Sorting by Reversals is Difficult

Alberto Caprara *

Abstract

We prove that the problem of sorting a permutation by the minimum number of reversals is NP-hard, thus answering a major question on the complexity of a problem which has widely been studied in the last years. The proof is based on the strong relationship between this problem and the problem of finding the maximum number of edge-disjoint alternating cycles in a suitably-defined bicolored graph. For this latter problem we derive a number of structural properties, that can be used for showing its NP-hardness.

Key words: sorting by reversals, alternating cycle decomposition, complexity.

AMS subject classifications: 68Q25, 68R10, 05C45.

1 Introduction

Let be given a permutation $\pi = (\pi_1 \ \dots \ \pi_n)$ of $\{1,\dots,n\}$, and denote by ι the identity permutation $(1\ 2\ \dots\ n-1\ n)$. A *reversal* of the interval (i,j) is an inversion of the subsequence $\pi_i \ \dots \ \pi_j$ of π, represented by the permutation $\rho = (1 \ \dots \ i-1 \quad j \ \dots \ i \quad j+1 \ \dots \ n)$. Composition of π with ρ yields $\pi\rho = (\pi_1 \ \dots \ \pi_{i-1} \ \pi_j \ \dots \ \pi_i \ \pi_{j+1} \ \dots \ \pi_n)$ where elements $\pi_i,\dots,\pi_j$ have been reversed. The problem of *sorting a permutation by the minimum number of reversals* (MIN-SBR) is defined as follows:

MIN-SBR: given a permutation π, find a shortest sequence of reversals $\rho_1,\dots,\rho_{d(\pi)}$ such that $\pi\rho_1\dots\rho_{d(\pi)} = \iota$, where $d(\pi)$ denotes the *reversal distance* of π.

MIN-SBR was inspired by computational biology applications, in particular genome rearrangements, and has widely been studied in the last years, among others, by Kececioglu and Sankoff [9, 8], Bafna and Pevzner [1], Hannenhalli and Pevzner [4, 5], Caprara, Lancia and Ng [2], Irving and Christie [7]. A major open question about MIN-SBR is its complexity: the problem was conjectured to be NP-hard by Kececioglu and Sankoff [9], and to our knowledge no proof of this result can be found in the literature, see also [10]. A stronger conjecture of Kececioglu and Sankoff [9] claimed that the special case of MIN-SBR where, for a given π, one wants to check whether $d(\pi)$ is equal to one half times the number of breakpoints of π (see below), was NP-complete. In fact Irving and Christie [7] recently disproved this latter conjecture, showing a polynomial time algorithm for solving this special case.

In this paper we prove that MIN-SBR is indeed NP-hard. Our result is derived by considering the strong relationship between MIN-SBR and the problem of finding the maximum number of edge-disjoint alternating cycles in a suitably-defined bicolored graph. This relationship was first pointed out by Bafna and Pevzner [1] and Kececioglu and Sankoff [9], and further illustrated by Hannenhalli and Pevzner [4]. Our starting point is the fact that partitioning the edge set of a Eulerian graph into the maximum number of cycles is NP-hard, see Holyer [6].

We next give the basic definitions and previous results used in the sequel.

Consider a permutation $\pi = (\pi_1 \ \dots \ \pi_n)$ of $\{1,\dots n\}$. A *long strip* of π is a subsequence $\pi_i \ \dots\pi_j$ of π such that $j > i+1$ and either $\pi_k = \pi_{k-1}+1$ for $k = i+1,\dots,j$, or $\pi_k = \pi_{k-1}-1$ for $k = i+1,\dots,j$. In other words, a long strip of π corresponds to three or more elements which appear in the same, or reverse, order in π and ι. As far as MIN-SBR is concerned, Hannenhalli and Pevzner [5] proved that one can assume without loss of generality that π contains no long strip; we therefore make this assumption in the following. We also assume without loss of generality $\pi_1 \neq 1$ and $\pi_n \neq n$. All these assumptions are not strictly necessary to derive our results, but lead to a simplified exposition of some parts.

Following the description in [1], define the *breakpoint graph* $G(\pi) = (V, R\cup B)$ of π as follows. Ideally add to π the elements $\pi_0 := 0$ and $\pi_{n+1} := n+1$, re-defining $\pi := (0 \ \pi_1 \ \dots \ \pi_n \ n+1)$. Also, let the *inverse permutation* π^{-1} of π be defined by $\pi^{-1}_{\pi_i} := i$ for $i = 0,\dots,n+1$. Each node $v \in V$ represents an element of π, therefore we let $V := \{0,\dots,n+1\}$. Graph $G(\pi)$ is *bicolored*, i.e. its edge set is partitioned in two subsets, each represented by a different color. R is the set of *red* (say) edges, each of the form (π_i, π_{i+1}), for all $i \in \{0,\dots,n\}$ such that $|\pi_i - \pi_{i+1}| \neq 1$, i.e. elements which are in consecutive positions in π but not in the identity permutation ι. Such a pair π_i, π_{i+1} is called a *breakpoint* of π. Let us indicate with $b(\pi) := |R|$ the number of breakpoints of π. B is the set of *blue* (say) edges, each of the form $(i, i+1)$, for all $i \in \{0,\dots,n\}$ such that $|\pi^{-1}_i - \pi^{-1}_{i+1}| \neq 1$, i.e. elements which are in consecutive positions in ι but not in π. Notice that each node $i \in V$

*DEIS, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy — e-mail: acaprara@deis.unibo.it

RECOMB 97, Santa Fe New Mexico USA

has either degree 2 or 4, and has the same number of incident blue and red edges. Therefore, $|R| = |B|(= b(\pi))$. The fact that $G(\pi)$ has no nodes of degree 0 follows from the assumption that π contains no long strip. Figure 1 depicts the breakpoint graph associated with permutation (4 2 1 3).

An *alternating cycle* of $G(\pi)$ is a sequence of edges $r_1, b_1, r_2, b_2, \ldots, r_m, b_m$, where $r_i \in R$, $b_i \in B$ for $i = 1, \ldots, m$; r_i and b_j have a common node for $i = j = 1, \ldots, m$ and for $i = j + 1$, $j = 1, \ldots, m$ (where $r_{m+1} := r_1$); and $r_i \neq r_j$, $b_i \neq b_j$ for $1 \leq i < j \leq m$. An *alternating path* is a subsequence of edges of some alternating cycle. For example, edges $(0,4),(4,3),(3,1),(1,0)$ and $(4,2),(2,3),(3,5),(5,4)$ form alternating cycles in the graph of Figure 1. It is sometimes convenient to assign each edge $(\pi_i, \pi_{i+1}) \in R$ an orientation from π_i to π_{i+1}, i.e. to orient the red edges of G from the endpoint which appears first in π to the endpoint which appears second. An alternating cycle of $G(\pi)$ is then called *directed* with respect to π if it is possible to walk along the whole cycle traversing each red edge in the direction of its orientation. An alternating cycle of $G(\pi)$ is called *undirected* with respect to π if it is not directed. For example, in Figure 1 alternating cycle $(4,2),(2,3),(3,5),(5,4)$ is directed with respect to (4 2 1 3), whereas cycle $(0,4),(4,3),(3,1),(1,0)$ is undirected with respect to (4 2 1 3).

An *alternating cycle decomposition* of $G(\pi)$ is a collection of edge-disjoint alternating cycles, such that every edge of G is contained in exactly one cycle of the collection. It is easy to see that $G(\pi)$ always admits an alternating cycle decomposition. In the graph of Figure 1, cycles $(0,4),(4,3),(3,1),(1,0)$ and $(4,2),(2,3),(3,5),(5,4)$ form an alternating cycle decomposition. Bafna and Pevzner [1] (see also Kececioglu and Sankoff [9]) proved the following property. For a given π let $c(\pi)$ be the maximum cardinality of an alternating cycle decomposition of $G(\pi)$.

Theorem 1 *([1], [9]) For every π, $d(\pi) \geq b(\pi) - c(\pi)$.*

Therefore $b(\pi) - c(\pi)$ gives a valid lower bound on the optimal solution value to MIN-SBR. In practical cases this bound turns out to be very tight, and is frequently equal to the optimum, as observed by the extensive experiments of Kececioglu and Sankoff [9] and Caprara, Lancia and Ng [2]. This empirical observation was formalized by the spectacular results of Hannenhalli and Pevzner [4], who were able to prove that the signed version of MIN-SBR is solvable in polynomial time.

The *signed* version of MIN-SBR is the problem of sorting a permutation by the minimum number of reversals, with a *parity* assigned to each element of the permutation, specifying that the number of reversals in a solution involving the element must be either even or odd. A permutation with a parity on each element is called *signed*, and denoted by $\vec{\pi}$. Let $d(\vec{\pi})$ denote the optimal solution value to signed MIN-SBR on $\vec{\pi}$, and $b(\vec{\pi})$ the number of breakpoints of $\vec{\pi}$ (see e.g. [1] for the definition of breakpoint of a signed permutation). Bafna and Pevzner [1] introduced the notion of breakpoint graph for a signed permutation $\vec{\pi}$, $G(\vec{\pi})$, which turns out to be a breakpoint graph where all nodes have degree 2, hence having a unique alternating cycle decomposition of cardinality $c(\vec{\pi})$. Furthermore, Bafna and Pevzner showed that Theorem 1 also applies to signed permutations, i.e. $d(\vec{\pi}) \geq b(\vec{\pi}) - c(\vec{\pi})$. The orientation of the red edges of $G(\vec{\pi})$ with respect to $\vec{\pi}$ is defined in the same way as in the unsigned case. Hannenhalli and Pevzner [4] proved that $d(\vec{\pi}) = b(\vec{\pi}) - c(\vec{\pi})$ if every cycle of $G(\vec{\pi})$ is undirected. Indeed in this case $\vec{\pi}$ contains no *hurdles*, and therefore is not a *fortress* as well; see [4]. We stress that the condition is only sufficient, and that undirected cycles are called *oriented* in [4], and directed cycles *unoriented*.

Given a (unsigned) permutation π and an alternating cycle decomposition of $G(\pi)$ into $c(\pi)$ cycles, it is easy to define a signed permutation $\vec{\pi}$ such that $b(\vec{\pi}) = b(\pi)$, $c(\vec{\pi}) = c(\pi)$, $d(\vec{\pi}) \geq d(\pi)$, and every cycle of $G(\vec{\pi})$ is undirected with respect to $\vec{\pi}$ if and only if its counterpart in the decomposition of $G(\pi)$ is undirected with respect to π (see e.g. [5]). If every cycle in this latter decomposition is in fact undirected with respect to π it follows that

$$d(\pi) \leq d(\vec{\pi}) = b(\vec{\pi}) - c(\vec{\pi}) = b(\pi) - c(\pi) \leq d(\pi)$$

hence implying

Theorem 2 *([1], [4, 5]) Let be given a permutation π and a maximum decomposition of the associated breakpoint graph into $c(\pi)$ alternating cycles. If every cycle in the decomposition is undirected with respect to π, then $d(\pi) = b(\pi) - c(\pi)$.*

The above discussion motivates the study of the following *maximum alternating cycle decomposition* (MAX-ACD) problem:

MAX-ACD: given the breakpoint graph $G(\pi)$ of some permutation π, find a maximum-cardinality alternating cycle decomposition of $G(\pi)$.

As one can observe, MAX-ACD is somehow related with the following problem, called *maximum eulerian cycle decomposition* (MAX-ECD):

MAX-ECD: given a Eulerian graph $H = (W, E)$, find a maximum-cardinality cycle decomposition of H, i.e. partition E into the maximum number of cycles.

In the early 80's Holyer [6] proved that checking whether the edge set of a given graph H can be partitioned into cliques of size k, is NP-complete for every $k \geq 3$. In particular, for $k = 3$ one wants to check whether the edge set of H can be partitioned into triangles. In this case H can be assumed Eulerian without loss of generality, the answer being clearly no otherwise. So the problem of finding, if any, a partition of the edge set of a Eulerian graph into triangles is NP-complete. This immediately implies

Theorem 3 *([6]) MAX-ECD is NP-hard.*

The paper is organized as follows. Section 2 gives a complete characterization of the bicolored graphs which are breakpoint graphs of some permutation. This characterization is used in Section 3 to prove that MAX-ACD is NP-hard, showing a polynomial transformation from MAX-ECD. Finally, in Section 4 we prove that MIN-SBR is NP-hard by showing a polynomial transformation from MAX-ACD.

2 A Characterization of Breakpoint Graphs

Next section is devoted to showing a polynomial-time transformation from MAX-ECD to MAX-ACD, thus proving the latter problem is NP-hard. (In fact, there would also be an easy transformation from MAX-ACD to MAX-ECD, which is omitted since it is useless for our purposes.) Such a transformation should pick any Eulerian graph $H = (W, E)$, and define a permutation π such that, by solving MAX-ACD on the associated breakpoint graph $G(\pi)$, one gets a partitioning of E into the maximum number of cycles. In fact, we bypass the explicit definition of π, and use the following characterization of breakpoint graphs to construct a suitable bicolored graph from the given Eulerian graph H. To simplify

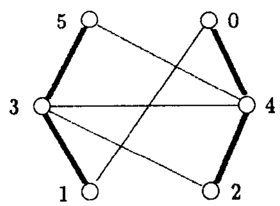

Figure 1: The breakpoint graph $G(\pi)$ associated with $\pi = (4\ 2\ 1\ 3)$ — blue edges are drawn as thin lines, red edges as thick lines.

the notation, given a (possibly bicolored) graph $G = (V, E)$ and any set $F \subseteq V \times V$ (possibly $F \not\subseteq E$), we let $G(F)$ be the graph defined by node set V and edge set F.

Theorem 4 *A bicolored graph $G = (V, R \cup B)$ is the breakpoint graph of some permutation π if and only if*

(i) each connected component of the subgraphs of G induced by edge set R, $G(R)$, and by edge set B, $G(B)$, is a simple path;

(ii) each node $i \in V$ has the same degree (1 or 2) in $G(R)$ and $G(B)$;

(iii) for each pair of nodes $i, j \in V$ connected by an edge in $G(R)$, i and j are not connected by an edge in $G(B)$, and viceversa, i.e. if $(i, j) \in R$ then $(i, j) \notin B$ and viceversa.

The necessity part follows immediately from the definition of breakpoint graph. For proving the sufficiency part, we show an algorithm to construct from G a permutation π such that $G(\pi) = G$. Let be given $G = (V, R \cup B)$ satisfying (i)-(iii). Consider the set U of the nodes of degree 2 in G, i.e. the set of nodes of degree 1 in $G(R)$ and $G(B)$. Notice that $|U|$ is even and ≥ 2. A *perfect matching* of the nodes in U is a set $M \subset U \times U$ such that every $i \in U$ is contained in exactly one pair in M. Such a perfect matching is called *Hamiltonian* if the graphs $G(R \cup M)$ and $G(B \cup M)$ are Hamiltonian circuits — notice that M can be Hamiltonian only if $M \cap (R \cup B) = \emptyset$. The importance of Hamiltonian matchings of U is motivated by the following property.

Claim 1 *Every Hamiltonian matching of U defines up to $|U|$ permutations π such that $G(\pi) = G$. Conversely, every π such that $G(\pi) = G$ defines a Hamiltonian matching of U.*

Proof. Given a Hamiltonian matching M of U, $|U|$ (possibly different) permutations of $|V|$ elements (including the dummy elements $\pi_0 := 0$ and $\pi_{|V|} := |V| - 1$) having G as breakpoint graph can be constructed as follows; see also procedure GET_PERMUTATION below. Choose any node $i \in U$ and number it 0. Consider then the Hamiltonian circuit $G(B \cup M)$ and walk along it starting from i and traversing first the blue edge incident to i. Number the nodes $1, 2, \ldots, |V| - 1$ according to the order in which they are visited by the walk. Consider now the Hamiltonian circuit $G(R \cup M)$ and walk along it starting from node i, letting $\pi_0 := 0$, and traversing first the red edge incident to i. Let elements $\pi_1, \pi_2, \ldots, \pi_{|V|}$ correspond to the numbers assigned to the nodes which are in turn visited by the walk. It is easy to check that $G(\pi) = G$.

Conversely, for any given permutation π of $\{1, \ldots n\}$, it is immediate to verify that the set $M := (\{(i, i+1) : 0 \leq i < n\} \setminus B) \cup \{(0, n+1)\} = (\{(\pi_i, \pi_{i+1}) : 0 \leq i < n\} \setminus R) \cup \{(0, n+1)\}$ defines a Hamiltonian matching of $G(\pi)$. □

procedure GET_PERMUTATION;
input: a bicolored graph $G = (V, R \cup B)$ satisfying (i)-(iii), and a Hamiltonian matching M of the nodes of degree 2 in G;
output: a permutation π such that $G(\pi) = G$;
begin
  let $h_0, h_1, h_2, \ldots, h_{|V|-1}$ be the Hamiltonian circuit $G(B \cup M)$ where h_0 has degree 2 in G and $(h_0, h_1) \in B$;
  for $i := 0$ **to** $|V| - 1$ **do** $\nu_{h_i} := i$;
  let $g_0, g_1, g_2, \ldots, g_{|V|-1}$ be the Hamiltonian circuit $G(R \cup M)$ where $g_0 = h_0$ and $(g_0, g_1) \in R$;
  for $i := 0$ **to** $|V| - 1$ **do** $\pi_i := \nu_{g_i}$
end.

For example, the Hamiltonian matching of $G(\pi)$ in Figure 1 defined by π is $M = \{(1, 2), (0, 5)\}$. Matching M is also associated with the permutation $(2\ 4\ 3\ 1)$, having $G(\pi)$ as breakpoint graph.

Notice that, given a breakpoint graph G and an associated Hamiltonian matching M, the Hamiltonian circuit $G(R \cup M)$ uniquely determines which alternating cycles of G are directed or undirected with respect to every permutation associated with G and M.

We show that every G satisfying (i)-(iii) has a Hamiltonian matching by giving a procedure to find such a matching. Construct the following *auxiliary graph* $A = (U, S \cup C)$, where nodes $i, j \in U$ are connected by a red edge $s \in S$ if i and j are the endpoints of a path in $G(R)$, and, symmetrically, nodes $i, j \in U$ are connected by a blue edge $c \in C$ if i and j are the endpoints of a path in $G(B)$. It is easy to check that each node of A has exactly 2 incident edges, one in S and one in C. (Notice that A is not necessarily a breakpoint graph, since a red edge in S and a blue edge in C might have the same endpoints.) Figure 2 depicts the auxiliary graph associated with the breakpoint graph of Figure 1.

procedure HAMILTONIAN_MATCHING;
input: a bicolored graph $G = (V, R \cup B)$ satisfying (i)-(iii);
output: a Hamiltonian matching M of the nodes of degree 2 in G;
begin
  let U be the set of nodes of degree 2 of G and $A = (U, S \cup C)$ the corresponding auxiliary graph;
  initialize $M := \emptyset$;
  while $|U| > 2$ **do**
  begin
    choose any node pair $i, j \in U$ such that $(i, j) \notin S \cup C$;
    comment: shrink i and j;
    let a and b respectively be the nodes connected to i and j by a red edge in S, and c and d respectively be the nodes connected to i and j

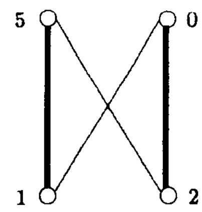

Figure 2: The auxiliary graph A associated with the breakpoint graph of Figure 1.

by a blue edge in C;
let $U := U \setminus \{i, j\}$,
$S := (S \setminus \{(i, a), (j, b)\}) \cup \{(a, b)\}$,
$C := (C \setminus \{(i, c), (j, d)\}) \cup \{(c, d)\}$,
$M := M \cup \{(i, j)\}$
end;
let i and j be the two nodes left in U;
let $M := M \cup \{i, j\}$
end.

Claim 2 *The set M constructed by the above procedure is a Hamiltonian matching.*

Proof. In the while-do loop it is easy to see that, as long as $|U| > 2$, i.e. $|U| \geq 4$, there always exists a node pair i, j such that $(i, j) \notin S \cup C$, since each node of A has degree 2. The fact that M is a perfect matching of the nodes of degree 2 in G such that $M \cap (R \cup B) = \emptyset$ is immediate. To see that M is Hamiltonian, it is sufficient to show that $G(R \cup M)$ and $G(B \cup M)$ contain no subcircuit. Indeed, at each execution of the while-do loop, there is a one-to-one correspondence between maximal paths in $G(R \cup M)$ (resp. in $G(B \cup M)$) and edges in S (resp. in C), and therefore the only way to introduce a subcircuit would be matching two nodes connected by an edge in S (resp. in C), which is avoided. □

In the example of Figure 1, $M = \{(1, 2), (0, 5)\}$ is the only possible Hamiltonian matching of $G(\pi)$; see Figure 2. With Claim 2, the proof of Theorem 4 is complete.

Given a bicolored graph $G = (V, R \cup B)$, the *subdivision* of an edge $e = (i, j) \in R \cup B$ is obtained by adding two new nodes, say a, b, to V, and by replacing e with the 3 edges (i, a), (a, b) and (b, j), where (i, a) and (b, j) have the same color as e, while (a, b) has a different color from e; see Figure 3.

Remark 1 *There is a one-to-one correspondence between alternating cycles and alternating cycle decompositions of a bicolored graph G and any bicolored graph G' obtained from G by subdividing edges.*

The above Remark immediately leads to the following

Remark 2 *Given a bicolored graph $G = (V, R \cup B)$ satisfying (ii) in Theorem 4 and*

(i') each connected component of the subgraphs of G induced by edge set R, $G(R)$, and by edge set B, $G(B)$, is either a simple path or a simple cycle;

one can define a graph $G' = (V', R' \cup B')$ satisfying (i)-(iii) in Theorem 4 whose size is bounded by a constant times the size of G and such that there is a one-to-one correspondence between alternating cycles and alternating cycle decompositions of G and G'.

Proof. G' can be obtained from G by first subdividing all the red edges $r = (i, j) \in R$ such that $(i, j) \in B$ — so as to satisfy (iii) — and then by subdividing one edge for each cycle in $G(B)$ or in $G(R)$ — so as to satisfy (i). □

3 The Complexity of MAX-ACD

This section is devoted to proving the following

Theorem 5 *MAX-ACD is NP-hard.*

The proof is based on a polynomial transformation from MAX-ECD to MAX-ACD. Given a Eulerian graph $H = (W, E)$, we construct a bicolored graph G satisfying (i)-(iii) in Theorem 4, whose size is polynomial in that of H and such that there is a one-to-one correspondence between cycles of H and alternating cycles of G, and between cycle decompositions of H and alternating cycle decompositions of G. The proof then follows from Theorems 3 and 4.

The construction is based on the replacement of each node v of H of degree d with a bicolored graph $G(v)$, uniquely determined by d. In particular $G(v)$ has d nodes of degree 1, each with an incident blue edge, and other nodes of degree 2 or 4, each having the same number of incident blue and red edges. Furthermore, $G(v)$ contains no alternating cycle, and for any partition of its degree-1 nodes into pairs, its edge set can be decomposed into $d/2$ alternating paths, each connecting a different node pair in the partition. The whole G is then obtained by interconnecting the $G(v)$'s. For each edge $e = (u, v) \in E$, a pair of degree-1 nodes in the graphs $G(v)$ and $G(u)$ is connected by a red edge, so as to ensure that every node of G has the same number of incident blue and red edges; see Figure 4.

We complete the proof by showing the structure of each graph $G(v)$ and proving the properties mentioned above.

For each pair of integers d, m, d even, let the bicolored graph $G(d, m)$ be defined as follows. $G(d, m)$ is planar, and is depicted in Figure 5 for $d = 8$ and $m = 2$. Let $s := d/2$ and $r := \lceil d/4 \rceil$. $G(d, m)$ is subdivided into m *levels*, which are all equal to each other with the exception of the last one.

Each level l, $l = 1, \ldots, m$ contains $2s + 1$ nodes; $s + 1$ of them are the *upper nodes* of the level, indicated with $q^l_1, \ldots, q^l_{s+1}$, and the other s are the *lower nodes* of the level, indicated with $p^l_1, \ldots, p^l_s$. Nodes $q^l_1, \ldots, q^l_{s+1}$ are connected to $p^l_1, \ldots, p^l_s$ by the d red edges $(q^l_i, p^l_i), (q^l_{i+1}, p^l_i)$, $i = 1, \ldots, s$. Furthermore, for $l = 1, \ldots, m - 1$ nodes $q^l_1, \ldots, q^l_{s+1}$ are connected to the lower nodes of level $l+1$ by the d blue edges $(p^{l+1}_i, q^l_i), (p^{l+1}_i, q^l_{i+1})$, $i = 1, \ldots, s$. Finally, the upper nodes of the last level m are connected to each other by the s blue edges (q^m_i, q^m_{i+1}), $i = 1, \ldots, s$.

$G(d, m)$ also contains d *bottom nodes*, indicated with $b_1, \ldots, b_d$, connected to the lower nodes of level 1 by the d blue edges $(b_{2i-1}, p^1_i), (b_{2i}, p^1_i)$, $i = 1, \ldots, s$.

It is easy to check that the following holds

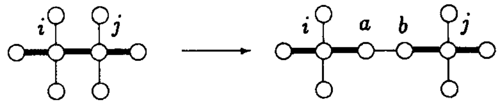

Figure 3: The subdivision of edge (i, j).

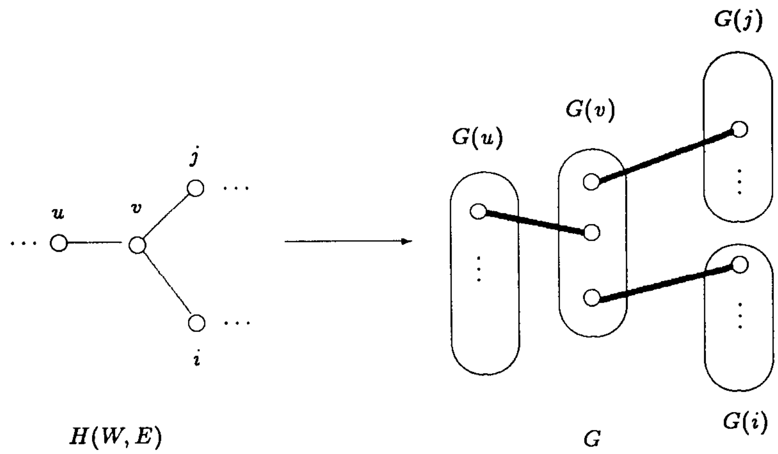

Figure 4: Outline of the transformation from MAX-ECD to MAX-ACD — the circles inside each $G(h)$ represent the associated degree-1 nodes.

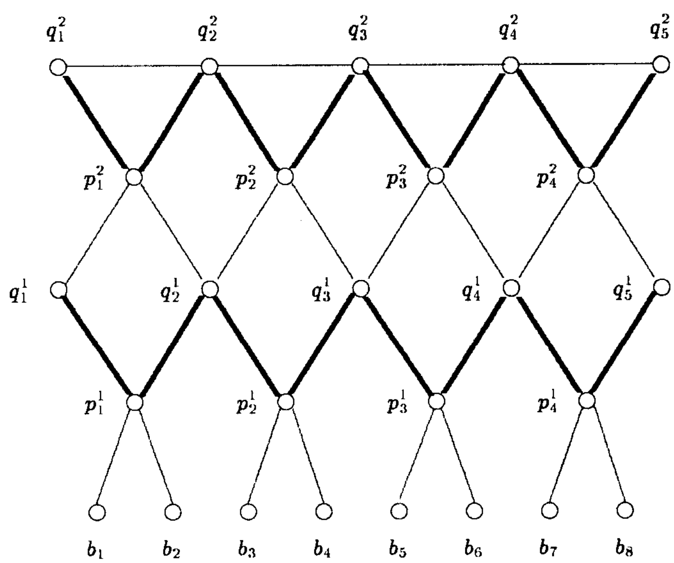

Figure 5: The bicolored graph $G(d, m)$ for $d = 8$ and $m = 2$.

Remark 3 *All nodes of $G(d,m)$ with the exception of the bottom ones have the same number of incident red and blue edges. Furthermore, $G(d,m)$ contains no alternating cycle, and every alternating path connecting a pair of bottom nodes is formed by $4m+1$ edges, the $(2m+1)$-th being a blue edge which connects two upper nodes of level m. Therefore, any s edge-disjoint alternating paths connecting bottom node pairs of $G(d,m)$ form a decomposition of the edge set of $G(d,m)$.*

In order to show the key property of $G(d,m)$, we first prove a technical intermediate result. To this purpose, we need some definitions.

The *zig-zag path* connecting node q_h^l to q_{h+1}^l in $G(d,m)$, $1 \leq l \leq m; 1 \leq h \leq s$, is the alternating path defined by edges $(q_h^{k-1}, p_h^k), (p_h^k, q_h^k)$, $k = l+1, \ldots, m-1$; (q_h^m, q_{h+1}^m); $(q_{h+1}^k, p_h^k), (p_h^k, q_{h+1}^{k-1})$, $k = m, \ldots, l+1$. In Figure 5, a zig-zag path can be seen as a path which goes from a level l to the last level, "turning right" at q-nodes and "left" at p-nodes; then "turns right" at a q-node of the last level; and finally goes from the last level back to level l, "turning left" at q-nodes and "right" at p-nodes. (See e.g. the path $(q_1^1, p_1^2), (p_1^2, q_1^2), (q_1^2, q_2^2), (q_2^2, p_1^2), (p_1^2, q_2^1)$ in Figure 5.)

Consider t edge-disjoint alternating paths $P_1, \ldots, P_t$ connecting upper nodes of level l, and a node q_h^l having an incident edge not contained in $P_1, \ldots, P_t$. The zig-zag path starting from node q_h^l *shifted* with respect to $P_1, \ldots, P_t$, is the alternating path defined as follows, having no edge in common with $P_1, \ldots, P_t$. The path first goes from level l to the last level m trying to use edges of the form (q_i^k, p_i^{k+1}) and (p_i^k, q_i^k). If edge (q_i^k, p_i^{k+1}) or (p_i^k, q_i^k) is already contained in some of $P_1, \ldots, P_t$, the path uses edge (q_i^k, p_{i-1}^{k+1}) or (p_i^k, q_{i+1}^k), respectively. Once a node q_j^m has been reached, the path uses edge (q_j^m, q_{j+1}^m) if it is contained in none of $P_1, \ldots, P_t$, and edge (q_j^m, q_{j-1}^m) otherwise. Then the path goes from level m back to level l trying to use edges of the form (q_i^k, p_{i-1}^{k-1}) and (p_i^k, q_{i+1}^{k-1}), and using edges of the form (q_i^k, p_i^{k-1}) and (p_i^k, q_i^{k-1}), respectively, if this is not possible without intersecting paths $P_1, \ldots, P_t$. The final node of the path, say q_a^l, clearly depends on $P_1, \ldots, P_t$. In Figure 5, a zig-zag path shifted with respect to a set of paths can be seen as a path which follows the same rules for "turning" as a zig-zag path, as long as it does not meet an edge in the other paths; in this latter case, the path "turns" on the opposite direction. In Figure 5, if $P_1 = (q_1^1, p_1^2), (p_1^2, q_2^2), (q_2^2, q_3^2), (q_3^2, p_3^2), (p_3^2, q_4^1)$ and $P_2 = (q_1^2, p_2^2), (p_2^2, q_3^2), (q_3^2, q_4^2), (q_4^2, p_4^2), (p_4^2, q_5^1)$; the zig-zag path starting from q_2^1 shifted with respect to P_1 and P_2 is (q_2^1, p_1^2), $(p_1^2, q_1^2), (q_1^2, q_2^2), (q_2^2, p_2^2), (p_2^2, q_3^1)$.

The *up-right* path connecting node q_h^l to q_{h+k}^{l+k} in $G(d,m)$, $1 \leq l \leq m-k; 1 \leq h \leq s+1-k$, is the alternating path defined by the edges $(q_{h+i}^{l+i}, p_{h+i}^{l+i+1}), (p_{h+i}^{l+i+1}, q_{h+i+1}^{l+i+1})$, $i = 0, \ldots, k-1$ (see e.g. the path $(q_1^1, p_1^2), (p_1^2, q_2^2)$ in Figure 5). Similarly, the *up-left* path connecting node q_h^l to q_{h-k}^{l+k} in $G(d,m)$, $1 \leq l \leq m-k; k+1 \leq h \leq s+1$, is the alternating path defined by the edges $(q_{h-i}^{l+i}, p_{h-i-1}^{l+i+1}), (p_{h-i-1}^{l+i+1}, q_{h-i-1}^{l+i+1})$, $i = 0, \ldots, k-1$ (see e.g. the path $(q_5^1, p_4^2), (p_4^2, q_4^2)$ in Figure 5).

Claim 3 *Consider the $s+1$ upper nodes of level l in $G(d,m)$, $q_1^l, \ldots, q_{s+1}^l$. If $m \geq l + r$, for any pair i, j, $1 \leq i < j \leq s$, there exist s edge-disjoint alternating paths connecting node pairs q_h^l, q_{h+1}^l, $h \in \{1, \ldots, s\} \setminus \{i, j\}$, and either pairs q_i^l, q_j^l; q_{i+1}^l, q_{j+1}^l, or q_i^l, q_{j+1}^l; q_{i+1}^l, q_j^l.*

Proof. If $j = i+1$ the proof is trivial, since all node pairs can be connected by zig-zag paths. (In fact, if q_i^l has to be connected to $q_{j+1}^l = q_{i+2}^l$ and q_{i+1}^l to $q_j^l = q_{i+1}^l$, i.e. to itself, the corresponding paths can be immediately derived from the zig-zag paths connecting q_i^l to q_{i+1}^l and q_{i+1}^l to q_{i+2}^l.) Otherwise, observe that $i = 1$ and $j = s$ can be assumed without loss of generality. Indeed, if $i \neq 1$ or $j \neq s$, all pairs q_h^l, q_{h+1}^l, $h = 1, \ldots, i-1$ and $h = j+1, \ldots, s$, can be connected by zig-zag paths, and the proof follows from the validity of the claim for $s' = j - i + 1$, $i = 1$, and $j = s'$.

Assuming therefore $i = 1$, $j = s$ and $s > 2$, we show the alternating paths connecting each pair; see also Figure 6. For the paths connecting q_1^l to either q_s^l or q_{s+1}^l, say P_1, and the path connecting q_2^l to either q_{s+1}^l or q_s^l, say P_2, we distinguish between the cases of s even and s odd.

If s is even, and therefore $r = s/2 (= d/4)$, P_1 is defined by the up-right path from q_1^l to q_{r+1}^{l+r}, by the zig-zag path from q_{r+1}^{l+r} to q_{r+2}^{l+r}, by the up-left path from q_{r+2}^{l+r} to q_{s+1}^{l+1}, by edge $(q_{s+1}^{l+1}, p_s^{l+1})$, and either by edge (p_s^{l+1}, q_s^l) or (p_s^{l+1}, q_{s+1}^l), depending on the case. Symmetrically, P_2 is defined either by edge (q_{s+1}^l, p_s^{l+1}) or (q_s^l, p_s^{l+1}), by edge (p_s^{l+1}, q_s^{l+1}), by the up-left path from q_s^{l+1} to q_{r+1}^{l+r}, by the zig-zag path from q_{r+1}^{l+r} to q_r^{l+r}, by the up-right path from q_r^{l+r} to q_1^{l+1}, and by the path $(q_1^{l+1}, p_1^l), (p_1^l, q_2^l)$.

If s is odd, and therefore $r = (s+1)/2 (= (d+2)/4)$, P_1 is defined by the up-right path from q_1^l to q_{r+1}^{l+r}, by the zig-zag path from q_{r+1}^{l+r} to q_{r+2}^{l+r}, by the up-left path from q_{r+2}^{l+r} to q_{s+1}^{l+2}, by the path $(q_{s+1}^{l+2}, p_s^{l+2}), (p_s^{l+2}, q_{s+1}^{l+1}), (q_{s+1}^{l+1}, p_s^{l+1})$, and either by edge (p_s^{l+1}, q_s^l) or (p_s^{l+1}, q_{s+1}^l). Symmetrically, P_2 is defined either by edge (q_{s+1}^l, p_s^{l+1}) or (q_s^l, p_s^{l+1}), by edge (p_s^{l+1}, q_s^{l+1}), by the up-left path from q_s^{l+1} to q_r^{l+r}, by the zig-zag path from q_r^{l+r} to q_{r-1}^{l+r}, by the up-right path from q_{r-1}^{l+r} to q_1^{l+2}, and by the path $(q_1^{l+2}, p_1^{l+2}), (p_1^{l+2}, q_1^{l+1}), (q_1^{l+1}, p_1^{l+1})$, (p_1^{l+1}, q_2^l).

In both cases it is easy to check that, for $h = 2, \ldots, s-1$, the zig-zag path starting from node q_h^l shifted with respect to P_1 and P_2, has q_{h+1}^l as final node. Furthermore, all these shifted zig-zag paths have no edge in common with each other, and therefore, together with P_1 and P_2, form the desired set of paths. □

The key property of $G(d,m)$ is shown by the following

Claim 4 *Consider any partition of the bottom nodes of $G(d,m)$ into pairs. If $m \geq r(s-1)+1$, the edge set of $G(d,m)$ can be decomposed into s alternating paths, each connecting a different node pair in the partition.*

Proof. Consider any partition τ of the bottom nodes of $G(d,m)$ into pairs $\{b_{t_1}, b_{t_2}\}, \{b_{t_3}, b_{t_4}\}, \ldots, \{b_{t_{d-1}}, b_{t_d}\}$. Notice that τ can be obtained from the partition $\sigma := \{b_1, b_2\}, \{b_3, b_4\}, \ldots, \{b_{d-1}, b_d\}$ by exchanging k times, with $k \leq s-1$, elements in two different pairs. We show by induction on k that the paths exist in fact for $m \geq rk+1$. Observe that by using zig-zag paths it is trivial to show that if the paths exist for a given value of m, they also do for higher values of m.

If $k = 0$ and $m = 1$ the paths clearly exist. Suppose now they exist for $k = z-1$ and $m = r(z-1)+1$. We show that they also do for $k = z$ and $m = rz - 1$.

Consider a partition τ of the bottom nodes, obtainable from σ by z exchanges. By the induction hypothesis, τ is also obtainable by one exchange from a partition ω for which there exist edge-disjoint alternating paths in $G(d,m')$, with $m' := r(z-1)+1$, connecting all the corresponding node pairs.

Without loss of generality, let $\{b_{t_1}, b_{t_2}\}, \{b_{t_3}, b_{t_4}\}$ be the new pairs in τ after the exchange, and $\{b_{t_1}, b_{t_3}\}, \{b_{t_2}, b_{t_4}\}$ be the corresponding pairs in ω. Also, let $q_a^{m'}$ and $q_{a+1}^{m'}$ be the top nodes such that the path connecting b_{t_1} to b_{t_3} in $G(d, m')$ is defined by an alternating path connecting b_{t_1} to $q_a^{m'}$, the blue edge $(q_a^{m'}, q_{a+1}^{m'})$, and an alternating path connecting $q_{a+1}^{m'}$ to b_{t_3}. Similarly, let $q_b^{m'}$ and $q_{b+1}^{m'}$ be the top nodes such that the path connecting b_{t_2} to b_{t_4} in $G(d, m')$ is defined by an alternating path connecting b_{t_2} to $q_b^{m'}$, the blue edge $(q_b^{m'}, q_{b+1}^{m'})$, and an alternating path connecting $q_{b+1}^{m'}$ to b_{t_4}; see Figure 6.

(The only other case that one should take into account is that in which the path connecting b_{t_2} to b_{t_4} is defined by an alternating path connecting b_{t_2} to $q_{b+1}^{m'}$, the blue edge $(q_{b+1}^{m'}, q_b^{m'})$, and an alternating path connecting $q_b^{m'}$ to b_{t_4}. Nevertheless, this latter case can be treated exactly in the same way as the one considered.)

When one goes from $G(d, m')$ to $G(d, m' + r)$, i.e. when the number of levels is increased by r, Claim 3 shows that each blue edge $(q_c^{m'}, q_{c+1}^{m'})$, $c \in \{1, \dots, s\} \setminus \{a, b\}$ can be replaced by alternating paths connecting its endpoints, while blue edges $(q_a^{m'}, q_{a+1}^{m'})$ and $(q_b^{m'}, q_{b+1}^{m'})$ can be replaced by alternating paths connecting $q_a^{m'}$ to $q_b^{m'}$ and $q_{a+1}^{m'}$ to $q_{b+1}^{m'}$, respectively; see Figure 6.

Therefore, if the number of levels is increased by r, there exist paths connecting all the new pairs in the partition. This proves the inductive step and hence the claim. □

For any even integer d, let $m(d) := \lceil d/4 \rceil (d/2 - 1) + 1$.

The transformatiom from MAX-ECD to MAX-ACD is then the following. Given any Eulerian graph $H = (W, E)$, we define a graph G from H by replacing each $v \in W$ of degree d with the graph $G(v) := G(d, m(d))$, and then, for each $(v, u) \in E$, by connecting two bottom nodes of $G(v)$ and $G(u)$ by a red edge. This connection is easily made so that all bottom nodes of $G(v)$, $v \in W$, have 1 incident red edge; see again Figure 4. It is then clear that G has size polynomial in that of H, and, by Claim 4, there is a one-to-one correspondence between cycles of H and alternating cycles of G, as well as between cycle decompositons of H and alternating cycle decompositions of G. Finally, G is a breakpoint graph since it satisfies the conditions of Theorem 4, so the proof of Theorem 5 is complete.

4 The Complexity of MIN-SBR

A possible way of reading Theorem 2 is the following. Given a breakpoint graph $G(\pi)$ of some permutation π, it would be possible to compute $c(\pi)$ by solving MIN-SBR on π, if there existed an optimal cycle decomposition of $G(\pi)$ made up of undirected (with respect to π) cycles only. Unfortunately, this is not always the case. Nevertheless, given any breakpoint graph $G(\pi)$, we show how to construct another breakpoint graph $\tilde{G}$ having size bounded by a constant times the size of $G(\pi)$ and such that there is a one-to-one correspondence between alternating cycles and alternating cycle decompositions of $G(\pi)$ and $\tilde{G}$. Furthermore, we show how to define a permutation $\tilde{\pi}$ such that $G(\tilde{\pi}) = \tilde{G}$ and every alternating cycle of $\tilde{G}$ is undirected with respect to $\tilde{\pi}$. In this case, $d(\tilde{\pi}) = b(\tilde{\pi}) - c(\tilde{\pi})$ by Theorem 2, $c(\pi) = c(\tilde{\pi})$, and $b(\tilde{\pi})$ is trivially determined. So the construction allows the computation of the optimal value of MAX-ACD for any generic breakpoint graph $G(\pi)$, which is clearly NP-hard, by solving a suitably-defined MIN-SBR instance, whose size is polynomial (in fact, proportional) to that of G. This leads to the main result of this paper, namely

Theorem 6 *MIN-SBR is NP-hard.*

The remainder of this section is devoted to showing the polynomial-time transformation from MAX-ACD to MIN-SBR mentioned above. Consider a breakpoint graph $G(\pi) = (V, R \cup B)$ associated with some permutation π, and let M be the Hamiltonian matching of $G(\pi)$ determined by π (see Section 2).

Construct the breakpoint graph $\tilde{G}$ by applying the following procedure to $G(\pi)$, which replaces every red edge of G by an alternating path of 5 edges, by subdividing "twice" the original edge.

procedure DOUBLE_SUBDIVISION;
input: a breakpoint graph $G(\pi) = (V, R \cup B)$ and
  an associated Hamiltonian matching M of the nodes
  of degree 2 in $G(\pi)$;
output: a breakpoint graph $\tilde{G} = (\tilde{V}, \tilde{R} \cup \tilde{B})$ and
  an associated Hamiltonian matching $\tilde{M}$ of the nodes
  of degree 2 in $\tilde{G}$, such that there is a one-to-one
  correspondence between alternating cycles and
  alternating cycle decompositions of G and $\tilde{G}$, and
  all alternating cycles of $\tilde{G}$ are undirected with
  respect to every permutation $\tilde{\pi}$ associated with
  $\tilde{G}$ and $\tilde{M}$;
begin
  initialize $\tilde{V} := V$, $\tilde{R} := R$, $\tilde{B} := B$, $\tilde{M} := M$;
  let $n := |V|$, and denote the nodes in V by $\{1, \dots, n\}$;
  for each $r \in R$ **do**
  begin
    let i and j be the endpoints of r, and $e = (a, b)$
    any edge of M, where a and b are such that the
    Hamiltonian circuit $\tilde{G}(\tilde{R} \cup \tilde{M})$ is the union of
    edge e, a path from a to i, edge r, and a path
    from j to b;
    comment: add new nodes $n+1, n+2, n+3$,
      $n+4$ to $\tilde{V}$, and replace edge (i, j) with
      the alternating path $(i, n+1), (n+1, n+2)$,
      $(n+2, n+3), (n+3, n+4), (n+4, j)$;
    let $\tilde{V} := \tilde{V} \cup \{n+1, n+2, n+3, n+4\}$,
    $\tilde{R} := (\tilde{R} \setminus \{(i, j)\}) \cup \{(i, n+1), (n+2, n+3)$,
      $(n+4, j)\}$,
    $\tilde{B} := \tilde{B} \cup \{(n+1, n+2), (n+3, n+4)\}$;
    comment: update $\tilde{M}$ so as to define a
      Hamiltonian matching on the modified $\tilde{G}$,
      such that all cycles containing the new red
      edges are undirected with respect to every
      permutation associated with $\tilde{G}$ and $\tilde{M}$;
    let $\tilde{M} := (\tilde{M} \setminus \{(a, b)\}) \cup \{(n+1, n+4)$,
      $(n+2, a), (n+3, b)\}$;
    let $n := n + 4$
  end
end.

Figure 7 shows the effect of the replacement of the red edge (i, j) on graphs $\tilde{G}(\tilde{R} \cup \tilde{M})$ and $\tilde{G}(\tilde{B} \cup \tilde{M})$.

The fact that the dimension of $\tilde{G}$ is proportional to that of $G(\pi)$, and there is a one-to-one correspondence between alternating cycles and alternating cycle decompositions of $G(\pi)$ and $\tilde{G}$ is clear from the construction: $\tilde{G}$ is obtained from $G(\pi)$ by subdividing "twice" each red edge; see Remark 1.

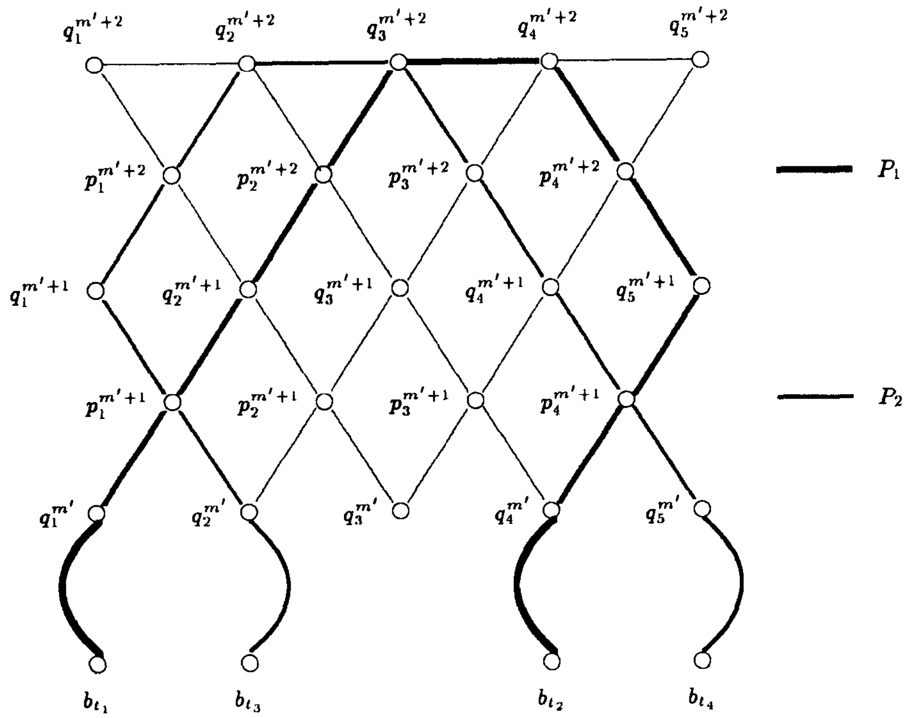

Figure 6: The inductive step in the proof of Claim 4, for $d = 8$ (and hence $s = 4, r = 2$) and $a = 1, b = 4$ — the thickest lines represent the path P_1 from b_{t_1} to b_{t_2}, and the remaining thick lines the path P_2 from b_{t_3} to b_{t_4}.

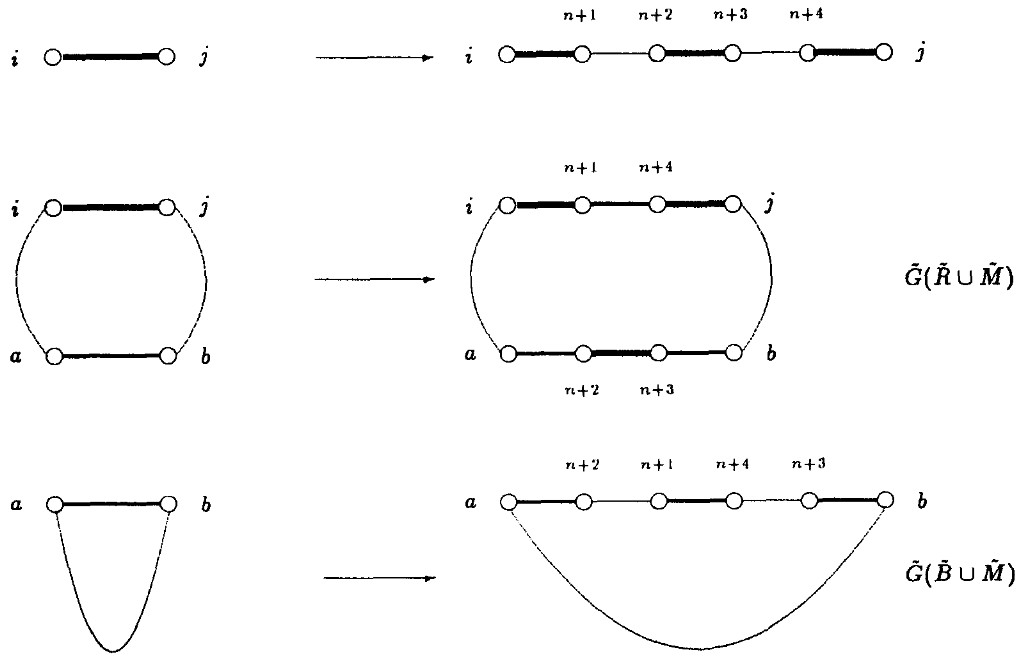

Figure 7: Replacement of the red edge (i, j) in procedure DOUBLE_SUBDIVISION — edges in $\tilde{B}$, $\tilde{M}$ and $\tilde{R}$, respectively, are drawn as lines of increasing thickness.

Claim 5 *The set $\tilde{M}$ is a Hamiltonian matching of $\tilde{G}$, and every cycle of $\tilde{G}$ is undirected with respect to every permutation associated with $\tilde{G}$ and $\tilde{M}$.*

Proof. First of all, at the end of each execution of the for each-do loop the updated $\tilde{M}$ is a Hamiltonian matching of the new $\tilde{G}$. The updating formula of $\tilde{M}$ corresponds indeed to an insertion of the new degree-2 nodes $n+1, n+2, n+3, n+4$ in the previous Hamiltonian circuits $\tilde{G}(\tilde{R} \cup \tilde{M})$ and $\tilde{G}(\tilde{B} \cup \tilde{M})$. In $\tilde{G}(\tilde{R} \cup \tilde{M})$, edges (i,j) and (a,b) are replaced by the paths $(i,n+1),(n+1,n+4),(n+4,j)$ and $(a,n+2),(n+2,n+3),(n+3,b)$, respectively, whereas in $\tilde{G}(\tilde{B} \cup \tilde{M})$ edge (a,b) is replaced by the path $(a,n+2),(n+2,n+1),(n+1,n+4),(n+4,n+3),(n+3,b)$; see Figure 7.

Every permutation $\tilde{\pi}$ associated with $\tilde{G}$ and $\tilde{M}$ defines an orientation of the edges in $\tilde{R}$. In particular, each edge is oriented according to the direction it is traversed by the Hamiltonian circuit $\tilde{G}(\tilde{R} \cup \tilde{M})$ starting from the node corresponding to $\tilde{\pi}_0$, and traversing a red edge first; see Section 2.

After the replacement of edge $r = (i,j)$, to every alternating cycle of $G(\pi)$ which contains this edge, there corresponds a cycle of $\tilde{G}$ that contains edges $(i,n+1),(n+1,n+2),(n+2,n+3),(n+3,n+4),(n+4,j)$. Furthermore, with respect to every $\tilde{\pi}$ associated with $\tilde{G}$ and $\tilde{M}$, either edge $(i,n+1)$ is oriented from i to $n+1$, edge $(n+4,j)$ from $n+4$ to j, and edge $(n+2,n+3)$ from $n+3$ to $n+2$, or, conversely, edge $(i,n+1)$ is oriented from $n+1$ to i, edge $(n+4,j)$ from j to $n+4$, and edge $(n+2,n+3)$ from $n+2$ to $n+3$. This property is mantained throughout the procedure, since the new Hamiltonian circuits $\tilde{G}(\tilde{R} \cup \tilde{M})$ are obtained from the previous ones by replacing edges with paths.

The above discussion shows that at the end of the procedure, when all red edges have been replaced, every cycle of $\tilde{G}$ is undirected with respect to any permutation associated with $\tilde{G}$ and $\tilde{M}$. □

Therefore, by Theorem 2, it is possible to compute the optimal solution value of MAX-ACD on $G(\pi)$ by solving the MIN-SBR instance defined by any $\tilde{\pi}$ associated with $\tilde{G}$ and $\tilde{M}$; namely $c(\pi) = c(\tilde{\pi}) = b(\tilde{\pi}) - d(\tilde{\pi})$. (As one might expect, it is also easy to determine the corresponding optimal alternating cycle decomposition of $G(\pi)$ from an optimal sequence of reversals needed to sort $\tilde{\pi}$ — we skip this part since it is not strictly necessary.) This completes the proof of Theorem 6.

5 Conclusions

We have shown that the problem of sorting a permutation by the minimum number of reversals is NP-hard, thus proving a conjecture made by Kececioglu and Sankoff [9] a few years ago, as soon as the problem came to the attention of computer scientists. It is therefore very unlikely that there exists a polynomial-time algorithm for the problem, despite its signed case was recently shown to be efficiently solvable by Hannenhalli and Pevzner [4]. As an immediate consequence, the problem of sorting *words* by reversals (see e.g. [10]) is NP-hard. This latter problem calls for a shortest sequence of reversals transforming a string $w_1 \ldots w_n$, such that $w_i \in \{1,\ldots,m\}$ for $i = 1,\ldots,n$ and $n \geq m$, into a sorted string $y_1 \ldots y_n$ where $y_i \leq y_{i+1}$ for $i = 1,\ldots,n$, and therefore is a clear generalization of MIN-SBR.

Our onstruction can easily be adapted to the case of *circular* permutations (see e.g. [8]), where also reversals of intervals of the type (j,i), $j > i$, are allowed, transforming $\pi = (\pi_1 \ldots \pi_n)$ into $(\pi_{i+j-1} \ldots \pi_{j+1}\, \pi_j\, \pi_{i+1} \ldots \pi_{j-1}\, \pi_{i+n}\, \pi_{i+n-1} \ldots \pi_{i+j})$, where all indexes must be taken modulo n.

Other extensions are probably limited to problems having a strong relationship with some graph decomposition problem of the same flavour as MAX-ACD. We do not know whether this is the case, e.g., for the problem of sorting a permutation by *prefix reversals* (also known as the *pancake flipping problem*), studied by Gates and Papadimitriou [3].

Acknowledgments

I am grateful to Giuseppe Lancia for helpful discussions on the subject. I also thank Sridhar Hannenhalli for pointing out reference [7], and Rob Irving and David Christie for sending me the paper.

References

[1] V. Bafna and P.A. Pevzner, "Genome Rearrangements and Sorting by Reversals", *SIAM Journal on Computing* 25 (1996) 272–289.

[2] A. Caprara, G. Lancia and S.K. Ng, "A Column-Generation Based Branch-and-Bound Algorithm for Sorting by Reversals", Working Paper (1995), DEIS, University of Bologna.

[3] W.H. Gates and C.H. Papadimitriou, "Bounds for Sorting by Prefix Reversals", *Discrete Mathematics* 27 (1979) 47–57.

[4] S. Hannenhalli and P.A. Pevzner, "Transforming Cabbage into Turnip (Polynomial Algorithm for Sorting Signed Permutations by Reversals)", *Proceedings of the 27th Annual ACM Symposium on the Theory of Computing* (1995) 178-187, ACM Press.

[5] S. Hannenhalli and P.A. Pevzner, "Reversals Do Not Cut Long Strips", Technical Report CSE-95-006, Department of Compter Science and Engineering, The Pennsylvania State University, February 1995.

[6] I. Holyer, "The NP-Completeness of Some Edge-Partition Problems", *SIAM Journal on Computing* 10 (1981) 713–717.

[7] R.W. Irving and D.A. Christie, "Sorting by Reversals: a Conjecture of Kececioglu and Sankoff", Working Paper (1996), Dept. of Computer Science, University of Glasgow.

[8] J. Kececioglu and D. Sankoff, "Efficient Bounds for Oriented Chromosome Inversion Distance", *Proceedings of 5th Annual Symposium on Combinatorial Pattern Matching*, Lecture Notes in Computer Science 807 (1994) 307–325, Spinger Verlag.

[9] J. Kececioglu and D. Sankoff, "Exact and Approximation Algorithms for Sorting by Reversals, with Application to Genome Rearrangement", *Algorithmica* 13 (1995) 180–210.

[10] P.A. Pevzner and M.S. Waterman, "Open Combinatorial Problems in Computational Molecular Biology", *Proceedings of the 3rd Israel Symposium on the Theory of Computing and Systems* (1995) 158–172, IEEE Computer Society Press.