Methodology for Realizing Graphs in Stable Polycyclic Unsaturated Hydrocarbons

Aaron Germuth and Alex Aravind

Department of Computer Science
University of Northern British Columbia
E-mail: (germuth,csalex)@unbc.ca

July 8, 2013

Abstract

This report outlines the methodology used to find realistic graphs for classified Kekulé cells.

1 Introduction

Certain conjugated hydrocarbons exhibit an electrical switching behavior which could be useful in molecular circuitry. This behavior is completely decided by the Kekulé Cell of the molecule's graph. In such graphs, atoms are abstracted as nodes, and chemical covalent bonds as edges. Ports are special nodes which do not need to participate in conjugation, and can connect to the outside world.

All possible Kekulé cells of rank [1] <7 can be simplified to a cell from Hesselink's [1] classification. He also went on to discover graphs for (nearly) every cell in his classification. However, most graphs do not correspond to realistic molecules. We wish to improve upon the classification by providing more accurate graphs.

In order for a graph to be considered 'realistic' we have composed several restrictions:

1. Interval vertices are restricted to a degree of <4

2. Ports are restricted to a degree of <3

3. Cycles within the graph must be of size >4

---

[1]Rank is defined as the amount of ports within a single graph or cell

4. All Graphs must be connected

5. No intersection of edges

With these restrictions we have found a graph for every classification of rank 4 and 5 using methods outlined below: exhaustive search, template molecules, and the editing of existing graphs.

2   Exhaustive Search

Hesselink (Section 4.2) [1] started with the observation that Kekulé cells are monotonic (preserve order) in the set of edges and therefore:

$E_0 \subseteq E \implies K(V, E_0) \subseteq K(V, E)$

Where V = set of vertices, E = set of edges, and K(V, E) = Kekulé cell of graph G = (V,E).

Every port assignment found in K(V, E_0) is also found in K(V, E). This implies the possibility that we can start with a smaller graph G_0, and slowly add edges to approximate K(V, E). If at any point $E_0 \not\subseteq E$, we have failed to find a graph and the process is repeated with more possible vertices. In order to get a terminating procedure, we fix a finite set V of vertices and a set of edges L, which can be added to the graph.

This method relies on starting with a smaller subset graph to search from. This graph is obtained by including only the ports as vertices and allowing 'border channels' between the ports. These are edges which have been proven to not effect the Kekulé Cell, and therefore can be included. From this small graph edges are slowly added.

Since we consider all sets E where:

$$E_0 \subseteq E \subseteq E_0 \cup L$$

the time complexity is already quite large. All possible internal vertex arrangements are added by considering each graph of the isomorphic class of matched graphs. When 8 vertices are used, this class is 10413 graphs long. Considering there are 2^{6*8} relations between a set of six vertices and a set of 8, and over 10000 graphs to try, it is unfeasible to attempt to search for any graphs with 8 or more internal vertices.[2]

Instead of simply terminating after finding the first applicable graph, we can continue searching until the vertex limit is reached and retrieve all graphs which satisfy the cell. We then prune out graphs with unrealistic degrees on their vertices. This may still result graphs which are unrealistic (such as disjoint), but in most cases we can alleviate this by editing the graph (see Section 4).

3   Using Known Aromatic Hydrocarbons as Templates

The greatest disadvantage of last method is its limitation in the amount of vertices. Many poly-cyclic polyunsaturated hydrocarbons contain upwards of 10 atoms (pyrene, pyracylene, see Figure

---

[2]Only even sets of vertices are considered due to Lemma 2 of Hesselink Section 3.2 [1]

1). Not only are molecules such as these stable, but they may be necessary. Adding more nodes to a graph while preserving the cell tends to lower the degree of the nodes, and so larger graphs may be needed to fit all of our restrictions.
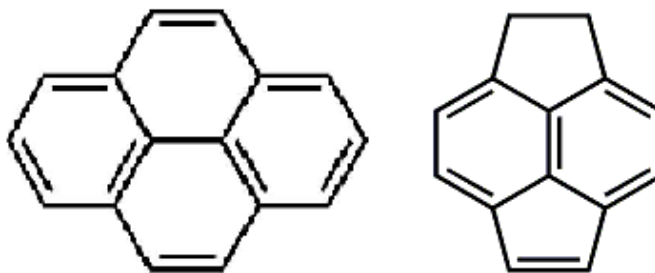

Figure 1: Two large polycylic aromatic hydrocarbons. From left to right, pyrene, and pyracylene.

In this approach we take commonly known conjugated hydrocarbons, and use them as a template. We consider all possible port locations and classify the molecule for each combination. For example, Hesselink [1] looked at all 14 possible combinations of 5 ports on pyracylene and classified 8 different cells [1].

In order to do this efficiently, we must consider symmetry. Consider the example given above, pyracylene with 5 ports. Since there are 8 nodes in pyracylene which can support a port due to the laws of carbon chemistry (degree of <3), there are 8 positions to put 5 ports. This gives $\binom{8}{5}$ or 46 possible combinations of port locations. However, this can be reduced to only 14 if we consider the symmetry groups of the nodes and trim out combinations which are symmetric.

The advantage to this approach is that one molecule may result in multiple classifications, and we know the molecule is stable. However, we could never look at every one, and it is some what of a matter of luck in picking the right molecule.

4   Editing structures

There has been some literature on methods to change a graph without changing its cell. Hesselink et al [2] show a way to split a node of high degree into multiple nodes of lower degree. M.H Van der Veen [3] classifies a list of operations which can be used to edit structures. Of course, in order to edit a structure you have to have a graph to edit, so this method is normally used in addition to one of the above methods.

4.1   Joining Disconnected Graphs

Some graphs obtained from section 2 are disjoint, as connectedness is not a requirement in the recursive step. Molecules however, must be connected. If an atom is not directly bonded within the molecule, it is not part of that molecule. Luckily, we can use usually connect two disjoint

graphs without affecting the cell. If we add internal vertices which must always conjugate 'within' themselves, they will not interfere with the cell. For example, see Figure 1.
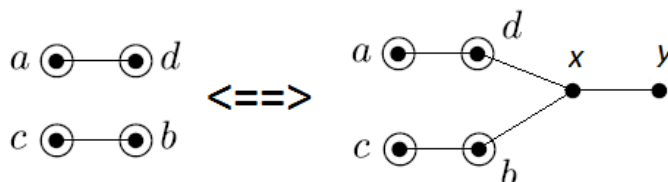

Figure 2: Transforming a disjoint graph into a connected one. Ports are represented by encircled nodes.

In the right-hand graph of Figure 2, y is not a port and so must always contain a double bond. The only other node which is available to form a double bond with y is x, so in every resonance structure of this molecule x and y will contain a double bond with each other. Therefore d or b can never form a double bond with their new edge, and the Kekulé cell has not been changed. This now resembles a stable connected molecule.

This procedure is similar to *Operation iii* and *Operation vi* of [3].

4.2 Reducing the degree on all ports

Sometimes the best graph you can find still has a high degree at one or more nodes. One common strategy is to 'extend' all the ports away by one bond, and leave a new internal vertex at their old position. It is unclear whether this always holds true in all cases, but regardless is a common strategy to reduce the degree.

Hesselink et al. [2] show a technique to converge 3 nodes into one. If the process is done backwards, it can split a node of high degree into two nodes of lesser degree. It can even be done twice, by first merging and then re-splitting, in order to shuffle edges across nodes, similar to *Operation vii* of [3].

4.3 Removing Steric Hinderance

Another factor which causes results to not correspond to stable molecules is steric hinderance between carbons atoms. Carbon is tetrahedral and prefers to bond to 4 identities at approximately $109.5°$. However, in triangular arrangements such as cyclopropene, the carbon bonds are forced to be $60°$. In cyclobutadiene, the bonds are $90°$. However, we can easily avoid such hinderance by again adding two internal vertices which are conjugated to everything (similiar to *Operation vi* of [3]). In the case of cyclopropene, we add two vertices to create cyclopentene. With cyclobutadiene, we get benzene (see Figure 3). This is why we restrict cycles within the graph to be of size >4. Within a molecule such as cyclopentene, the angles are $108°$, explaining it's stability.
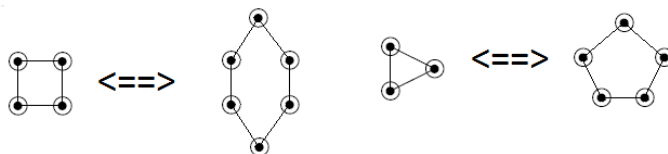
4

Figure 3: Removing steric hinderance by adding additional internal vertices. From left to right, cyclobutadiene, benzene, cyclopropane, cyclopentane

4.4   Adding or Removing Edges

Based on the nature of Hesselink's classifications, as the Kekule classification 'number' increases (ex. K3 as opposed to K7) the complexity of the cell increases. The number of resonance states and therefore port assignments increases until the last classification which contains every assignment of even length. We can use this fact to help us search for missing graphs. For example, assume you had a graph for K21 and were searching for K23. Since K23 has more port assignments than K21, we can try adding edges to our graph of K21 in order to find K23. This approach is by no means a guarantee, but has been useful in the past.

5   Current Results

Insert 7 graphs for rank 4 and 24 graphs for rank 5 here:

6   Putting it Together

All the above logic has been combined into one application which can find stable graphs for nearly every classification of rank 5. The procedure is as follows. First, Hesselink's approach is performed. Instead of simply returning the first graph which satisfies the cell, we keep searching until we have tried all possibilities within the vertex limit. We then parse through our list of applicable graphs. First all graphs are checked for their highest degree. Graphs with a degree more than 3 on normal vertices or more than 2 on ports are removed.

Next, graphs which are disjoint are fixed. This procedure is outlined in section 4.2. The graph is parsed for detection of 3 cycles (cycles of length 3). The steric hinderance of a 3 cycle is too unstable to be present within a real molecule. For this reason, we extend such cycles to 5 cycles, which are commonly seen in aromatic hydrocarbons.

We then add to the long list of trimmed graphs based on all the template molecules. If one section has not found an applicable graph at this point, we attempt to add edges. We take every graph from every classification below the missing one, and try to add possible edges to the graph. Only edges which will exceed the maximum degree of the graph are considered. In the case of adding an edge between the ports, the ports can often be 'extended' first to prevent the degree from exceeding our limit. From this, we display the graph with the lowest amount of nodes to the user.

(currently in text form).

We want to put all these methods together into one application which can get you realistic graphs for every classification. I believe first it should run Hesselink's program and try to get graphs. If more than one graph is available, the first one found should be viewable, but ideally, all should be viewable somehow. Then, if some haven't been found, we try 10 or so built in molecules. Option should be there to add in your own custom molecule and see if it fits any classification. Finally, if some still aren't found, it should take classifications it already has and try adding/removing edges to get intermediate ones. Some editing will be done i imagine.

6.1   Current Questions

Can you get the rank of a cell from it's port assignments. Does every single possible cell with 5 ports have at least one port assignment with e? if not, you must tell the rank of your cell. How can a graph which only one internal node, which gets a cell like a b abc be normalized and centered and got a port assignment of 0 ab abcd. There is no way that 0 could ever be a valid Kekulé structure. I think the answer to this question, is that it can't but it is equivalent somehow in terms of electrical switching behaviour.

References

[1] W. H. Hesselink, Graph Theory for alternating hydrocarbons with attached ports. Indagationes Mathematicae, Elsevier, 24:115141, 2013.

[2] W.H. Hesselink, J.C. Hummelen, H.T. Jonkman, H.G. Reker, G.R. Renardel de Lavalette, M.H. van der Veen, Kekule Cells for Molecular Computation. Cornell University Library Online, 2013.

[3] M.H. van der Veen. π-Logic. PhD thesis, University of Groningen, May 2006.

[4] A.J. Heeger, S. Kivelson, J.R. Schrieffer, and W.-P Su. Solitons in conducting polymers, Rev. Mod. Phys.,60:781, 1988.