

Counterexamples to a proposed algorithm for Fries structures of benzenoids

Patrick W. Fowler · Wendy Myrvold ·
William H. Bird

Received: 1 April 2012 / Accepted: 4 June 2012 / Published online: 22 June 2012
© Springer Science+Business Media, LLC 2012

Abstract The *Fries number* of a benzenoid is the maximum number of benzenoid hexagons over all of its Kekulé structures (perfect matchings), and a *Fries canonical structure* is a perfect matching that realises this maximum. A recently published algorithm claims to determine Fries canonical structures of benzenoids via iterated Hadamard products based on the adjacency matrix (Ciesielski et al. in *Symmetry* 2:1390–1400, 2010). This algorithm is re-examined here. Convergence is typically rapid and often yields a single candidate perfect matching, but the algorithm can give an exponential number of choices, of which only a small number are canonical. More worryingly, the algorithm is found to give incorrect results for the Fries number for some benzenoids with as few as seven hexagonal faces. We give a combinatorial reformulation of the algorithm in terms of linear combinations of perfect matchings (with weights at each stage proportional to the products of weights of the edges included in a matching). In all the cases we have examined, the algorithm converges to a maximum-weight matching (or combination of maximum-weight matchings), and where the algorithm fails, either no best Fries matching is of maximum weight, or a best Fries matching is of maximum weight but a sub-optimal matching of the same weight is chosen.

Keywords Benzenoids · Fries structure · Clar structure · Kekule structure · Perfect matching · Hadamard product · Graph algorithms · Pauling Weight

P. W. Fowler (✉)
Department of Chemistry, University of Sheffield, Sheffield S3 7HF, UK
e-mail: P.W.Fowler@sheffield.ac.uk

W. Myrvold · W. H. Bird
Department of Computer Science, University of Victoria, Victoria,
BC V8W 3P6, Canada

1 Introduction

Benzenoid hydrocarbons form an important subset of π -electron systems, with a vast literature concerning their physical, chemical and biological properties. Historically, their study has played a central role in the development of theories of organic chemistry, in particular in connection with the concept of aromaticity. This background is well reviewed in Ref. [1], where it is pointed out that two related ideas, those of Fries structures [2] and Clar Sextets [3] have helped to give a qualitative understanding of π -electronic structure and stability of benzenoids and similar compounds. Both are based on Kekulé structures of the molecule (perfect matchings of the molecular graph G).

The early work of Fries [2] associated stability with the existence of Kekulé structures that achieve large numbers of benzenoid hexagons, i.e., hexagons that contain three double bonds (not necessarily switchable). The cardinality of the maximum set of benzenoid hexagons over all Kekulé structures is the Fries number, $F(G)$.

Clar's more sophisticated version of this idea [3] was that stability is conferred by the existence of Kekulé structures with large numbers of hexagonal rings with three resonating double bonds (i.e., the number of Clar sextets). On the Clar view, the figure of merit is the Clar Number, $C(G)$, i.e., the cardinality of the maximum *independent* set of hexagons, each carrying three switchable double bonds. Structures that share the maximum number of isolated Clar sextets may then be ranked more finely in stability using other empirical considerations [1,4].

Both $F(G)$ and $C(G)$ can be found exactly by several methods, e.g., by one approach that takes linear time and uses linear programming [5–7]. In Ref. [1], the authors present a matrix algorithm based on Hadamard products that is intended to find what they term a 'Fries canonical structure' for benzenoid hydrocarbons i.e., a perfect matching that contains the maximum number $F(G)$ of benzenoid hexagons. Convergence is typically rapid and the algorithm often yields a single candidate perfect matching, but the algorithm sometimes gives an exponential number of choices, of which only a small number are canonical. The authors of the algorithm do not suggest a way of dealing with this problem, other than brute-force inspection, which can require exponential time for some quite simply defined families of benzenoids, as we will show below. A more serious problem is that the algorithm does not always deliver a Fries canonical structure, again as will be shown below.

Canonical matchings are proposed in Ref. [1] both as targets of interest in themselves, and as intermediates in the generation of Clar structures with the maximum number of sextets. An implicit assumption [1] is that a Kekulé structure with $F(G)$ benzenoid hexagons can always be used to generate a structure with $C(G)$ Clar sextets by selection of a maximum subset of independent hexagons.

In the present paper, we re-examine the procedure described in Ref. [1], which we will call the CKC algorithm, test it on complete sets of small benzenoids, and show that it does not always lead to a canonical Fries structure, a fact which at least casts doubt on its proposed use for finding Kekulé structures that realise the Clar number. The smallest (i.e., those with fewest hexagons) counterexamples to the algorithm have seven hexagons. A combinatorial reformulation is then used to show that the algorithm can be understood as working with convex combinations of perfect matchings, with

the weight of each matching computed from products of edge weights of included double-bond edges, and where the edge weights change at each iteration of the algorithm. In all the cases of failure of the algorithm that we have examined, it is converging to a matching of maximum weight matching or a combination of such matchings, where either a canonical Fries matching is not of maximum weight, or is of maximum weight but is not the one chosen by CKC.

2 The CKC algorithm

The authors of Ref. [1] do not give an explicit statement of their working definition of the class of benzenoids, but we assume that they intend to include at least the molecules that correspond to simple planar 2-connected graphs embedded in the plane such that all internal faces are hexagons, all vertices not on the external face have degree 3 and vertices on the external face have degree 2 or 3. This will be our working definition of benzenoids. We will show that the algorithm sometimes fails even within this restricted definition. The algorithm in Ref. [1] uses the Hadamard product [8]. The Hadamard product of two matrices \mathbf{A} and \mathbf{B} , indicated by a circle, is the matrix \mathbf{H} with entries $h_{u,v}$ that are the products of the corresponding entries in \mathbf{A} and \mathbf{B} , i.e.,

$$\mathbf{H} = \mathbf{A} \circ \mathbf{B} \Leftrightarrow h_{u,v} = a_{u,v} b_{u,v}. \quad (1)$$

The CKC algorithm is then defined as follows [1]. For a given benzenoid with n vertices (carbon centres), first construct the $n \times n$ adjacency matrix, $\mathbf{A} = \mathbf{P}_0$, with $a_{u,v} = 1$ if vertices u and v are connected by an edge, and 0 otherwise. Then for $k = 1, 2 \dots$ the matrix \mathbf{P}_k is defined by

$$\mathbf{P}_k = \mathbf{P}_{k-1} \circ \mathbf{P}_{k-1}^{-1}. \quad (2)$$

It is assumed in Ref. [1] that the procedure will converge to some limiting matrix \mathbf{P}_∞ , and in practice convergence was found to be rapid for the examples tested there and for the many cases that we have examined (see below).

This mysterious-looking procedure has some obvious properties.

- (i) The construction can proceed only if \mathbf{A} has an inverse, i.e., if the molecular graph is non-singular. Appearance of a singular matrix \mathbf{P}_k at any later stage will also terminate the process.
- (ii) The starting adjacency matrix \mathbf{A} is symmetric, so \mathbf{A}^{-1} is also symmetric, and therefore \mathbf{P}_1 is symmetric. Hence, each subsequent iterate \mathbf{P}_k is symmetric in the sense that the entries in positions u, v and v, u are equal. Furthermore, the matrix \mathbf{P}_k is also symmetric in the sense in that it is transformed into itself under all automorphisms of the graph, a property that it inherits from \mathbf{A} .
- (iii) The nature of the Hadamard product is that each multiplicand acts as a mask for the other, leading to a zero entry in the product matrix if and only if there is a zero entry in that position in at least one of the matrices being multiplied. In the first iteration, the adjacency matrix has, by definition, non-zero entries in positions corresponding to bonds (edges of the molecular graph) and nowhere

else. Hence \mathbf{P}_1 and all subsequent \mathbf{P}_k have the property that non-zero entries do not occur for non-bonded pairs u, v . Likewise, all \mathbf{P}_k matrices have zero entries on the diagonal.

The entries in the matrix \mathbf{A}^{-1} for a non-singular benzenoid graph have a well-understood chemical significance. In chemistry, the Pauling Bond Order [9] of a bond (u, v) is the fraction of the full set of Kekulé structures of the graph in which that edge (u, v) carries a double bond. This is equal to the ratio $m(G-u-v)/m(G)$ where $m(G)$ is the number of perfect matchings of G and $m(G-u-v)$ is the number of perfect matchings of the graph with vertices u and v and their incident edges deleted. The entry in \mathbf{A}^{-1} for a pair of vertices that are connected by an edge is equal to the Pauling Bond Order [1]. The entries for pairs of non-adjacent vertices are, apart from sign, also given by the ratio $m(G-u-v)/m(G)$, which is then known as the Pauling Long Bond Order [10], a quantity that plays a part in the theory of molecular conduction of graphene fragments [11]. The matrix \mathbf{P}_1 for a benzenoid therefore has entries that are equal to the Pauling Bond Orders for each edge pair, and otherwise are zero. At each iteration, the matrix \mathbf{P}_k ($k > 0$), if it exists, is a symmetric, weighted form of the adjacency matrix of G , with each row sum equal to 1. (For a justification of this, see the combinatorial reformulation of the algorithm described below.)

It is not obvious from this formulation that a non-singular \mathbf{P}_0 guarantees the existence of a sequence of \mathbf{P}_k matrices with $k > 0$ for benzenoids, as in general the Hadamard product of non-singular matrices may be singular. Nor is it clear that the iterative process will always converge to a limit. Our alternative formulation (see later) shows that neither of these potential problems can occur for benzenoids. However, application of the CKC algorithm to non-benzenoid bipartite graphs could easily lead to the first problem: for example, one exotic isomer of naphthalene has a molecular graph consisting of a square fused to an octagon; the graph is non-singular, but \mathbf{P}_1 is singular, and the iteration breaks down immediately.

In discussing the algorithm, the authors of Ref. [1] utilise an adjacency matrix \mathbf{K}_i for each perfect matching $i = 1, \dots, m(G)$ of a graph G : the matrix \mathbf{K}_i has entries 1 in positions (u, v) and (v, u) for all pairs connected by a double bond in the i th matching, and 0 elsewhere. Each matrix \mathbf{K}_i is symmetric, self-inverse and has a determinant $(-1)^{n/2}$, where n is the number of vertices in the graph. The assumption is made [1] that, for a non-singular starting benzenoid, the iteration process will always converge to a \mathbf{K} matrix that represents a perfect matching with $F(G)$ simultaneously benzenoid hexagons, or to a linear combination of \mathbf{K} matrices from which a perfect matching with $F(G)$ simultaneously benzenoid hexagons can be deduced. This assumption is in fact incorrect (see below).

The CKC algorithm is not difficult to program. Some experimental observations from the small number of calculations described in Ref. [1] and calculations with our own programs on datasets including all Kekulean benzenoids with at most 13 hexagonal faces (1,671,018 graphs, generated with CaGe [12]) are listed below.

- (i) In all the cases examined, the iterative process converges smoothly within a few iterations. (For example, $n/2$ iterations are sufficient to give rms average deviation per element between successive Hadamard product matrices of at most 10^{-12} for cases with up to 9 hexagons.) Convergence is immediate in

some cases, with $\mathbf{P}_1 = \mathbf{P}_2$, as for benzene (see below). The determinant of \mathbf{P}_k evolves during the iterative process: for benzenoids, the initial determinant is $|\det \mathbf{A}| = |\det \mathbf{P}_0| = [m(G)]^2$, which is typically a large number, but the Hadamard product, \mathbf{P}_1 , has a determinant similar to that of \mathbf{A}^{-1} , i.e., of magnitude $\sim [m(G)]^{-2}$, which can be perilously close to zero for some classes of benzenoid [13]. The final determinant has a magnitude that depends on the case, as explained below.

- (ii) The entries in the converged Hadamard product are all drawn from the set $\{1, 1/2, 0\}$. For all the benzenoids examined, the converged \mathbf{P}_∞ has the same general structure: it has entries $1/2$ corresponding to the edge sets of h disjoint ‘ambiguous’ hexagons (for some $h \geq 0$), and entries 1 for a further set of $n/2 - 3h \geq 0$ edges that are disjoint from each other and from the hexagon edges. The entries 1 and $1/2$ are interpreted as fixed double bonds and potential double bonds, respectively.
- (iii) The determinant of \mathbf{P}_∞ is $(-1)^{n/2}/2^{4h}$, which arises by taking the overall product of $n/2 - 3h$ blocks of determinant -1 , and h blocks of determinant $-1/16$ (each hexagon has six edges of weight $1/2$ that can be factored into 2 different matchings). In the vast majority of benzenoids tested, the limiting case $h = 0$ is encountered, and \mathbf{P}_∞ is the \mathbf{K} matrix for a perfect matching of G . The other limiting case, where $h = n/6$, seems to be much less common. The spectrum of \mathbf{P}_∞ consists of eigenvalues $+1, +1/2, -1/2, -1$ with multiplicities $n/2 - 2h, 2h, 2h, n/2 - 2h$.
- (iv) The inverse of \mathbf{P}_∞ has entries $+1$ in all the positions where \mathbf{P}_∞ has a non-zero entry, and in addition it has entries -1 for those pairs that correspond to the three diagonals of each of the h ambiguous hexagons. It is easy to see that this structure implies that taking the Hadamard product of \mathbf{P}_∞ and its inverse will simply recover \mathbf{P}_∞ , and hence that the process has converged.

Figure 1 illustrates 3 cases satisfying, $h = n/6$, $h = 0$ and $0 < h < n/6$, giving a pictorial interpretation of the converged matrix \mathbf{P}_∞ and its inverse. The first possibility, where all vertices belong to ambiguous hexagons is illustrated by benzene, which has an adjacency matrix of circulant form, with $a_{u,v} = 1$ for $u - v = 1 \pmod 6$, and 0 otherwise. The inverse, \mathbf{A}^{-1} , has entries $1/2$ for $u - v = 1 \pmod 6$, 0 for $u - v = 2 \pmod 6$, $-1/2$ for $u - v = 3 \pmod 6$, and 0 otherwise (or, in chemical nomenclature, $1/2, 0, -1/2$ for ortho, meta and para pairs). The Hadamard product \mathbf{P}_1 has entries $+1/2$ for all ortho pairs, and zero otherwise. The inverse \mathbf{P}_1^{-1} has entries $+1$ for ortho pairs, -1 for para pairs, and zero otherwise. Convergence is therefore reached on the first cycle. Small benzenoids with $h = n/6$ are rare. Apart from benzene itself, no examples exist for benzenoids with 12 or fewer hexagons. Figure 2 shows the only other example with $h = n/6$ that we have found, the 13-hexagon hexabenzocoronene.

The second row in Fig. 1 illustrates the second limiting case with the molecular graph of naphthalene, which has the more typical behaviour of rapid but not immediate convergence of the Hadamard product to a \mathbf{K} matrix. (Entries in the bond-order matrix converge to either 1 or 0.) For example, the four classes of symmetry-distinct edges labelled a, b, c, d in Fig. 1b bear entries $1/3, 2/3, 1/3, 1/3$. During the iterations, these follow trajectories

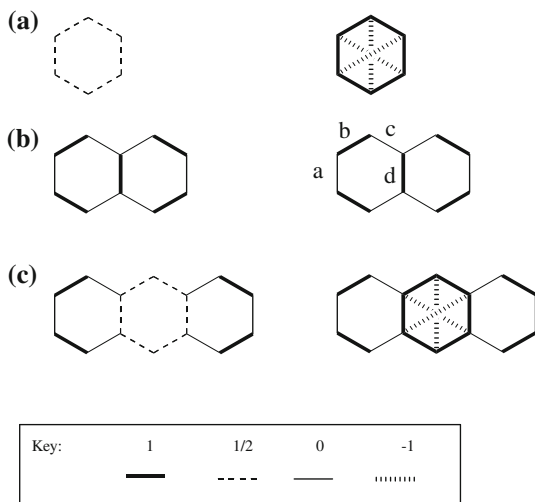


Fig. 1 Three types of outcome of the CKC algorithm for a benzenoid. Rows **a**, **b** and **c** illustrate cases with $h = n/6$ (in this case, $h = 1$), $h = 0$ and $0 < h < n/6$, where h is the number of hexagons with all edge weights equal to $1/2$ (the ‘ambiguous’ hexagons). The pictures represent the converged Hadamard product (left) and its inverse (right), with line style indicating values of entries (see Key). Vertices joined by *hatched lines* are not adjacent in the graph. As can be seen from the positions of $+1$ (bold) edges, each converged Hadamard product matrix is self-replicating under Hadamard multiplication with its inverse. Letters *a*, *b*, *c*, *d* on the naphthalene structure refer to the discussion in the text

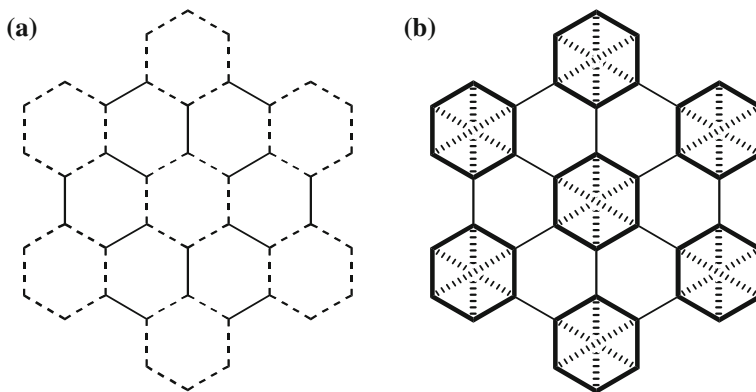


Fig. 2 **a** Molecular graph of hexabenzocoronene, an example where in the limit the vertices of the graph are covered by a set of hexagons with edge entries equal to $1/2$. **b** In the inverse matrix, the entries are $+1$ for the edges of those hexagons and -1 for their diagonals

$$\begin{aligned}
 a : & 1/3 \rightarrow 1/6 \rightarrow 1/102 \rightarrow 1/1020102 \rightarrow \dots \rightarrow 0; \\
 b : & 2/3 \rightarrow 5/6 \rightarrow 101/102 \rightarrow 1020101/1020102 \rightarrow \dots \rightarrow 1; \\
 c : & 1/3 \rightarrow 1/6 \rightarrow 1/102 \rightarrow 1/1020102 \rightarrow \dots \rightarrow 0; \\
 d : & 1/3 \rightarrow 2/3 \rightarrow 100/102 \rightarrow 1020100/1020102 \rightarrow \dots \rightarrow 1;
 \end{aligned}$$

which can be understood in detail in terms of the combinatorial reformulation of the algorithm discussed below. For naphthalene, the limit is the matrix for the unique fully symmetric perfect matching shown in Fig. 1b, which realises $F(G) = 2$.

The third row in Fig. 1c illustrates an intermediate case ($0 < h < n/6$) with the molecular graph of anthracene. The proposed [1] interpretation of \mathbf{P}_∞ is that the matrix is the average of \mathbf{K} matrices for two perfect matchings (see Fig. 3), each of which realises the Fries number of 2 for this linear polyacene graph.

A point which is not discussed in detail in Ref. [1] is how to construct a Fries structure from a Hadamard product when $h \neq 0$. Each hexagon with six entries $1/2$ leads to two possible assignments of formal double bonds within that face, implying an exponential algorithm with 2^h choices to determine $F(G)$, if indeed $F(G)$ can be reached in this way. Kekulene (Fig. 4) is an example where many of the 2^h choices turn out not to yield the full Fries number; in this case only 2 of the 64 choices give $F(G)$. Maximum values of h for the benzenoids (obtained by the direct search to be described in the next section), and hence worst cases for the exponential algorithm are given in Table 1. It is easy to construct infinite series in which h grows linearly with n : one such is shown in Fig. 5, where each increase of n by 10 adds an extra ambiguous hexagon to the CKC result, but a correct canonical Fries structure is obtained for only 2 of the 2^h possible assignments of double bonds within the ambiguous hexagons. Without some way to avoid the exponential step, the CKC algorithm would rapidly become impractical for general sets of larger benzenoids.

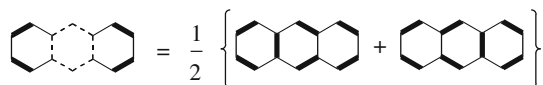


Fig. 3 A Hadamard product matrix resulting from the CKC algorithm. Here $h = 1$, and this is interpreted as a superposition of two Kekulé structures [1]

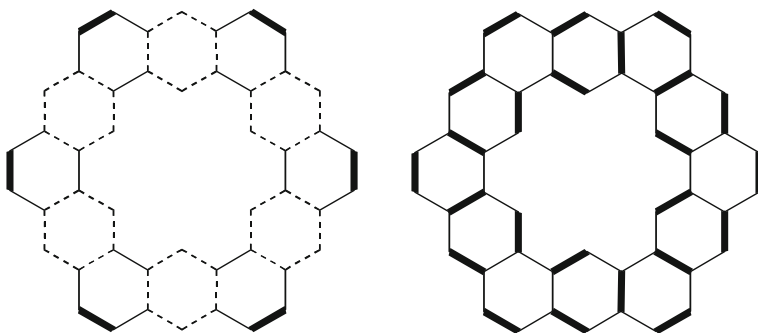


Fig. 4 The converged Hadamard product matrix resulting from application of the CKC algorithm to kekulene is shown on the left. Here $h = 6$, and there are 64 perfect matchings that can be constructed by assignment of three double bonds within each dotted hexagon. Of the 64 choices: 2 give 6 benzenoid hexagons, 30 give 7, 30 give 8 and only 2 give the 9 benzenoid hexagons corresponding to a Fries canonical structure. One such choice is shown on the right; the other can be found by a concerted switch of bonds in all formerly dotted hexagons

Table 1 Applications of the CKC algorithm to benzenoids with H hexagonal faces: h_{\max} is the largest number of faces with entries $1/2$ for all edges appearing in the converged structures

H	1	2	3	4	5	6	7	8	9	10	11	12	13
h_{\max}	1	0	1	2	1	2	3	3	4	4	4	5	7

Note that kekulene (Fig. 4), with 12 hexagonal faces and $h = 6$, is not a benzenoid on the present definition, as it has a hole

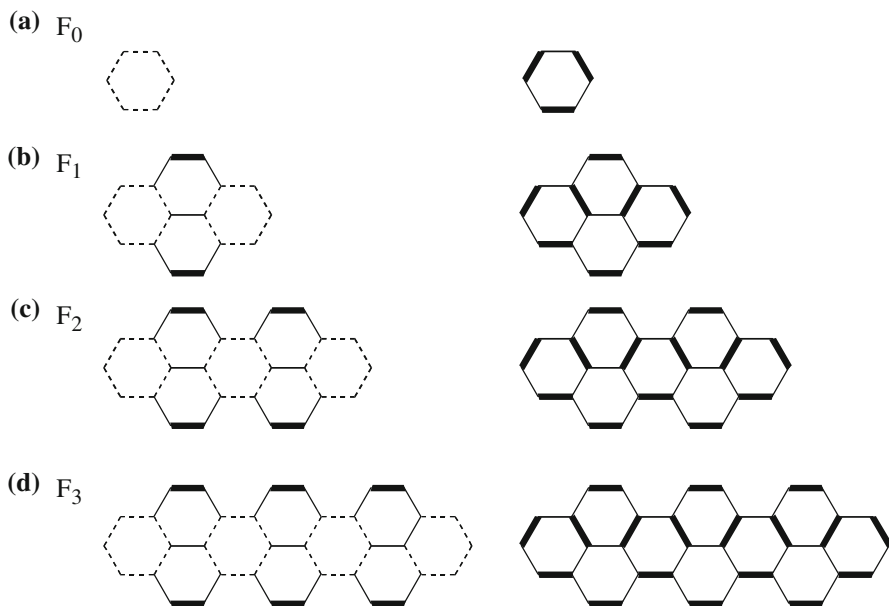


Fig. 5 An infinite family series illustrating the worst-case exponential nature of the CKC algorithm. The graph F_{k+1} is created by fusion of a phenylene fragment onto the perimeter of F_k , as shown. F_k has $3k + 1$ hexagonal faces, and 2×3^k perfect matchings, and the CKC algorithm selects a structure with $k + 1$ ambiguous hexagons and $2k$ double bonds, as shown on the *left*. The Fries number is $2k + 1$, which is realised by exactly 2 of the 2^{k+1} assignments of double bonds in the hexagons, one of which is shown on the *right*. The cost of brute-force checking of all possibilities is therefore exponential

3 Does the CKC algorithm always give the Fries number?

For the CKC algorithm, we have not found any instance of non-convergence for Kekulean benzenoids, but the question remains as to whether, when the algorithm converges, it always gives a correct result. From this point of view, failure of the CKC algorithm might occur in one of two ways. First, the algorithm might lead to a unique perfect matching that has fewer than $F(G)$ simultaneous benzenoid hexagons. Secondly, the algorithm might lead to a Hadamard product matrix that generates 2^h choices of perfect matchings, with $h \neq 0$, where no choice realises the Fries number of the given benzenoid. In fact, failures of both kinds occur. A search of small benzenoids was made. For each graph, each matching realising the Fries number was checked to see whether it had the property that its edges formed a subset of the edges with non-zero

Table 2 Failures of the CKC algorithm

H	1	2	3	4	5	6	7	8	9	10	11	12	13
N_K	1	1	2	6	15	51	190	764	3,223	14,107	62,879	284,918	1,304,861
N_F	0	0	0	0	0	0	3	13	80	421	2,034	10,451	52,065
N_0	0	0	0	0	0	0	3	12	79	419	2,005	10,283	50,693
N_h	0	0	0	0	0	0	0	1	1	2	29	168	1,372

N_K is the number of Kekulean benzenoids with H hexagonal faces, and N_F is the number of cases where the algorithm fails to deliver the correct Fries number. In the final two rows, N_F is broken down into N_0 , the number of cases where the algorithm gives a unique matching with incorrect Fries number, and N_h , the number of cases where the algorithm gives two or more choices of matchings, all with incorrect Fries number

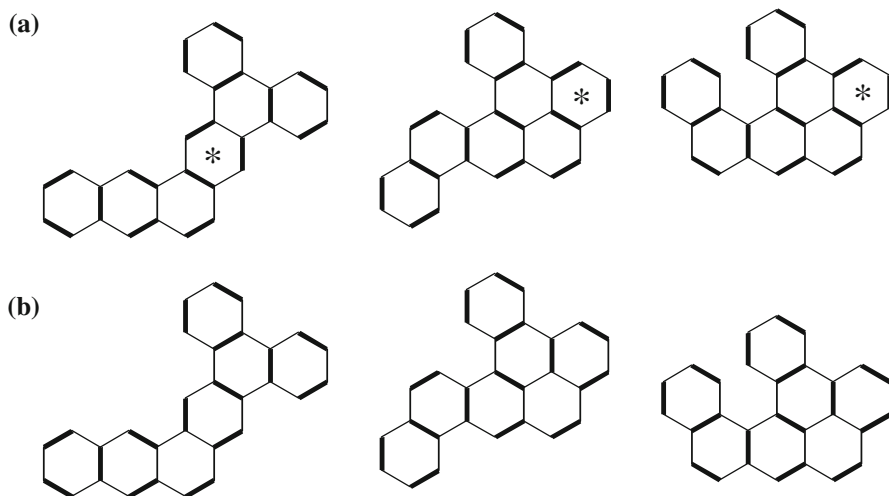


Fig. 6 The three smallest counterexamples to the CKC algorithm. These are all of the first kind, in which CKC converges to a single sub-optimal perfect matching. **a** The converged perfect matchings for three 7-hexagon benzenoids produced by the algorithm. Each has only 5 benzenoid hexagons. **b** Fries canonical matchings for the same benzenoids, showing that $F(G) = 6$ can be achieved in each case. The star shows how the sub-optimal result of the CKC algorithm could be converted to optimal by an ad hoc rotation of the bonds within the starred hexagon

weight given by the CKC algorithm. This search revealed the smallest counterexamples to the conjecture that the CKC algorithm for a benzenoid always yields a Fries structure.

Table 2 gives some statistics for the search. It can be seen that the algorithm fails in a small proportion of cases, from seven hexagons onwards, and that, at least initially, the great majority of those failures are of the first kind, where the converged matrix represents a single perfect matching with fewer than $F(G)$ benzenoid hexagons. Figure 6 shows the three smallest examples of failure of the algorithm. In each case, \mathbf{P}_∞ has $h = 0$ and has five benzenoid hexagons, but the true value is $F(G) = 6$. Figure 7 shows the two smallest examples of failures of the second kind, those where

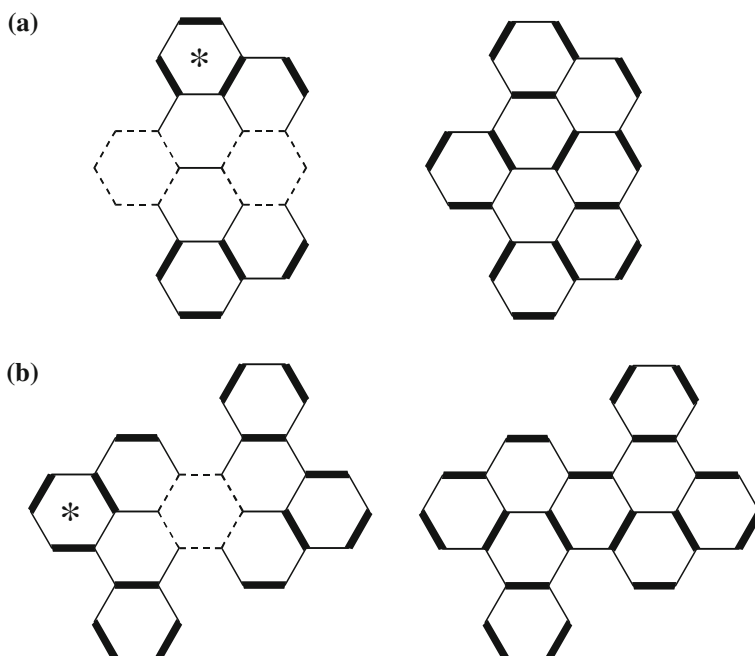


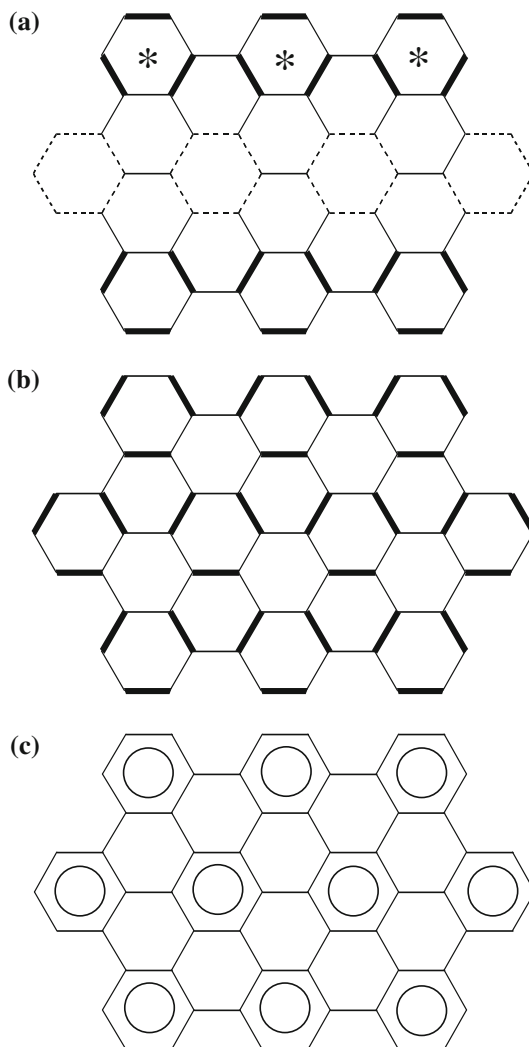
Fig. 7 The two smallest counterexamples of the second kind (with $h \neq 0$ hexagons composed of half-double bonds) to the CKC algorithm. **a** The smallest benzenoid counterexample with $h = 2$ has a total of 8 hexagonal faces (*left*). All four assignments of the ambiguous hexagons yield 5 benzenoid hexagons, but (*right*) the correct total of $F(G) = 6$ benzenoid hexagons can be achieved by reversing the CKC choice of fixed double bonds in the *starred hexagon*. **b** The counterexample with $h = 1$ has 9 hexagonal faces, and $F(G) = 7$, whereas the CKC algorithm finds only 6 benzenoid hexagons (*left*). Ad hoc reassignment of fixed double bonds in the *starred hexagon* would again be needed to achieve the correct Fries number

h is non-zero; in each case the best assignment of the ambiguous hexagons recovers only $F(G) - 1$ benzenoid hexagons. Figure 8 shows a highly symmetrical counterexample where the shortfall between the CKC result and the true Fries number is 3. This molecular graph is also interesting because it is a leapfrog benzenoid [14], and hence totally resonant (i.e., has Clar number $C = n/6$, so that every vertex is present in one sextet). The leapfrog is constructed from a parent benzenoid graph by placing a new $\pi/6$ -rotated hexagon within each hexagonal face, adding an edge to cross at right angles any parent edge that is common to two hexagonal faces, and deleting all edges and vertices of the parent. A leapfrog benzenoid has a single canonical Clar structure in which a sextet is located on each face that arose from a parent face. In the case of the graph in Fig. 8, the parent is ovalene.

4 Combinatorial reformulation of the CKC algorithm

Given that we now know that the CKC algorithm does not invariably produce a perfect matching or linear combination of perfect matchings that yields $F(G)$ benzenoid hexagons, a question arises: given that the algorithm appears always to converge, at

Fig. 8 A symmetrical counterexample to the CKC algorithm. **a** The converged CKC structure has full D_{2h} symmetry, which is broken on assigning the four ambiguous hexagons, to give, at best, C_{2v} symmetry and 12 benzenoid hexagons. **b** The Fries canonical structure also has C_{2v} symmetry. (Even with best possible assignment of the ambiguous hexagons of the CKC structure, reassignment of fixed double bonds in either the three *starred hexagons* of the *top row* in **(a)** or their symmetry equivalents in the *bottom row* is required to achieve the full count of $F(G) = 15$). **c** The Clar structure for this leapfrog benzenoid, with $C(G) = 10$



least for Kekulean benzenoids, what characterises the matchings or combinations of matchings that it does produce?

The progression of the CKC algorithm for benzenoids through matrices \mathbf{P}_0 , $\mathbf{P}_1, \dots, \mathbf{P}_k$ can be understood in terms of iterative refinement of convex combinations of perfect matchings. Given a set of k matrices $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_k$ and k probabilities p_1, p_2, \dots, p_k ($0 \leq p_i \leq 1$) such that $p_1 + p_2 + \dots + p_k = 1$, the matrix

$$\mathbf{M} = \sum_{i=1}^k p_i \mathbf{M}_i \quad (3)$$

is called a *convex combination* of the matrices $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_k$. If, as before, $\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_{m(G)}$ are the $m(G)$ matrices corresponding to the perfect matchings of G , then each matrix \mathbf{P}_k of the CKC algorithm for $k \geq 1$ corresponds to a convex combination of the \mathbf{K}_i matrices computed as follows. In computing \mathbf{P}_{k+1} from \mathbf{P}_k , the weight w_i of matching M_i is the product

$$w_{M_i} = \prod_{(u,v)} p_{uv}^{(k)}, \quad (4)$$

where (u, v) is an edge of the matching and $p_{uv}^{(k)}$ is the entry in \mathbf{P}_k corresponding to that edge. It is possible to prove for benzenoids that \mathbf{P}_{k+1} is then the convex combination

$$\mathbf{P}_{k+1} = \sum_{i=1}^{m(G)} \frac{w_i}{S} \mathbf{K}_i \quad (5)$$

where S is the scaling factor used to convert the weights to probabilities, i.e.,

$$S = \sum_{i=1}^{m(G)} w_i. \quad (6)$$

The proof of the exact equivalence step-by-step of this combinatorial formulation with the Hadamard-matrix formulation is given in the “[Appendix](#)”.

The iteration begins with $\mathbf{P}_0 = \mathbf{A}$, which gives equal weight to all edges, and produces as the first guess a linear combination in which all perfect matchings appear with equal probability $1/m(G)$. The matrix \mathbf{P}_1 coincides with the matrix of Pauling Bond Orders for edges and has zero entries elsewhere, i.e., exactly as given by the Hadamard product of \mathbf{A} and \mathbf{A}^{-1} . The new convex combination of matchings has probabilities proportional to products of Pauling Bond Orders. We call these products, for reference, the *Pauling Weights* (see Fig. 9 for examples). The convex combination determines the new edge weights, and so on. At convergence, the edge weights and the weights of the matchings are self-consistent, each set leading to the other. From this description it is easy to see why the CKC algorithm converges to matrices with row sums equal to 1.

The key characteristic of benzenoid graphs that makes it possible to reformulate the Hadamard matrix method in these combinatorial terms is that in the expansion of the determinant of a weighted adjacency matrix (e.g. a matrix \mathbf{P}_k), the sign of every term is the same and is equal to $(-1)^{n/2}$. It is this property that provides the well-known relationship between the adjacency matrix and the number of perfect matchings of a benzenoid [15, 16] ($\det \mathbf{A} = (-1)^{n/2} [m(G)]^2$), and is also the reason that $\mathbf{P}_1 = \mathbf{A} \circ \mathbf{A}^{-1}$ gives the Pauling Bond Orders. It also leads to a single sign for all weights in the convex combination of \mathbf{K} matrices. A sufficient condition for bipartite graphs for all terms in the determinant to have uniform sign is that for each cycle C such that the graph $G-C$ has a perfect matching, C is an aromatic cycle, i.e., $|C| = 4k + 2$. This *cycle condition* is obeyed by several wider classes of graphs,

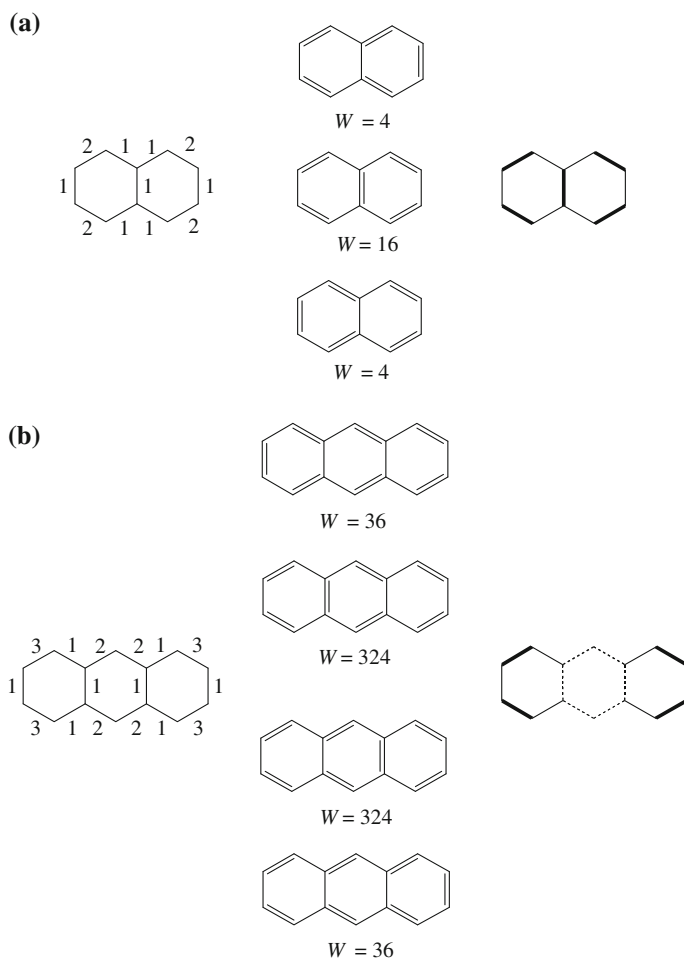


Fig. 9 Weightings of perfect matchings and the results of the CKC algorithm: **a** naphthalene; **b** anthracene. In each case, the *left column* shows the molecular graph annotated with edge weights proportional to Pauling Bond Orders. The *centre column* shows the full set of perfect matchings (Kekulé structures) with their relative *Pauling Weights*. The *righthand column* shows the final result of the CKC algorithm, which picks out the unique perfect matching of maximum Pauling Weight in **(a)** and a linear combination of the pair of equivalent perfect matchings with maximum Pauling Weight in **(b)**

including hexagonal systems with $4k + 2$ -sized ‘holes’, such as kekulene. Where this cycle condition is not obeyed, the CKC Hadamard-matrix algorithm can have terms of different sign in the expansion of the determinants, and the equivalence of the two formulations is destroyed (see Fig. 10). The combinatorial approach appears to us to be more likely to provide chemically relevant results in these cases because it avoids cancellations in the weights.

An implementation of this combinatorial version of the algorithm was made, and was shown to give complete agreement with the Hadamard formulation at all stages of the iteration, as it should. Explicit construction of matrix inverses is not needed, nor is the Hadamard masking procedure of the previous formulation. However, the

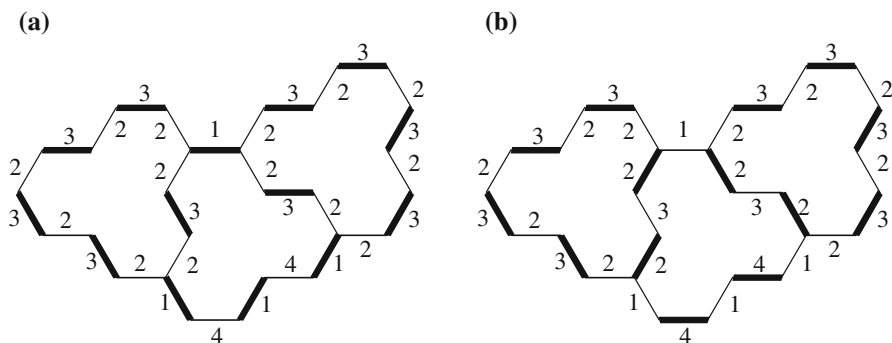


Fig. 10 Divergence of the combinatorial and Hadamard versions of the CKC algorithm. This hypothetical structure is formally constructed by fusing three phenalene units and deleting their central atoms to make ‘holes’ of size 12. The structure is labeled with relative Pauling Bond orders. The perfect matchings produced by application of **a** the Hadamard version and **b** the combinatorial version differ in Pauling Weights by a factor of 9:256 and are matchings of minimum and maximum weight, respectively. The matching in **(b)** arising from the combinatorial procedure uniquely preserves conjugation in all three of the phenalene units

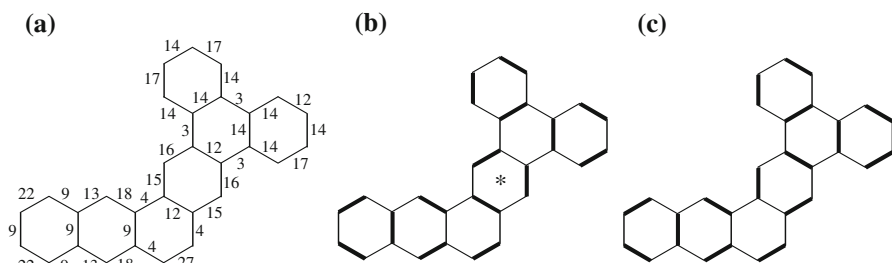


Fig. 11 Weight analysis for a counterexample to the CKC algorithm. This 7-hexagon benzenoid is the first counterexample from Fig. 6. It has edge weights as marked in **(a)**, and gives matching in **(b)** as the result of the CKC algorithm (5 benzenoid hexagons), but can achieve $F = 6$ benzenoid hexagons, as shown in **(c)**. The two choices of fixed double bonds in the *starred hexagon* give Pauling Weights for the respective matchings that are in ratio $12 \times 16 \times 16:12 \times 15 \times 15$, with CKC choosing the matching with the larger, and in fact maximum, product of bond orders

combinatorial version of the method requires construction of the complete set of perfect matchings of G , which unfortunately can take exponential time. The point of the second method is, however, not computational efficiency, but clarity of interpretation, especially in the case of understanding counterexamples to the CKC method.

A plausible conjecture, based on our examination with the new program of the complete sets of benzenoids described earlier, is that the end result of the CKC algorithm for a Kekulean benzenoid graph is either a perfect matching of maximum Pauling Weight for the particular graph, or an average structure representing the superposition of a set of perfect matchings of maximum Pauling Weight. Figure 11 analyses the first counterexample to the CKC procedure from this point of view. In this case, there is no perfect matching of maximum Pauling Weight that realizes the Fries number, and

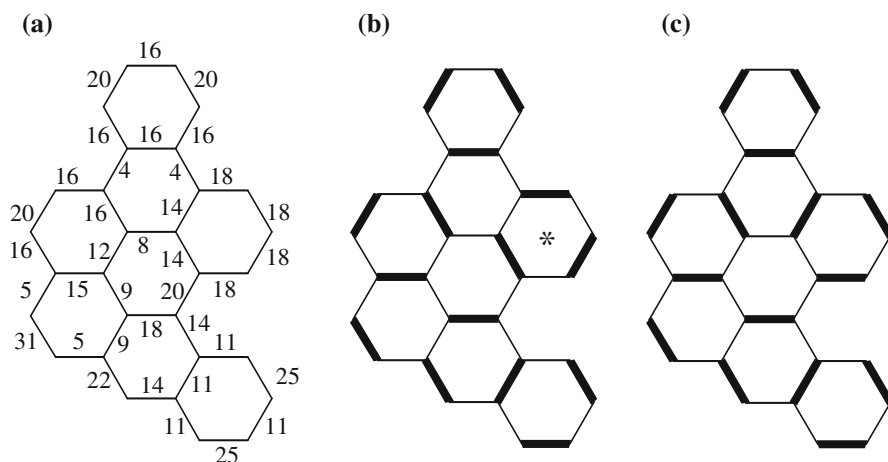


Fig. 12 Weight analysis for a counterexample to the CKC algorithm. This 8-hexagon benzenoid has edge weights as marked in (a), and gives the matching in (b) as the result of the CKC algorithm (5 benzenoid hexagons), but $F = 6$ benzenoid hexagons can be achieved, as shown in (c). *Rotation* of the matched edges of the *starred hexagon* inter-converts the matchings but does not change the Pauling Weight, as the relevant edge weights are 14, 18 and 18 in both cases. Both matchings achieve the maximum Pauling Weight for this benzenoid, but CKC converges to the matching with fewer benzenoid hexagons

the CKC result corresponds to a matching with the maximum weight rather than maximum number of benzenoid hexagons. As the Figure shows, switching the matched edges in one hexagon would simultaneously lower the Pauling Weight and increase the number of benzenoid hexagons. A more interesting example is given in Fig. 12, where there is a Fries canonical structure that has the maximum Pauling Weight, but the CKC algorithm fails to find it, choosing instead a matching that is also of maximum Pauling Weight but has only $F(G) - 1$ benzenoid hexagons. This separation between matchings that start out the iterative process with equal weight is understandable; the weight of a matching in subsequent iterations depends not only on its own Pauling Weight but also on the weights of the other matchings with which it has edges in common. As we have shown, this process can lead to convergence to a result that does not maximize the Fries number.

5 Conclusions

The CKC algorithm, although it often finds Fries canonical structures, sometimes fails. It seems that the algorithm emphasizes weight of perfect matchings, in the sense defined above, over the number of benzenoid hexagons they contain. There are reliable alternatives for determination of both Fries and Clar numbers of benzenoids [7]. As we have shown, the Hadamard matrix and combinatorial formulations are equivalent for benzenoids as defined here, although it could be argued that the latter gives a more transparent interpretation. A sufficient condition for equivalence of the formulations is the ‘sign condition’ set out in the previous section. For more general Kekulean

graphs, the combinatorial formulation is still applicable, because it treats the signs of the matrix entries correctly, and hence avoids the singularities and cancellations that would be introduced by naïve application of the Hadamard matrix method. An example of a hypothetical conjugated structure for which the two formulations are inequivalent is shown in Fig. 10; in this case the combinatorial method yields a chemically plausible matching of maximum Pauling Weight, with fully conjugated phenalene circuits, whereas the Hadamard formulation converges to an unlikely-looking matching that is actually of *minimum* weight.

Finally, we return to the proposed [1] use of Fries canonical structures to find one or more perfect matchings that realise the Clar number of the graph. Our computer search revealed that for small benzenoids (with up to 13 hexagons), every matching realising the Fries number had at least one choice for independent benzenoid hexagons realising the Clar number. An implicit assumption in Ref. [1] is that a Clar structure could always be determined from a canonical Fries structure (or at least, from one of the type produced by the CKC algorithm). In order to evaluate this approach to Clar structures further, it would be interesting to determine whether the following two conjectures are true or not:

Conjecture 1 For all benzenoids G , there exists some perfect matching with $F(G)$ benzenoid hexagons for which there is a choice of $C(G)$ independent hexagons.

Conjecture 2 For all benzenoids G , for every perfect matching with $F(G)$ benzenoid hexagons there is a choice of $C(G)$ independent hexagons

Acknowledgments WM thanks NSERC for a Discovery Grant. PWF thanks Prof. Barry T. Pickup (University of Sheffield) for useful discussions. Claire Theuret (École Nationale de Chimie Physique et Biologie de Paris) is thanked for assisting with calculations as part of an undergraduate project.

Appendix

This appendix explains why the combinatorial approach proposed in the main text is equivalent to the CKC algorithm for benzenoids. It also provides greater insight about what the CKC algorithm is actually computing. Use of a determinant computation to determine the number of matchings of a benzenoid is a standard approach (see, e.g., Lovasz and Plummer's classic textbook on graph matching theory [16]). A combinatorial explanation is included here, as it is required for understanding the combinatorics of the CKC algorithm.

One definition of the determinant of an $n \times n$ matrix \mathbf{A} is as

$$\det(\mathbf{A}) = \sum_{\pi} \text{sign}(\pi) a_{1,\pi_1} a_{1,\pi_2}, \dots, a_{1,\pi_n} \quad (\text{A.1})$$

where the sum is over all permutations π of 1 to n . If the matrix \mathbf{A} is the adjacency matrix of a benzenoid, it is well known that the non-zero terms in this expansion all have the same sign (equal to $(-1)^{n/2}$). The edges $(1, \pi_1), (2, \pi_2), \dots, (n, \pi_n)$ induce a 2-regular subgraph that can be interpreted as the union of an ordered pair of perfect

matchings. The components of this 2-regular subgraph are all even cycles of the benzenoid. If both (u, v) and (v, u) are included then this is interpreted as a component which is a 2-cycle.

To find the ordered pair of perfect matchings (M_1, M_2) corresponding to a non-zero term in the determinant, first select the smallest numbered vertex of each cycle. If u is the smallest vertex of a cycle and the determinant expansion includes the term $a_{u,v}$, then place edge (u, v) in the matching M_1 . To complete the assignment of edges to M_1 and M_2 , alternate between the two matchings on each cycle. This shows that in the determinant expansion of the adjacency matrix of a benzenoid, the non-zero terms are in one-to-one correspondence to ordered pairs of perfect matchings. The non-zero entries of the adjacency matrix are all ones and the sign for each is $(-1)^{n/2}$. Hence for a benzenoid, the number of perfect matchings is the square root of the absolute value of the determinant of the adjacency matrix. For graphs that are not benzenoids, this approach to enumerating matchings does not always work because the signs of the terms can differ and can result in some cancellation of terms. (An example of this is given in Fig. 10).

If the initial matrix corresponds to one of the \mathbf{P}_k matrices, then each edge (u, v) can be considered to have a weight $w_{u,v}$ equal to the u, v entry in \mathbf{P}_k , i.e., $p_{uv}^{(k)}$. Define the *weight* of a perfect matching M to equal the product of the weights of its edges, and denote it by w_M . The determinant of \mathbf{P}_k for a benzenoid is equal to

$$(-1)^{n/2} \sum_{(M_1, M_2)} w_{M_1} w_{M_2} \quad (\text{A.2})$$

where the sum is over all ordered pairs (M_1, M_2) of perfect matchings.

Jacobi's theorem states that

$$\mathbf{A}^{-1} = \text{Adj}(\mathbf{A}) / \det(\mathbf{A}) \quad (\text{A.3})$$

where entry i, j of $\text{Adj}(\mathbf{A})$ is equal to

$$(-1)^{i+j} \det(\mathbf{A}^{[j,i]}) \quad (\text{A.4})$$

where $\mathbf{A}^{[j,i]}$ is matrix \mathbf{A} with row j and column i removed. In considering the application of this result to the matrix \mathbf{P}_k , note that the only important entries of $(\mathbf{P}_k)^{-1}$ for our purpose are those which correspond to edges in the original graph since all others are zeroed out when taking the Hadamard product.

The non-zero terms that contribute to $\mathbf{A}^{[u,v]}$ for an edge (u, v) can be interpreted as subgraphs consisting of an ordered pair of matchings M_1 and M_2 where M_1 is a matching containing the edge (u, v) and M_2 is an arbitrary matching. But there is no term in the determinant expansion for $\mathbf{A}^{[u,v]}$ for the edge (u, v) as row u and column v are deleted (although there can be a term for M_2 which corresponds to (v, u) because row v and column u are still included).

The weight in the determinant expansion for $\mathbf{A}^{[u,v]}$ for a given choice of M_1 and M_2 is therefore equal to

$$\frac{1}{\det(\mathbf{A})} \left[\frac{w_{M_1}}{w_{u,v}} \right] w_{M_2}. \tag{A.5}$$

When the Hadamard product $\mathbf{P}_k \circ \mathbf{P}_k^{-1}$ is taken, the contribution has magnitude

$$w_{u,v} \frac{1}{\det(\mathbf{A})} \left[\frac{w_{M_1}}{w_{u,v}} \right] w_{M_2} \tag{A.6}$$

and so the final contribution for the edge is equal to the sum over all ordered pairs (M_1, M_2) of matchings where M_1 contains the edge (u, v) of

$$\frac{w_{M_1} w_{M_2}}{\det(\mathbf{A})}. \tag{A.7}$$

For benzenoids, the numerator and the denominator have the same sign and so the resulting entries are non-negative. In this expression, there is a common factor

$$S = \sum_{i=1}^{m(G)} w_{M_i} \tag{A.8}$$

in both the numerator (where it arises from making each possible choice for a second matching M_2) and the denominator (which is equal to $(-1)^{n/2} S^2$). Cancel this out on top and bottom to show that the entry for edge (u, v) in \mathbf{P}_{k+1} after the Hadamard product has been made is

$$\frac{1}{S} \sum_M w_M \tag{A.9}$$

where the sum is over all the matchings M that contain the edge (u, v) .

A mathematically equivalent formula for the CKC matrix \mathbf{P}_{k+1} for a benzenoid is therefore

$$\mathbf{P}_{k+1} = \frac{1}{S} \sum_{i=1}^{m(G)} w_{M_i} \mathbf{K}_i \tag{A.10}$$

as indicated in the main text.

References

1. A. Ciesielski, T.M. Krygowski, M.K. Cyrański, *Symmetry*, **2**, 1390–1400 (2010) (www.mdpi.com/journal/symmetry; doi:10.3390/sym2031390)

2. K. Fries, *Justus Liebigs Ann. Chem.* **454**, 121–324 (1927)
3. E. Clar, *The Aromatic Sextet* (Wiley, New York, 1972)
4. A.T. Balaban, D.J. Klein, *J. Phys. Chem. C* **113**, 19123–19133 (2009)
5. P. Hansen, M. Zheng, *J. Math. Chem.* **15**, 93–107 (1994)
6. H. Abeledo, G.W. Atkinson, *Linear Algebra Appl.* **420**, 441–448 (2007)
7. H. Abeledo, G.W. Atkinson, in *The Clar and Fries problems for benzenoid hydrocarbons are linear programs*, ed. by P. Hansen, P.W. Fowler, M. Zheng *Discrete Mathematical Chemistry*, vol 51, in series (American Mathematical Society, Providence, 2000), pp. 1–8
8. R.A. Horn, C.R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985)
9. L. Pauling, L.O. Brockway, J.Y. Beach, *J. Am. Chem. Soc.* **57**, 2705–2709 (1935)
10. T. Morikawa, S. Narita, D.J. Klein, *Chem. Phys. Lett.* **402**, 554–558 (2004)
11. P.W. Fowler, B.T. Pickup, T.Z. Todorova, W. Myrvold, *J. Chem. Phys.* **131**, 244110 (2009)
12. G. Brinkmann, O. Delgado Friedrichs, S. Liskén, A. Peeters, N. Van Cleemput, CaGe—a virtual environment for studying some special classes of plane graphs—an update. *MATCH Commun. Math. Comput. Chem.* **63**, 533–552, (2010). Program obtainable from <http://caagt.ugent.be/CaGe>
13. A.T. Balaban, X. Liu, S.J. Cyvin, D.J. Klein, *J. Chem. Inf. Comput. Sci.* **33**, 429–436 (1993)
14. P.W. Fowler, S.J. Austin, O.J. Dunning, J.R. Dias, *Chem. Phys. Lett.* **224**, 123–130 (1994)
15. S. Fajtlowicz, P.E. John, H. Sachs, *Croat. Chem. Acta* **78**, 195–201 (2005)
16. L. Lovász, M.D. Plummer, *Matching Theory, North-Holland Mathematics Studies*, vol. 121 (North-Holland, Amsterdam, 1986)