Computing the Local Aromaticity of Benzenoids Thanks to Constraint Programming

Yannick Carissan, Chisom-Adaobi Dim, Denis Hagebaum-Reignier, Nicolas Prcovic, Cyril Terrioux, Adrien Varet

HAL Id: hal-02931928
https://amu.hal.science/hal-02931928

Submitted on 16 Feb 2021

Computing the Local Aromaticity of Benzenoids Thanks to Constraint Programming[⋆,⋆⋆]

Yannick Carissan[1][0000−0002−9876−0272], Chisom-Adaobi Dim[1],
Denis Hagebaum-Reignier[1][0000−0001−8761−1047], Nicolas Prcovic[2],
Cyril Terrioux[2][0000−0002−9779−9108], and Adrien Varet[2]

[1] Aix Marseille Univ, CNRS, Centrale Marseille, ISM2, Marseille, France
[2] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
{firstname.name}@univ-amu.fr

Abstract. Benzenoids are a subfamily of hydrocarbons (molecules that are only made of hydrogen and carbon atoms) whose carbon atoms form hexagons. These molecules are widely studied in theoretical chemistry. Then, there is a lot of problems relative to this subject, like the benzenoid generation or the enumeration of all its Kekulé structures (i.e. all valid configurations of double bonds). In this context, the computation of the local aromaticity of a given benzenoid is an important problematic since the aromaticity cannot be measured. Nowadays, computing aromaticity requires quantum chemistry calculations that are too expensive to be used on medium to large-sized molecules. But, there exist some methods related to graph theory which can allow us to compute it. In this article, we describe how constraint programming can be useful in order to compute the aromaticity of benzenoids. Moreover we show that our method is much faster than the reference one, namely NICS.

Keywords: Constraint programming · Modeling · Graph variables and constraints · Chemistry

1 Introduction

Polycyclic aromatic hydrocarbons (PAHs) are hydrocarbons whose carbons are forming cycles of different sizes. The properties of these molecules depend on their aromaticity, which is a fundamental concept in chemistry (defined in Section 2). Since the discovery of graphene by Andre Geim and Konstantin Novoselov awarded with the 2010 Nobel price in physics, the interest in the aromaticity concept vividly revives due to its potential importance in nanoelectronics: aromaticity favors electronic flow through molecules, thus aromatic compounds are of interest for the design of nanoelectronic compounds. *Benzenoids* are a subfamily of PAHs made of 6-membered carbon rings (i.e. each cycle of six carbon

---
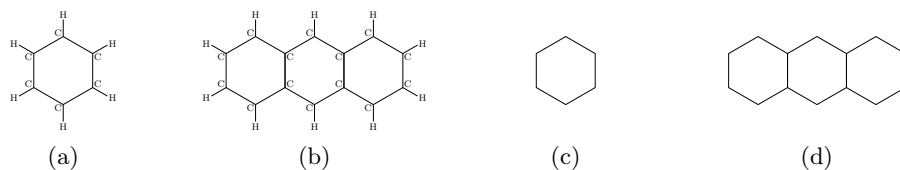
Fig. 1. Two small benzenoids: benzene (a) and anthracene (b) with their graphical representations (c) and (d).

atoms forms a hexagon). To fill carbon valency, each atom of carbon is bonded to either two other carbons and one hydrogen or three carbons. For example, Figures 1(a)-(b) are representing two benzenoids: benzene and anthracene.

PAHs are well-studied in various domains because of their energetic stability, molecular structures or optical properties. In natural environment, these molecules are created by the incomplete combustion of carbon contained in combustibles [11]. PAHs are also studied in interstellar chemistry because they are suspected to be present in interstellar clouds and are believed to act as catalysts for chemical reactions taking place in space [3]. They are also intensively studied in other domains like molecular nanoelectronics [24] or organic synthesis [20,13].

In this context, *aromaticity* is very important. It allows chemists to link the energetic stability of a molecule to its molecular structure [2]. The stability of a molecule is a measure of the energy needed to break all chemical bonds and separate all the atoms of the molecule apart. Because of aromaticity, some molecules have an extra term in this energy: breaking them apart requires more energy than for non aromatic molecules with the same number of atoms. Recently, some methods using quantum chemistry were established in order to compute the aromaticity of a given molecule. The most popular one called *NICS* (Nuclear Independent Chemical Shift [1]) consists of applying a magnetic field perpendicular to the molecular plane and observe the behavior of its electrons. Analyzing the response of the electronic density allows chemists to quantify the aromaticity of the molecule. However, this method has a very high cost and computing the aromaticity of large molecules can easily take a few days. This large computational cost is due to the fact that quantum chemistry calculations require many steps involving iterative procedures before doing the actual calculation of aromaticity. To circumvent this drawback, some methods using graph theory were proposed in the 1990s [19,10,9], which roots can be tracked back to the work of Hückel in the 1930s [7]. They will be presented in the following parts.

In this paper, we present a new method based on constraint programming in order to compute the local aromaticity of benzenoids. For example, this method needs to enumerate particular cycles or count the number of Kekulé structures. Such tasks can be modeled as CSP instances and solved efficiently thanks to constraint solvers like Choco [4] while requiring a reduced implementation effort unlike usual methods from theoretical chemistry or any bespoke methods based on algorithm engineering.
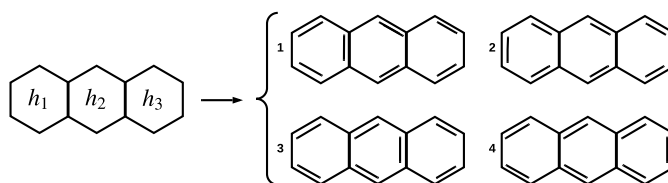
Fig. 2. Kekulé structures of anthracene.

The paper is organized as follows. Section 2 recalls some basic notions about chemistry and constraint programming. Then, Section 3 introduces some existing methods to compute the aromaticity of benzenoids. In Section 4, we describe a new method which exploits constraint programming in order to compute the local aromaticity. In Section 5, we present some experimental results which show the interest of our approach. Finally, we conclude and give some perspectives in Section 6.

2   Preliminaries

2.1   Theoretical Chemistry

Benzene (represented in Figure 1(a)) is a molecule made of 6 carbon atoms and 6 hydrogen atoms. Its carbon atoms form a hexagon (also called *benzenic cycle* or *benzenic ring*) and each of them is linked to a hydrogen atom. *Benzenoids* are a subfamily of PAHs containing all molecule which can be obtained by aggregating benzenic rings. For example, Figure 1(b) shows anthracene, which contains three benzenic rings.

Before going further, we recall some basic definitions of chemistry. Firstly, the *valence* of an atom is the number of bonds that it can build with its electrons (one electron per bond). Carbon and hydrogen atoms have a valence of 4 and 1 respectively. As in a benzenoid, each carbon atom is linked either to two other carbon atoms and one hydrogen atom or to three other carbon atoms, we can easily deduce that one of its electrons is not used. These electrons are called π-*electron* and can be used to enhance one bond by establishing *double bonds* (i.e. a bond involving two electrons per atom). Therefore, each carbon is involved in a double bond and two single bonds.

A *Kekulé structure* of a benzenoid is a valid configuration of its double bonds (i.e. a configuration in which each carbon atom is involved in exactly one double bond). Figure 2 depicts all the Kekulé structures of anthracene. A benzenoid can have several Kekulé structures or none (Figure 3(a) depicts an example of benzenoid which has no Kekulé structure). We denote $\mathcal{K}(B)$ the set of all Kekulé structures of a benzenoid B. Note that the number of Kekulé structures of a benzenoid can be exponential. Therefore, given a benzenoid, generating all its Kekulé structures is a hard problem. A benzenoid continually alternates
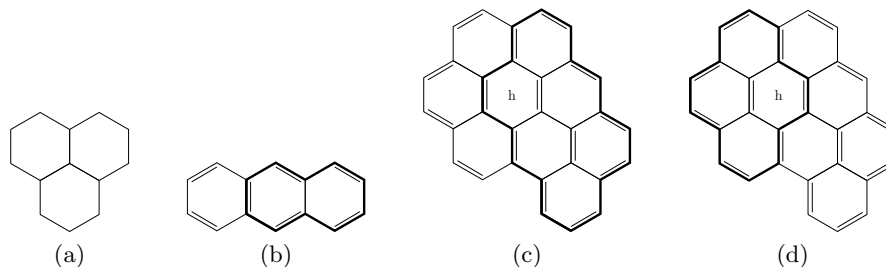
Fig. 3. A benzenoid having no Kekulé structure (a), a conjugated circuit in thick line (b) and an example of redundant circuits (c) and (d).

between its Kekulé structures. This dynamic is at the origin of the notion of aromaticity. There exist some methods based on graph theory which allow to compute the resonance energy of a given benzenoid (i.e. the energy induced by its aromaticity) and these methods require to be able to enumerate all its Kekulé structures [15].

Aromaticity is a concept built by chemists in the early 20th century in order to account for the surprising chemical stability of the benzene molecule. In this molecule, after making a single bond to each of its three neighbors (two carbon and one hydrogen), each carbon of the hexagonal geometry carries one extra electron. Electrons tend to form bonds (i.e. pair with another electron) whenever possible. Thus, this electron forms a molecular bond with the electron of a neighboring carbon atom. When all six electrons do the same, the electronic structure, first proposed by Kekulé [8], is obtained. Yet, two such structures exist as the pairing for one carbon can be with any of its two neighbors. The interaction or resonance of these two coexisting solutions is described by quantum physics and leads to an over-stabilization energy called aromaticity. This concept can be extended to fused benzene rings. It turns out that aromatic molecules often have a characteristic smell and/or taste, hence the name of the concept.

Due to the physical nature of aromaticity, hydrogen atoms do not play any role in its determination. Thus, it is custom not to take them into account in connectivity based methods. So, we do not represent them afterwards. Therefore, a benzenoid can be represented as an undirected graph $B = (V, E)$, with V the set of vertices and E the set of edges. Every vertex in V represents a carbon atom and every edge of E represents a bond between the two corresponding carbons. Moreover, this kind of graph, is connected, planar and bipartite. Figures 1(c) and (d) represent the graphs related to the molecules of benzene and anthracene. Finally, we can remark that the set of double bonds of a Kekulé structure is nothing more than a perfect matching on the benzenoid. As a reminder, a *perfect matching* of an undirected graph $G = (V, E)$ is a set of edges $E' \subseteq E$ such that $\forall (e_1, e_2) \in E' \times E', e_1 \neq e_2, e_1 \cap e_2 = \emptyset$ and $\bigcup_{e \in E'} e = V$.

2.2   Constraint Programming

An instance P of the *Constraint Satisfaction Problem (CSP)* is a triplet (X, D, C). $X = \{x_1, \ldots, x_n\}$ is a set of n *variables*. For each variable x_i of X, there exists an associated domain $D_{x_i} \in D = \{D_{x_1}, \ldots, D_{x_n}\}$ which represents the values that x_i can take. $C = \{c_1, ..., c_e\}$ represents a set of e *constraints*. Constraints represent the interactions between the variables and describe the allowed combinations of values.

Solving a CSP instance $P = (X, D, C)$ amounts to find an assignment of all the variables of X with a value contained in their associated domain which satisfies all the constraints of C. Such an assignment is called a *solution*. This problem is NP-hard [22].

Many libraries are available to represent and solve CSP instances. In this paper, we exploit the open-source Java library *Choco* [4]. This choice is highly guided by our need to be able to define *graph variables* and directly apply graph-related constraints (e.g. connected or cyclic constraints). Graph variables have as domain a set of graphs defined by a lower bound (a sub-graph called GLB) and an upper bound (a super-graph called GUB). Moreover, Choco implements the usual global constraints which make the modeling easier and its solver is efficient and configurable.

3   Computing Resonance Energy of a Benzenoid

3.1   Definitions

Resonance energy is used to quantify the energy induced by the aromaticity of a benzenoid. It allows us to get information about its stability (the lower the energy, the greater the stability). It is possible to compute this energy globally (on the entire molecule), which is called *global aromaticity* or locally (by assigning an energy to each hexagon), which is called *local aromaticity*. The latter is the most interesting since it allows us to identify the least stable parts of the molecule. Knowing that chemical reactions are more likely to occur on these parts, it can be used to predict the location of a reaction.

Randić presented a method which approximates the resonance energy of a given benzenoid by enumerating all the linearly independent Minimal Conjugated Circuits (later called h-MCCs) of each of its Kekulé structures [17]. Before going further, we have to introduce some definitions:

Definition 1 ([14]). *Let B be a benzenoid and K one of its Kekulé structures. A **conjugated circuit** $\mathcal{C}$ of K is a cycle of B whose edges correspond alternately to single and double bonds in K. The size of $\mathcal{C}$ (noted $|\mathcal{C}|$) is the integer i such that $\mathcal{C}$ contains $4i + 2$ edges.*

So, a conjugated circuit is a cycle alternating between single and double bonds. For example, the cycle in thick line in Figure 3(b) is a conjugated circuit of size 2.

Given a benzenoid B and one of its cycles $\mathcal{C}$, we call *interior of $\mathcal{C}$* the subgraph induces by all the edges and vertices which are in the interior of C.

Now, let us introduce the covering of a hexagon by a conjugated circuit:

Definition 2 ([18]). *Let B be a benzenoid and K one of its Kekulé structure. A conjugated circuit $\mathcal{C}$ of K covers a hexagon h if and only if h is contained into the interior of $\mathcal{C}$. Two conjugated circuits are said **redundant circuits** if they cover the same hexagon.*

For example, let us consider the Kekulé structure in Figure 3(c). The hexagon h is covered by two conjugated circuits: a first one of size 4 (Figure 3(c)) and a second one of size 3 (Figure 3(d)). In this case, we pay more attention to the circuit with the smallest size:

Definition 3 ([18]). *Let B be a benzenoid and K one of its Kekulé structure. $\mathcal{C}$ is a **minimal conjugated circuit** of the hexagon h of B (also called h-MCC) w.r.t. K if $\mathcal{C}$ is one of the covering circuits of h having the smallest size.*

So, if we look at Figures 3(c)-(d), the circuit of size 3 is a h-MCC for the hexagon h.

An energy is given to each of these conjugated circuits depending on their size (a smaller circuit has a higher energy). Thereafter, we denote R_i the energy induced by a conjugated circuit of size i. Initially, these values were calculated using the formula $R_i = \frac{1}{i^2}$, but optimized values were established by linear regression: $R_1 = 0.869, R_2 = 0.246, R_3 = 0.100$ and $R_4 = 0.041$ [16]. These values make it possible to compute the energy induced for a given Kekulé structure:

Definition 4 ([18]). *Let B be a benzenoid and K one of its Kekulé structure. The energy $R(K)$ induced by the minimal circuits of K is defined as follows:*

$$R(K) = \sum_{i \in \{1,2,\dots\}} r_i(K) \times R_i$$

where $r_i(K)$ is the number of minimal circuits of size i in K.

By extension, one defines the energy induced by a benzenoid.

Definition 5 ([18]). *Let B be a benzenoid. The energy $R(B)$ induced by B is defined as follows:*

$$R(B) = \sum_{K \in \mathcal{K}(B)} R(K)$$

Finally, one can define the resonance energy of a benzenoid:

Definition 6 ([18]). *Let B be a benzenoid. The **resonance energy** $E(B)$ of B is defined as follows:*

$$E(B) = \frac{R(B)}{|\mathcal{K}(B)|}$$

For example, if we look at the Kekulé structures of anthracene in Figure 2 and consider all their h-MCCs, we can easily see that its resonance energy is $\frac{6R_1+4R_2+2R_3}{4} = 1.57$ if we take the optimized R_i values.

To conclude with this part, we call *local aromaticity* of a benzenoid B on the hexagon h the energy obtained by using the previous formula, but with only looking at conjugated circuits which are minimal circuits covering h.

Definition 7. *Let h be a hexagon of a benzenoid B. The local resonance energy $E(B,h)$ of h is defined as follows:*

$$E(B,h) = \frac{\sum\limits_{K\in\mathcal{K}(B),\ \mathcal{C}\ a\ h-MCC\ w.r.t.\ K} R_{|\mathcal{C}|}}{|\mathcal{K}(B)|}$$

For example, if we consider Figure 2, the local aromaticity of hexagon h_1 of anthracene is $\frac{2R_1+R_2+R_3}{4} = 0.521$. Indeed, if we look at hexagon h_1, we have:

- A circuit of size 1 for Kekulé structure 1 (the circuit covering h_1).
- A circuit of size 1 for structure 2 (the same circuit as before).
- A circuit of size 2 for structure 3 (the circuit covering h_1 and h_2).
- A circuit of size 3 for structure 4 (the circuit covering h_1, h_2 and h_3).

Note that the sum of $E(B,h)$ over all the hexagons h of B is equal to the energy $E(B)$.

3.2   Computing the Resonance Energy

Lin and Fan [10] proposed a method based on the definition of the resonance energy. Given a benzenoid B, this method first enumerates all minimal conjugated circuits (i.e. computes a h-MCC for all its hexagons h) for each Kekulé structure. Then, it deduces the energy induced by each minimal conjugated circuit and adds them up. Finally, it divides the obtained sum by the number of Kekulé structures to obtain the resonance energy. Such a method was implied in Randić's work. The main contribution of Lin and Fan consists in describing how to compute the h-MCCs. For that, they identify the specific forms of the h-MCC depending on the location of the double bonds of h. Therefore this method requires to analyze the double bonds of all the hexagons.

The main drawback of this method is that it requires to generate all the Kekulé structures of the benzenoid. As the number of Kekulé structures may be exponential, this method is clearly inefficient in practice and can only be used for small benzenoids.

To overcome this, Lin [9] proposed another method that is able to compute all the minimal circuits of a given benzenoid without generating its Kekulé structures. This method only considers circuits having a size at most 4. So, it provides an approximation of the resonance energy. Before describing this method, let us introduce some needed definitions.
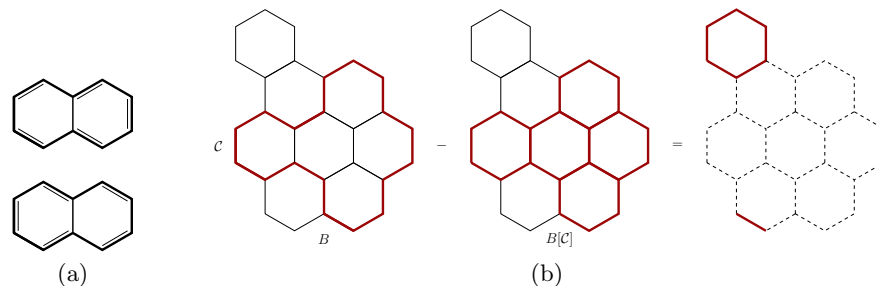
Fig. 4. A cycle $\mathcal{C}$ such as $M(\mathcal{C}) = 2$ (a) and an example of computation of the number of occurrences of a cycle (b).

Definition 8. *Let B be a benzenoid and $\mathcal{C}$ a cycle of B (with $4i + 2$ edges $\forall i \in \mathbb{N}$). $M(\mathcal{C})$ is the number of perfect matchings of $\mathcal{C}$ and its interior inducing a minimal circuit for at least one of the hexagons it covers.*

For example, let us consider $\mathcal{C}$ as being the cycle of size 2 represented in Figure 4(a). $\mathcal{C}$ can induce two different conjugated circuits that are clearly minimal and cover the two hexagons. So we have $M(\mathcal{C}) = 2$.

Definition 9 ([19]). *Let $B = (V, E)$ be a benzenoid and $\mathcal{C}$ a cycle of B (with $4i + 2$ edges $\forall i \in \mathbb{N}$). $B[\mathcal{C}]$ is the sub-graph of B induced by $\mathcal{C}$ and its interior.*

The method presented by Lin [9] relies on the following theorem:

Theorem 1 ([19]). *Let B be a benzenoid and $\mathcal{C}$ a cycle of B (with $4i + 2$ edges $\forall i \in \mathbb{N}$). $\mathcal{C}$ is a h-MCC in $|\mathcal{K}(B - B[\mathcal{C}])| \times M(\mathcal{C})$ Kekulé structures of B where $B - B[\mathcal{C}]$ is the sub-graph induced by the removal of the vertices belonging to $\mathcal{C}$ and its interior.*

Let us consider the benzenoid B described in Figure 4(b) and the cycle $\mathcal{C}$ depicted in red thick line. So, $B[\mathcal{C}]$ corresponds exactly to all the hexagons in the interior of $\mathcal{C}$, namely the hexagons depicted in red thick line in the middle figure. To compute the number of occurrences of $\mathcal{C}$ as a h-MCC, we have to compute the number of perfect matchings of the sub-graph induced by $B - B[\mathcal{C}]$. In this example, this sub-graph (depicted in red thick solid line in the rightmost figure) has two perfect matchings. Moreover, we have $M(\mathcal{C}) = 1$ because if we consider the two Kekulé structures of $\mathcal{C}$ which allow it to be a conjugated circuit, there is only one of them for which $\mathcal{C}$ is a minimal circuit for at least one of its hexagon. So, we can conclude that $\mathcal{C}$ appears twice as an h-MCC in all the Kekulé structures of B.

To sum up, the method presented by Lin [9] (described in Algorithm 1) takes as input a benzenoid B and a base containing all the cycles of size at most 4 which can induce at least one h-MCC, and another base containing all the redundant circuits of the same sizes. On Line 1, it generates the set of cycles of B which belongs to the first base (we denote this set $\mathcal{C}^*$). Then, for each cycle

---

Algorithm 1: *Compute_Resonance_Energy*

Input: a benzenoid B, a base of h-MCC, a base of redundant circuits
Output: the resonance energy $E(B)$

1 $\mathcal{C}^* \leftarrow generate\_circuits(B, 1, 4)$

2 $energy \leftarrow 0$

3 **foreach** $\mathcal{C} \in \mathcal{C}^*$ **do**

4 $\quad \big\lfloor\; energy \leftarrow energy + R_{|\mathcal{C}|} \times |K(B - B[\mathcal{C}])| \times M(\mathcal{C})$

5 **foreach** $(\mathcal{C}_1, \mathcal{C}_2) \in \mathcal{C}^* \times \mathcal{C}^*$ **do**

6 $\quad$ **if** $redundant(\mathcal{C}_1, \mathcal{C}_2)$ **then**

7 $\quad\quad$ $size \leftarrow \max(|\mathcal{C}_1|, |\mathcal{C}_2|)$

8 $\quad\quad$ $energy \leftarrow energy - R_{size} \times |K(B - B[\mathcal{C}_1 \cup \mathcal{C}_2])|$

9 **return** $\frac{energy}{|\mathcal{K}(B)|}$

---

in $\mathcal{C}^*$, it counts how many h-MCC are induced by this cycle in all the Kekulé structures of B (Lines 3-4), as shown in Figure 4(b). To conclude, it needs to find all couple of cycles of B which can produce one of the redundant circuits described in the second base and to take care not to count the cycle having the largest size (Lines 5-8).

The principal interest of this method is that it does not require to enumerate all the Kekulé structures of the given benzenoid. The only problem it has to solve is counting the number of perfect matchings in a graph, what was proved to be polynomial for benzenoids [6].

4 The Proposed Method

4.1 Preliminary Definitions

In this part, we propose a new method, using constraint programming, which refines method propose by Lin [9] by computing local aromaticity. Remind that local aromaticity is more useful than global one since it helps to predict the parts of molecules where chemical reactions may take place while leading to global information like global aromaticity. Before going into details, we have to introduce some definitions. First, we need to handle coordinates:

Definition 10. *Let $B = (V, E)$ be a benzenoid. A **coordinate function** $c :$ $V \to \mathbb{Z}^2$ of B is a function that maps a couple of integers $(c(v).x, c(v).y)$ (i.e. an abscissa and an ordinate in the Cartesian coordinate plane) to each vertex v of B such that if $(v_0, v_1, v_2, v_3, v_4, v_5)$ are the vertices forming a hexagon (given clockwise) with v_0 the vertex having the largest ordinate, we have:*

$$\begin{cases} c(v_1) = (c(v_0).x + 1, c(v_0).y - 1) \\ c(v_2) = (c(v_0).x + 1, c(v_0).y - 2) \\ c(v_3) = (c(v_0).x, c(v_0).y - 3) \\ c(v_4) = (c(v_0).x - 1, c(v_0).y - 2) \\ c(v_5) = (c(v_0).x - 1, c(v_0).y - 1) \end{cases}$$
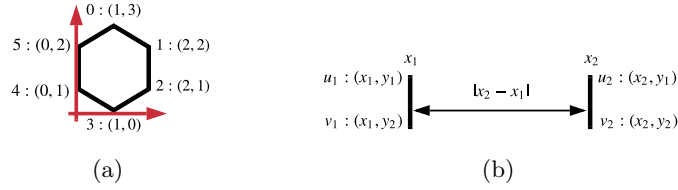
Fig. 5. Benzene with coordinates (a) and example of interval (b).

Figure 5(a) describes a simple example of coordinates for benzene.

Then, we consider some particular edges:

Definition 11. *Let B be a benzenoid and c a coordinate function. An edge $e = (u,v) \in E$ is a **vertical edge** of B if and only if $c(u).x = c(v).x$.*

The vertical edges of the benzenoid depicted in Figure 5(a) are $\{1,2\}$ and $\{4,5\}$. We now introduce the notion of interval related to vertical edges:

Definition 12. *Let B be a benzenoid and c a coordinate function. An **interval** I of B is a couple $I = (e_1, e_2)$ of vertical edges such as:*

$$\begin{cases} e_1 = (u_1, v_1) \in E \\ e_2 = (u_2, v_2) \in E \\ c(u_1).y = c(u_2).y \\ c(v_1).y = c(v_2).y \end{cases}$$

We denote:

$$\begin{cases} I.x_1 = c(u_1).x \\ I.y_1 = c(u_1).y \\ I.x_2 = c(u_2).x \\ I.y_2 = c(v_1).y \end{cases}$$

*We denote $|I| = |I.x_2 - I.x_1|$ the **size** of I.*

To sum up, an interval represents the space contained between two vertical edges that have the same ordinate. Figure 5(b) shows an example of interval.

4.2   Method Description

In this part, we describe our refined algorithm (see Algorithm 2) based on constraint programming. This method takes as input a benzenoid $B = (V, E)$, a coordinate function c, a base containing all the cycles of size at most 4 that can induce at least one h-MCC, and another one containing all the couples of cycles of the first base which can form redundant circuits and it returns an approximation of the local energy $E(B, h)$ for each hexagon h. With this aim in view, we first compute the set $\mathcal{C}^*$ of all the cycles of B whose size is at most 6 (Line 1). Then for each cycle $\mathcal{C}$ of $\mathcal{C}^*$ (Line 2), we first identify the cycle by a collection of

---

Algorithm 2: *Compute_Resonance_Energy_CP*

Input: a benzenoid B, a coordinate function c, a base of h-MCC, a base of
        redundant circuits

Output: the local resonance energy $E(B,h)$ for each hexagon h of B

1 $\mathcal{C}^* \leftarrow generate\_cycles\_choco(B, 1, 6)$

2 foreach $\mathcal{C} \in \mathcal{C}^*$ **do**

3 $id \leftarrow identify\_cycle(\mathcal{C})$

4   **if** $in\_minimal\_circuits\_base(id)$ **then**

5   **foreach** $\mathcal{C}_m \in minimal\_circuits(\mathcal{C})$ **do**

6 $h \leftarrow$ hexagon s.t. $\mathcal{C}_m$ is a h-MCC

7 $energy[h] \leftarrow energy[h] + R_{|\mathcal{C}_m|} \times |\mathcal{K}(B - B[\mathcal{C}_m])|$

8   **else if** $in\_redudant\_circuits\_base(id)$ **then**

9   **foreach** $(\mathcal{C}_1, \mathcal{C}_2)$ *s.t.* $\mathcal{C} = \mathcal{C}_1 \bigcup \mathcal{C}_2$ *and* $redundant(\mathcal{C}_1, \mathcal{C}_2)$ **do**

10 $h \leftarrow$ hexagon s.t. $\mathcal{C}_1$ and $\mathcal{C}_2$ are redundant over h

11 $\mathcal{C}' \leftarrow circuit\_with\_max\_size(\mathcal{C}_1, \mathcal{C}_2)$

12 $energy[h] \leftarrow energy[h] - |\mathcal{K}(B - B[\mathcal{C}])| \times R_{|\mathcal{C}'|}$

13 foreach h *of* B **do** $energy[h] \leftarrow energy[h]/|\mathcal{K}(B)|$

14 return $energy$

---

intervals (Line 3). Then, if $\mathcal{C}$ appears in the base of minimal circuits (Line 4), we enumerate all of the minimal circuits it induces and for each of them (Line 5), we add its contribution to the local resonance energy of the hexagon for which it is minimal (Lines 6-7). Please note that for each cycle $\mathcal{C}$, we treat separately each of its minimal circuits, so we do not have to consider $M(\mathcal{C})$ anymore. However, $\mathcal{C}$ can also correspond to the contouring of the union of two redundant circuits over a hexagon h (Line 8-9-10). If so, we have not to take into account the contribution of the largest circuit (Lines 11-12). Finally, we divide the contribution of each hexagon by the number of Kekulé structures of B (Line 13).

Now, we detail below the main steps of Algorithm 2.

Enumeration of All the Cycles We need to identify all the cycles which correspond to either a h-MCC of size at most 4 (Line 4 of Algorithm 2) or a union of two h-MCCs (Line 6). As the union of two h-MCCs of size 4 is at most of size 6, we have to enumerate all the cycles of size at most 6.

In order to enumerate all the cycles of size at most 6, we model this problem as a CSP instance $P_1 = (X_1, D_1, C_1)$. First, we consider a graph variable x_G whose domain is all the possible graphs between the empty graph and the graph B. This variable models the cycle we look for. To ensure that the value of this variable is a cycle, we impose the graph constraint cycle [5] on x_G. It remains to be ensured that the size of this cycle is at most 6. For this, we introduce a Boolean variable x_e per edge e of B. x_e is set to 1 if the edge e appears in the graph depicted by x_G, 0 otherwise. Then, we use a collection of channeling constraints in order to link the variables x_e and the variable x_G. More precisely,

for each edge e, we use a channeling constraint between x_e and x_G which imposes $x_e = 1 \iff e$ *appears in* x_G. Finally, we add a global constraint `sum` over all the variables x_e to impose $\sum\limits_{x_e | e \in E} x_e \in \{6, 10, 14, 18, 22, 26\}$ because we consider circuits of size at most 6 and a circuit of size i has $4i + 2$ edges. The channeling and sum constraints make it possible to ensure that the size of the built cycle is suitable. At the end, we obtain the following instance P_1:
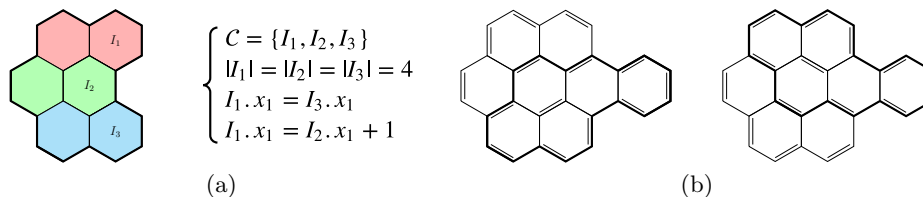
$$
\begin{cases}
X_1 = \{x_G\} \cup \{x_e | e \in E\} \\
D_1 = \{D_{x_G}\} \cup \{D_{x_e} | e \in E\} \text{ with } D_{x_G} = \{g | \emptyset \subseteq g \subseteq B\} \text{ and } D_{x_e} = \{0, 1\} \\
C_1 = \{cycle(x_G), \sum\limits_{x_e | e \in E} x_e \in \{6, 10, 14, 18, 22, 26\}\} \cup \{channeling(x_e, x_G) | e \in E\}
\end{cases}
$$

As Choco implements graph variables and offers a large amount of graph-related constraints and global constraints, this model can be easily expressed with Choco.

Counting the Number of Kekulé Structures For each cycle $\mathcal{C}$ of $\mathcal{C}^*$, we need to count the number of Kekulé structures of $\mathcal{C}$ (Line 5) or one of $B - B[\mathcal{C}]$ (Lines 5 and 10). In 2001, Rispoli presented a method which can count the number of Kekulé structures of a benzenoid [21]. The main idea of this method is to transform the given benzenoid B into a specific matrix whose determinant is the number of Kekulé structures of B. Unfortunately, although this task is polynomial, it is too time-consuming. For instance, for a molecule having 19 hexagons and 54 carbon atoms, the method proposed by Rispoli requires more than 15 minutes while the approach we propose only needs a few seconds. It seems that the time-expensive step is the computation of the determinant. So, we consider an alternative solution based again on constraint programming. We model this problem as a CSP instance $P_2 = (X_2, D_2, C_2)$ for which every solution corresponds to a Kekulé structure. As any benzenoid $B = (V, E)$ is a bipartite graph, we can divide V into two disjoint sets V_1 and V_2 such that every edge of E links a vertex of V_1 to one of V_2. We consider a variable y_v per vertex v of V_1 whose domain contains every vertex w from V_2 such that $\{v, w\} \in E$. By so doing, if the variable y_v is assigned with value w, it means that the edge $\{v, w\}$ corresponds to a double bond. By definition of a solution, this ensures that there is a single double bond for any carbon atom of V_1. It remains to ensure the same property for the vertices of V_2. This can be achieved by considering an `all-different` constraint involving all the variables of X_2. So we obtain the following instance P_2:

$$
\begin{cases}
X_2 = \{y_v | v \in V_1\} \\
D_2 = \{D_{y_v} | v \in V_1\} \text{ with } D_{y_v} = \{w | w \in V_2, \ \{v, w\} \in E\} \\
C_2 = \{\texttt{all-different}(X_2)\}
\end{cases}
$$

Clearly, the solutions of P_2 correspond to the Kekulé structures of B and so to perfect matchings of B. Regarding the filtering of the `all-different` constraint, Regin proposed an efficient algorithm based on the matchings of a particular

$$\begin{cases} \mathcal{C} = \{I_1, I_2, I_3\} \\ |I_1| = |I_2| = |I_3| = 4 \\ I_1.x_1 = I_3.x_1 \\ I_1.x_1 = I_2.x_1 + 1 \end{cases}$$

(a)   (b)

Fig. 6. A cycle and the relations between its intervals (a), and a second example of redundant circuits (b).

graph, called the *value graph* [23]. We can then note that, for our instance P_2, the value graph related to the `all-different` constraint we use is exactly the graph B. Moreover, any solver enforcing this filtering at each step of the search is able to count efficiently the number of solutions since only assignments leading to solutions are explored. Note that another model was proposed [12]. It considers binary variables and sum global constraints, but does not provide any theoretical guarantee about the efficiency, unlike the model we propose. Finally, the model we describe applies for a benzenoid B and can be easily specialized to apply to any part of B (e.g. any cycle $\mathcal{C}$ or induced sub-graph $B - B[\mathcal{C}]$).

Identification of Cycles Once Choco returns a cycle, we need to determine if this cycle belongs to the base of h-MCCs (Line 4) or one of redundant circuits (Line 6). For achieving this task, we first represent any cycle by the bias of a set of intervals and some relations between these intervals. The purpose of these relations is to represent the distance between two vertical edges of each couple of intervals (either the two left edges or the two right edges). So, with a set of intervals with this kind of relations, we are able to build the associated cycle and vice versa. Figures 6(a) shows an example of such a representation. Accordingly, we construct each base by describing every cycle (h-MCC or redundant cycle) identified by Lin [9] by a set of intervals and some relations between them. Now, each time a new cycle is returned by Choco, we translate it into a set of intervals and their relations, and check if it belongs to one of the two bases by simply comparing sets and relations. So, we are able to determine if this cycle can induce h-MCC or if it can be obtained by the union of redundant circuits.

Furthermore, the second base contains, for each cycle $\mathcal{C}$, a set of a couple of cycles whose union forms $\mathcal{C}$. This allows us to remove the energy that we have over-counted due to redundant circuits (Line 10). For example, let us consider $\mathcal{C}$ as being the union of the two circuits represented in Figure 6(b). It can be obtained either by the union of the cycles of the leftmost figure (of sizes 3 and 4), or by the cycles of the rightmost one (of sizes 3 and 4 too). So, each time Choco finds $\mathcal{C}$ in a benzenoid, we need to remove the energy associated to two circuits of size 4.

5   Experimentations

In this part, we provide an experimental comparison between our CP-based method (denoted CRECP for *Compute_Resonance_Energy_CP*) and a reference method, namely NICS. In our comparison, we do not consider the method proposed by Lin since this latter is not available and re-implementing it is not a trivial task. Moreover chemists are more interested by local aromaticity than global one [1].

About the experimental protocol, the CRECP method was implemented in Java and compiled with Java SE 1.8. It relies on Choco Solver version 4.2.3. Note that we used the default settings of Choco. For NICS, we exploited the implementation provided in a commercial program (see http://gaussian.com/). Both CRECP and NICS are run on server with 2.20 GHz Intel Xeon Gold processor and 256 Gb under CentOS Linux release 8.1.1911. We consider as benchmark a set of 28 benzenoids of various and reasonable size so that NICS can be executed within a reasonable amount of time. The sources of CRECP and benchmark are available at https://github.com/AdrienVaret/CPLocalAromaticity.

First, Table 1 shows a comparison between the runtimes of NICS and CRECP method. Clearly, the CRECP method is much faster than the NICS method. Indeed, for the considered benzenoids, the runtime of CRECP does not exceed one minute while NICS may requires several hours. For instance, for the benzenoid 28, NICS takes about 14 hours to compute the local aromaticity, while CRECP only needs 43 seconds.

Then, an important question from a chemical viewpoint is the quality of computed values. Remind that, if both approaches give a description of local aromaticity, CRECP and NICS cannot lead to similar numbers. Indeed, the circuit approach of CRECP is an attempt to describe the behavior of the electronic structure of the molecule as a superposition of closed electronic circuits whereas the NICS approach measures how much the electronic structure would be distorted by an external magnetic field. However, their trends should coincide. Figure 7 presents the values of the local energy of some hexagons for some considered benzenoids. Blue and red values are respectively ones produced by CRECP and NICS. The values of unlabeled hexagons can be deduced by symmetry. As the values produced by the two methods are incomparable in nature, our comparison must focus on the ordering that these values induce on the hexagons. As we can see, both methods lead to similar orderings, what shows that our approach may constitute an interesting and faster alternative to assess the local aromaticity of benzenoid.

6   Conclusions and Perspectives

In this paper, we have presented a new method based on constraint programming for computing the local aromaticity of benzenoids. This method refines the method proposed by Lin by dealing with local aromaticity instead of global one. In practice, we have shown that it turns to be significantly faster than NICS
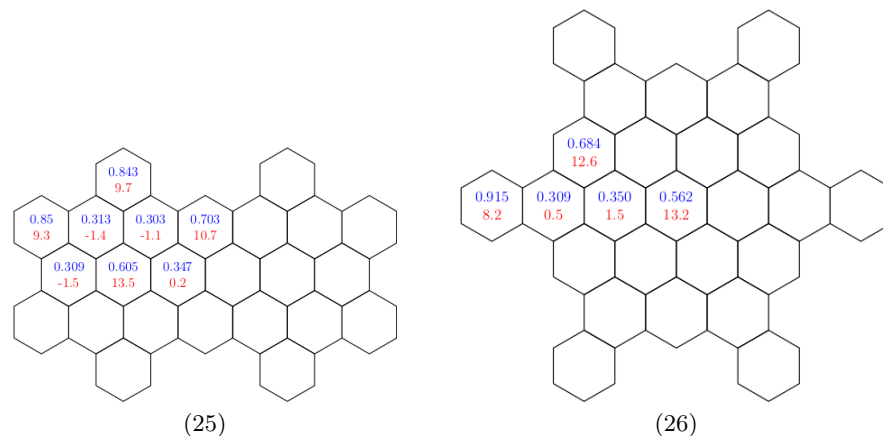
Fig. 7. Results of NICS (red/bottom values) and CRECP (blue/top values) methods.

(which is considered as a reference by theoretical chemists) while providing similar results.

This work is a preliminary step in which we only consider conjugated circuit of size at most 4. By so doing, we make an approximation which may be of poor quality for some classes of benzenoids. An extension of this work consists in identifying all the h-MCC of size greater than 4 and the corresponding union of redundant circuits. If this work was done by hand by Lin [9] for h-MCCs of size at most 4, constraint programming will be of great help here to cope with the combinatorial explosion. Beyond, many problems about benzenoids in theoretical chemistry may fall within the scope of constraint programming. For instance, when no Kekulé structure exists for a given benzenoid, chemists are interested in finding the structure that comes closest to it, what may be expressed as a constraint optimization problem.

References

1. Chen, Z., Wannere, C.S., Corminboeuf, C., Puchta, R., von Ragué Schleyer, P.: Nucleus-Independent Chemical Shifts (NICS) as an Aromaticity Criterion. Chem Rev **105**, 3842–3888 (2005)
2. Clar, E., Schoental, R.: Polycyclic Hydrocarbons Volume 1. Springer Berlin, Berlin (1964)
3. Draine, B.T.: Astronomical Models of PAHs and Dust. EAS Publications Series **46**, 29–42 (2011). https://doi.org/10.1051/eas/1146003
4. Fages, J.G., Lorca, X., Prud'homme, C.: Choco solver user guide documentation. https://choco-solver.readthedocs.io/en/latest/
5. Fages, J.: Exploitation de structures de graphe en programmation par contraintes. Ph.D. thesis, École des mines de Nantes, France (2014)
6. Hansen, P., Zheng, M.: A linear algorithm for perfect matching in hexagonal systems. Discrete Math. **122**, 179–196 (1993)

Table 1. Runtimes of NICS and CRECP for the considered set of benzenoids.

Benzenoid			Runtime	
id.	#carbons	#hexagons	NICS	CRECP
1	6	1	1 min 37 s	0.219 s
2	10	2	5 min 56 s	0.239 s
3	14	3	32 min 38 s	0.257 s
4	14	3	15 min 12 s	0.254 s
5	18	4	21 min 19 s	0.273 s
6	18	4	33 min 50 s	0.276 s
7	18	4	44 min 36 s	0.292 s
8	16	4	16 min 8 s	0.279 s
9	18	4	1 h 5 min 58 s	0.275 s
10	22	5	30 min 13s	0.336 s
11	22	5	1 h 35 min 55 s	0.301 s
12	22	5	1 h 29 min 32 s	0.302 s
13	22	5	59 min 46 s	0.309 s
14	22	5	1 h 42 min 51 s	0.302 s
15	22	5	1 h 45 min 40 s	0.306 s
16	22	5	55 min 47 s	0.329 s
17	22	5	4 h 29 min 7 s	0.339 s
18	20	5	42 min 16 s	0.336 s
19	20	5	1 h 22 min 30 s	0.312 s
20	20	5	24 min 21 s	0.313 s
21	24	7	42 min 42 s	0.454 s
22	42	13	3 h 42 min 54 s	1.929 s
23	60	20	7 h 4 min 58 s	8.912 s
24	60	19	8 h 43 min 33 s	11.382 s
25	72	24	11 h 54 min 4 s	39.825 s
26	78	25	14 h 12 min 44 s	43.655 s
27	36	11	2 h 39 min 5 s	1.044 s
28	54	17	4 h 33 min 25 s	7.05 s

7. Hückel, E.: Quantentheoretische Beiträge zum Benzolproblem. Zeitschrift für Physik **70**, 204–286 (1931)
8. Kekulé, A.: Untersuchungen über aromatische verbindungen ueber die constitution der aromatischen verbindungen. Justus Liebigs Annalen der Chemie **137**(2), 129–196 (1866). https://doi.org/10.1002/jlac.18661370202, https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/jlac.18661370202
9. Lin, C.: Efficient Method for Calculating the Resonance Energy Expression of Benzenoid Hydrocarbons Based on the Enumeration of Conjugated Circuits. J. Chem. Inf. Comput. Sci. **40**, 778–783 (2000)
10. Lin, C., Fan, G.: Algorithms for the Count of Linearly Independent and Minimal Conjugated Circuits in Benzenoid Hydrocarbons. J. Chem. Inf. Comput. Sci **39**, 782–787 (1999)
11. Luch, A.: The Carcinogenic Effects of Polycyclic Aromatic Hydrocarbons. Imperial College Press, London (2005), https://www.worldscientific.com/

worldscibooks/10.1142/p306

12. Mann, M., Thiel, B.: Kekulé Structures Enumeration Yields Unique SMILES. In: Proceedings of Workshop on Constraint Based Methods for Bioinformatics (2013)
13. Narita, A., Wang, X.Y., Feng, X., Müllen, K.: New advances in nanographene chemistry. Chemical Society Reviews **44**(18), 6616–6643 (2015). https://doi.org/10.1039/C5CS00183H
14. Randić, M.: Conjugated circuits and resonance energies of benzenoid hydrocarbons. Chemical Physics Letters **38**, 68–70 (1976). https://doi.org/10.1016/0009-2614(76)80257-6
15. Randić, M.: Aromaticity of Polycyclic Conjugated Hydrocarbons. Chemical Reviews **103**(9), 3449–3606 (2003). https://doi.org/10.1021/cr9903656
16. Randić, M.: Benzenoid Rings Resonance Energies and Local Aromaticity of Benzenoid Hydrocarbons. Journal of Computational Chemistry **40(5)**, 753–762 (2019)
17. Randić, M., Balaban, A.T.: Local Aromaticity and Aromatic Sextet Theory beyond Clar. Int. J. Quantum Chem. **108(17)** (2018)
18. Randić, M., Guo, X.: Recursive Method for Enumeration of Linearly Independent and Minimal Conjugated Circuits of Benzenoid Hydrocarbons. Journal of Chemical Information and Modeling **34(2)** (1994)
19. Randić, M., Guo, X., Klein, D.J.: Analytical Expressions for the Count of LM-Conjugated Circuits of Benzenoid Hydrocarbons. Int. J. Quantum Chem. **60**, 943–958 (1996)
20. Rieger, R., Müllen, K.: Forever young: Polycyclic aromatic hydrocarbons as model cases for structural and optical studies. Journal of Physical Organic Chemistry **23**(4), 315–325 (2010). https://doi.org/10.1002/poc.1644
21. Rispoli, F.J.: Counting perfect matchings in hexagonal systems associated with benzenoids. Mathematics Magazine **14**, 194–200 (2001)
22. Rossi, F., van Beek, P., Walsh, T.: Handbook of Constraint Programming. Elsevier (2006)
23. Régin, J.C.: A filtering algorithm for constraints of difference in CSPs. In: Proceedings of AAAI. pp. 362–367 (1994)
24. Wu, J., Pisula, W., Müllen, K.: Graphenes as Potential Material for Electronics. Chemical Reviews **107**(3), 718–747 (2007). https://doi.org/10.1021/cr068010r