Available online at www.sciencedirect.com

SciVerse ScienceDirect

Indagationes Mathematicae 24 (2013) 115–141

indagationes mathematicae

www.elsevier.com/locate/indag

Graph theory for alternating hydrocarbons with attached ports

Wim H. Hesselink*

Department of Computing Science, University of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands

Received 28 November 2011; accepted 4 July 2012

Communicated by J.W. Klop

Abstract

Properties of molecules of certain hydrocarbons give rise to difficult questions in graph theory. This paper is primarily devoted to the graph theory, but the physico-chemical motivation, which is somewhat speculative, is also presented.

Molecules of unsaturated hydrocarbons exhibit alternating paths with single and double bonds. Such alternating paths have been postulated to be electrically conductive. When used to conduct, however, such a path is toggled: the single and double bonds are interchanged. This can imply that other alternating paths appear or disappear. In this way, switching behavior arises. This suggests a possibility of molecular computing.

Molecules are represented by graphs where certain nodes, called ports, are chosen as connectors to the outside world. At these ports other chemical groups can be attached to observe and influence the behavior. A choice of single and double bonds in the molecule is represented by an almost-perfect matching in the graph—almost, in the sense that the ports and only the ports are allowed to have no double bond attached to them. The corresponding graph theory is a qualitative idealization of the molecules.

It turns out that the switching behavior is completely determined by sets of ports, called cells. The paper is devoted to the question which cells are Kekulé cells, i.e., correspond to almost-perfect matchings in graphs. We prove that every Kekulé cell is what is known as a linkable Δ-matroid (it appears that this was known).

An anonymous referee showed us the existence of a linkable Δ-matroid with 7 ports that is not a Kekulé cell. The argument is presented. We classify the linkable cells with ≤ 5 ports and show that they all are

* Tel.: +31 505264649.
E-mail address: w.h.hesselink@rug.nl.

doi:10.1016/j.indag.2012.07.002

Kekulé cells. We also classify the linkable cells with 6 ports. There are 214 classes; 210 classes contain Kekulé cells; only 4 classes are undecided.

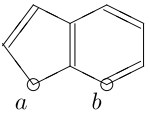
Keywords: Hydrocarbon; Undirected graph; Matching; Delta-matroid; Matroid; Group action; Classification

1. Introduction

In this paper, we propose and investigate discrete Kekulé Theory, which is a graph theory for alternating hydrocarbons with attached ports. It is a theory within discrete mathematics and graph theory, part of the theory of Δ-matroids [1,2] and close to matching theory [10]. It is inspired by chemistry [12] and first studied in [7]. The aim is to investigate and systematize qualitative switching behavior in polycyclic polyunsaturated hydrocarbons, with molecular computing as an ultimate goal.

Consider the indene molecule sketched below (indene is a constituent of indigo). In such hydrocarbons, the bonds between the carbon atoms can be single bonds and double bonds, but *ideally* every atom has precisely one double bond. In the indene molecule, however, this is not possible. By attaching other groups to the molecule at specific atoms, which are called *ports*, the rule can be weakened in the sense that at these ports the number of double bonds is *at most* one.

In the diagram, we have chosen ports at a and b, and indicated them by the little circles. Now there are three solutions, two states with no double bond at a (one of which is sketched), and one with no double bond at b.

Indene: 

The reader should realize that the single and double bonds mentioned form an idealization. When, for fixed port behavior, there are several solutions to the combinatorial problem of placing the single and double bonds, the solutions have no independent physical meaning, and are called *resonance structures*. The real quantum-mechanical state is a superposition of these resonance structures. Even worse, the graph itself is an idealization: some bonds in the graph are stronger than others, and atoms that are not connected in the graph do influence each other. In discrete Kekulé theory, however, we form a qualitative abstraction that only to some extent corresponds to the physical or chemical reality.

In a given molecule, it is possible to give every carbon atom precisely one double bond if and only if its graph has a so-called *perfect matching*. The number of perfect matchings is an indicator of the stability of the molecule [10, Section 8.7]. Our subject matter is thus an extension of classical matching theory [10]. Our matchings need not be perfect, but imperfections are only permitted at the ports.

1.1. Switching behavior

It has been postulated [13] that there is an electrically open channel through the molecule between two ports if and only if there is an alternating path between the ports, i.e., a path that alternately consists of single and double bonds (this is slightly inaccurate, see Section 2.2 for the correct definition). By sending a so-called *soliton* over the alternating path, the single and double bonds along the path are toggled [5]. In the case of indene drawn above, there is such a

path between a and b: it goes from a to the northeast and then around the righthand hexagon. By toggling this path, one gets a state with no double bond at b.

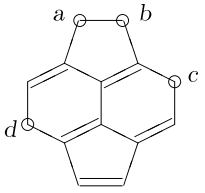
The reader may object that the alternating path is a mathematical concept without physical meaning. Yet, if there is an alternating path between two ports in one of the resonance structures, each of the other physically indistinguishable resonance structures also has an alternating path between these ports, and toggling any of these gives the same class of new resonance structures. This follows from the Theorems 1 and 2 in Section 2. Therefore, even though a specific alternating path has no meaning, it does make physical sense to state that a pair of ports is connected by some alternating path.

A *channel* is defined to be a pair of ports. We say that the channel is *open* if there is some alternating path between its ports. The above postulate says that the channel is electrically conductive if and only if it is open.

The switching behavior alluded to is the fact that by toggling an open channel, other channels can be opened or closed. Following [7], we show below that the switching behavior is completely determined by the so-called Kekulé cell of the graph/molecule. A pair (G, P) consisting of a graph G and a set P of ports is called a ported graph. Given a ported graph (G, P), every resonance structure induces a *port assignment*, which is the set that consists of the ports with a double bond. The Kekulé cell $Kp(G, P)$ is then the set of the port assignments of the resonance structures. As it does not depend on a particular choice of resonance structure, the Kekulé cell can have physical meaning. The first aim of this paper is to prove these results. This is done in Section 2.

Example. Let us attach four ports a, b, c, d to a pyracylene molecule at the positions indicated in the graph below. The resonance structure shown is the only one in which the ports have no double bonds, i.e., that corresponds to the empty port assignment, which is denoted by $\emptyset$. It is easy to see that the graph as drawn has alternating paths between a and b, and between b and d. There is also a long alternating path between b and c (use the lower pentagon). There is no alternating path between a and c. The switching behavior is illustrated by the fact that the channel between a and c is opened by toggling the open channel between b and d.

Pyracylene:



The Kekulé cell $Kp(G, P)$ for this graph consists of the port assignments $\emptyset, ab, ad, bc, bd, cd, abcd$. Here, a string like ab represents the set $\{a, b\}$, the port assignment obtained by adding a double bond between a and b. The port assignment $abcd = \{a, b, c, d\}$ can be obtained by toggling independent alternating paths on the channels ab and cd. According to Theorem 2, the observation that there is no alternating path between a and c corresponds to the fact that ac is not in the Kekulé cell.

It thus turns out that the switching behavior of the graph or molecule only depends on the cell. If we want an alternating hydrocarbon with a certain switching behavior, we can therefore split the search into the following three parts.

1. Search for cells K with the required switching behavior.
2. Search for ported graphs (G, P) with $K = Kp(G, P)$.
3. Search for suitably alternating hydrocarbon molecules that have such a ported graph.

Treatment of the first point is premature, because we do not yet know what kind of switching behavior will be required, or what will be available. With respect to the third point, when we have a ported graph (G, P) with $K = Kp(G, P)$, the papers [7,12] give rather far-reaching methods to change the graph without changing the cell. Therefore, finding any graph with cell K is a major step towards finding a suitable molecule. The question how to find or synthesize such a molecule, however, is beyond the scope of this paper.

We therefore focus on the second point: given a cell K over P, is there a graph G with $K = Kp(G, P)$? In other words, is K a Kekulé cell?

1.2. Positioning

There is a vast literature in chemistry on alternating hydrocarbons, but for our current questions this is largely irrelevant because of the qualitative abstraction we make here, and the fact that, for now, we ignore all questions of realizability of the graphs and stability of the molecules.

In comparison with classical matching theory [10], we have turned the questions inside out: the matchings are allowed to be nonperfect at specific locations, viz., at the ports. Just as our earlier paper [7], the present paper is inspired by the thesis π-Logic [12] of Marleen van der Veen.

An anonymous referee pointed out strong ties and some overlap with the theory of Δ-matroids [1,2]. We had indeed unknowingly reinvented several concepts from this theory. In most cases, we present these concepts now under the names used there.

1.3. Overview

In Section 2, we develop the graph theory needed, we define Kekulé states, alternating paths, and port restrictions, and establish their relationships. Section 3 contains the structure theory of Kekulé cells and shows the relationship with the theory of Δ-matroids. The theory has four important operations: port renaming, translation, contraction, and fusion. We present linkability of cells which is a necessary (but not sufficient) condition for a cell to be a Kekulé cell.

In Section 4, we describe algorithms to determine the Kekulé cell of a given graph, and to search for a graph that has a given cell as its Kekulé cell. In Section 5, we develop an algorithm to classify the linkable cells of a given order (number of ports involved). This concerns the action of the affine group generated by the port renamings and the translations. The classification shows that all linkable cells of order ≤ 5 are Kekulé cells. There are 214 classes of linkable cells of order 6, of which 210 are Kekulé cells. Conclusions are drawn in Section 6. Here, we raise the question of decidability: it is not clear whether the class of the Kekulé cells is decidable.

For most of the combinatorial problems in this paper, we have developed computer programs in the language C to search for solutions. This may be superfluous for the simplest cases, as indene above, but, even for a moderate case as pyracylene above, the program is more reliable than most humans, and for more complicated cases the program is indispensable. In most of the examples below, computer programs were used to obtain the solutions. The programs themselves, however, are based on the theory presented in this paper. The programs with some documentation and data files are available on our website [6] for inspection or use.

In comparison with [7], we have straightened the formalism to make the graphs smaller and to simplify some of the constructions. Section 2 and the first part of Section 3 are adaptations of material from [7] to the present formalism. Almost all of the remainder of the paper is new. We focus on linkable cells (in [7], such cells were called “connected by channels”) and, in particular

on the question whether every linkable cell is a Kekulé cell. For application to carbon chemistry, it would be interesting to see whether all Kekulé cells of Section 5.3 can be realized in stable molecules.

2. Graphs and discrete Kekulé Theory

For the sake of flexibility, we use definitions that are somewhat more liberal than those of [7]. As before, our graphs are undirected and without self-loops. The main difference with [7] is that the ports now are marked vertices with an arbitrary number of incident edges, whereas in [7] a vertex is a port if and only if it has one incident edge.

Notations. If V is a finite set, we write $\#V$ for its number of elements. We write $\mathbb{P}(V)$ to denote the power set of V, i.e., the set of the subsets of V. For a function $f : V \to W$, we write f^* for the direct image function $f^* : \mathbb{P}(V) \to \mathbb{P}(W)$ given by $f^*(A) = \{f(v) \mid v \in A\}$. If d is a natural number, we write $V^{(d)}$ for the set of subsets of V with d elements. In particular, $V^{(2)}$ is the set of all subsets of V with two elements (doubletons). We use $\square$ to indicate the end of a proof, remark, or example.

2.1. Graphs and Kekulé states

A *graph* is defined to be a pair $G = (V, E)$ where V is a finite set of vertices and E is a set of edges, i.e. doubletons of vertices, and therefore a subset of $V^{(2)}$. The elements of V and E are called the *vertices* and *edges* of G. Vertices that do not occur in edges (elements of $V \setminus \bigcup E$) are the *isolated* vertices of G.

We define a *ported graph* to be a pair (G, P) where G is a graph and P is a finite set. The elements of P are called *ports*. Ports that are not vertices are not used (one might regard them as additional isolated vertices). They are allowed for flexibility. In this way, we do not need to change the set of ports when considering a subgraph. Vertices that are not ports are called *internal*. Ported graphs are our formalization of molecules with ports.

We use the set of double bonds to characterize a resonance state of the molecule represented by the ported graph; this set is called a Kekulé state of the ported graph.

Definition 1. A *Kekulé state* W of a ported graph (G, P) with $G = (V, E)$ is defined to be a set $W \subseteq E$ of pairwise disjoint edges such that $V \subseteq P \cup (\bigcup W)$.

Here $\bigcup W$ is the union of all edges in W, i.e., the set of all vertices that occur in some edge of W.

A Kekulé state W represents the set of edges that carry double bonds in the molecule. The edges in W are required to be pairwise disjoint because every atom has at most one double bond. The inclusion $V \subseteq P \cup (\bigcup W)$ means that every atom without some double bond is a port.

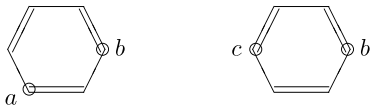
2.2. Alternating paths and curves

In the presence of ports with several incident edges, the definition of alternating paths in Section 1 is inaccurate. We need to sharpen it in the following way.

Definition 2. Given a ported graph and a Kekulé state W, a path between two ports is *alternating* if its steps alternate between single and double bonds and if the end points hold no double bonds outside of the path.

Note that intermediate ports on the path are allowed, but any intermediate port on the path needs one double bond on the path.

Example. The graph of benzene with two ports a and b with distance 2 (at a so-called meta position) has the same two Kekulé states as benzene without ports. One of them is depicted on the lefthand diagram below. Despite appearances, there is no alternating path between a and b, because both candidate paths have an end point with a double bond outside the path. The same holds for the other Kekulé state of this graph.

Benzene: 

In the righthand diagram, the ports b and c have distance 3 (a para position) and there is one alternating path between them (over the north). Toggling the path gives a Kekulé state that does not meet the ports.

Paths between ports, possibly alternating, are generalized to (alternating) curves, defined as follows.

For a graph $G = (V, E)$ and a vertex $v \in V$, we define the *degree* to be the number of incident edges: $deg(v, G) = \#\{e \in E \mid v \in e\}$. More generally, for any set of edges W, we define $deg(v, W) = \#\{e \in W \mid v \in e\}$.

Definition 3. A *curve* C in ported graph (G, P) with $G = (V, E)$ is a set of edges C of G with $deg(v, C) \leq 2$ for all vertices $v \in V$, and with $deg(v, C) \neq 1$ unless $v \in P$. If C is a curve, its *set of end points ends*(C) is defined as the set of ports p with $deg(p, C) = 1$.

Let $W \subseteq E$. A curve C is *W-alternating* if C is a curve in G, and for all $v \in V$,

(a1) $deg(v, C) = 1 \Rightarrow deg(v, W \setminus C) = 0$,
(a2) $deg(v, C) = 2 \Rightarrow deg(v, W \cap C) = 1$.

In words, condition (a1) says that an end point of the curve has no double bonds outside the curve, and condition (a2) says that an internal point of the curve has precisely one double bond in the curve.

Two curves are called disjoint when they have no common vertices. The union of two disjoint curves is again a curve. Conversely, every curve is a union of disjoint connected curves, and a connected curve is a cycle or a path between two ports.

We use the operator $\oplus$ for symmetric set differences: $A \oplus B = (A \setminus B) \cup (B \setminus A)$. This operator is commutative and associative, and satisfies $\emptyset \oplus A = A$ and $A \oplus A = \emptyset$. It makes the sets $\mathbb{P}(V)$ and $\mathbb{P}(V^{(2)})$ additive groups, and even vector spaces over the field with 2 elements.

The relevance of alternating curves and the operator $\oplus$ is shown in the following result, which was proved in a more restricted setting in [7].

Theorem 1. (a) *Let W be a Kekulé state of a ported graph (G, P) and let C be a W-alternating curve. Then $W \oplus C$ is a Kekulé state of (G, P).*
(b) *Let W and W' be Kekulé states of (G, P). Then $C = W \oplus W'$ is a W-alternating curve in (G, P).*

Proof. (a) It suffices to prove that $deg(v, W \oplus C) \leq 1$ for all $v \in V$, with equality unless $v \in P$. We consider three cases: first $deg(v, C) = 2$. Then $deg(v, C \cap W) = 1$ and hence

$deg(v, C \setminus W) = 1$, and $deg(v, W \setminus C) = 0$ because W is a Kekulé state. This implies $deg(v, W \oplus C) = 1$. Second, $deg(v, C) = 1$. Then $v \in P$ and $deg(v, W \setminus C) = 0$. Therefore $deg(v, W \oplus C) = deg(v, C \setminus W) \leq deg(v, C) = 1$. Finally, if $deg(v, C) = 0$ then $deg(v, W \oplus C) = deg(v, W)$.

(b) For every $v \in V$, because W and W' are Kekulé states, we have

$$deg(v, C) = deg(v, W \setminus W') + deg(v, W' \setminus W) \leq 2.$$

Moreover, equality clearly implies that $deg(v, W \cap C) = 1$. For $v \in V \setminus P$, we have $deg(v, W) = deg(v, W') = 1$ and hence $deg(v, C) \neq 1$. This proves that C is a curve. Finally, assume that $deg(v, W \setminus C) \neq 0$. Then $deg(v, W \cap W') \neq 0$ and hence $deg(v, C) = 0$ because $deg(v, W) \leq 1$ and $deg(v, W') \leq 1$. □

2.3. Port restriction and Kekulé cells

Kekulé states are supposed to be physically not observable. The only thing observable of a Kekulé state is its port restriction, the set of ports with a double bond. In chemistry, the (indistinguishable) Kekulé states with the same port restriction are called *resonance structures*.

Set-theoretically, the *port restriction* of a Kekulé state W is defined to be $W|P = P \cap (\bigcup W)$. This is the set of ports p with $deg(p, W) = 1$.

Lemma 1. *Let W be a Kekulé state of a ported graph (G, P) and let C be a W-alternating curve. Then $(W \oplus C)|P = (W|P) \oplus ends(C)$.*

Proof. Consider $p \in P$. First assume $p \in ends(C)$, i.e., $deg(p, C) = 1$. Then $deg(p, W \setminus C) = 0$ by (a1), and hence $deg(p, W \oplus C) = deg(p, C \setminus W) = 1 - deg(p, W)$. This implies $p \in (W \oplus C)|P$ if and only if $p \notin W|P$, as required.

Otherwise, we have $deg(p, C) = 0$ or 2. If $deg(p, C) = 0$, taking the symmetric difference with C does not touch p, so that $p \in (W \oplus C)|P$ if and only if $p \in W|P$, as required. If $deg(p, C) = 2$, then $deg(p, W \cap C) = 1$ by (a2), and hence $deg(p, W \setminus C) = 0$ and $deg(p, C \setminus W) = 1$. It follows that $deg(p, W \oplus C) = 1 = deg(p, W)$, so that $p \in (W \oplus C)|P$ and $p \in W|P$. □

Definition 4. A *port assignment* is a set of ports. For a ported graph (G, P), the *Kekulé cell* $Kp(G, P)$ is the set of port assignments k for which the ported graph G has a Kekulé state W with $k = W|P$.

Because the definition of $Kp(G, P)$ only depends on the set of all Kekulé states and not on a specific one, we can attach physical meaning to it.

In the example of Section 1.1, the graph of pyracylene has 4 ports and 10 internal nodes. Its Kekulé cell is $\{\emptyset, ab, ad, bc, bd, cd, abcd\}$, where, for brevity, we use strings to represent port assignments: $abcd$ stands for the set $\{a, b, c, d\}$.

We now transfer Theorem 1 to the port assignments.

Theorem 2. *Let (G, P) be a ported graph and $k \in Kp(G, P)$, say $k = W|P$ for Kekulé state W of (G, P). Let g be a port assignment. There is a W-alternating curve C with $g = ends(C)$ if and only if $k \oplus g \in Kp(G, P)$.*

Proof. Assume that $g = ends(C)$ for some W-alternating curve C. Then $W \oplus C$ is a Kekulé state by Theorem 1(a). Lemma 1 implies that $(W \oplus C)|P = k \oplus g$. This proves that $k \oplus g \in Kp(G, P)$.

Conversely, let $k \oplus g \in Kp(G, P)$. Then $k \oplus g = (W'|P)$ for some Kekulé state W'. By Theorem 1(b), there is a W-alternating curve C with $C = W \oplus W'$. Lemma 1 yields $k \oplus g = (W'|P) = (W \oplus C)|P = k \oplus ends(C)$. This implies that $g = ends(C)$. □

2.4. Switching behavior of cells

Recall that, in [13], the molecule corresponding to the graph G in Kekulé state W is postulated to be electrically conductive between two ports p and q if and only if there is an alternating path between them. There is an alternating path between p and q if and only if there is an alternating curve C with $\{p, q\} = ends(C)$. Theorem 2 therefore implies that the molecule is conductive if and only if $k = W|P$ satisfies $k \oplus \{p, q\} \in Kp(G, P)$.

We therefore define a *channel* to be a doubleton of ports, i.e., an element of $P^{(2)}$. A channel c is defined to be *open* in a port assignment $k \in Kp(G, P)$ if and only if $k \oplus c \in Kp(G, P)$.

Sending a soliton over an alternating path C corresponds to replacing the Kekulé state W by $W \oplus C$. When it is a path for channel c, this corresponds to replacing the port assignment k by $k \oplus c$. Whether the channel is open or closed, and how the port assignment is modified by using the channel, only depends on the current port assignment k and the set $Kp(G, P)$. We can therefore attach physical meaning to openness of channels and toggling of channels. The switching behavior of the molecule is completely determined by $Kp(G, P)$, and independent of other aspects of the graph. We can therefore abstract from the graphs.

We define a *cell* K on a finite set P to be a subset of $\mathbb{P}(P)$. Given a cell K, we define a channel c to be *open* in state $k \in K$ iff $k \oplus c \in K$. In this way, arbitrary cells have switching behavior. The *order* of cell K is the number of ports involved, i.e. $\#(\bigcup K)$.

We define a cell K on P to be a *Kekulé cell* if and only if there is a graph G such that $K = Kp(G, P)$. The set of all cells on P is $\mathbb{P}(\mathbb{P}(P))$. We write $KC(P)$ to denote the set of Kekulé cells on P. Clearly, $KC(P) \subseteq \mathbb{P}(\mathbb{P}(P))$.

3. Structure of Kekulé cells

In Section 3.1, we give some easy results about the class of all Kekulé cells on a given set of ports. Section 3.2 presents a dichotomy: every Kekulé cell is even or odd. We define a distance between port assignments. Section 3.3 exhibits some symmetric Kekulé cells. In Section 3.4, we introduce the operations on cells of port renaming and translation, both of which preserve Kekulé cells.

Section 3.5 introduces linkability of cells and proves that every Kekulé cell is linkable. At this point, we enter the field of Δ-matroids [1]. In Section 3.6, we introduce the operations on cells of contraction and fusion, which again preserve Kekulé cells. In Section 3.7, we show that the set of all Kekulé cells can be inductively defined from single edges, by means of fusion and contraction. In Section 3.8, it is proved that not every linkable cell is a Kekulé cell. In Section 3.9, we investigate regularity properties of the fusion operator.

3.1. The class of Kekulé cells

Recall that $KC(P)$ is the set of all Kekulé cells on P. If $K = Kp(G, P)$ for some graph $G = (V, E)$ with $P \not\subseteq V$, then we can add the nonvertex ports of $P \setminus V$ as isolated vertices to G and form a graph $G' = (V \cup P, E)$ with $K = Kp(G', P)$ such that $P \subseteq V'$.

If $P \subseteq P'$, every cell K on P is a cell on P'. Whether the cell is a Kekulé cell, does not depend on the port set considered.

Proposition 1. *Assume $P \subseteq P'$. Then $KC(P) = \{K \in KC(P') \mid \bigcup K \subseteq P\}$.*

Proof. Let $K \in KC(P)$. Clearly, $\bigcup K \subseteq P$. On the other hand, there is a graph $G = (V, E)$ with $K = Kp(G, P)$. We may assume that $V \cap P' \subseteq P$. Then $K = Kp(G, P') \in KC(P')$.

Conversely, let $K \in KC(P')$ with $\bigcup K \subseteq P$. There is a graph G with $K = Kp(G, P')$, say $G = (V, E)$. We remove all ports outside P from the graph by defining the vertex set $V_1 = V \setminus (P' \setminus P)$. Every Kekulé state of (G, P') is contained in V_1 because $Kp(G, P') = K \subseteq \mathbb{P}(P)$. Therefore $K = Kp(G_1, P) \in KC(P)$ for the graph $G = (V_1, E_1)$ given by $E_1 = E \cap V_1^{(2)}$. □

3.2. Parity and distance

Let $K = Kp(G, P)$ for ported graph (G, P) with $G = (V, E)$. Let U be the set of internal nodes of G. For any Kekulé state W of G, we have $U \subseteq \bigcup W$ and hence

$$\#(W|P) = \#\left(\left(\bigcup W\right) \setminus U\right) = \#\left(\bigcup W\right) - \#U = 2 \cdot \#W - \#U.$$

This implies the following lemma.

Lemma 2. *Let U be the set of internal nodes of a ported graph (G, P). If U has an even (odd) number of elements, every $k \in Kp(G, P)$ has an even (odd) number of elements.*

We therefore define a cell K to be *even* (*odd*) if every $k \in K$ has $\#k$ even (odd).

When one wants to visualize a (Kekulé) cell, it is natural to define the *distance* d between its port assignments by counting the number of ports in which they differ, i.e., $d(h, k) = \#(h \oplus k)$. Lemma 2 implies that, in a Kekulé cell, all distances between its port assignments are even.

This may suggest to regard the cell K as an undirected graph in which the edges are pairs $\{g, h\}$ with distance 2. If one does this, Theorem 3 in Section 3.5 shows that every Kekulé cell is connected in a very strong sense.

3.3. Kekulé cells for some symmetric graphs

We construct a few Kekulé cells to get a feeling what kind of Kekulé cells can occur. The cells in this section are special in that they are symmetric under all permutations of the ports.

Let P be a finite set of ports. Let $Even(P) = \bigcup_{j \geq 0} P^{(2j)}$ be the set of all subsets of P with an even number of elements. This is the largest even cell with ports in P. This cell turns out to be a Kekulé cell and correspond to the complete graph on the set P. Indeed, let G be the complete graph on P, i.e. $G = (P, P^{(2)})$. Then every set W of pairwise disjoint edges is a Kekulé state because every vertex is a port. It follows that $Kp(G, P) = Even(P)$.

Next, let U be a finite set of internal vertices with $\#U = d$. Let G_0 be the bipartite graph obtained by linking all elements of P to all elements of U, so that $G_0 = (V, E_0)$ where $V = P \cup U$ and $E_0 = \{\{p, u\} \mid p \in P, u \in U\}$. A set W of pairwise disjoint edges of G_0 is a Kekulé state of G_0 iff $U \subseteq \bigcup W$. It follows that $Kp(G_0, P) = P^{(d)}$. Therefore, for any $d \geq 0$, the cell $P^{(d)}$ is a Kekulé cell. Note that $P^{(0)} = \{\emptyset\}$ and that $P^{(d)} = \emptyset$ for $d > \#P$.

Let us now add some disjoint internal edges to G_0, i.e., choose a set E' of $e \leq \frac{1}{2}d$ disjoint doubletons in U, and replace G_0 by $G_1 = (V, E_1)$ where $E_1 = E_0 \cup E'$. As the elements of E may appear in Kekulé states of G_1, we get

$$Kp(G_1, P) = \bigcup_{i=0}^{e} P^{(d-2i)}.$$

Therefore these cells are also Kekulé. In particular, the cells $Even(P)$ and $Odd(P) = \bigcup_{j\geq 0} P^{(2j+1)}$ are Kekulé. This also shows that, in general, many different graphs may induce the same Kekulé cell.

3.4. Operations on Kekulé cells

There are two kinds of operations on (Kekulé) cells that preserve all information and functionality.

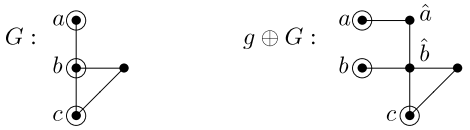
Renaming of ports. A Kekulé cell remains a Kekulé cell when its ports are renamed. More formally, let $f : P \to P'$ be a bijective function (i.e., one-to-one and onto). For any cell K on P, we define the *f-renaming* to be the cell $f^{**}(K)$ on P'. This operation is invertible: if $g \circ f = id$, then $g^{**}(f^{**}(K)) = K$. It is clear that renaming of ports is innocent: $f^{**}(K)$ is a Kekulé cell on P' if and only if K is a Kekulé cell on P, because one can transfer the renaming to the graphs.

Translation. For $g \in \mathbb{P}(P)$ and a cell K on P, we define the *translated* cell $g \oplus K = \{g \oplus k \mid k \in K\}$. This operation is its own inverse because $g \oplus (g \oplus K) = K$. Translation of cells distributes over unions and intersections of cells: $g \oplus (K \cup K') = (g \oplus K) \cup (g' \oplus K)$ and $g \oplus (K \cap K') = (g \oplus K) \cap (g' \oplus K)$. Translation is an action on cells in the sense that $\emptyset \oplus K = K$ and $g \oplus (g' \oplus K) = (g \oplus g') \oplus K$. Translation is called "twisting" in [2]. We prefer "translation" because it is translation of subsets of the vector space $\mathbb{P}(P)$ over the field with two elements with $\oplus$ as vector addition.

The distance is translation invariant because $d(g \oplus h, g \oplus k) = \#(g \oplus h \oplus g \oplus k) = \#(h \oplus k) = d(h, k)$.

If K is a Kekulé cell on P, then $g \oplus K$ is also a Kekulé cell on P. This is shown as follows. For a ported graph (G, P) with $G = (V, E)$ and a set $g \in \mathbb{P}(P)$, we define the graph $g \oplus G$ by the following construction, in which, roughly speaking, an extra edge is inserted between every port $p \in g$ and the graph. We first form the graph $\hat{G}$, isomorphic to G in which every port $p \in g$ is renamed to an internal vertex $\hat{p}$. Then graph $\hat{G}$ is extended to a graph $g \oplus G$ by adding edges $\{p, \hat{p}\}$ for all $p \in g$.

Example. The lefthand graph G below with three ports and one internal node has the Kekulé cell $K = \{b, c, abc\}$ because, in every Kekulé state of the graph, the one internal node needs to be covered by an edge to b or c. If the edge to c is chosen, the edge ab can be added to the Kekulé state.



The righthand graph is $g \oplus G$ where $g = ab$. The ports a and b have been replaced by $\hat{a}$ and $\hat{b}$, and new edges $\{a, \hat{a}\}$ and $\{b, \hat{b}\}$ have been added. It is easy to verify that its Kekulé cell equals $g \oplus K = \{a, abc, c\}$.

Proposition 2. *Let (G, P) be a ported graph and let $g \in \mathbb{P}(P)$. Then $Kp(g \oplus G, P) = g \oplus Kp(G, P)$.*

Proof. For any set of edges W of G, define $\hat{W}$ to consist of the edges of the graph $g \oplus G$ obtained from W by the renaming $p \mapsto \hat{p}$ for all $p \in g$.

Let W be a Kekulé state of G with $k = (W|P)$. For every port $p \in g$, we have $p \in k$ if and only if $\hat{p} \in \bigcup \hat{W}$. We therefore define $W' = \hat{W} \cup \{\{p, \hat{p}\} \mid p \in g \setminus k\}$. The set W' is a Kekulé state of $g \oplus G$, and $(W'|P) = (g \setminus k) \cup (k \setminus g) = g \oplus k$. This gives us a mapping from the Kekulé states of G to the Kekulé states of $g \oplus G$, which reduces to $(g\oplus)$ on the port restrictions.

Conversely, if W' is a Kekulé state of $g \oplus G$ with $k' = (W'|P)$, then W' contains all edges $\{p, \hat{p}\}$ with $p \in g \cap k'$. It follows that there is a unique set W of edges of G, such that $\hat{W} = W' \setminus \{\{p, \hat{p}\} \mid p \in g \cap k'\}$. Because W' consists of pairwise disjoint edges, so do $\hat{W}$ and W. W is a Kekulé state of G because every internal node of G is an internal node of $g \oplus G$ that is not of the form $\hat{p}$ with $p \in g$. This shows that the mapping $W \mapsto W'$ of the previous paragraph is surjective. Therefore $Kp(g \oplus G) = g \oplus Kp(G)$. □

In the technical sense of group actions, the operator $\oplus$ is not an action on graphs: $g \oplus (g' \oplus G) \neq (g \oplus g') \oplus G$ when $g \cap g' \neq \emptyset$.

3.5. Linkability of cells

The binary operator $\oplus$ of symmetric set difference on $\mathbb{P}(P)$ is commutative, associative and has a neutral element $\emptyset$. We can therefore extend it to an operator $\bigoplus$ for repeated operations. This operator is defined on sets $K \in \mathbb{P}(\mathbb{P}(P))$ by $\bigoplus \emptyset = \emptyset$ and $\bigoplus K = k \oplus \bigoplus(K \setminus \{k\})$ for every $k \in K$.

Definition 5. Let K be a cell on P, and let $h, k \in K$. Following [2], we define a *linking* between h and k in K to be a partition D of $h \oplus k$ in channels such that $h \oplus \bigoplus D' \in K$ for every subset D' of D. Cell K is defined to be *linkable* if every pair $h, k \in K$ has a linking in K.

If $d(h, k) = 2$, then $c = h \oplus k$ is a channel and the singleton $\{c\}$ is a linking between h and k.

If D is a linking between h and k, then $d(h, k) = 2 \cdot \#D$ because D is a partition in channels. We have $\bigoplus D' = \bigcup D'$ for every subset D' of D because D is a partition. In particular, $\bigoplus D = h \oplus k$, so that $h \oplus \bigoplus D = k$. It follows that D is also a linking between k and h. Every enumeration of D induces a different path of channels from h to k within the cell. Linkability is thus a much stronger condition on cells than connectedness on graphs.

Example. In the Kekulé cell $\{\emptyset, ab, ad, bc, bd, cd, abcd\}$ of pyracylene, there are three pairs of port assignments with distance 4, viz. $(\emptyset, abcd)$, (ab, cd), (bc, ad). The other distances are 2. The cell therefore looks like an octahedron with bd in the center. The cell is linkable. For example, the set $D = \{ab, cd\}$ is a linking between $\emptyset$ and $abcd$. D is also a linking between ab and cd in K.

Theorem 3. *Every Kekulé cell is linkable.*

Proof. Let K be a Kekulé cell, say $K = Kp(G, P)$ for some ported graph (G, P). Let $h, k \in K$. We can take Kekulé states W and W' so that $h = (W|P)$ and $k = (W'|P)$. Let $C = W \oplus W'$ be the W-alternating curve of Theorem 1(b). Curve C is the disjoint union of simple paths between ports, and cycles. Assume that $C_0, \ldots, C_{d-1}$ are the components of C that are the simple paths. Then C_i is a W-alternating path between ports for every $i < d$.

For each index i, let c_i be the channel that consists of the endpoints of path C_i. We take $D = \{c_i \mid i < d\}$. Since the paths C_i are components of C, they are disjoint. Therefore, the channels in D are disjoint. Lemma 1 implies that $k = h \oplus \bigoplus D$. Repeated application of Lemma 1 and Theorem 1(a) yields that $h \oplus \bigoplus D' \in Kp(G) = K$ for every subset D' of D. □

Theorem 3 thus constrains the class of Kekulé cells rather heavily. In Section 5.1, we shall see that all linkable cells that have not more than 5 ports, are Kekulé cells. This suggested to us the conjecture that every linkable cell is a Kekulé cell, which is not true; see Section 3.8.

As was pointed out to us by an anonymous referee, at this point we have entered the field of the Δ-matroids introduced by Bouchet [1]. The name Δ-matroid is taken from the operator Δ, used by Bouchet where we use $\oplus$.

A cell K is called a *Δ-matroid* [1] if, for every $h, k \in K$ and $p \in h \oplus k$, there is some $q \in h \oplus k$ with $q \neq p$ and $h \oplus \{p, q\} \in K$.

It is easy to see that every linkable cell is a Δ-matroid. Theorem 3 seems to be known to the authors of [2]. In any case, the theorem is a generalization of Theorem 2.1 of [1], which states (in our terms) that $Kp(G, P)$ is a Δ-matroid for every ported graph with port set P equal to the vertex set V of G.

3.6. Contraction and fusion of (Kekulé) cells

Next to renaming and translation, there are two other constructions on cells and Kekulé cells worth studying.

Contraction. For a cell K on P and a set $u \in \mathbb{P}(P)$, the *contraction* K/u is defined as the cell $K/u = \{k \setminus u \mid k \in K \wedge u \subseteq k\}$ on the port set $P \setminus u$. For instance, the cell on $abcd$ with the elements $\emptyset, ab, ad, bc, bd, cd, abcd$ has $K/b = \{a, c, d, acd\}$ on acd.

The associated construction on graphs is trivial. Given a ported graph with $K = Kp(G, P)$ and $u \subseteq V \cap P$, it suffices to make all ports in u internal.

Proposition 3. *Let (G, P) be a ported graph, and let $u \subseteq V \cap P$. Then $Kp(G, P)/u = Kp(G, P \setminus u)$ on $P \setminus u$.*

Note that K/u is empty when u contains ports that do not occur in K.

Fusion. Given two cells K_0 and K_1 on P, the *fusion* is defined by

$$K_0 \odot K_1 = \{h \cup k \mid h \in K_0, k \in K_1 : h \cap k = \emptyset\}.$$

The associated construction on graphs G_0 and G_1 yields a graph $G_0 \odot G_1$ constructed as follows. If necessary, rename internal nodes of the graphs in such a way that G_0 and G_1 have no common internal nodes, i.e., $G_0 = (V_0, E_0)$ and $G_1 = (V_1, E_1)$ with $V_0 \cap V_1 \subseteq P$. Then $G_0 \odot G_1 = (V_0 \cup V_1, E_0 \cup E_1)$. In other words, the graphs are joined by identifying the common ports (edges in $E_0 \cap E_1$ are between common ports).

Proposition 4. *Let G_0 and G_1 be ported graphs with a common port set P. Then $Kp(G_0 \odot G_1, P) = Kp(G_0, P) \odot Kp(G_1, P)$.*

Proof. Put $G = G_0 \odot G_1$. Let W be a Kekulé state of (G, P). Then $W_0 = W \cap E_0$ is a Kekulé state of (G_0, P) and $W_1 = W \setminus E_0$ is a Kekulé state of (G_1, P). Putting $k_i = (W_i|P)$, we have $k_0 \cap k_1 = \emptyset$ and $k_0 \cup k_1 = (W|P)$. This proves that $Kp(G, P) \subseteq Kp(G_0, P) \odot Kp(G_1, P)$. The proof of the other inclusion is similar. □

In the theory of matroids and Δ-matroids, the fusion operator is known as the "union" operator [1, p. 62]. It seems confusing to use "union" in this way.

The definition of $\odot$ immediately implies that it is a commutative operator on cells. It is also associative. Indeed, $K_0 \odot K_1 \odot K_2$ consists, in both possible interpretations, of the unions $g \cup h \cup k$ where (g, h, k) range over the pairwise disjoint triples in $K_0 \times K_1 \times K_2$. The element $\{\emptyset\}$ is the neutral element for $\odot$ in the sense that $K \odot \{\emptyset\} = K$ for all cells K. This shows that the set of

all cells with the operator $\odot$ is a commutative monoid. We can therefore use $\bigodot$ as an operator for repeated fusion.

3.7. Kekulé cells inductively defined

For a channel $c \in P^{(2)}$, we define the cell $Ed(c) = \{\emptyset, c\}$ on P. These cells serve as the base case for the following inductive characterization of $KC(P)$.

Theorem 4. (a) *For every channel c, we have $Ed(c) \in KC(P)$.*
(b) *For every pair $K_0, K_1 \in KC(P)$, we have $K_0 \odot K_1 \in KC(P)$.*
(c) *For every $K \in KC(P)$ and $u \in \mathbb{P}(P)$, we have $K/u \in KC(P \setminus u)$.*
(d) *$K \in KC(P)$ holds only if this can be inferred by a finite number of applications of* (a)–(c).

Proof. (a) The cell $Ed(c)$ is the Kekulé cell of the graph with the single edge c, where all ports that do not occur in c are isolated.
(b) This follows from Proposition 4.
(c) This follows from Proposition 3.
(d) Assume $K \in KC(P)$. Then there is a graph $G = (V, E)$ with $K = Kp(G, P)$. For the edges $c \in E$, we have the cells $Ed(c)$ on the port set V. The repeated fusion $K' = \bigodot_c Ed(c)$ where c ranges over all edges of G is therefore a cell on V. The cell K' consists of the unions of the subsets $W \subseteq E$ that are pairwise disjoint. Take $u = V \setminus P$, the set of internal nodes of G. Then $K'/u = K$ on P.

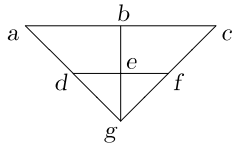
Finally, by (a), the cells $Ed(c)$ are Kekulé cells on V. By (b), K' is a Kekulé cell on V, and by (c) K itself is a Kekulé cell on P. □

3.8. Linkable does not imply Kekulé

As we thought for some time that every linkable cell might be a Kekulé cell, we define a cell K to be an *lk-refutation* iff it is linkable and not a Kekulé cell. The referee mentioned in Section 3.5 showed however that not every linkable cell is a Kekulé cell. The argument is as follows.

One may recall that a cell K is called a (base set of a) *matroid* if, for every $h, k \in K$ and $p \in h \setminus k$, there is some $q \in k \setminus h$ with $h \oplus \{p, q\} \in K$. A cell is a matroid iff it is Δ-matroid and all its members have the same cardinality [1,2]. A matroid is called "transversal" when it can be constructed in a specific way. A "gammoid" is a contraction of a transversal matroid [9]. A Kekulé cell in which all members have the same cardinality is by construction a contraction of a matching matroid. Every matching matroid is transversal (see e.g. [3]). Therefore, every Kekulé cell in which all members have the same cardinality is a gammoid.

Consider the cell K on the set $P = abcdefg$ that consists of all triples from P except abc, def, adg, beg, cfg. The exceptions are the lines in the drawing:



Cell K has 30 members. It is the matroid used by Ingleton [9] to show that a strongly base-orderable matroid need not be a gammoid. As all its members have the same cardinality, it is therefore not a Kekulé cell. On the other hand, for matroids, strongly base-orderable is the same as linkable [2]. Indeed, our C program confirms that cell K is linkable. The cell is therefore an *lk*-refutation.

3.9. The Kekulé monoid

The monoid of the cells on P is ordered by inclusion, and $\odot$ is monotonic with respect to this order.

For any pair of cells K_0, K_1, we have

$$\emptyset \in K_0 \odot K_1 \quad \equiv \quad \emptyset \in K_0 \wedge \emptyset \in K_1. \tag{0}$$

In fact, $g \cup h = \emptyset$ if and only if both $g = \emptyset$ and $h = \emptyset$, and the latter two sets are disjoint. We also have

$$K_0 \subseteq K_0 \odot K_1 \quad \equiv \quad K_0 = \emptyset \vee \emptyset \in K_1. \tag{1}$$

If $K_0 = \emptyset$, the lefthand side holds trivially. If $\emptyset \in K_1$, the lefthand side holds because of $K_0 = K_0 \odot \{\emptyset\} \subseteq K_0 \odot K_1$. Conversely, assume $K_0 \neq \emptyset$ and $K_0 \subseteq K_0 \odot K_1$. Choose $k \in K_0$ with $\#k$ minimal. As $k \in K_0 \odot K_1$, there are $g \in K_0$ and $h \in K_1$ with $g \cap h = \emptyset$ and $g \cup h = k$. This implies that $\#g + \#h = \#k$. By minimality of $\#k$, it follows that $g = k$ and $\emptyset = h \in K_1$.

In view of the formulas (0) and (1), we define a cell K to be *regular* if $\emptyset \in K$. For every cell K and every $k \in K$, the translation $k \oplus K$ is regular and $k \oplus (k \oplus K) = K$. Therefore, every nonempty cell is the translation of a regular cell.

Let K be a regular cell. If K has a decomposition $K = K_0 \odot K_1$, it follows from (0) that K_0 and K_1 are both regular. It then follows from (1) that K_0 and K_1 are both contained in K.

We define a linkable regular cell K to be *decomposable* iff it is the fusion of two linkable proper subcells. We define K to be *indecomposable* iff it is not decomposable. We define a linkable cell K to be *original* iff $k \oplus K$ is indecomposable for every $k \in K$. Let a cell K be called a *minimal lk-refutation* iff it is an lk-refutation and $\#K \leq \#K'$ for every (other) lk-refutation K'.

Proposition 5. *Every minimal lk-refutation is original.*

Proof. Let K be a minimal lk-refutation. Assume $k \oplus K$ is decomposable for some $k \in K$, say $k \oplus K = K_0 \odot K_1$ with K_0 and K_1 linkable proper subcells of $k \oplus K$. As K_0 and K_1 are linkable and have fewer elements than K, they are Kekulé cells. Therefore $k \oplus K = K_0 \odot K_1$ is a Kekulé cell. Therefore $K = k \oplus (k \oplus K)$ is a Kekulé cell, contradicting the assumption that it is a lk-refutation. □

4. From graph to cell, and back again

In the remainder of this paper, we fix the port set P. We therefore omit P as an argument of Kp.

In this section, we treat the problems of computing the Kekulé cell of a given graph, and the much more difficult, converse problem of searching for a graph with a given cell as a Kekulé cell.

The first problem is solved in Section 4.1. Section 4.2 presents our main idea for the converse problem. In Section 4.3, we show that a small part of the graph can be read from the cell. In Section 4.4, we combine this idea with a treatment of the internal vertices to an algorithm that is productive for most linkable cells of order ≤6. Section 4.5 presents an alternative approach that is useful for four remaining cases.

4.1. From graph to the Kekulé cell

One of the first questions of the theory is how to compute the Kekulé cell $Kp(G)$ of a given graph G. We know two completely different ways to do this.

The first algorithm [7, Section 3] reduces the problem to a system of linear equations over the field with two elements to determine the so-called semi-Kekulé states, followed by a search to find the Kekulé states themselves, followed by a projection to get the Kekulé port assignments.

We here present a different solution based on recursion. The idea is to express the Kekulé cell of the graph in terms of the Kekulé cells of smaller graphs. It is here that we use the flexibility of Section 2.1 that ports need not be vertices. There are two possibilities for getting smaller graphs: either remove a single edge, or remove a vertex together with its incident edges. We use the second possibility.

Fix a graph $G = (V, E)$. For a set of vertices $U \subseteq V$, we define the induced subgraph $G \mid U = (U, E \cap U^{(2)})$ and its Kekulé cell $M(U) = Kp(G \mid U)$. It is clear that $M(U) = Kp(G)$ when $U = V$. It therefore suffices to be able to compute $M(U)$ for arbitrary U. The base case of the recursion with empty U is easy.

In the recursive step, we remove a vertex u from U. We make a case distinction between an interior vertex u and a port u, and we let edge e range over $nbh(u, U) = \{e \in E \cap U^{(2)} \mid u \in e\}$, the neighborhood of vertex u in $G|U$. We need here translation (see Section 3.4) over the port assignments $e \cap P$.

Proposition 6. (a) $M(\emptyset) = \{\emptyset\}$.
(b) *Let* $u \in U \setminus P$. *Then* $M(U) = \bigcup_e((e \cap P) \oplus M(U \setminus e))$.
(c) *Let* $u \in U \cap P$. *Then* $M(U) = M(U \setminus \{u\}) \cup \bigcup_e((e \cap P) \oplus M(U \setminus e))$.

Proof. (a) The emptyset $\emptyset$ is the only Kekulé state of the empty graph, and $\emptyset$ is its port restriction.
(b) Given U and $u \in U \setminus P$, a set of edges W is a Kekulé state of $G|U$ if and only if there is a (necessarily unique) $e \in nbh(u, U) \cap W$ such that $W' = W \setminus \{e\}$ is a Kekulé state of $G|(U \setminus e)$. In that case, $W|P = (e \cap P) \cup (W'|P) = (e \cap P) \oplus (W'|P)$.
(c) If $u \in U \cap P$, we have also the possibility that $u \notin \bigcup W$, so that W is a Kekulé state of $G|(U \setminus \{u\})$. □

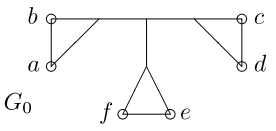
Proposition 6 suggests two possible strategies: either remove the elements of $U \cap P$ as soon as possible, or remove the interior vertices from U as soon as possible. The first strategy is probably the best one, because of the implication

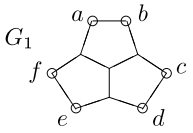
$$U \cap P = \emptyset \Rightarrow M(U) = (m(G|U)?\{\emptyset\} : \emptyset),$$

where $m(H)$ is defined to mean that the graph H has a perfect matching. The article [11] gives an algorithm to find a perfect matching in time complexity $\mathcal{O}(e\sqrt{v})$ where e is the number of edges and v is the number of vertices; $m(H)$ can also be computed by a determinant criterion [10, Theorem 8.2.3].

We have used Proposition 6 to construct a computer program to construct the Kekulé cell of an arbitrary graph [6].

Example. The diagram below contains two symmetric graphs, each with 6 ports and 4 internal nodes. In either case, one can use the symmetry to compute the Kekulé cell by hand. This, however, is cumbersome and error prone. We therefore prefer to use the computer program.





The lefthand graph G_0 has a symmetry group of 48 automorphisms. Its Kekulé cell is

$$K_0 = \{ac, bc, ad, bd, ae, be, ce, abce, de, abde, acde, bcde, af, bf, cf, abcf, df, abdf, acdf, bcdf, acef, bcef, adef, bdef\}.$$

The righthand graph G_1 has a symmetry group of 6 automorphisms with a transitive action on the ports. Its Kekulé cell is

$$K_1 = \{ab, ac, ad, bd, cd, abcd, ae, be, ce, abce, acde, bcde, bf, cf, abcf, df, abdf, bcdf, ef, abef, acef, adef, bdef, cdef, abcdef\}.$$

In either case, the symmetry of the cell is most easily analyzed by taking the complement within $Even(Q)$. These complements are

$$\overline{K_0} = \{\emptyset, ab, cd, ef, abcd, abef, cdef, abcdef\},$$
$$\overline{K_1} = \{\emptyset, af, bc, de, abde, acdf, bcef\}.$$

4.2. *How to find a graph for a given cell?*

Given a cell K, how to find a graph G with $K = Kp(G)$? In view of Proposition 1, we may assume that $P = \bigcup K$. As we now want to change the set of edges, we write $Kp(G) = Kp(V, E)$ when $G = (V, E)$.

Our starting point is the observation that the function Kp is monotonic in the set of edges, in the sense that, if $E_0 \subseteq E$, then $Kp(V, E_0) \subseteq Kp(V, E)$. Therefore, given K, we can begin searching for a graph $G_0 = (V, E_0)$ with $Kp(G_0) \subseteq K$ and then add edges to E_0 to approximate K.

Of course, the initial graph should be such that there can be hope of reaching K. So, e.g., in view of Lemma 2, if the cell is even (odd), we should give G_0 a even (odd) number of internal nodes.

The growth of $Kp(V, E)$ when E grows, is determined in the next result, which is similar to Proposition 6.

Proposition 7. *Let $G = (V, E)$. Assume $e \in E$. Write $E_0 = E \setminus \{e\}$, and $U = V \setminus e$, and $E_1 = E \cap U^{(2)}$. Then $Kp(G) = Kp(V, E_0) \cup ((e \cap P) \oplus Kp(U, E_1))$.*

Proof. First, every Kekulé state of (V, E_0) is a Kekulé state of G. Conversely, every Kekulé state of G that does not contain e is a Kekulé state of (V, E_0). Recall that edge e is a doubleton of vertices and $e \subseteq V$. The graph $G' = (U, E_1)$ is therefore well-defined. If W is a Kekulé state of the graph G that contains e, then $W \setminus \{e\}$ is a Kekulé state of the graph G'. Conversely, if W' is a Kekulé state of G' then $W' \cup \{e\}$ is a Kekulé state of G. This implies that $Kek(G)$ is the union

$$Kek(V, E_0) \cup \{W' \cup \{e\} \mid W' \in Kek(U, E_1)\}.$$

It remains to take the port restrictions and to observe that $(W' \cup \{e\})|P = (e \cap P) \oplus (W'|P)$ whenever $W' \subseteq U^{(2)}$. □

In order to get a terminating procedure, we fix a finite set V of vertices that are allowed to participate, and a set of edges L that can be added to the graph. We start with a small initial graph (V, E_0) with a Kekulé cell $K_0 \subseteq K$, and we try to add edges from L to E_0. This is done in the recursive procedure *findGraph* of Fig. 1.

Precondition: $Kp(V, E_0) = K_0 \subseteq K$ and $L \subseteq V^{(2)} \setminus E_0$.
Action: if possible, return E with $E_0 \subseteq E \subseteq E_0 \cup L$ and $Kp(V, E) = K$, else return $\perp$.

$findGraph(V, E_0, K_0, K, L) =$
if $K_0 = K$ **then return** E_0 **end** ;
if $L = \emptyset$ **then return** $\perp$ **end** ;
choose $e \in L$; $L' := L \setminus \{e\}$;
$E := findGraph(V, E_0, K_0, K, L')$;
if $E \neq \perp$ **then return** E **end** ;
$U := V \setminus e$;
$K' := (P \cap e) \oplus Kp(U, E_0 \cap U^{(2)})$;
if $K' \subseteq K$
then return $findGraph(V, E_0 \cup \{e\}, K_0 \cup K', K, L')$
else return $\perp$ **end** .

Fig. 1. Method to find an edge set E with $E_0 \subseteq E \subseteq E_0 \cup L$ and $Kp(V, E) = K$.

For a call of *findGraph* that satisfies the precondition, the first recursive call in the body obviously also satisfies the precondition. The second recursive call in the body satisfies the precondition because of $Kp(V, E_0 \cup \{e\}) = K_0 \cup K'$ by Proposition 7. It follows from the precondition that, when the procedure terminates, it delivers a correct answer. The recursion terminates because L' is smaller than L.

Unfortunately, the time complexity of this procedure is large. In essence, we consider all sets E with $E_0 \subseteq E \subseteq E_0 \cup L$. The only optimization is that we approach E from below, and stop when $Kp(V, E) \not\subseteq K$.

In the remainder of this section, we discuss how to make the initial choice of E_0 and L.

4.3. Border channels

From the given cell K, we know the set of ports involved. A natural starting point is therefore to ask which edges can or should be placed between the ports. The answer is captured in the concept of border channels.

Let a channel c be called a *border channel* of cell K iff $k \cup c \in K$ for all $k \in K$ with $k \cap c = \emptyset$.

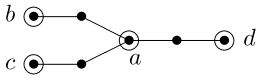
Lemma 3. *Let $K = Kp(V, E)$ and let c be a channel. Then c is a border channel of K iff $K = Kp(V, E \cup \{c\})$.*

Proof. Put $E' = E \cup \{c\}$ and $K' = Kp(V, E')$. Then $K \subseteq K'$ because every Kekulé state of $G = (V, E)$ is a Kekulé state of $G' = (V, E')$.

First, assume that c is not a border channel. Then there is $k \in K$ with $k \cap c = \emptyset$ and $k \cup c \notin K$. Then G has a Kekulé state W with $k = W|P$. Since $k \cap c = \emptyset$ and $c \subseteq P$, $W' = W \cup \{c\}$ is a Kekulé state of G'. It follows that $k \cup c = W'|P \in K'$, proving that $K \neq K'$.

Conversely, assume $K \neq K'$. Then there is $k' \in K' \setminus K$. There is a Kekulé state W' of G' with $W'|P = k'$. Since $k' \notin K$, W' is not a Kekulé state of G. Therefore $c \in W'$. It follows that $W := W' \setminus \{c\}$ is a Kekulé state of G. The set $k := k' \setminus c$ satisfies $k = W|P \in K$ and $k \cup c = k' \notin K$. This proves that c is not a border channel. □

Example. The graph sketched below has four ports and three internal vertices. Its Kekulé cell is $\{abc, abd, acd, bcd\}$.



In this case, all six channels (ab, etc.) are border channels, but none of them is an edge of the graph. One can add them to the graph without changing the cell.

Let $bc(K)$ denote the set of the border channels of the cell K. This is a finite set. If $K = Kp(V, E)$, one can add the elements of $bc(K)$ to E without changing $Kp(V, E)$ one by one. In this way, we get the following corollary.

Corollary 1. *Let $K = Kp(V, E)$. Then $K = Kp(V, E \cup bc(K))$.*

Corollary 2. *Let K be a Kekulé cell. Then there is a graph (V, E) with $K = Kp(V, E)$ and $E \cap P^{(2)} = bc(K)$.*

This result means that, when we search for a graph for a given cell K, we can begin with the ports occurring in K, connect them according to $bc(K)$, and then add some internal vertices, and edges from ports to internal vertices and between internal vertices.

4.4. Using the best border graph

Given a nonempty linkable cell K, we can make optimal use of the result of the previous section by choosing a port assignment $k \in K$ with the maximal number $\#bc(k \oplus K)$. The cell $K' = k \oplus K$ is regular. If we can find a graph G' with $K' = Kp(G')$, backward translation yields $K = Kp(k \oplus G')$ by Proposition 2.

Let P be the set of ports of K. We regard the graph $G_0 = (P, bc(K'))$ as the "best border graph" of K. It satisfies $Kp(G_0) \subseteq K'$. The graph G_0 has no internal vertices and is therefore usually not big enough to hold a graph that solves our problem. We may need to add some internal vertices.

When we add internal vertices, we also need to add internal edges in such a way that the graph of the internal vertices together with the internal edges has a perfect matching, because the cell K' is regular and therefore needs a Kekulé state W with $W|P = \emptyset$.

We therefore construct a list of representatives (U, F) of the isomorphism classes of graphs with perfect matchings, and we try the algorithm of Fig. 1 with $V = P \cup U$ and $E_0 = bc(K') \cup F$ and $L = \{\{q, u\} \mid q \in Q, u \in U\}$. In other words, we consider all possible edges between the ports (in P) and the internal vertices (in U), but we add no new internal edges.

Up to isomorphism, there are 8 matched graphs of 0, 2, or 4 vertices, and there are 101 matched graphs with 6 vertices, but we have no results where the graphs with 6 vertices can be used.

4.5. Using decomposition

When the method of the previous section does not yield a result, there is another possible approach based on the ideas of Section 3.9.

In view of Theorem 3, we may assume that K is linkable. For every $k \in K$, the translation $k \oplus K$ is a linkable regular cell. This cell may have a decomposition $k \oplus K = K_0 \odot K_1$. By

Section 3.9, the cells K_0 and K_1 are both regular and contained in K. If we can find graphs G_0 and G_1 with $K_i = Kp(G_i)$ for $i = 0, 1$, then $K = Kp(k \oplus (G_0 \odot G_1))$ is a solution of the problem.

Of course, this can help only when K_0 and K_1 are linkable, and proper subsets of K. In particular, this approach fails when cell K is original.

5. Classification of Kekulé cells

The aim is to determine the equivalence classes of the linkable cells. Equivalence is meant up to port renaming and translation; these operations are both defined in Section 3.4. In fact, if a cell K is linkable, all its renamings $f^{**}(K)$ and its translations $g \oplus K$ are also linkable.

Classification of objects with respect to some equivalence relation usually works by means of a normal form, such that every object has a unique normal form, and that the verification that some object is in normal form does not require a comparison with other objects in its equivalence class. In the present case, we have no normal form concept with these properties.

It is useful to drastically generalize the setting. Cells are subsets of the set $\mathbb{P}(P)$, which is a finite vector space over the field with two elements. Translation of cells is translation in this vector space. Port renamings are linear transformations in this vector space.

In Section 5.1, we therefore first treat the problem of classification of subsets of a vector space under the action of an affine group. This is specialized to our Kekulé setting in Section 5.2. In Section 5.3, we give the classification of the linkable cells of order ≤ 5. This is also the classification of the Kekulé cells of order ≤ 5. In Section 5.4, we discuss the classification of the linkable cells of order 6. As there are 214 classes, we do not give the classification in full. It can be obtained from [6].

5.1. Classification for an action of an affine group

Let V be a vector space and let G be a group of linear transformations of V. Let us define subsets K and K' of V to be similar (notation $K \sim K'$) iff $K' = v + gK$ for some $v \in V$ and $g \in G$. Similarity is an equivalence relation and the aim is to determine the equivalence classes.

In other words, we use the semidirect product group $V \times G$, which acts on V by $(v, g)x = v + gx$ because $(v, g)((w, h)x) = v + g(w + hx) = (v + gw, gh)x$. The equivalence classes for similarity are the orbits under the obvious induced action of $V \times G$ on $\mathbb{P}(V)$. Groups like $V \times G$ are called affine groups.

As we want to classify not all cells but only the linkable ones, we assume given a subset $W \subseteq \mathbb{P}(V)$ that consists of finite subsets of V, and is invariant under the action of $V \times G$. The aim is to determine the orbits in W.

The empty set $\emptyset$ remains empty under renaming and translations. It thus forms a class by itself. We may therefore assume that all elements of W are nonempty.

Now the first possible normalization of the elements $K \in W$ is translation of K to contain the origin. In fact, if $k \in K$ then $0 \in -k + W$. We therefore call an element $K \in W$ *regular* if $0 \in K$.

To translate the origin to K still leaves too much freedom. We would prefer to put the origin in the "center" of K, but what is the "center"? As we have no definition of a single center of K, we define it as a subset of K.

In order to define the center, we need some kind of preference. We therefore assume that some totally ordered set Z is given and that $h : W \to Z$ is a G-invariant function, i.e., with

$h(gK) = h(K)$ for all $K \in W$. For $K \in W$, we define the *center* $C(K)$ of K (with respect to h) to be the set of elements $x \in K$ with $h(-x+K)$ minimal, i.e., with $h(-x+K) \leq h(-y+K)$ for all $y \in K$. Every $K \in W$ has a nonempty center because Z is totally ordered and K is nonempty and finite.

We define $K \in W$ to be *centered* (with respect to h) if $0 \in C(K)$. The set of centered elements is denoted $Cl_0 = \{K \in W \mid 0 \in C(K)\}$. It is easy to verify the following lemma.

Lemma 4. *Let $K \in W$.*

(a) *The center translates: $C(v + K) = v + C(K)$ for every $v \in V$.*
(b) *$v + K \in Cl_0$ if and only if $-v \in C(K)$.*
(c) *The center is G-covariant: $C(gK) = gC(K)$ for all $g \in G$.*
(d) *If $K \in Cl_0$, then $gK \in Cl_0$ for every $g \in G$.*

By (b), we have that every $K \in W$ is by translation similar to an element of Cl_0. This is our first normalization step. Note however that $C(K)$ may have several elements.

Next comes normalization with respect to G. For this purpose, we assume that there is a second function $m : W \to Z$. We define K to be *optimal* iff K is centered and $m(K) \leq m(gK)$ for all $g \in G$. It is easy to verify that every element $K \in W$ is similar to some optimal element. The set of optimal elements of W is called the *raw classification* Cl_1.

It remains to classify the optimal elements of Cl_1 up to similarity. For this purpose, we assume that the set W itself has a total order, and we use it to define the final classification by

$$Cl = \{K \in Cl_1 \mid \forall K' \in Cl_1 : K \sim K' \Rightarrow K \leq K'\}.$$

For $K \in Cl_1$, we need to determine the set $H(K) = \{K' \in Cl_1 \mid K' \sim K\}$, in order to remove the bigger elements from Cl_1. As a first approximation, we define $H_1(K) = Cl_1 \cap \{gK \mid g \in G\}$. We claim that, for every $K \in W$,

$$H(K) = \bigcup_{v \in C(K)} H_1(-v + K).$$

This is proved as follows.

$$\begin{aligned} K' \in H(K) &\equiv \{\text{definition } H(K)\}\ K' \in Cl_1 \wedge \exists v, g : K = v + gK' \\ &\equiv \{\text{Lemma 4(d)}\}\ K' \in Cl_1 \wedge \exists v, g : -v + K \in Cl_0 \wedge K' = g^{-1}(-v + K) \\ &\equiv \{\text{Lemma 4(b)}\}\ \exists v : v \in C(K) \wedge K' \in H_1(-v + K). \end{aligned}$$

This reduces the computation of function H on Cl_1 to the computations of C and H_1 on Cl_0.

As this computation takes work, and the removal of the elements $K' \in H(K)$ with $K < K'$ also takes work, it is important to choose the order $\leq$ on W in such a way that the elements of $H(K)$ are close together. We therefore observe that, for $K' \in H(K)$, say with $K = v + gK'$ as above, we have $h(K') = h(-v + K) = h(K)$ because h is G-invariant and $v \in C(K)$. This proves that function h is constant on $H(K)$. We now choose the order on W in such a way that

$$h(K) < h(K') \Rightarrow K < K'.$$

This is realized by taking a lexicographical order, with comparison of h values first.

5.2. Specialization to the Kekulé setting

In the Kekulé application, V is the power set $\mathbb{P}(P)$ of a finite set P of *ports*, G is the permutation group of the ports, and W is the set of the linkable cells.

In this situation, we need a G-invariant function from W to some ordered set, which is easily computable but preserves as much information as possible. The requirement of ordering suggests counting. As the ports are equivalent under permutation, the first idea would be to count all ports as 1. Given an element $K \in \mathbb{P}(V)$, however, we can retain more information when we let the weight of a port depend on K. We decided to let the weight of a port be high when that port occurs often in the larger elements of K. We thus define:

$$w(p) = \sum_{k \in K} (p \in k ?\ \#k + 1 : 0). \tag{2}$$

We then define the weight of an element $k \in K$ to be the sum $w(k) = \sum_{p \in k} w(p)$, and the *weightlist* $h(K)$ of K to be the sequence of all weights $w(k)$ of $k \in K$, in ascending order. Note that $h(K)$ does not change when the ports are permuted.

Example. Assume $K = \{\emptyset, bc, ad, abcd, ae, abce\}$, for ports a, b, etc. We have $w(a) = 16$, $w(b) = w(c) = 13$, and $w(d) = w(e) = 8$. It follows that $w(\emptyset) = 0$, $w(bc) = 26$, $w(ad) = w(ae) = 24$, $w(abcd) = w(abce) = 50$. The weightlist is therefore $(0, 24, 24, 26, 50, 50)$. Ports f, g, etc. do not occur in the cell and can be ignored.

We thus take for Z the set of all finite sequences of natural numbers, represented as infinite sequences of natural numbers with zeros from some point onward. We use on Z the lexicographical order with $xs < ys$ given by

$$\exists i \in \mathbb{N} : xs(i) < ys(i) \wedge \forall j \in \mathbb{N} : i < j \Rightarrow xs(j) = ys(j).$$

The function $h : W \to Z$ assigns to a cell its weightlist.

For the second function $m : W \to Z$, we take the sequence of port weights, assuming that the ports are numbered by natural numbers. For instance, in the above example, using the alphabetical order for the ports, we have $m(K) = (16, 13, 13, 8, 8)$ (ignoring ports $f, g \cdots$). The sequence $m(K)$ is called the *weighted histogram*. Permutations of the ports just permute the elements of the sequence. Therefore the minimal m-value is obtained when the sequence is descending, as is the case in the example. In other words, a centered cell is optimal iff its weighted histogram is descending. This implies that the occurring ports are concentrated in the beginning of the port enumeration, just as one would want.

The raw classification Cl_1 for a given order thus consists of all optimal linkable cells of that order.

Remark. The port weight definition (2) is chosen with the aim to make the raw classification small. For order 5, it yields a raw classification of 32 cells. If we replace $\#k + 1$ by $\#k$ in formula (2), we get a raw classification of 38 cells. If one replaces the definition by $w(p) = 1$, a raw classification of 103 cells appears.

Having obtained Cl_1, we have to determine the sets $H_1(K)$. This is easy because the group $\{g \in G \mid m(gK) = m(K)\}$ consists of the permutations that preserve the sequence of port lists $m(K)$. In the case of $m(K) = (16, 13, 13, 8, 8)$, one can permute b and c, and d and e, without changing the histogram. In the special case of the above example, these permutations do not

change K itself. In general, however, there is no guarantee that the permutations that preserve $m(K)$, also preserve K itself.

5.3. Small order

As announced above, we wrote a computer program, see [6], to determine the equivalence classes of linkable cells, and a program that, given a cell K, tries to find a graph G with $Kp(G) = K$.

In Section 3.3, we have seen that the cells $\emptyset$ and $\{\emptyset\}$ are both Kekulé cells. These are the only cells that need no ports. For every $k \in \mathbb{P}(P)$, the translated cell $\{k\} = k \oplus \{\emptyset\}$ is equivalent with $\{\emptyset\}$. For nonempty k, this shows that the order of a cell is not invariant under translation.

As above, a cell K is *regular* iff it contains the empty port assignment $\emptyset$. All regular cells in a similarity class have the same order. We use this as the order of the class.

We write $Even(r) = Even(P)$ where P is the set of the first r letters of the alphabet. This cell is always centered, and all its elements are centers. We can take the full graph on P as the graph with Kekulé cell $Even(P)$.

For the cases of order $\leq$5, the computer program almost instantaneously yields a classification and corresponding graphs. The results are presented below. For each order, we number the classes from 1. For each class, we give one representative cell K_i, with its set of centers C_i, and a graph G_i. The cells are sets of port assignments; the port assignments are sets of ports, and we represent them as strings of ports in alphabetical order, with the empty one represented by $\emptyset$. In the graphs, the vertices are represented by bullets and the ports are encircled.

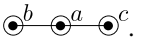
In each case, the graph presented is the first one found by our computer program. We made no effort to find graphs that fit the requirements of carbon chemistry. Recall from [7,12] that there are rather far-reaching methods to transform a graph without changing the Kekulé cell. It is beyond the scope of this paper to search for graphs that correspond to stable molecules of alternating hydrocarbons.

The classes of order $\leq$3

There are two classes of cells of order 0: the class of the empty cell $\emptyset$, and the class of the singleton cell $\{\emptyset\}$.

There are no regular cells of order 1.

In order 2 there is only one class, namely the one of $Even(2)$.

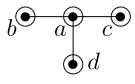
Order 3 has two classes: $K_1 = \{\emptyset, ab, ac\}$ with $C_1 = K_1$. This is the Kekulé cell of the graph 

.

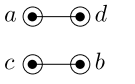
The other class is the one of $K_2 = Even(3) = C_2$.

The 7 *classes of order* 4

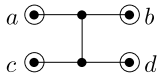
$K_1 = \{\emptyset, ab, ac, ad\} = C_1$

G_1 : 

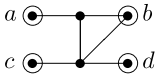
$K_2 = \{\emptyset, bc, ad, abcd\} = C_2$

G_2 : 

$K_3 = \{\emptyset, ac, bc, ad, bd\}$
$C_3 = \{\emptyset\}$

G_3 : 

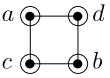
$K_4 = \{\emptyset, ab, ac, bc, ad, bd\}$
$C_4 = \{\emptyset, ab\}$

G_4 : 

The graph G_3 links the ports a, b, c, d, via two anonymous internal vertices. We have $\emptyset \in Kp(G_3)$ because it is the port restriction of the Kekulé state that only holds the edge between these internal vertices.

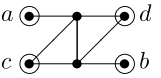
$K_5 = \{\emptyset, ac, bc, ad, bd, abcd\}$
$C_5 = K_5$

G_5 :



$K_6 = \{\emptyset, ab, ac, bc, ad, bd, cd\}$
$C_6 = \{\emptyset\}$

G_6 :

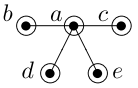


$K_7 = \{\emptyset, ba, ca, cb, da, db, dc, dcba\} = Even(4) = C_7$.

The 24 *classes of order* 5

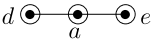
$K_1 = \{\emptyset, ab, ac, ad, ae\} = C_1$

G_1 :



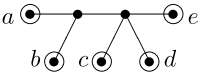
$K_2 = \{\emptyset, bc, ad, abcd, ae, abce\}$
$C_2 = K_2$

G_2 :



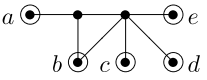
$K_3 = \{\emptyset, ac, bc, ad, bd, ae, be\}$, $C_3 = \{\emptyset\}$

G_3 :



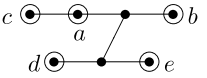
$K_4 = \{\emptyset, ab, ac, bc, ad, bd, ae, be\}$
$C_4 = \{\emptyset, ab\}$

G_4 :



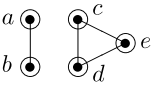
$K_5 = \{\emptyset, ac, ad, bd, abcd, ae, be, abce\}$
$C_5 = \{\emptyset, ac, ad, ae\}$

G_5 :



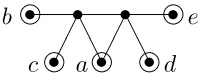
$K_6 = \{\emptyset, ab, cd, abcd, ce, abce, de, abde\} = C_6$

G_6 :



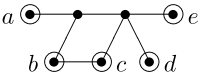
$K_7 = \{\emptyset, ab, ac, ad, bd, cd, ae, be, ce\}$
$C_7 = \{\emptyset\}$

G_7 :



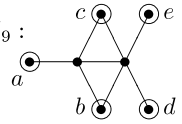
$K_8 = \{\emptyset, ac, bc, ad, bd, abcd, ae, be, abce\}$
$C_8 = \{\emptyset, ac, bc\}$

G_8 :



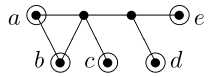
$K_9 = \{\emptyset, ab, ac, bc, ad, bd, cd, ae, be, ce\}$
$C_9 = \{\emptyset\}$

G_9 :



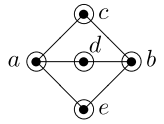
$K_{10} = \{\emptyset, ab, ad, bd, cd, abcd, ae, be, ce, abce\}$
$C_{10} = \{\emptyset, ab\}$

G_{10} :



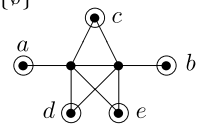
$K_{11} = \{\emptyset, ac, bc, ad, bd, abcd, ae, be, abce, abde\} = C_{11}$

G_{11} :



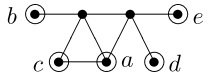
$K_{12} = \{\emptyset, ab, ac, bc, ad, bd, cd, ae, be, ce, de\}$
$C_{12} = \{\emptyset\}$

G_{12} :



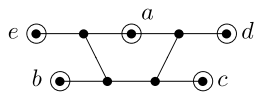
$K_{13} = \{\emptyset, ab, ac, ad, bd, cd, abcd, ae, be, ce, abce\}$
$C_{13} = \{\emptyset, ab, ac\}$

G_{13} :



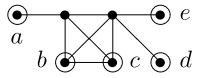
$K_{14} = \{\emptyset, ab, ac, ad, cd, abcd, ae, be, abce, de, bcde\}$
$C_{14} = \{\emptyset, ab, ac, ad, ae\}$

G_{14} :



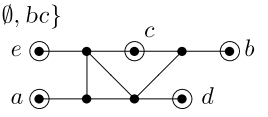
$K_{15} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce\}$
$C_{15} = \{\emptyset, ab, ac, bc\}$

G_{15} :



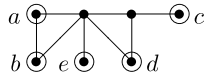
$K_{16} = \{\emptyset, ab, ac, bc, bd, cd, abcd, ae, be, ce, abde, acde\}$
$C_{16} = \{\emptyset, bc\}$

G_{16} :



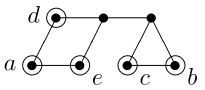
$K_{17} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ce, abce, de, abde\}$
$C_{17} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd\}$

G_{17} :



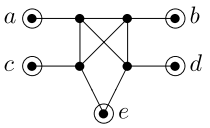
$K_{18} = \{\emptyset, bc, ad, bd, cd, abcd, ae, be, ce, abce, abde, acde\} = C_{18}$

G_{18} :



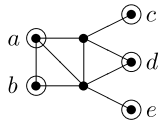
$K_{19} = \{\emptyset, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde\}$
$C_{19} = \{\emptyset\}$

G_{19} :



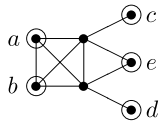
$K_{20} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, ce, abce, de, abde\}$
$C_{20} = \{\emptyset, ab, ac, ad\}$

G_{20} :



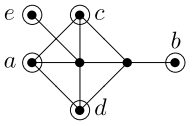
$K_{21} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde\}$
$C_{21} = \{\emptyset, ab\}$

G_{21} :



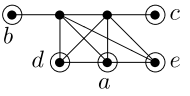
$K_{22} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, be, ce, abce, de, abde, acde\}$
$C_{22} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd\}$

G_{22} :



$K_{23} = \{\emptyset, ab, ac, bc, ad, bd, cd, abcd, ae, be, ce, abce, de, abde, acde\}$
$C_{23} = \{\emptyset, ab, ac, ad, ae\}$

G_{23} :



$K_{24} = Even(5) = C_{24}$.

This classification shows that, up to order 5, all linkable cells are Kekulé cells. In three cases of order 5, we use more than two internal vertices, namely for the classes of K_{14}, K_{16}, and K_{19}.

The pyracylene molecule (see Section 1.1) has eight atoms where ports can be located according to the rules of carbon chemistry. Up to symmetry, there are 14 ways to choose 5 ports at these locations. It turns out that the resulting 14 graphs induce 8 different types of Kekulé cells, viz. the types of K_9, K_{13}, K_{15}, K_{16}, K_{19}, K_{20}, K_{21}, K_{22}; see [6].

5.4. Classification for order 6

The classification program is able to classify the cells of order 6 in around one hour. This yields 214 classes. Using the method in Section 4.4, we find graphs for 206 classes. Using

decomposition as sketched in Section 4.5, we found graphs for 4 of the 8 remaining cells. The resulting classification can be obtained at [6].

Allowing six internal vertices to the graph searching program does not yield new graphs. Allowing eight internal vertices is computationally prohibitive because there are $2^{6\cdot 8}$ relations between a set of six elements and a set of eight elements, and there are 10413 isomorphism classes of matched graphs on eight vertices.

The 4 cells that remain without graphs have 24, 24, 25, 25 elements, respectively. They are subsets of $Even(6)$, which has 32 elements. We number them as K_i according to the enumeration generated by the program.

$$\begin{aligned}
K_{132} &= \{\emptyset, bc, ad, cd, abcd, ae, be, abce, de, abde, acde, bcde, af, bf,\\
&\quad cf, abcf, abdf, acdf, bcdf, ef, acef, bcef, adef, bdef\},\\
K_{133} &= \{\emptyset, ab, bd, cd, abcd, ae, ce, abce, de, abde, acde, bcde, af, bf,\\
&\quad cf, abcf, abdf, acdf, bcdf, ef, acef, bcef, adef, bdef\},\\
K_{144} &= \{\emptyset, ac, bc, ad, bd, ae, be, ce, abce, de, abde, acde, bcde, af, bf,\\
&\quad cf, abcf, df, abdf, acdf, bcdf, acef, bcef, adef, bdef\},\\
K_{152} &= \{\emptyset, ac, bc, ad, bd, abcd, ae, be, abce, de, abde, acde, bcde, af,\\
&\quad bf, cf, abcf, abdf, acdf, bcdf, ef, abef, bcef, adef, cdef\}.
\end{aligned}$$

Cell K_{144} is especially interesting. It is invariant under a group of 48 port renamings. If we remove the empty port assignment from it, we get the Kekulé cell K_0 of the example in Section 4.1. It is not clear, however, whether the graph G_0 drawn there can be modified to yield K_{144}.

We used our computer program to determine the original cells of order ≤ 6. Up to port renaming and translation, $Even(2)$ is the only original cell of order ≤ 5. As $Even(2)$ is a Kekulé cell, it is not a lk-refutation. The only original cells of order 6, are the four cells mentioned above. This proves that minimal lk-refutations have order ≥ 6. Section 3.8 gives an lk-refutation of order 7.

6. Conclusions

The study of cells and Kekulé cells is motivated by explorations in theoretical chemistry that seem to open up the potential of molecular computation. Independent of speculations about applications, however, they also form a beautiful combinatorial structure with challenging questions.

We introduced Kekulé cells as combinatorial objects that determine the qualitative switching behavior of alternating hydrocarbons (Theorem 2). In an effort to characterize Kekulé cells, we obtained *linkability* as a necessary condition (Theorem 3), which is not sufficient.

We do not know whether the class KC of the Kekulé cells is *decidable* [4,8], i.e., whether one can make a computer program that reads (an encoding of) a cell K from its input and then computes whether or not K is a Kekulé cell. It is easy to make a computer program that enumerates all graphs, and verifies for each of them whether it has K as its Kekulé cell. This program, however, would not terminate if K is not a Kekulé cell. The program therefore only shows that the class KC is *semi-decidable*.

We showed in Theorem 4 that the class KC can be defined inductively, without mentioning graphs. This does not imply decidability because the contraction operator used makes cells smaller instead of larger.

We obtained a classification of the Kekulé cells of order ≤ 5, and of the linkable cells of order ≤ 6. We found 2 classes of Kekulé cells of order 0, 1 class of order 2, 2 classes of order 3, 7 classes of order 4, and 24 classes of order 5. In order 6, we found 214 classes of linkable cells. Of these, 210 are Kekulé cells while for 4 of the classes this is presently unknown.

Acknowledgment

We are grateful to the anonymous referee who showed the overlap with the theory of Δ-matroids and proved that linkable cells need not be Kekulé cells.

References

[1] A. Bouchet, Matchings and Δ-matroids, Discrete Appl. Math. 24 (1989) 55–62.
[2] A. Bouchet, W.H. Cunningham, Delta-matroids, jump systems and bisubmodular polyhedra, SIAM J. Discrete Math. 8 (1995) 17–32.
[3] V. Bryant, A characterisation of strict matching matroids, Combin. Probab. Comput. 1 (1992) 209–217.
[4] D. Harel, Y. Feldman, Algorithmics, the Spirit of Computing, third ed., Pearson, 2004.
[5] A.J. Heeger, S. Kivelson, J.R. Schrieffer, W.-P. Su, Solitons in conducting polymers, Rev. Modern Phys. 60 (1988) 781.
[6] W.H. Hesselink, Graph theory for alternating hydrocarbons, 2011, http://www.cs.rug.nl/~wim/kekule/index.html.
[7] W.H. Hesselink, J.C. Hummelen, H.T. Jonkman, H.G. Reker, G.R. Renardel de Lavalette, M.H. van der Veen, Kekulé cells for molecular computation, 2007. See: http://arxiv.org/abs/0704.2282.
[8] J.E. Hopcroft, J.D. Ullman, Introduction to Automata Theory, Languages and Computation, Addison-Wesley, 1979.
[9] A.W. Ingleton, Transversal matroids and related structures, in: Higher Combinatorics, in: NATO Advanced Study Inst. Ser. C: Math. Phys. Sci., vol. 31, 1977, pp. 117–131.
[10] L. Lovász, M.D. Plummer, Matching Theory, in: Annals of Discrete Mathematics, vol. 29, North-Holland, 1986.
[11] S. Micali, V.V. Vazirani, An $\mathcal{O}(\sqrt{|\ v\ |}\cdot |\ E\ |)$ algorithm for finding maximum matching in general graphs, in: 21st Annual Symposium on Foundations of Computer Science, IEEE, 1980, pp. 17–27.
[12] M.H. van der Veen, π-Logic, Ph.D. Thesis, University of Groningen, May 2006.
[13] S.N. Yalirahi, M.A. Ratner, Interplay of topology and chemical stability on the electronic transport of molecular junctions, Ann. New York Acad. Sci. 960 (2002) 153.