$ zahtevamo še, da je število n sodo. Med drugim pokažemo, da ima vsaka molekula, ki vsebuje vsaj 3 nekolinearne atome, končno grupo simetrij.

V razdelku 2.6 predstavimo nekaj ključnih rezultatov spektralne teorije grafov. Ta igra ključno vlogo pri Hücklovi teoriji. Pri spektralni teoriji grafov študiramo lastne vrednosti matrike sosednosti grafa. Zanima nas predvsem, kaj nam lahko lastne vrednosti povedo o lastnostih pripadajočih grafov. Znano je na primer, da je graf dvodelen natanko tedaj, ko je njegov spekter simetričen (glede na število 0). Za nekatere družine grafov (polne grafe, polne dvodelne grafe, poti in cikle) izračunamo lastne vrednosti in pripadajoče lastne vektorje. Predstavimo še izrek o prepletanju, ki nam pove, v kakšni relaciji so lastne vrednosti nekega

grafa in njegovega induciranega podgrafa. Prav tako pomemben je Perron-Frobeniusov izrek, ki pove, da ima vsak graf neko realno lastno vrednost, ki je po absolutni vrednosti večja ali enaka od vseh ostalih lastnih vrednosti, njen pripadajoči lastni vektor pa ima same nenegativne komponente.

V tretjem poglavju vpeljemo kemijske koncepte preko teorije grafov. Poglavje je namenjeno matematikom, ki želijo bolje razumeti kemijsko ozadje problemov, ki jih obravnavamo v tem delu. Najprej na kratko predstavimo teorijo valenčne vezi, kjer se seznanimo s σ in π vezjo. Razdelek 3.2 je nekoliko daljši in je posvečen Hücklovi teoriji molekulskih orbital. To je bržkone najpreprostejša teorija molekulskih orbital, ki je nastala na podlagi kvantne mehanike. Omogoča izračun energij molekulskih orbital nenasičenih ogljikovodikov oz. konjugiranih π sistemov. Začnemo s Schrödingerjevo valovno enačbo

$$\mathcal{H}\Psi = E\Psi. \tag{8.2}$$

in naredimo celo vrsto poenostavitev oz. aproksimacij. Osrednjega pomena v kvantni kemiji je Born-Oppenheimerjeva aproksimacija. Ta pravi, da lahko atomom priredimo fiksne koordinate v prostoru, saj se v primerjavi z elektroni gibljejo zelo počasi. Molekulske orbitale izrazimo kot linearne kombinacije atomskih orbital, tj.

$$\phi_i = \sum_{r=1}^{n} c_r^{(i)} \chi_r, \tag{8.3}$$

kjer je ϕ_i i-ta molekulska orbitala, $\{\chi_r\}_{r=1}^n$ pa je v naprej dana baza atomskih orbital. Ker je izraz (8.3) le aproksimacija, zelo natančnih rezultatov niti ne pričakujemo. Namen Hücklove teorije ni, da bi zelo natančno izračunali lastnosti ene same molekule, ampak nam omogoči predvsem, da lahko med seboj primerjamo cele družine molekul. Po daljši izpeljavi dobimo matrično enačbo

$$(\mathbf{H} - \varepsilon I_{n\times n})\mathbf{c} = 0_{n\times 1}, \tag{8.4}$$

kjer je $\mathbf{H}$ hermitska matrika, ki ji pravimo Hücklov Hamiltonian. V enačbi (8.4) je ε energija molekulske orbitale, elementi matrike $\mathbf{H} = [H_{ij}]_{i,j=1}^n$ pa so definirani kot integrali

$$H_{ij} \coloneqq \int \chi_i \mathcal{H} \chi_j \,\mathrm{d}\tau. \tag{8.5}$$

Hückel je naredil še eno aproksimacijo in sicer je vse vrednosti H_{ii}, $1 \leq i \leq n$, postavil na α, vrednost H_{ij}, $i \neq j$, $1 \leq i, j \leq n$ pa je enaka β, če sta i-ti in j-ti atom povezana s σ vezjo, in 0 sicer. To pomeni, da energije molekulskih orbital določimo tako, da poiščemo lastne vrednosti matrike sosednosti Hücklovega grafa molekule. V Hücklovem grafu vsako vozlišče predstavlja en ogljikov atom konjugiranega π sistema, povezave pa σ vezi med njimi. Če so $\lambda_1, \ldots, \lambda_n$ lastne vrednosti Hücklovega grafa, potem so energije molekulskih orbital

$$\varepsilon_i = \alpha + \lambda_i \beta. \tag{8.6}$$

Vsaki lastni vrednosti torej pripada ena molekulska orbitala, ki lahko gosti do dva elektrona. Večkratnim lastnim vrednostim pripadajo tako imenovane degenerirane orbitale. Elektroni polnijo orbitale v takem vrstnem redu, da se najprej napolnijo tiste z nižjo energijo. Pri degeneriranih orbitalah prejme vsaka po en elektron, še preden prejme katera

od njih drugega. Če poznamo število π elektronov neke molekule, lahko zapišemo njeno elektronsko konfiguracijo, tj. vektor števil $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$, kjer je

$$v_i = \begin{cases} 2, & \text{če ima orbitala } \phi_i \text{ dva elektrona;} \\ 1, & \text{če ima orbitala } \phi_i \text{ en elektron;} \\ 0, & \text{če je orbitala } \phi_i \text{ prazna.} \end{cases} \tag{8.7}$$

Skupna energija π elektronov sistema je

$$E^\pi = \sum_{i=1}^{n} v_i \varepsilon_i = \sum_{i=1}^{n} v_i(\alpha + \lambda_i \beta). \tag{8.8}$$

Definirajmo še π naboj atoma r:

$$Q_r = \sum_{i=1}^{n} v_i \left(c_r^{(i)}\right)^2, \tag{8.9}$$

kjer je v_i število elektronov v i-ti orbitali, $c_r^{(i)}$ pa r-ta komponenta i-tega lastnega vektorja. Če sta atoma r in s povezana s σ vezjo, potem je red π vezi med njima

$$P_{rs}^\pi = \sum_{i=1}^{n} v_i c_r^{(i)} c_s^{(i)}. \tag{8.10}$$

Definiramo lahko še skupni red vezi $P_{rs}^{\sigma+\pi} = P_{rs}^\sigma + P_{rs}^\pi$, ki poleg prispevka π vezi vključuje še prispevek σ vezi. Red σ vezi med atomoma r in s, P_{rs}^σ, je enak 1, če sta sosednja v Hücklovem grafu. Številu

$$N_r^\pi = \sum_{s \in G(r)} P_{rs}^\pi \tag{8.11}$$

pravimo π vezno število atoma r. To je vsota vseh redov π vezi, ki dani atom r povezujejo z ostalimi. Podobno je

$$N_r^{\sigma+\pi} = \sum_{s \in G(r)} P_{rs}^{\sigma+\pi} \tag{8.12}$$

skupno vezno število atoma r. V Hücklovi teoriji poznamo tudi količino, ki ji pravimo prosta valenca. Prosta valenca atoma r je

$$F_r = N_{\max}^{\sigma+\pi} - N_r^{\sigma+\pi}, \tag{8.13}$$

kjer je $N_{\max}^{\sigma+\pi}$ največje vezno število, ki ga je teoretično mogoče doseči. Število F_r interpretiramo kot neizkoriščen potencial, ki ga ima atom r za tvorbo kemijskih vezi. V literaturi zasledimo, da je vrednost $N_{\max}^{\sigma+\pi}$ enaka $3+\sqrt{3}$, a se izkaže, da to ni bilo nikoli dokazano. Tej domnevi pravimo Coulsonova domneva o maksimalnem veznem številu in se ji posvetimo v 4. poglavju.

Razdelek 3.3 je posvečen fulerenom. Ogljik ima mnogo alotropov, med katerimi sta diamant in grafit znana že dolgo. Raziskovanje fulerenov se je začelo v 80. letih, ko je Kroto s sodelavci napovedal obstoj fulerena C_{60}, ki je znan kot buckminsterfuleren in ima grupo simetrij ikozaedra. Po tem odkritju je postalo raziskovanje fulerenov zelo popularno, Kroto in sodelavci pa so leta 1996 prejeli Nobelovo nagrado za kemijo. Fulerene lahko predstavimo s konveksnimi poliedri ali pa, ekvivalentno, z grafi.

Definicija 1. Fulerenski graf *je kubičen ravninski 3-povezan graf, ki ima sama petkotniška in šestkotniška lica.*

Da sta predstavitvi z grafi in s konveksnimi poliedri zares ekvivalentni, nam pove Steinitzev izrek, ki pravi, da je skelet vsakega konveksnega poliedra v $\mathbb{R}^3$ ravninski 3-povezan graf. Tudi obratno je res, tj. vsak ravninski 3-povezan graf je skelet nekega konveksnega poliedra v $\mathbb{R}^3$. Preprosta posledica Eulerjeve formule je, da ima vsak fuleren natanko 12 petkotniških lic. Za vsako sodo število $n \geq 20$, razen za $n = 22$, obstaja vsaj en fuleren. Poseben razred fulerenov so tako imenovani IPR fulereni, pri katerih nobena dva petkotnika nimata skupne povezave. V tem delu definiramo petkotniško incidenčno particijo:

Definicija 2. *Petkotniška incidenčna particija fulerena $\mathcal{F}$, $PIP(\mathcal{F})$, je particija števila 12, ki jo dobimo tako, da preštejemo vozlišča v vsaki povezani komponeti grafa, ki je induciran na vozliščih stopnje 5 v ravninskem dualu fulerena $\mathcal{F}$.*

Z uporabo petkotniške incidenčne particije lahko definiramo IPR fulerene:

Definicija 3. *Fuleren $\mathcal{F}$ je IPR fuleren, če je*

$$PIP(\mathcal{F}) = \underbrace{1 + 1 + \cdots + 1}_{12}. \tag{8.14}$$

Izkaže se, da obstaja le 41 fulerenov $\mathcal{F}$, za katere velja $PIP(\mathcal{F}) = 12$. Zelo pomemben je tudi naslednji izrek:

Izrek 1 (Mani)**.** *Naj bo G 3-povezan ravninski graf. Potem v $\mathbb{R}^3$ obstaja konveksen polieder P, za katerega velja, da je graf G 1-skelet poliedra P in polieder P realizira grupo simetrij grafa G.* □

To med drugim pomeni, da fulereni nimajo "skritih simetrij" in da je njihove točkovne grupe mogoče poiskati na kombinatoričen način. Znano je, da pride v poštev samo 28 točkovnih grup, ki so prikazane v tabeli 3.5.

Generiranje fulerenskih grafov omogoča program **fullgen**, ki sta ga leta 1995 razvila Brinkmann in Dress. Leta 2012 pa so Brinkmann, Goedgebeur in McKay razvili še hitrejši program, ki se imenuje **buckygen**.

V razdelku 3.4 spoznamo Kekuléjeve strukture. To so pravzaprav strukturne formule, pri katerih je vsak ogljikov atom povezan z natanko eno dvojno vezjo z nekim drugim ogljikovim atomom, ostale vezi pa so enojne. Temu kemijskemu konceptu v teoriji grafov ustrezajo popolna prirejanja. Prirejanje je podmnožica F povezav grafa G, kjer nobeni dve povezavi iz F nista incidenčni. Če je vsako vozlišče grafa G krajišče kake povezave, potem je prirejanje popolno. Pravimo, da je graf Kekuléjev, če ima vsaj eno popolno prirejanje. Tega lahko poiščemo v polinomskem času. Preštevanje Kekuléjevih struktur (oz. popolnih prirejanj) pa je v splošnem NP-težek problem. Za nekatere posebne družine grafov lahko Kekuléjeve strukture še vedno preštejemo v polinomskem času. Kasteleynov izrek nam pove, da so ravninski grafi (na primer fulereni in koronoidi) takšna družina.

Ob koncu poglavja si pogledamo še metodo konjugiranih tokokrogov, s pomočjo katere lahko napovemo obroče tokove oz. tokove π elektronov. Spoznamo še Paulingov red vezi, tj.

razmerje med številom Kekuléjevih struktur, ki vsebujejo neko dano povezavo, in številom vseh Kekuléjevih struktur grafa.

V razdelku 3.5 obravnavamo matematični model za opis samosestavljanja poliedrov iz verig peptidov. Leta 2013 so Gradišar in soavtorji zasnovali polipeptid TET12, ki se uspešno samosestavi v tetraeder. Veriga vsebuje 12 peptidov in je zasnovana tako, da se določeni pari peptidov "zlepijo skupaj". Končni rezultat je tetraeder, ki ima 6 povezav. Polipeptid TET12 modeliramo z usmerjeno potjo $\vec{P}_{13}$, ki ima oznake na svojih povezavah. Vsaka povezava namreč predstavlja en peptid, oznake pa povedo, katere peptide uporabimo. Poleg tega podamo še lepilno preslikavo, ki nam pove, kateri peptidi (povezave) se zlepijo skupaj.

Matematični model za načrtovanje polipeptidov, ki se zložijo v želeno strukturo, sta opisala Rus in Klavžar. Naj bo

$$W = w_0\, w_1\, w_2 \ldots w_{2m} \tag{8.15}$$

dvojni obhod grafa G (tj. vsako povezavo prečkamo natanko dvakrat). Izkaže se, da ima vsak graf dvojni obhod. Če peptide izberemo tako, kot nam veleva dvojni obhod, se bo morda polipeptid sestavil v želeni skelet poliedra. Klavžar in Rus sta poiskala dodatne pogoje, ki jih mora izpolnjevati dvojni obhod, da se bo polipeptid gotovo sestavil v želeni skelet. V tem primer mu pravimo stabilen obhod. Pokazala sta, da stabilni obhod grafa G obstaja, če in samo če ima graf G minimalno stopnjo vsaj 3.

Kasneje so Fijavž, Pisanski in Rus koncept stabilnega obhoda nadgradili in definirali krepek obhod grafa. Dokazali so, da v vsakem grafu obstaja krepek obhod.

Slike krepkih obhodov običajno rišemo tako, kot je prikazano na sliki 3.37. V tem delu se osredotočimo na grafe, ki so vloženi v ploskve. Vložitev predstavimo z zemljevidom in definiramo obhod zemljevida. Obhodi zemljevida so v tesnem sorodstvu s krepkimi obhodi. Predstavljajmo si, da zemljevid razrežemo po povezavah. Tako dobimo "koščke sestavljanke", ki ustrezajo licem. Obhod zemljevida inducira delne obhode na svojih licih. Izpeljali smo algoritem, ki uporablja tehniko dinamičnega programiranja za preštevanje (oz. enumeracijo) obhodov zemljevida. Obhod zemljevida lahko sestavimo iz delnih obhodov tako, da lica lepimo skupaj in spajamo združljive delne obhode.

V četrtem poglavju se posvetimo Coulsonovi domnevi o maksimalnem veznem številu.

Domneva 1. *Za vse kemijske grafe G velja*

$$N^{\pi}(G) \leq \sqrt{3}. \tag{8.16}$$

□

Domneva 2. *Zvezda $K_{1,3}$ je edini kemijski graf, ki doseže enakost v enačbi (8.16), tj. $N^{\pi}(K_{1,3}) = \sqrt{3}$ in $N^{\pi}(G) < \sqrt{3}$, če je $G \ncong K_{1,3}$ katerikoli drug kemijski graf.* □

Najprej pokažemo, da je vezno število omejeno navzgor:

Trditev 2. *Naj bo G Hücklov graf. Potem velja*

$$N_r^{\pi} \leq 2\lambda_1(G). \tag{8.17}$$

□

Za dvodelne grafe lahko zgornjo mejo nekoliko izboljšamo:

Trditev 3. *Naj bo G dvodelen Hücklov graf. Potem velja*

$$N_r^\pi \leq \lambda_1(G). \tag{8.18}$$

□

Uspelo nam je pokazati, da domneva velja za grafe, ki imajo natanko eno pozitivno lastno vrednost. V dokazu smo uporabili Smithovo karakterizacijo grafov z natanko eno pozitivno lastno vrednostjo. Z uporabo Mooreove meje in izreka o prepletanju smo pokazali:

Izrek 4. *Obstaja končno mnogo (neizomorfnih) grafov, ki imajo maksimalno stopnjo omejeno z danim številom Δ in natanko n_+ pozitivnih lastnih vrednosti.* □

Iz tega izreka direktno sledi:

Posledica 5. *Obstaja končno mnogo (neizomorfnih) kemijskih grafov, ki imajo natanko n_+ pozitivnih lastnih vrednosti.* □

To pomeni, da bi lahko domnevo potrdili za grafe, ki imajo $n_+ > 1$ pozitivnih lastnih vrednosti tako, da bi jih poiskali s pomočjo računalnika.

Za vsak $n \in \{2, 3, \ldots, 20\}$ smo poiskali graf, ki ima maksimalno π vezno število med vsemi kemijskimi grafi na n vozliščih. Pri tem nastopata dve družini grafov:

Definicija 4. Pohabljena gosenica *na n vozliščih je graf G, ki ima množico vozlišč*

$$V(G) = \{v_0, v_1, \ldots, v_{n-1}\}$$

in množico povezav

$$E(G) = \{v_{i-1}v_i \mid 1 \leq i \leq n-1 \wedge i \bmod 3 \neq 0\} \cup \{v_{i-2}v_i \mid 2 \leq i \leq n-1 \wedge i \bmod 3 = 0\}.$$

Definicija 5. Zmaj *na $n \geq 5$ vozliščih je graf G, ki ima množico vozlišč*

$$V(G) = \{v_0, v_1, \ldots, v_{n-1}\}$$

in množico povezav

$$E(G) = \{v_0v_2, v_0v_3, v_1v_2, v_1v_3, v_2v_4, v_3v_4\} \cup \{v_{i-1}v_i \mid 5 \leq i \leq n-1 \wedge i \bmod 3 \neq 1\} \cup \{v_{i-2}v_i \mid 7 \leq i \leq n-1 \wedge i \bmod 3 = 1\}.$$

Pohabljeno gosenico z n vozlišči označimo s $\mathrm{CriCat}(n)$, zmaja z n vozlišči pa s $\mathrm{Kite}(n)$.

Trditev 6. *Med vsemi kemijskimi grafi z $2 \leq n \leq 20$ vozlišči nobeno π vezno število ne preseže meje $\sqrt{3}$ iz domneve 1.* □

Glede na rezultate, ki smo jih dobili z računalnikom, smo postavili naslednjo domnevo:

Domneva 3. *Kemijski graf, ki doseže največje π vezno število med vsemi grafi z n vozlišči, je* Kite(n) *v primeru, ko je $n \geq 14$ in $n \bmod 3 = 2$. V vseh ostalih primerih največje π vezno število doseže graf* CriCat(n). □

Če se omejimo le na kemijska drevesa, se izkaže, da za $2 \leq n \leq 20$ največje π vezno število doseže pohabljena gosenica CriCat(n). To nas napeljuje k naslednji domnevi:

Domneva 4. *Kemijsko drevo, ki doseže največje π vezno število med vsemi kemijskimi drevesi z n vozlišči, je* CriCat(n). □

Preiskali smo tudi vse fulerene, ki imajo največ 100 vozlišč. Rezultati so prikazani v tabeli 4.5, njihove slike pa najdete v dodatku B. Izkazalo se je, da noben fuleren z $n \leq 100$ vozlišči ne preseže meje $\sqrt{3}$ iz domneve 1. Opazili smo, da v teh fulerenih največje π vezno število vedno realizira vozlišče, ki je incidenčno trem šestkotniškim licem, razen če so petkotniška lica razporejena tako, da ne dopuščajo vozlišč, ki bi bila incidenčna zgolj s šestkotniškimi lici. Postavili smo naslednjo domnevo:

Domneva 5. *Naj bo $\mathcal{F}$ fuleren in $\mathcal{V} \subset V(\mathcal{F})$ množica vozlišč, ki so incidenčna trem šestkotniškim licem. Če je množica $\mathcal{V}$ neprazna, potem je vsako vozlišče, ki realizira maksimalno π vezno število v fulerenu $\mathcal{F}$, nujno element množice $\mathcal{V}$.* □

Poiskali smo tudi vse fulerene, ki dosežejo minimalno vrednost $\max_r N_r$ med vsemi fulereni z n vozlišči. Rezultati so prikazani v tabeli 4.6, slike pa najdete v dodatku B. Če primerjamo fulerene, ki maksimizirajo vrednost $\max_r N_r$, s tistimi, ki to vrednost minimizirajo, opazimo, da je v večini primerov (toda ne vedno) grupa simetrij fulerena, ki minimizira $\max_r N_r$, večja od grupe simetrij fulerena, ki to vrednost maksimizira. Opazimo tudi, da se v tabeli 4.6 pogosto pojavljajo nanocevke, katerih kapice so obliži na sliki 4.6.

Simetrije nam torej niso v pomoč, če iščemo kemijske grafe, ki dosežejo velike vrednosti $\max_r N_r$. Intuicija pravi, da se mora mnogo vozlišč žrtvovati, če želimo, da eno med njimi zbere dovolj naboja. Temu bi lahko rekli tudi princip mušketirjev ("Vsi za enega, eden za vse!").

Preiskali smo tudi benzenoide, ki imajo največ $n \leq 32$ vozlišč. Iskali smo tako tiste, ki minimizirajo vrednost $\max_r N_r$, kot tudi tiste, ki to vrednost maksimizirajo. Tudi benzenoidi, ki smo jih preiskali, ne presežejo meje $\sqrt{3}$ iz domneve 1. Če nas zanimajo kemijski grafi, ki imajo velike vrednosti $\max_r N_r$, jih gotovo ne bomo našli med benzenoidi in fulereni.

Benzenoidi so pomembna družina ogljikovodikov. Obravnavamo jih v petem poglavju. V monografiji [102] jih Gutman in Cyvin definirata kot kondenzirane policiklične nenasičene konjugirane ogljikovodike, ki so sestavljeni zgolj iz šestkotniških obročev. V razdelku 5.2 natančno spoznamo neskončno šestkotniško mrežo, ki jo označimo s $\mathcal{H}$. V naslednjem razdelku predstavimo klasični pristop k teoriji benzenoidov. V razdelku 5.4 pa obravnavamo nov pristop k teoriji heksagonalnih sistemov, ki temelji zgolj na incidencah med šestkotniki, ki tvorijo neskočno šestkotniško mrežo $\mathcal{H}$.

Na $\mathcal{H}$ poglejmo kot na neskončno množico šestkotnikov. Heksagonalni sistem $\mathcal{K} \subseteq \mathcal{H}$ je poljubna podmnožica množice $\mathcal{H}$. Če imata šestkotnika a in b skupno povezavo, pravimo, da sta sosedna in pišemo $a \sim b$. Šestkotnika a in b sta v relaciji $\equiv_{\mathcal{K}}$, $a \equiv_{\mathcal{K}} b$, če obstaja zaporedje

$$c_0 = a, c_1, c_2, \ldots, c_m = b, \tag{8.19}$$

kjer je $c_{i-1} \sim c_i$ za vse $i = 1, 2, \ldots, m$ in $c_i \in \mathcal{K}$. Jasno je, da je $\equiv_{\mathcal{K}}$ ekvivalenčna relacija, heksagonalni sistem $\mathcal{K}$ pa razpade na ekvivalenčne razrede $\{\mathcal{C}_i\}_{i \in C(\mathcal{K})}$, ki jih imenujemo povezane komponente. Pravimo, da je heksagonalni sistem povezan, če ima eno samo povezano komponento. Benzenoide in koronoide definiramo takole:

Definicija 6. *Koronoid $\mathcal{K}$ je končen povezan heksagonalni sistem. Če je tudi njegov komplement $\mathcal{K}^{\complement}$ povezan, mu pravimo benzenoid.*

Nato dokažemo zelo uporabno lemo:

Lema 7. *Naj bo $\mathcal{K}$ končni heksagonalni sistem. Njegov komplement $\mathcal{K}^{\complement}$ ima $d+1$, $d \geq 0$, (tj. končno mnogo) povezanih komponent:*

$$\mathcal{K}^{\complement} = \mathcal{C}_\infty \sqcup \mathcal{C}_1 \sqcup \mathcal{C}_2 \sqcup \ldots \sqcup \mathcal{C}_d. \tag{8.20}$$

Vsaka od teh komponent razen ene, ki je označena s $\mathcal{C}_\infty$, je končna. Vsaka končna komponenta $\mathcal{C}_i$, $1 \leq i \leq d$, je koronoid. □

Zgornja lema seveda ne velja za neskončne heksagonalne sisteme. S pomočjo zgornje leme smo pokazali naslednji izrek:

Izrek 8. *Naj bo $\mathcal{K}$ koronoid. Njegov komplement $\mathcal{K}^{\complement}$ ima končno število $b(\mathcal{K}) = d+1$, $d \geq 0$, povezanih komponent $\mathcal{B}_\infty, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_d$ ter velja:*

$$\mathcal{K}^{\complement} = \mathcal{B}_\infty \sqcup \mathcal{B}_1 \sqcup \mathcal{B}_2 \sqcup \ldots \sqcup \mathcal{B}_d. \tag{8.21}$$

Natanko ena od teh komponent, ki je označena z $\mathcal{B}_\infty$, je neskončna, ostale komponente pa so benzenoidi. Tudi $\mathcal{B}_\infty^{\complement}$ je benzenoid. □

Benzenoidom $\mathcal{B}_1, \ldots, \mathcal{B}_d$ v zgornjem izreku pravimo koronalne odprtine.

Definirali smo benzenoidno zaprtje koronoida $\mathcal{K}$ in sicer kot presek vseh benzenoidov, ki vsebujejo $\mathcal{K}$ kot podmnožico. Benzenoidno zaprtje koronoida $\mathcal{K}$ označimo s $\overline{\mathcal{K}}$. Pokazali smo, da je $\overline{\mathcal{K}} = \mathcal{B}_\infty^{\complement}$, kjer je heksagonalni sistem $\mathcal{B}_\infty$ definiran kot v izreku 8. Pokazali smo tudi, da je presek dveh benzenoidov končen heksagonalni sistem, čigar povezane komponente so benzenoidi.

Avtomorfizmi neskončne šestkotniške mreže $\mathcal{H}$ so bijekcije $\phi\colon \mathcal{H} \to \mathcal{H}$, za katere velja, da je $\phi(a) \sim \phi(b)$ natanko tedaj, ko je $a \sim b$. Grupo vseh avtomorfizmov neskončne šestkotniške mreže $\mathcal{H}$ označimo z $\mathrm{Aut}(\mathcal{H})$. Heksagonalna sistema $\mathcal{H}$ in $\mathcal{L}$ sta ekvivalentna, če obstaja $\psi \in \mathrm{Aut}(\mathcal{H})$, da je $\psi(\mathcal{K}) = \mathcal{L}$. Koronoid je degeneriran, če ima koronalno odprtino, ki jo sestavlja en sam šestkotnik. Nedegenerirano zaprtje koronoida $\mathcal{K}$, *NonDeg*$(\mathcal{K})$, je presek vseh nedegeneriranih koronoidov, ki vsebujejo $\mathcal{K}$ kot podmnožico.

Z $G(\mathcal{K})$ označimo 1-skelet koronoida $\mathcal{K}$. Grafu $G(\mathcal{K})$ pravimo koronoidni graf. Velja naslednje:

Izrek 9. *Naj bo $\mathcal{K}$ koronoid in naj bo $C \subseteq G(\mathcal{K})$ cikel dolžine 6. Potem lahko $C \hookrightarrow G(\mathcal{H})$ razširimo do vložitve $G(\mathcal{K}) \hookrightarrow G(\mathcal{H})$ na en sam način.* □

Dokaz je konstruktiven in porodi algoritem za vlaganje poljubnega koronoidnega grafa v neskončno heksagonalno mrežo $\mathcal{H}$. Pokazali smo naslednjo lemo:

Lema 10. *Naj bo $\mathcal{K}$ koronoid in naj bo G graf, za katerega velja $G \cong G(\mathcal{K})$. Potem obstaja (do simetrije natančno) natanko en nedegeneriran koronoid $\mathcal{N}$, da je $G \cong G(\mathcal{N})$. Še več, $\mathcal{N} \cong NonDeg(\mathcal{K})$.* □

Izkaže se tudi, da rob vsake koronalne odprtine nedegeneriranega koronoida vsebuje vsaj dve vozlišči stopnje 2.

V razdelku 5.5 se seznanimo s kodo robnih povezav benzenoida. To je en od popularnejših kombinatoričnih opisov benzenoidov. Kodo robnih povezav dobimo tako, da obhodimo rob benzenoida (začenši pri vozlišču stopnje 3) in štejemo robne povezave med dvema zaporednima vozliščema stopnje 3. Koda lahko vsebuje števila med `1` in `5`. Edina izjema je benzen, ki ima kodo `6` (saj nima nobenega vozlišča stopnje 3). Če na kodi robnih povezav naredimo ciklični pomik ali obrat, dobimo ekvivalentno kodo. Leksikografsko največji med njimi pravimo kanonična koda robnih povezav. Če podamo kodo robnih povezav, potem je benzenoid enolično določen (do simetrije natančno).

Generiranje benzenoidov omogoča programski paket **CaGe**, ki je brezplačen in javno dostopen na naslovu `https://caagt.ugent.be/CaGe/`. V tem delu opišemo zelo preprost algoritem za generiranje benzenoidov, ki uporablja robno kodo povezav.

V razdelku 5.6 obravnavamo konveksne benzenoide. Naša definicija konveksnega benzenoida je:

Definicija 7. *Benzenoid $\mathcal{B}$ je konveksen, če je za vsak par šestkotnikov $a, b \in \mathcal{B}$ cel interval $I_{\mathcal{H}}(a, b)$ vsebovan v $\mathcal{B}$.*

To je običajna (geodezična) definicija konveksnosti v metrični teoriji grafov.

Definicija 8. *Benzenoid $\mathcal{B}$ zadošča pravilu majhnega paralelograma, če za vsak par šestkotnikov $a, b \in \mathcal{B}$, kjer je $d_{\mathcal{H}}(a, b) = 2$, velja*

$$d_{\mathcal{B}}(a, b) = 2 \implies I_{\mathcal{H}}(a, b) \subseteq \mathcal{B}. \tag{8.22}$$

Nato pokažemo:

Trditev 11. *Benzenoid $\mathcal{B}$ je konveksen, če in samo če:*

(a) je $\mathcal{B}$ povezan in

(b) $\mathcal{B}$ zadošča pravilu majhnega paralelograma.

□

Iz kode robnih povezav lahko enostavno razberemo, če je benzenoid konveksen:

Trditev 12. *Benzenoid $\mathcal{B}$ je konveksen natanko tedaj, ko njegova koda robnih povezav ne vsebuje simbola `1`.* □

Neskončen benzenoid je neskončen heksagonalni sistem $\mathcal{K}$, katerega komplement $\mathcal{K}^{\complement}$ je bodisi povezan in neskončen bodisi prazna množica. Primer neskončnega benzenoida je benzenoid, ki ga imenujemo polravnina ($\mathcal{K}_3$ na sliki 5.9). Dokažemo naslednji izrek:

Izrek 13. *Obstaja neštevno mnogo (neekvivalentnih) neskončnih benzenoidov.* □

Konveksne benzenoide je mogoče karakterizirati še na en način:

Trditev 14. *Benzenoid $\mathcal{B}$ (končen ali neskončen) je konveksen, če in samo če ga lahko dobimo kot presek polravnin.* □

Izkaže se, da obstaja (do simetrije natančno) le števno mnogo neskončnih konveksnih benzenoidov in sicer trije sporadični primeri, tri enoparametrične družine in ena dvoparametrična družina. Končne konveksne benzenoide smo klasificirali v 6 družin, ki jim pravimo *fundamentalne družine konveksnih benzenoidov* in sicer so to: linearna veriga, enakostranični trikotnik, enakokraki trapez, romboid, petkotniški benzenoid in šestkotniški benzenoid. Za vse fundamentalne družine smo poiskali enolično parametrizacijo in kode robnih povezav. Konveksnost je mogoče opisati še na en način:

Trditev 15. *Končen benzenoid $\mathcal{B}$ je konveksen, če in samo če je ekvivalenten nekemu pripadniku ene od fundamentalnih družin.* □

Ta karakterizacija je omogočila izpeljavo hitrega algoritma za enumeracijo konveksnih benzenoidov.

Razdelek 5.7 je posvečen Kekuléjevim strukturam benzenoidov. Najprej definiramo popoln sistem poti in predstavimo izrek, ki pove, da so popolni sistemi poti v povratnoenolični korespondenci s Kekuléjevimi strukturami v koronoidih. Nato za vsako fundamentalno družino konveksnih benzenoidov podamo eksplicitno formulo za izračun števila Kekuléjevih struktur. Na koncu predstavimo še Sachsev algoritem, ki v linearnem času preveri, če je benzenoid Kekuléjev.

V šestem poglavju obravnavamo *obliže*. Obliž je (2-povezan) ravninski graf, v katerem lahko nastopajo lica različnih dolžin (ne le šestkotniška, tako kot pri benzenoidih). Notranja vozlišča so stopnje 3, robna pa so lahko stopnje 2 ali 3. Tudi na obližih lahko definiramo kodo robnih povezav, vendar ta koda obliža ne določa nujno enolično. Obliž, v katerem nastopajo le lica dolžine 5 ali 6, se imenuje *fulerenski obliž*. Obliž je *konveksen*, če se v njegovi kodi robnih povezav ne pojavi simbol `1`. Obliže in *naluknjane obliže* obravnavamo v tem delu na enak način, kot smo obravnavali benzenoide in koronoide v razdelku 5.4. Obliž je posplošitev benzenoida, naluknjan obliž pa je posplošitev koronoida. Razvijemo teorijo, ki je v analogiji s teorijo iz razdelka 5.4, le da namesto neskončne heksagonalne mreže $\mathcal{H}$ uporabimo poljuben kubični graf, ki je vložen v ravnino. Na koncu nakažemo, kako lahko rezultate posplošimo na neskončne kubične grafe, in utemeljimo, da mora izbrana vložitev neskončnega kubičnega grafa zadoščati nekaterim dodatnim predpostavkam, sicer naletimo na nekatere težave, ki so topološke narave.

Sedmo poglavje je posvečeno altanom. To je družina kemijskih grafov, ki je bila v zadnjem času deležna precejšnje pozornosti večih teoretičnih kemikov in matematikov. Mallion, Dickens, Zanasi in njegovi soavtorji so jih obravnavali predvsem v povezavi z obročnimi tokovi. Sprva so altane definirali le za benzenoide in nekatere obliže, Gutman pa je definicijo posplošil na poljubne grafe.

Naj bo (G, S) urejeni par, kjer je G graf, S pa ciklično urejena podmnožica njegovih vozlišč, ki ji pravimo *periferni koren*. *Operacija altan* par (G, S) preslika v nov par (G_1, S_1),

ki ga dobimo na sledeč način: Predpostavimo, da je $V(G) = \{0, 1, \ldots, n-1\}$ in da ima množica S k elementov. Naj bo

$$S_0 = \{n, n+1, \ldots, n+k-1\} \quad \text{in} \quad S_1 = \{n+k, n+k+1, \ldots, n+2k-1\}. \tag{8.23}$$

Graf G_1 dobimo tako, da vozliščem grafa G dodamo še $S_0 \cup S_1$. Nato grafu G_1 dodamo še nove povezave. Vozlišča $S_0 \cup S_1$ povežemo v cikel C dolžine $2k$:

$$(n, n+k, n+1, n+k+1, \ldots, n+k-1, n+2k-1, n). \tag{8.24}$$

Poleg teh dodamo še povezave $(s_i, n+i)$, $0 \leq i < k$, kjer je s_i i-to vozlišče množice S. Vozlišča na ciklu C, ki so stopnje 2, tvorijo nov periferni koren S_1. Altan označimo z $A(G, S)$.

Altan dvodelnega grafa ni nujno dvodelni graf. Naslednji izrek pove, pod katerimi pogoji je $A(G, S)$ vendarle dvodelni graf:

Izrek 16. *Naj bo (G, S) par, kjer je G graf, množica S pa periferni koren. Altan $A(G, S)$ je dvodelen natanko tedaj, ko velja:*

(a) G je dvodelen graf in

(b) elementi množice S pripadajo isti podmnožici dvodelnega razbitja vozlišč grafa G.

□

Z $A^n(G, S)$ označimo *n-ti altan* od (G, S), tj.

$$\underbrace{A(A(\cdots A}_{n}(G, S)\cdots)). \tag{8.25}$$

Pravimo mu tudi *iterirani altan*. Izkaže se, da je $A^n(G, S)$ dvodelen graf, če in samo če (G, S) zadošča pogojem iz izreka 16. Definiramo še altane benzenoidov, obližev in fulerenskih obližev. Med drugim pokažemo naslednji izrek:

Izrek 17. *Naj bo $\mathcal{B}$ benzenoid. Altan $A^n(\mathcal{B})$ je nanocevka (s kapico na enem od koncev), če in samo če je $\mathcal{B}$ konveksen.* □

Z uporabo Gutmanovih rezultatov smo pokazali, da je število Kekuléjevih struktur altana $A^n(\mathcal{B})$ enako $2^n K(\mathcal{B})$. Ta formula velja tudi za obliže.

V razdelku 7.5 uvedemo posplošene altane in iterirane posplošene altane. Posplošeni altan ima lahko več perifernih korenov hkrati, vendar lahko vsako vozlišče stopnje 2 pripada kvečjemu enemu od perifernih korenov.

Posplošene altane študiramo še na koronoidih in naluknjanih obližih. Pravi posplošeni altan koronoida ne bo nikoli spet koronoid. Po drugi strani pa velja:

Izrek 18. *Naj bo $\mathcal{P}$ naluknjan obliž. Njegov posplošeni altan $A^n(\mathcal{P})$ je tudi naluknjan obliž. Še več, če je $G(\mathcal{P})$ 2-povezan, je tudi $G(A^n(\mathcal{P}))$ 2-povezan.* □

Na koncu pokažemo še, da lahko enostavno izračunamo število Kekuléjevih struktur posplošenega altana, če poznamo število Kekuléjevih struktur osnovnega grafa. Enostavno lahko izračunamo tudi Paulingove rede vezi posplošenega altana, če poznamo Paulingove rede vezi osnovnega grafa.