Binary coding of Kekulé structures of catacondensed benzenoid hydrocarbons

Sandi Klavžar,[a,*] Aleksander Vesel,[a] Petra Žigert[a], Ivan Gutman[b]

[a] *Department of Mathematics, PEF, University of Maribor, Koroška 160, 2000 Maribor, Slovenia*

[b] *Faculty of Science, University of Kragujevac, P. O. Box 60, YU–34000 Kragujevac, Yugoslavia*

Received July 2000
Revised January 2001

`Running Title:` Coding Kekulé structures of benzenoids

Key words: benzenoid hydrocarbons, benzenoid graph, resonance graph, Kekulé structure, algorithm, binary coding

telephone number of corresponding author: +386-1-721-04-02

fax number of corresponding author: +386-1-25-172-81

e-mail address of corresponding author: `sandi.klavzar@uni-lj.si`

Abstract

An algorithm is described by means of which the Kekulé structures of a catacondensed benzenoid molecule (with h hexagons) are transformed into binary codes (of length h). By this, computer–aided manipulations with, and memory–storage of Kekulé structures are much facilitated. Any Kekulé structure can easily be recovered from its binary code.

Key words: benzenoid hydrocarbons, benzenoid graph, resonance graph, Kekulé structure, algorithm, binary coding

1. Introduction

The fact that for some polycyclic conjugated molecules (benzenoid hydrocarbons in particular) more than one classical structural formula can be written fascinates chemists longer than a century. These structural formulae are usually referred to as "*Kekulé structures*" and a vast amount of theoretical work on them has been done. One major direction of research is finding the number K of Kekulé structures (Cyvin and Gutman, 1988) and trying to relate K with various physico–chemical properties of the underlying compounds (Gutman and Cyvin, 1989). Another direction is the study of individual Kekulé structures and finding relations between them. A famous problem along these lines (never satisfactorily solved) is the identification of the Kekulé structure that provides the most faithful representation of the true bonding in the respective molecule (Friess, 1927; Friess et al., 1935; Graovac et al., 1973). Some authors consider Kekulé structures as quantum–mechanical objects among which some (quantum–mechanical) interactions exist; for detail see in McWeeny and Sutcliffe (1969). In the case of benzenoid molecules the simplest model for such interactions is the following: if two Kekulé structures differ in the position of three double bonds, then they interact (by an unspecified, but constant amount); otherwise their interaction is neglected. Within this model the concept of the so–called "*resonance graph*" occurs naturally: The vertices of the resonance graph $R(G)$ of the benzenoid hydrocarbon B are the Kekulé structures of B; two vertices of $R(G)$ are adjacent if the corresponding Kekulé structures interact, that is if they differ in the position of just three double bonds.

This model, together with the concept of resonance graphs, was first put forward by Gründler (1982, 1983) and was then re-invented by El–Basil (1993a, 1993b) and, independently, by Randić (Randić et al., 1996, Randić, 1997). In addition to this, without any reference to quantum theory, Zhang, Guo and Chen introduced resonance graphs and established their basic mathematical properties (Zhang et al., 1988a, 1988b; Chen and Zhang, 1997).

For examples illustrating the construction and the basic structural features of resonance graphs see the references quoted as well as Figs. 6, 8, 9 and 10.

Two of the present authors (Klavžar and Žigert, 2000) discovered recently an intriguing property of the resonance graph of catacondensed benzenoid systems, namely that they belong to the class of median graphs. The result was later generalized by Klavžar, Žigert and Brinkmann to the so-called catacondensed even ring systems (Klavžar et. al., 2001). A far–from–obvious consequence of this result is that a unique (and quite short!) binary code can be associated to every Kekulé structure. In this paper we describe an algorithm for the construction of this binary code, as well as for the re-construction of the Kekulé structure from the code.

The advantage of this coding/decoding procedure for computer–aided manipulation with Kekulé structures (in particular, for the construction of the resonance graph) should be obvious. The possibility of storing the complete information on a Kekulé structure by means of a short binary string is another asset that may become indispensable when dealing with benzenoid systems possessing thousands and hundred thousands Kekulé structures (Cyvin and Gutman, 1987). Such benzenoid hydrocarbons have recently been both synthesized (Iyer et al., 1997) and detected in flame (Homann, 1998).

2. Mathematical background of the algorithm

The Cartesian product $G\Box H$ of graphs G and H (Imrich and Klavžar, 2000) has the vertex set $V(G) \times V(H)$, and vertices (a, x) and (b, y) are adjacent in $G\Box H$ whenever $ab \in E(G)$ and $x = y$, or $a = b$ and $xy \in E(H)$.

Hypercubes are of utmost importance for our algorithm and are defined as follows. The vertex set of the *n-cube* Q_n consists of all n-tuples $b_1 b_2 \dots b_n$ with $b_i \in \{0, 1\}$. Two vertices are adjacent if the corresponding tuples differ in precisely one place. Q_n is also called a *hypercube of dimension* n. Note that Q_1 is the complete graph on two vertices, that is, K_2, while Q_2 is the 4-cycle C_4 and Q_3 is the ordinary 8-vertex cube. A useful way to represent Q_n is to take two disjoint copies of Q_{n-1} and join by an edge pairwise identical vertices in the corresponding copies. In other words, $Q_n = Q_{n-1}\Box K_2$.

Recently Klavžar and Žigert (2000) proved that the resonance graph $R(G)$ of a cat-

acondensed benzenoid graph G can be isometrically embedded into the h-dimensional hypercube Q_h, where h is the number of hexagons of G. In fact, they showed an even stronger statement, namely that $R(G)$ is a median graph; for more details on median graphs see elsewhere (Mulder, 1978, 1980; Klavžar and Mulder, 1999; Imrich and Klavžar, 2000).

The proof outlined by Klavžar and Žigert (2000), as well as the one of Klavžar, Žigert and Brinkmann (2001), gives a (theoretical) indication how $R(G)$ can be embedded into Q_h. This embedding can be roughly described as follows.

A 1-factor (or perfect matching) of a graph G on n vertices is a selection of $n/2$ mutually independent edges of G. Only graphs with even number of vertices may have 1-factors, and then every vertex is the endpoint of exactly one of the selected edges. The fact that a Kekulé structure of a conjugated molecule is in a one–to–one correspondence with a 1-factor of the underlying molecular graph is well and long known (e. g., Cvetković et al., 1972). In the following considerations we shall, where appropriate, instead of "Kekulé structure" use the equivalent, but somewhat more precise, mathematical term "1-factor".

Let G be a catacondensed benzenoid graph that is obtained from a catacondensed benzenoid graph G' by attaching a hexagon H over an edge e. Suppose that G' contains $h-1$ hexagons and that we have already embedded $R(G)$ into Q_{h-1}. We partition the 1-factors of G into the sets $\mathcal{F}_e(G)$, $\mathcal{F}^2_{\overline{e}}(G)$, and $\mathcal{F}^3_{\overline{e}}(G)$, where

$\mathcal{F}_e(G)$ consists of the 1-factors containing the edge e;
$\mathcal{F}^2_{\overline{e}}(G)$ consists of the 1-factors not containing e, but containing (exactly) two edges of H;
$\mathcal{F}^3_{\overline{e}}(G)$ consists of the 1-factors not containing e, but containing three edges of H.

Then one can embed $R(G)$ into $Q_h = Q_{h-1} \Box K_2$ in such a way that the 1-factors of $\mathcal{F}_e(G) \cup \mathcal{F}^2_{\overline{e}}(G)$ lie in one copy of Q_{h-1}, the 1-factors of $\mathcal{F}^3_{\overline{e}}(G)$ in the other copy and the 1-factors of $\mathcal{F}_e(G)$ and $\mathcal{F}^3_{\overline{e}}(G)$ are pairwise joined by by an edge.

Let us explain this embedding in some more detail.

For a subset X of vertices of a graph G, the subgraph of G induced by X is

denoted by $G[X]$.

We have already observed that the 1-factors of G, that is the vertices of the resonance graph $R(G)$, can be partitioned as follows:

$$V(R(G)) = \mathcal{F}_e(G) \cup \mathcal{F}^2_{\overline{e}}(G) \cup \mathcal{F}^3_{\overline{e}}(G) .$$

There is a natural bijective correspondence between the 1-factors of $\mathcal{F}_e(G)$ and of $\mathcal{F}_e(G')$. In addition, a 1-factor from $\mathcal{F}^2_{\overline{e}}(G)$ corresponds to a unique 1-factor from $\mathcal{F}_{\overline{e}}(G')$. It follows that the subgraph $G_1 = R(G)[\mathcal{F}_e(G) \cup \mathcal{F}^2_{\overline{e}}(G)]$ of $R(G)$ is isomorphic to $R(G')$. Consider now a 1-factor F from $\mathcal{F}^3_{\overline{e}}(G)$. In $R(G)$ it is adjacent to a unique 1-factor $\overline{F}$ of G_1. We note next that two 1-factors F_1 and F_2 of $\mathcal{F}^3_{\overline{e}}(G)$ are adjacent if and only if the corresponding 1-factors $\overline{F_1}$ and $\overline{F_2}$ are adjacent. Therefore, $R(G)[\mathcal{F}_e(G)]$ and $R(G)[\mathcal{F}^3_{\overline{e}}(G)]$ are isomorphic and $R(G)[\mathcal{F}_e(G) \cup \mathcal{F}^3_{\overline{e}}(G)]$ is a subgraph of $R(G)$ isomorphic to $R(G)[\mathcal{F}_e(G)] \Box K_2$.

For further details of this embedding/proof we refer to Klavžar and Žigert (2000) or/and to Klavžar, Žigert and Brinkmann (2001).

3. The algorithm

In the previous section we pointed out that the resonance graph $R(G)$ of a catacondensed benzenoid system G can be (isometrically) embedded into Q_h. In other words, to each 1-factor of G we can uniquely assign a binary string of length h, such that each bit in the string corresponds to a unique hexagon. In particular, two strings imply the existence of an edge in $R(G)$ if and only if they differ in precisely one position.

Let H be a pendant hexagon of a catacondensed benzenoid graph G. Let G' be the graph obtained from G by removing H (but not e). Suppose that G' contains $h-1$ hexagons and that we have already embedded $R(G')$ into Q_{h-1}. Let $S(G')$ be the set of the binary strings of length $h-1$ corresponding to the embedding of $R(G')$.

In order to establish the embedding of G, we distinguish between three cases with regard to the position of the double bonds, see Fig. 1.

Fig. 1 comes about here

From the discussion in the previous section it follows that 1-factors of the cases a., b., and c. belong to $\mathcal{F}_e(G)$, $\mathcal{F}^2_{\overline{e}}(G)$, and $\mathcal{F}^3_{\overline{e}}(G)$, respectively.

Since $R(G)$ can be embedded into $Q_h = Q_{h-1}\Box K_2$ in such a way that the 1-factors of $\mathcal{F}_e(G) \cup \mathcal{F}^2_{\overline{e}}(G)$ lie in one copy of Q_{h-1} and the 1-factors $\mathcal{F}^3_{\overline{e}}(G)$ in the other copy of Q_{h-1}, the corresponding set of strings of length h can be obtained by concatenating the same digit (say 0) to each $x \in S(G')$ if cases a. and b. occur, and by concatenating the other digit (say 1) to each $x \in S(G')$ if case c. occurs. G' is obviously a catacondensed benzenoid graph, thus the method described above, can be applied as a recursive procedure that can be repeated till a single hexagon remains.

The two 1-factors (Kekulé structures) of a single hexagon (benzene) are shown in Fig. 2. Their set of strings obviously consists of digits 0 and 1, where one of them (say 0) pertains to the 1-factor on the left–hand side and the other (say 1) to the 1-factor at the right–hand side.

Fig. 2 comes about here

In Fig. 3 the three 1-factors (Kekulé structures) of naphthalene are depicted. Based on the procedure described above they are coded with the set of strings $S(G) = \{00, 01, 10\}$ as shown in the figure.

Fig. 3 comes about here

First, the right digit is assigned as 0, 0, and 1, respectively, according to the cases b., a., and c. of Fig. 1. Then, in G', that is, in the left hexagon, the remaining edges correspond to the right, to the left, and to the left case of Fig. 2. Hence the first digits are 1, 0, and 0, respectively.

However, the recursive procedure can be awkward if the graph G has more hexagons. Therefore, an iterative procedure would be more convenient. In order to develop such a procedure we first examine catacondensed benzenoids without branches. Two hexagons of a catacondensed benzenoid have either one common edge (and are then said to be *adjacent*) or have no common vertices. A *hexagonal chain* is an unbranched catacondensed benzenoid graph, that is, the catacondensed system in which no hexagon is adjacent to three other hexagons.

The inner dual of a hexagonal chain G with h hexagons is obviously the h-vertex path. Starting from a terminal hexagon H_1 we can assign consecutive numbers to each hexagon to get the ordering $H_1, H_2, \ldots, H_h$.

Let S_i be the set of strings corresponding to the embedding of the resonance graph that corresponds to a hexagonal chain consisting of the hexagons $H_1, H_2, \ldots, H_i$ and suppose that S_{i-1} is already known. Then the determination of S_i depends solely on the previous two hexagons H_{i-1} and H_{i-2}. The six possible cases are shown in Fig. 4.

Fig. 4 comes about here

According to the positions of hexagons H_{i-2}, H_{i-1}, and H_i we arrive at two major cases. The hexagons may be linearly arranged, see Fig. 4a or not (then a so called *kink* occurs), see Fig. 4b.

Furthermore, the labeling also depends on the position of the double bonds of H_{i-1}. Let $x \in S_{i-1}$. Then $x = y0$ or $x = y1$, where y is a pertinent string of length $i-2$.

According to the cases in Fig. 4 we get:

a. (Cases without a kink). S_i can be obtained from S_{i-1} by inserting strings $x0$ and $x1$ for each $x = y1$ and by inserting string $x0$ for each $x = y0$.

b. (Cases with a kink). S_i can be obtained from S_{i-1} by inserting strings $x0$ and $x1$ for each $x = y0$ and by inserting string $x0$ for each $x = y1$.

We are now ready to present the algorithm for the embedding of the resonance graph of s hexagonal chain G with at least two hexagons. Since two previous hexagons are needed at each step of the procedure, we start the algorithm with the embedding of naphthalene, which was already determined to be $\{00, 01, 10\}$. The algorithm at i-th step builds the new set of strings of length i (denoted by S') obtained from the strings of length $i-1$ (denoted by S).

Procedure LABELS(G);
begin
1. Consecutively number the hexagons of G to get the sequence H_1, H_2, ..., H_h;
$S := \{00, 01, 10\}$;
For $i := 3$ to h **do begin**
$S' := \emptyset$;
If H_i induces a kink **then begin**
For each $x \in S$ **do begin**
$S' := S' \cup \{x0\}$;
If $x = y0$ **then** $S' := S' \cup \{x1\}$; **end**;
end
else begin
For each $x \in S$ **do begin**
$S' := S' \cup \{x0\}$;
If $x = y1$ **then** $S' := S' \cup \{x1\}$; **end**;
end
$S := S'$;
end;
end.

Remark. In order to determine whether H_i induces a kink, the positions of H_i, H_{i-1}, and H_{i-2} have to be examined.

The procedure for a branched catacondensed benzenoid is a little (but not much!) more involved. The inner dual of a catacondensed benzenoid graph G is an h-vertex tree T. An arbitrary pendant hexagon of G is selected as the root of this tree. The hexagons of G are then numbered such that H_i is a predecessor of H_j in T if and only if $i < j$. Such a numbering can be obtained by the Depth–first search algorithm (DFS) or by the Breadth–first search algorithm (BFS) (Kozen, 1992). Note that the predecessor of every hexagon has to be known in order to determine whether H_i induces a kink or not.

Based on the discussion above, Step 1 of LABELS should be replaced by the following:

> Starting at any pendant hexagon apply the Depth-first search (DFS) algorithm to consecutively visit the hexagons of the graph. At each visited hexagon do the following:
>
> > assign a consecutive index to the hexagon,
> > record a predecessor of the hexagon H_i.

4. Two examples

In this section we present two examples of the algorithm described in the previous section. In addition, at the end we also show the reverse procedure, namely how one can construct the Kekulé structure that corresponds to a given binary code.

As the first example we consider a simple case of the resonance graph of naphthacene. The hexagons are numbered in a natural way as shown in Fig. 5.

Fig. 5 comes about here

In Fig. 6 the complete construction of the Kekulé structures, resonance graphs, and the codes for the Kekulé structures are shown, starting from the data for naphthalene.

Fig. 6 comes about here

In our second example we consider a more complex benzenoid hydrocarbon – benzo[s]picene, see Fig. 7.

Fig. 7 comes about here

In Fig. 8 the first part of the procedure is shown: the generation of the two–, three– and four–digits codes for the vertices of the resonance graphs and the corresponding Kekulé structures.

Fig. 8 comes about here

By considering the fifth hexagon of benzo[s]picene, we get the five–digits codes. These codes and the corresponding Kekulé structures are shown in Fig. 9.

Fig. 9 comes about here

Finally, considering the sixth hexagon, we derive altogether 22 Kekulé structures, consistent with the fact (Cyvin and Gutman, 1987) that $K(\text{benzo[s]picene}) = 22$. These are too numerous to be drawn. Instead, their codes are given in Table 1 whereas the corresponding resonance graph is shown Fig. 10.

Table 1 comes about here

Fig. 10 comes about here

At this point the natural question is how one can reconstruct the Kekulé structure from its binary code. This can be done just by reverting the algorithm. Of course, an input data for such a procedure is also the selected ordering of hexagons with which the codes were constructed.

Instead of describing the reverse procedure in all details, consider, for example, the Kekulé structure F_6 from Table 1. Its binary code is 100101. Recalling the ordering of hexagons from Fig. 7, we can construct the Kekulé structure corresponding to 100101 as it is shown in Fig. 11.

Fig. 11 comes about here

Acknowledgements

This work was supported in part by the Ministry of Science and Technology of Slovenia under the grant 101–504.

References

Chen, R., Zhang, F., 1997, Discr. Appl. Math. 74, 191.

Cvetković, D., Gutman, I., Trinajstić, N., 1972. Chem. Phys. Lett. 16, 614.

Cyvin, S. J., Gutman, I., 1988. Kekulé Structures in Benzenoid Hydrocarbons. Springer–Verlag, Berlin.

El–Basil, S., 1993a. J. Math. Chem. 14, 305.

El–Basil, S., 1993b. J. Mol. Struct. (Theochem) 288, 67.

Friess, K., 1927. Liebigs Ann. Chem. 454, 121.

Friess, K., Walter, R., Schilling, K., 1935. Liebigs Ann. Chem. 516, 248.

Graovac, A., Gutman, I., Randić, M., Trinajstić, N., 1973. J. Am. Chem. Soc. 95, 6267.

Gründler, W., 1982. Wiss. Z. Univ. Halle 31, 97.

Gründler, W., 1983. Z. Chem. (Leipzig) 23, 157.

Gutman, I., Cyvin, S. J., 1989. Introduction to the Theory of Benzenoid Hydrocarbons. Springer–Verlag, Berlin.

Homann, K. H., 1998. Angew. Chem. Int. Ed. 37, 2435.

Imrich, W., Klavžar, S., 2000. Product Graphs: Structure and Recognition. Wiley, New York.

Iyer, V. S., Wehmeier, M., Brand, J. D., Keegstra, M. A., Müllen, K., 1997. Angew. Chem. Int. Ed. 36, 1604.

Klavžar, S., Mulder, H. M., 1999. J. Combin. Math. Combin. Comp. 30, 103.

Klavžar, S., Žigert, P., 2000. Preprint Ser. Univ. Ljubljana IMFM 38, 1.

Klavžar, S., Žigert, P., Brinkmann, G., 2001. Discr. Appl. Math., to appear.

Kozen, D. C., 1992. The Design and Analysis of Algorithms. Springer–Verlag, New York.

McWeeny, R., Sutcliffe, B. T., 1969. Methods of Molecular Quantum Mechanics. Academic Press, London.

Mulder, H. M., 1978. Discrete Math. 24, 197.

Mulder, H. M., 1980. The Interval Function of a Graph. Math. Centrum, Amsterdam.

Randić, M., 1997. Int. J. Quantum Chem. 63, 585.

Randić, M., Klein, D. J., El–Basil, S., Calkins, P., 1996. Croat. Chem. Acta 69, 1639.

Zhang, F., Guo, X., Chen, R., 1988a. Discr. Math. 72, 405.

Zhang, F., Guo, X., Chen, R., 1988b. Acta Math. Appl. Sin. 4, 131.

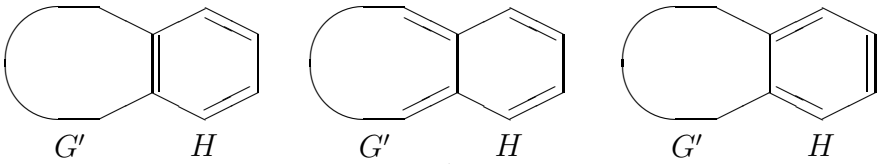

Figure 1: Possible arrangements of the double bonds in the hexagon H .

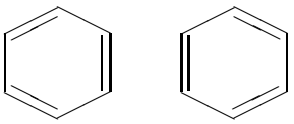

Figure 2: The two Kekulé structures of benzene.

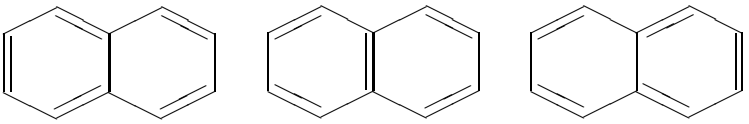

Figure 3: The three Kekulé structures of naphthalene.

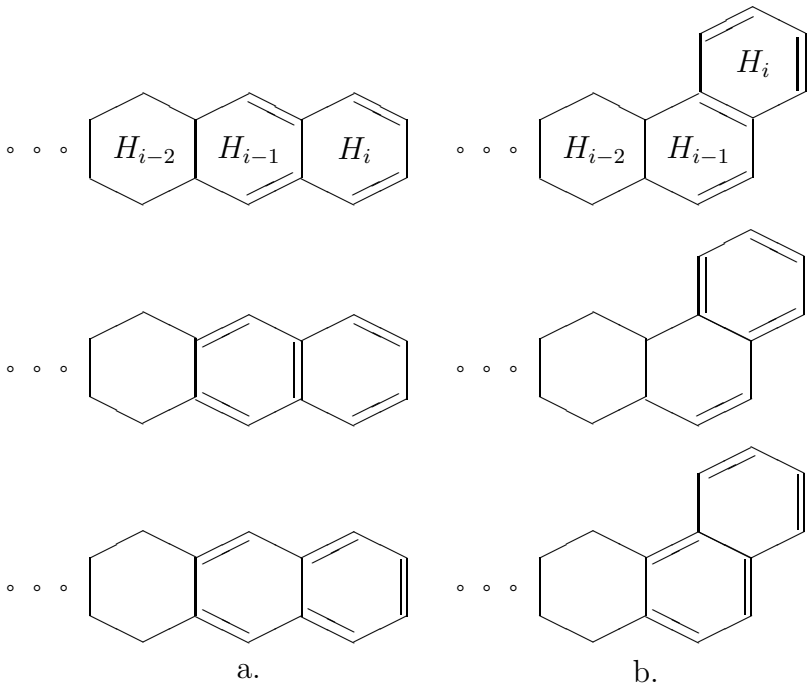

Figure 4: Labelling of three consecutive hexagons.

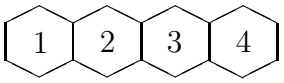

Figure 5: Naphthacene and the labelling of its hexagons.

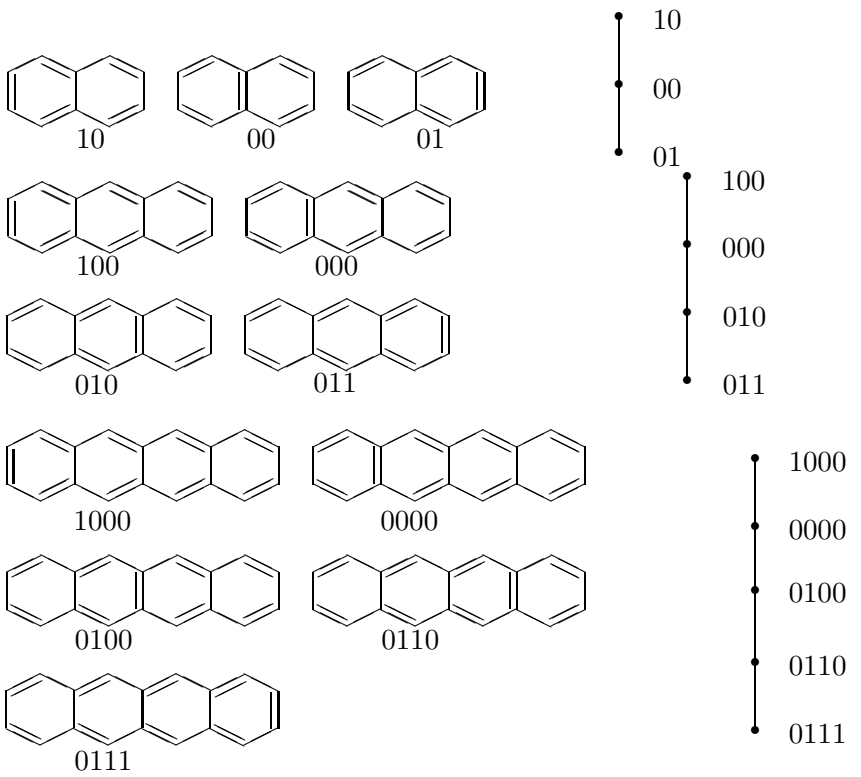

Figure 6: Kekulé structures, resonance graphs and vertex codes of naphthalene, anthracene, and naphthacene, cf. Fig. 5.

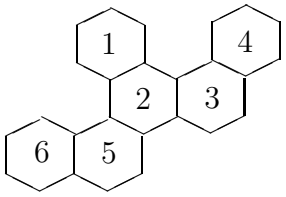

Figure 7: Benzo[s]picene and the labelling of its hexagons.

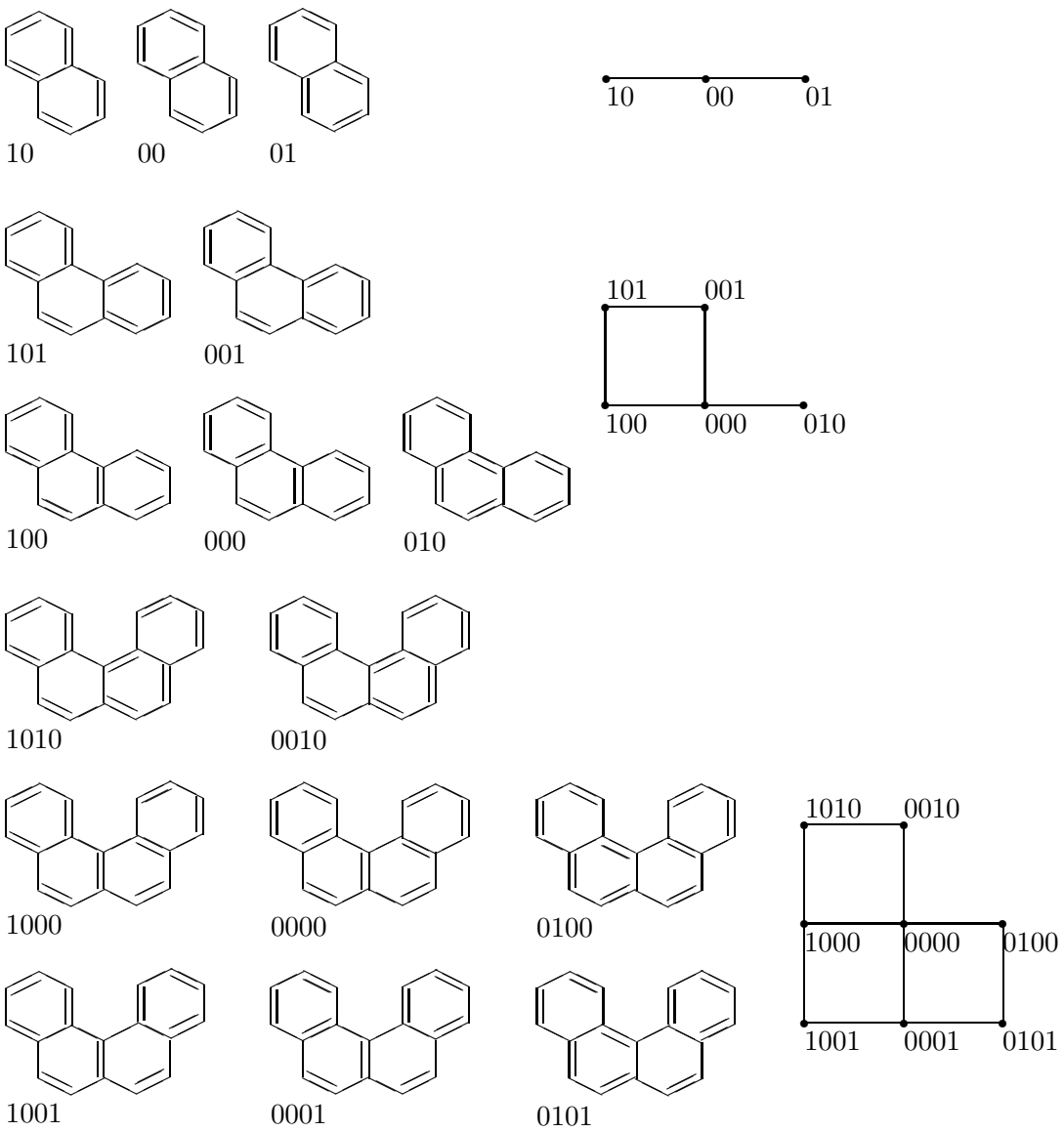

Figure 8: Kekulé structures, resonance graphs, and vertex codes of naphthalene, phenanthrene and benzo[c]phenanthrene; note that these consist of the first two, first three and first four hexagons, respectively, of benzo[s]picene, cf. Fig. 7.

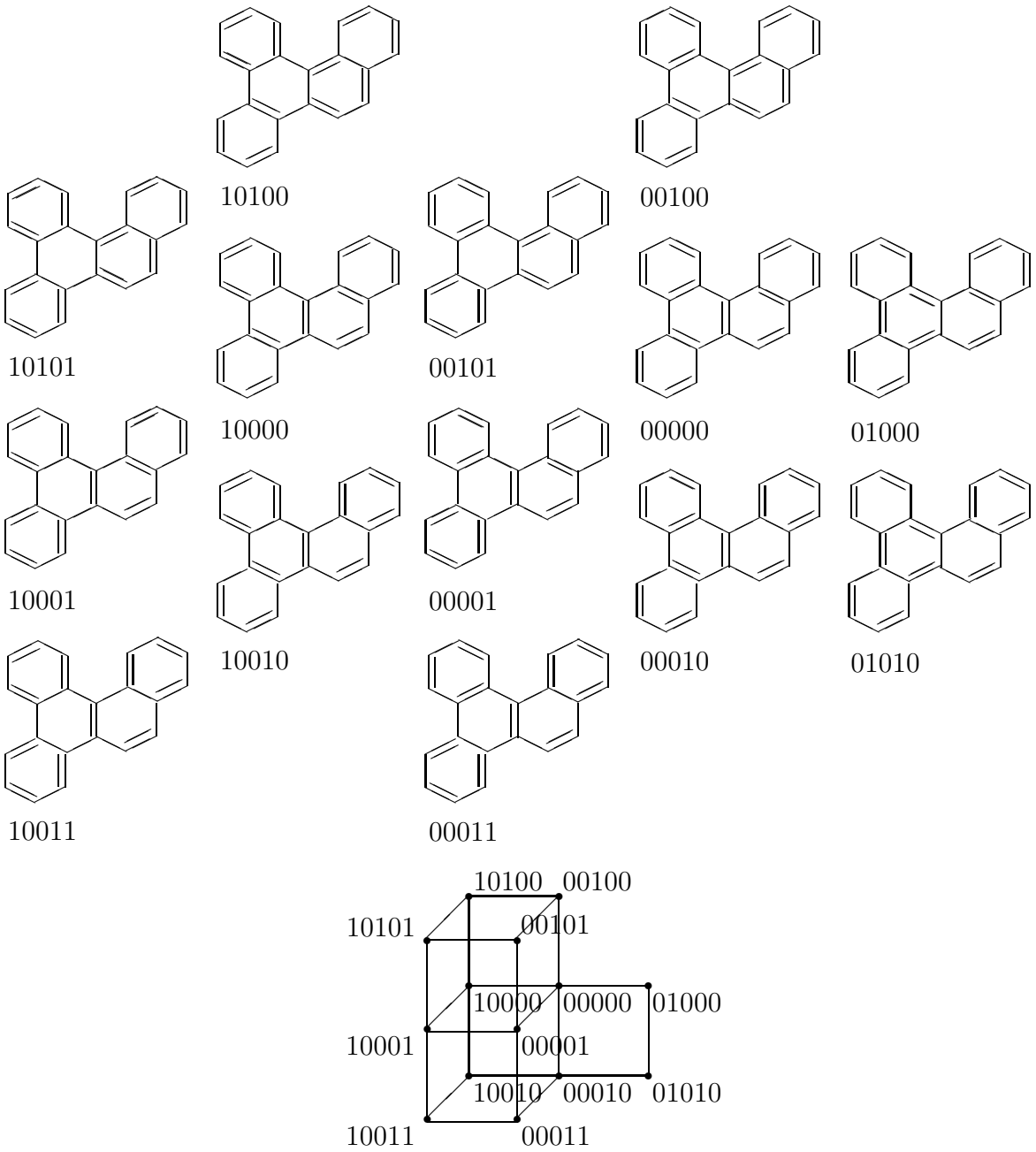

Figure 9: Kekulé structures, resonance graphs and vertex codes of benzo[g]chrysene; note that it consists of the first five hexagons of benzo[s]picene, cf. Fig. 7.

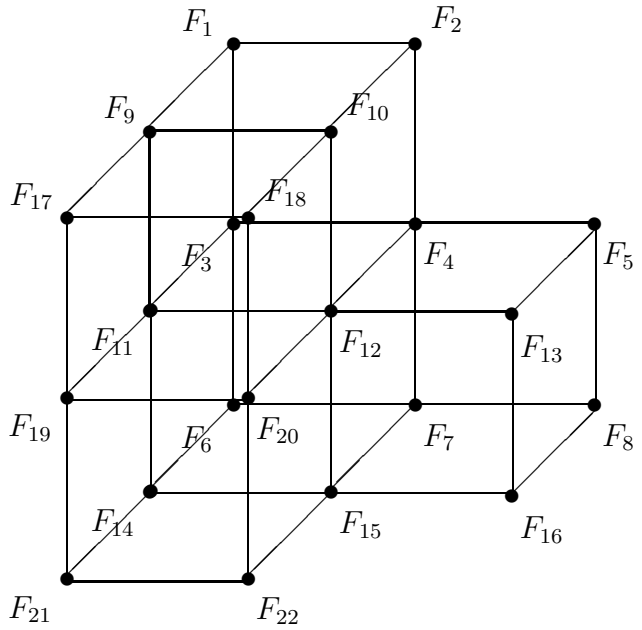

Figure 10: The resonance graph of benzo[s]picene, cf. Fig. 7.

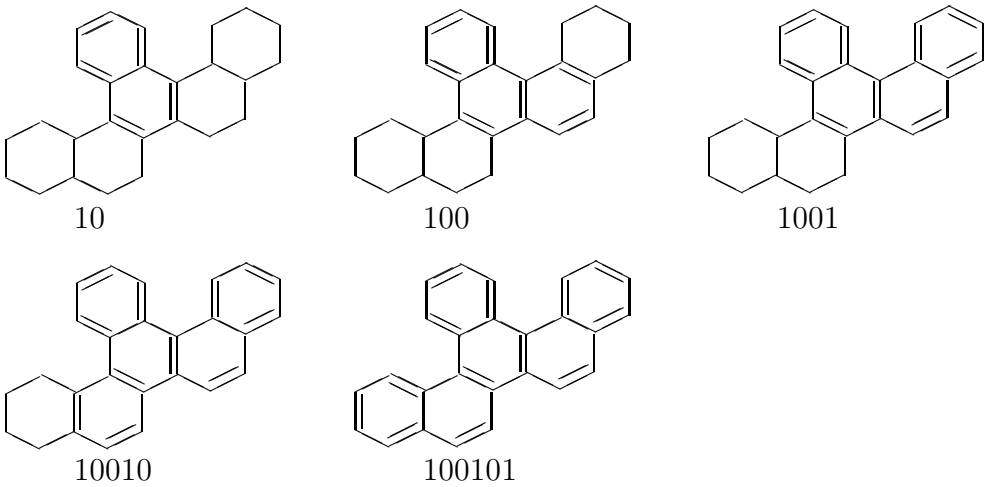

Figure 11: Constructing a Kekulé structure of benzo[s]picene from the respective binary code, cf. Fig. 7.

Kekulé structure	code	Kekulé structure	code	Kekulé structure	code
F_1	101001	F_9	101000	F_{17}	101010
F_2	001001	F_{10}	001000	F_{18}	001010
F_3	100001	F_{11}	100000	F_{19}	100010
F_4	000001	F_{12}	000000	F_{20}	000010
F_5	010001	F_{13}	010000	F_{21}	100110
F_6	100101	F_{14}	100100	F_{22}	000110
F_7	000101	F_{15}	000100		
F_8	010101	F_{16}	010100		

Table 1: The binary codes of the Kekulé structures of benzo[s]picene, cf. Figs. 7 and 10.

Figure and Table Captions

Fig. 1. Possible arrangements of the double bonds in the hexagon H.

Fig. 2. The two Kekulé structures of benzene.

Fig. 3. The three Kekulé structures of naphthalene.

Fig. 4. Labelling of three consecutive hexagons.

Fig. 5. Naphthacene and the labelling of its hexagons.

Fig. 6. Kekulé structures, resonance graphs and vertex codes of naphthalene, anthracene, and naphthacene, cf. Fig. 5.

Fig. 7. Benzo[s]picene and the labelling of its hexagons.

Fig. 8. Kekulé structures, resonance graphs, and vertex codes of naphthalene, phenanthrene and benzo[c]phenanthrene; note that these consist of the first two, first three and first four hexagons, respectively, of benzo[s]picene, cf. Fig. 7.

Fig. 9. Kekulé structures, resonance graphs and vertex codes of benzo[g]chrysene; note that it consists of the first five hexagons of benzo[s]picene, cf. Fig. 7.

Fig. 10. The resonance graph of benzo[s]picene, cf. Fig. 7.

Fig. 11. Constructing a Kekulé structure of benzo[s]picene from the respective binary code, cf. Fig. 7.

Table 1. The binary codes of the Kekulé structures of benzo[s]picene, cf. Figs. 7 and 10.