Improved Inapproximability Results for the Shortest Superstring and Related Problems

Marek Karpinski* Richard Schmied†

Abstract

We develop a new method for proving explicit approximation lower bounds for the Shortest Superstring problem, the Maximum Compression problem, the Maximum Asymmetric TSP problem, the $(1,2)$–ATSP problem and the $(1,2)$–TSP problem improving on the best known approximation lower bounds for those problems.

1 Introduction

In the **Shortest Superstring (SSP) problem**, we are given a finite set S of strings and we would like to construct their shortest superstring, which is the shortest possible string such that every string in S is a proper substring of it.
The task of computing a shortest common superstring appears in a wide variety of application related to computational biology [L90]. Vassilevska [V05] proved that approximating the SSP problem with less than $1217/1216$ is $\mathbf{NP}$-hard. The currently best known approximation algorithm is due to Mucha [M12] and yields an approximation factor of $2\frac{11}{23}$.

In this paper, we prove that the Shortest Superstring problem is $\mathbf{NP}$-hard to approximate within any constant approximation ratio better than $333/332$.

In the **Traveling Salesperson (TSP) problem**, we are given a metric space (V,d) and the task consists of constructing a shortest tour visiting each vertex exactly once.

The TSP problem in metric spaces is one of the most fundamental $\mathbf{NP}$-hard optimization problems. The decision version of this problem was shown early to be $\mathbf{NP}$-complete by Karp [K72]. Christofides [C76] gave an algorithm approximating the TSP problem within $3/2$, i.e., an algorithm that produces a tour with length

*Dept. of Computer Science and the Hausdorff Center for Mathematics, University of Bonn. Supported in part by DFG grants and the Hausdorff Center grant EXC59-1. Email: marek@cs.uni-bonn.de

†Dept. of Computer Science, University of Bonn. Work supported by Hausdorff Doctoral Fellowship. Email: schmied@cs.uni-bonn.de

being at most a factor $3/2$ from the optimum. As for lower bounds, a reduction due to Papadimitriou and Yannakakis [PY93] and the PCP Theorem [ALM+98] together imply that there exists some constant, not better than $1+10^{-6}$, such that it is **NP**-hard to approximate the TSP problem with distances either one or two. For discussion of bounded metrics TSP, see also [T00]. The best known approximation lower bound for the general version of this problem is due to Lampis [L12]. He proved that the TSP problem is **NP**-hard to approximate with an approximation factor less than $185/184$. The restricted version of the TSP problem, in which the distance function takes values in $\{1,\ldots,B\}$, is referred to as the $(1,B)$*–TSP problem*. The $(1,2)$–TSP problem can be approximated in polynomial time with an approximation factor $8/7$ due to Berman and Karpinski [BK06]. On the other hand, Engebretsen and Karpinski [EK06] proved that it is **NP**-hard to approximate the $(1,B)$–TSP problem with an approximation factor less than $741/740$ for $B=2$ and $389/388$ for $B=8$.

In this paper, we prove that it is **NP**-hard to approximate the $(1,2)$–TSP problem with an approximation factor less than $535/534$.

In the **Asymmetric Traveling Salesperson (ATSP) problem**, we are given an asymmetric metric space (V,d), i.e., d is not necessarily symmetric, and we would like to construct a shortest tour visiting every vertex exactly once. The best known algorithm for the ATSP problem approximates the solution within $O(\log n/\log\log n)$, where n is the number of vertices in the metric space [AGM+10]. On the other hand, Papadimitriou and Vempala [PV06] proved that the ATSP problem is **NP**-hard to approximate with an approximation factor less than $117/116$. It is conceivable that the special cases with bounded metric are easier to approximate than the cases when the distance between two points grows with the size of the instance. Clearly, the $(1,B)$–ATSP problem, in which the distance function is taking values in the set $\{1,\ldots,B\}$, can be approximated within B by just picking any tour as the solution. When we restrict the problem to distances one and two, it can be approximated within $5/4$ due to Bläser [B04]. Furthermore, it is **NP**-hard to approximate this problem with an approximation factor better than $321/320$ [EK06]. For the case $B=8$, Engebretsen and Karpinski [EK06] constructed a reduction yielding the approximation lower bound $135/134$ for the $(1,8)$–ATSP problem.

In this paper, we prove that it is **NP**-hard to approximate the $(1,2)$–ATSP problem with an approximation factor less than $207/206$.

In the **Maximum Compression (MAX–CP) problem**, we are given a collection of strings $S=\{s_1,\ldots,s_n\}$. The task is to find a superstring for S with maximum compression, which is the difference between the sum of the lengths of the given strings and the length of the superstring.

In the exact setting, an optimal solution to the Shortest Superstring problem is an optimal solution to this problem, but the approximate solutions can differ significantly in the sense of approximation ratio. The Maximum Compression problem arises in various data compression problems (cf. [S88]). The best known approximation upper bound is $3/2$ [KLS+05] by reducing it to the MAX-ATSP problem, which is defined below.

On the approximation lower bound side, Vassilevska [V05] proved that it is $\mathbf{NP}$-hard to approximate this problem with a constant approximation factor better than $1072/1071$.

In this paper, we prove that approximating the Maximum Compression problem with an approximation ratio less than $204/203$ is $\mathbf{NP}$-hard.

In the **Maximum Asymmetric Traveling Salesperson (MAX–ATSP) Problem**, we are given a complete directed graph G and a weight function w assigning each edge of G a nonnegative weight. The task is to find a tour of maximum weight visiting every vertex of G exactly once .

This problem is well-known and motivated by several applications (cf. [BGS02]). A good approximation algorithm for the MAX–ATSP problem yields a good approximation algorithm for many other optimization problems such as the Shortest Superstring problem, the Maximum Compression problem and the $(1,2)$–ATSP problem. In particular, an α–approximation algorithm for the Max–ATSP problem implies an α–approximation algorithm for the Maximum Compression problem (cf. [KLS+05]).

The MAX–$(0,1)$–ATSP problem is the restricted version of the MAX–ATSP problem, in which the weight function w takes values in the set $\{0,1\}$. Vishwanathan [V92] constructed an approximation preserving reduction proving that any $(1/\alpha)$–approximation algorithm for the MAX–$(0,1)$–ATSP problem transforms in a $(2-\alpha)$– approximation algorithm for the $(1,2)$–ATSP problem. Due to the explicit approximation lower bound for the $(1,2)$–ATSP problem given in [EK06], it is $\mathbf{NP}$-hard to approximate the MAX–$(0,1)$–ATSP problem with an approximation factor less than $320/319$.

The best known approximation algorithm for the restricted version of this problem is due to Bläser [B04] and achieves an approximation ratio $5/4$.

For the general problem, Kaplan et al. [KLS+05] designed an algorithm for the MAX–ATSP problem yielding the best known approximation upper bound of $3/2$. Elbassioni, Paluch and v. Zuylen [EPZ12] gave a simpler approximation algorithm for the problem with the same approximation ratio.

In this paper, we prove that approximating the MAX–ATSP problem with an approximation ratio less than $204/203$ is $\mathbf{NP}$-hard.

2 Preliminaries

Throughout, for $i \in \mathbb{N}$, we use the abbreviation $[i]$ for the set $\{1,\ldots,i\}$. Given an finite alphabet Σ, a string is an element of Σ^*. Given a string v, we denote the length of v by $|v|$. For two strings x and y, we define the overlap of x and y, denoted $ov(x,y)$, as the longest suffix of x that is also a prefix of y. Furthermore, we define the prefix of x with respect to y, denoted $pref(x,y)$, as the string u with $x = u\, ov(x,y)$.

In this paper, an instance (V, d) of the $(1,2)$–ATSP problem is specified by means of a directed graph $D_V = (V, A)$, where $(x, y) \in A$ if and only if $d(x, y) = 1$. In addition, we refer to an arc $(x, y) \in V \times V$ as a z-arc if $d(x, y) = z \in \{1, 2\}$. In order to specify an instance of the $(1,2)$-TSP problem, we will use undirected graphs.

3 Hybrid Problem

Berman and Karpinski [BK99] introduced the following Hybrid problem and proved that this problem is **NP**-hard to approximate with some constant.

Definition 1 (Hybrid problem)**.** *Given a system of linear equations mod 2 containing n variables, m_2 equations with exactly two variables, and m_3 equations with exactly three variables, find an assignment to the variables that satisfies as many equations as possible.*

The following result is due to Berman and Karpinski [BK99].

Theorem 1 ([BK99])**.** *For any constant $\delta \in (0, 1/2)$, there exists instances of the Hybrid problem $\mathcal{H}(\nu)$ with 42ν variables, 60ν equations with exactly two variables, and 2ν equations with exactly three variables such that: (i) Each variable occurs exactly three times. (ii) Either there is an assignment to the variables that leaves at most $\delta \cdot \nu$ equations unsatisfied, or else every assignment to the variables leaves at least $(1 - \delta)\nu$ equations unsatisfied. (iii) It is **NP**-hard to decide which of the two cases in item (ii) above holds. (iv) An optimal assignment to the variables in $\mathcal{H}(\nu)$ can be transformed in polynomial time into an optimal assignment satisfying all 60ν equations with two variables in $\mathcal{H}(\nu)$.*

The instances of the Hybrid problem produced in Theorem 1 have an even more special structure, which we are going to describe. The equations containing three variables are of the form $x \oplus y \oplus z = \{0, 1\}$. These equations stem from the Theorem of Håstad [H01] dealing with the hardness of approximating equations with exactly three variables. We refer to it as the MAX–E3–LIN problem, which can be seen as a special instance of the Hybrid problem.

Theorem 2 ([H01])**.** *For any constant $\delta \in (0, 1/2)$, there exists systems of linear equations mod 2 with $2 \cdot \nu$ equations and exactly three unknowns in each equation such that:*
(i) Each variable in the instance occurs a constant number of times, half of them negated and half of them unnegated.
(ii) Either there is an assignment satisfying all but at most $\delta \cdot \nu$ equations, or every assignment leaves at least $(1 - \delta)\nu$ equations unsatisfied.
*(iii) It is **NP**-hard to distinguish between these two cases.*

Let us describe briefly the reduction from the MAX–E3–LIN problem to the Hybrid problem. For a detailed description, we refer to [BK99], [BK03] and [K01]. For every variable x of the original instance I of the MAX–E3–LIN problem, we introduce a corresponding set of variables V_x. If the variable x occurs t_x times in I, then, V_x contains $n = 7t_x$ new variables $x_1, \dots, x_n$. The variables contained in $\{x_{7 \cdot i} \mid i \in [t_x]\}$ are called *contact variables*, whereas the remaining variables in V_x are called *checker variables*. All variables in V_x are connected by equations of the form $x_i \oplus x_{i+1} = 0$ with $i \in [n-1]$ (cirle equations) and $x_1 \oplus x_{7t_x} = 0$ (circle border equation). In addition, there exists equations of the form $x_i \oplus x_j = 0$ with $\{i,j\} \in M_x$ (matching equations), where the set M_x induces a perfect matching on the indexset of checker variables. In the remainder, we refer to this construction as the circle $\mathcal{C}_x$ containing the variables $x_i \in V_x$. Every occurrence of the variable x in an equation with three variables in I is replaced by a corresponding contact variable in V_x. Accordingly, every variable in the corresponding instance $I_{\mathcal{H}}$ of the Hybrid problem occurs exactly three times.

4 Our Results

We now formulate our results.

Theorem 3. *Let $\mathcal{H}$ be an instance of the Hybrid problem with n circles, 60ν equations with two variables and 2ν equations with exactly three variables satisfying the properties described in Theorem 1.*

1. *It is possible to construct in polynomial time an instance $(V_{\mathcal{H}}, d_{\mathcal{H}})$ of the $(1,2)$-ATSP problem such that:*
(i) *If there exists an assignment ϕ to the variables of $\mathcal{H}$ which leaves at most $\delta\nu$ equations unsatisfied for some $\delta \in (0,1)$, then, there exist a tour with length at most $3 \cdot 60\nu + 13 \cdot 2\nu + n + 1 + \delta\nu$.*
(ii) *From every tour in $(V_{\mathcal{H}}, d_{\mathcal{H}})$ with length $206 \cdot \nu + n + 1 + \delta\nu$, we can construct in polynomial time an assignment that leaves at most $\delta \cdot \nu$ equations in $\mathcal{H}$ unsatisfied.*

2. *It is possible to construct in polynomial time an instance $(V_{\mathcal{H}}, d_{\mathcal{H}})$ of the $(1,2)$-TSP problem such that:*
(i) *If there exists an assignment ϕ to the variables of $\mathcal{H}$ which leaves at most $\delta\nu$ equations unsatisfied for some $\delta \in (0,1)$, then, there exist a tour with length at most $8 \cdot 60\nu + 27 \cdot 2\nu + 3(n+1) + 1 + \delta\nu$.*
(ii) *From every tour σ in $(V_{\mathcal{H}}, d_{\mathcal{H}})$ with length $534 \cdot \nu + 3(n+1) + 1 + \delta\nu$, we can construct in polynomial time an assignment that leaves at most $\delta \cdot \nu$ equations in $\mathcal{H}$ unsatisfied.*

3. *It is possible to construct in polynomial time an instance $\mathcal{S}_{\mathcal{H}}$ of the Shortest Superstring problem such that:*
(i) *If there exists an assignment ϕ to the variables of $\mathcal{H}$ which leaves at most $\delta\nu$ equations unsatisfied for some $\delta \in (0,1)$, then, there exist a superstring s_ϕ for $\mathcal{S}_{\mathcal{H}}$ with length at most $5 \cdot 60\nu + 16 \cdot 2\nu + 7n + \delta\nu$.*
(ii) *From every superstring s for $S_{\mathcal{H}}$ with length $|s| = 332\nu + u + 7n + \delta\nu$, we can construct in polynomial time an assignment to the variables of $\mathcal{H}$ that leaves at most $\delta\nu$ equations in $\mathcal{H}$ unsatisfied.*

4. *It is possible to construct in polynomial time an instance $\mathcal{S}_{\mathcal{H}}$ of the Maximum Compression problem such that:*
(i) *If there exists an assignment ϕ to the variables of $\mathcal{H}$ which leaves at most $\delta\nu$*

equation unsatisfied for some $\delta \in (0,1)$, then, there exist a superstring s_ϕ for $\mathcal{S}_\mathcal{H}$ with compression at least $3 \cdot 60\nu + 12 \cdot 2\nu + 5n - \delta\nu$.
(ii) *From every superstring s for $S_\mathcal{H}$ with compression $204\nu+5n-\delta\cdot\nu$, we can construct in polynomial time an assignment to the variables of $\mathcal{H}$ that leaves at most $\delta \cdot \nu$ equations in $\mathcal{H}$ unsatisfied.*

The former theorem can be used to derive an explicit approximation lower bound for the $(1,2)$–ATSP problem.

Corollary 1. *For every $\epsilon > 0$, it is $\mathbf{NP}$-hard to approximate the $(1,2)$–ATSP problem within any constant approximation ratio better than $207/206 - \epsilon$.*

Proof. First of all, we choose $k \in \mathbb{N}$ and $\delta > 0$ such that $\frac{207-\delta}{206+\delta+12/k} \geq \frac{207}{206} - \epsilon$ holds. Given an instance $\mathcal{E}_3$ of the MAX-E3-LIN problem, we generate k copies of $\mathcal{E}_3$ and produce an instance $\mathcal{H}$ of the Hybrid problem. Then, we construct the corresponding instance $(V_\mathcal{H}, d_\mathcal{H})$ of the $(1,2)$–ATSP problem with the properties described in Theorem 3.1. We conclude according to Theorem 1 that there exist a tour in $(V_\mathcal{H}, d_\mathcal{H})$ with length at most $206\nu k + \delta\nu k + (n+1) \leq (206 + \delta + \frac{2n}{k\nu})\nu k \leq (206 + \delta + \frac{2 \cdot 6}{k})\nu k$ or the length of a tour in $(V_\mathcal{H}, d_\mathcal{H})$ is bounded from below by $206\nu k + (1-\delta)\nu k + n + 1 \geq (206 + (1-\delta))\nu k \geq (207-\delta)\nu k$. From Theorem 1, we know that the two cases above are $\mathbf{NP}$-hard to distinguish. Hence, for every $\epsilon > 0$, it is $\mathbf{NP}$-hard to find a solution to the Shortest Superstring problem with an approximation ratio $\frac{207-\delta}{206+\delta+12/k} \geq \frac{207}{206} - \epsilon$. □

Analogously, Theorem 3 can be used to derive approximation lower bounds for the other problems summarized in Figure 1. The explicit approximation lower bound for the Max–ATSP problem is obtained by using a well-known approximation preserving reduction from the Maximum Compression problem to the MAX–ATSP problem (cf. [KLS+05]).

Problem	Our Results	Previously known
$(1,2)$–ATSP	$207/206$	$321/320$ [EK06]
$(1,2)$–TSP	$535/534$	$741/740$ [EK06]
MAX–ATSP	$204/203$	$320/319$ [EK06]
MAX–CP	$204/203$	$1072/1071$ [V05]
SSP	$333/332$	$1217/1216$ [V05]

Figure 1: Comparison of our results to previously known explicit approximation lower bounds.

For other details and explicit approximation lower bounds for related problems, see [KS11] and [KS12].

5 The $(1,2)$–ATSP problem

Given an instance of the Hybrid problem $\mathcal{H}$, we want to transform $\mathcal{H}$ into an instance of the $(1,2)$–ATSP problem. Fortunately, the special structure of the linear equations in the Hybrid problem is particularly well-suited for our reduction, since a part of the equations with two variables form a cycle and every variable occurs exactly three times. The main idea of our reduction is to make use of the special structure of the circles in $\mathcal{H}$. Every circle $\mathcal{C}_l$ in $\mathcal{H}$ corresponds to a subgraph D_l in the instance $D_{\mathcal{H}}$ of the $(1,2)$-ATSP problem. Moreover, D_l forms almost a cycle. An assignment to the variable x^l will have a natural interpretation in this reduction. The parity of x^l corresponds to the direction of movement in D_l of the underlying tour. The circle graphs $D_1, \ldots, D_n$ of $D_{\mathcal{H}}$ are connected and build together the

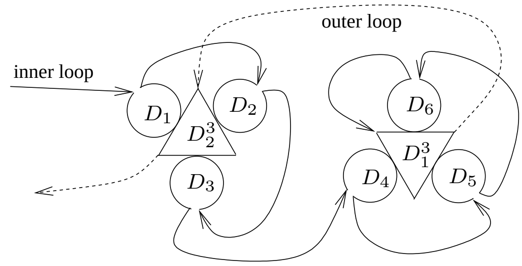

Figure 2: An illustration of $D_{\mathcal{H}}$ and a tour in $D_{\mathcal{H}}$.

inner loop of $D_{\mathcal{H}}$ (Figure 2). Every variable x_i^l in a circle $\mathcal{C}_l$ possesses an associated parity graph P_i^l (Figure 3(a)) in D_l as a subgraph. The two natural ways to traverse a parity graph will be called $0/1$-traversals (Figure 3(b)&(c)) and correspond to the parity of the variable x_i^l. Some of the parity graphs in D_l are also contained in graphs D_c^3 (Figure 5 and Figure 6 for a more detailed view) corresponding to equations with three variables of the form $g_c^3 \equiv x \oplus y \oplus z = 0$. (We may assume that equations with three variables are of the form $x \oplus y \oplus z = 0$ or $\bar{x} \oplus y \oplus z = 0$ due to the transformation $\bar{x} \oplus y \oplus z = 0 \equiv x \oplus y \oplus z = 1$.) These graphs are connected and build the *outer loop* of $D_{\mathcal{H}}$. The outer loop of the tour checks whether the $0/1$-traversals of the parity graphs correspond to a satisfying assignment of the equations with three variables. If an underlying equation is not satisfied by the assignment defined via $0/1$-traversals of the associated parity graphs, it will be punished by using a costly arc with distance 2.

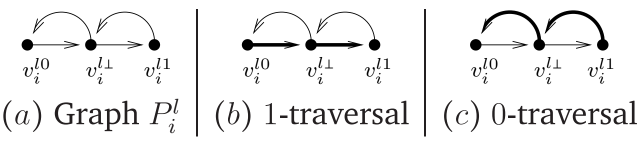

Figure 3: Traversals of the parity graph P_i^l. Traversed arcs are illustrated by thick arrows.

Constructing $D_{\mathcal{H}}$ from the Instance $\mathcal{H}$

Given a instance of the Hybrid problem $\mathcal{H}$, we are going to construct the corresponding instance $D_{\mathcal{H}}$ of the $(1,2)$-ATSP problem. For every type of equation in $\mathcal{H}$, we will introduce a specific graph or a specific way to connect the so far constructed subgraphs. In particular, we will distinguish between graphs corresponding to circle equations, matching equations, circle border equations and equations with three variables. First of all, we introduce graphs corresponding to the variables in $\mathcal{H}$.

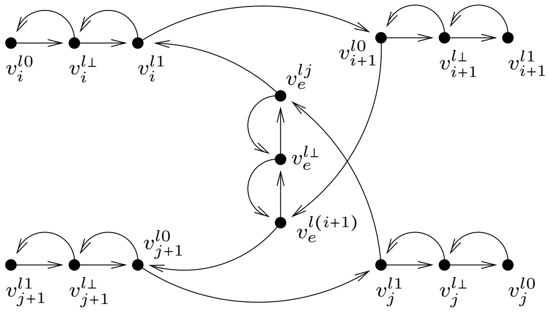

Figure 4: Connecting the parity graph P_e^l.

Variable Graphs: For every variable x_i^l in $\mathcal{H}$, we introduce the parity graph P_i^l consisting of the vertices $\{v_i^{l1}, v_i^{l\perp}, v_i^{l0}\}$ and is displayed in Figure 3(a).

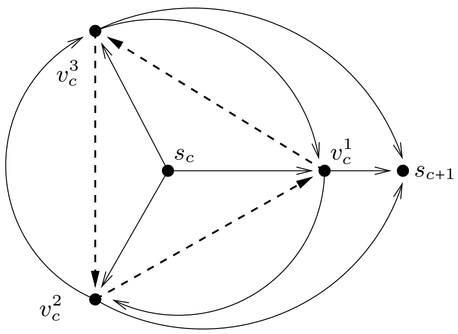

Figure 5: Gadget for $x \oplus y \oplus z = 0$.

Matching and Circle Equations: Let $\mathcal{H}$ be an instance of the hybrid problem, $\mathcal{C}_l$ a circle in $\mathcal{H}$ and M_l the associated perfect matching. Furthermore, let $x_i^l \oplus x_j^l = 0$ with $e = \{i, j\} \in M_l$ and $i < j$ be a matching equation. Due to the construction of $\mathcal{H}$, the circle equations $x_i^l \oplus x_{i+1}^l = 0$ and $x_j^l \oplus x_{j+1}^l = 0$ are both contained in $\mathcal{C}_l$. Then, we introduce the associated parity graph P_e^l consisting of the vertices v_e^{lj}, $v_e^{l\perp}$ and $v_e^{l(i+1)}$. In addition, we connect the parity graphs P_i^l, P_{i+1}^l, P_j^l, P_{j+1}^l and P_e^l as depicted in Figure 4.

Equations with Three Variables: Let $g_c^3 \equiv x_i^l \oplus x_j^s \oplus x_t^k = 0$ be an equation with three variables in $\mathcal{H}$. Then, we introduce the graph D_c^3 (Figure 5) corresponding to the equation g_c^3. The graph D_c^3 includes the vertices s_c, v_c^1, v_c^2, v_c^3 and s_{c+1}. Engebretsen and Karpinski [EK06] used this graph in their reduction and proved the following statement.

Proposition 1 ([EK06]). *There is a Hamiltonian path from s_c to s_{c+1} in the graph displayed in Figure 5 if and only if an even number of dashed arcs is traversed.*

This construction is extended by replacing the dashed arcs with the parity graphs P_e^l, P_b^s and P_a^k, where $e = \{i, i+1\}$, $b = \{j, j+1\}$ and $a = \{t, t+1\}$. In Figure 6, we display D^3_c with its connections to the graph corresponding to $x^l_i \oplus x^l_{i+1} = 0$. (In case of $g^3_c \equiv \bar{x}^l_i \oplus x^s_j \oplus x^u_k = 0$, we create (v^{l1}_i, v^{l1}_e), (v^{l0}_{i+1}, v^{l1}_i) and (v^{l0}_e, v^{l0}_{i+1}) instead.)

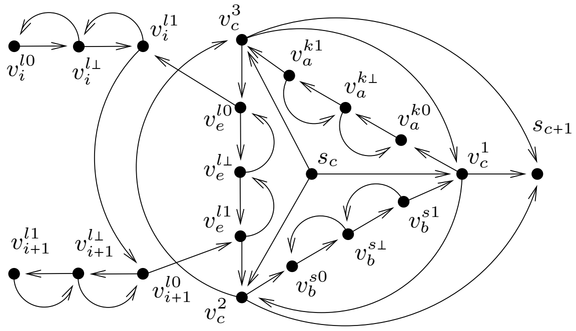

Figure 6: The graph D^3_c corresponding to $g^3_c \equiv x^l_i \oplus x^s_j \oplus x^u_t = 0$ connected to graphs corresponding to $x^l_i \oplus x^l_{i+1} = 0$.

Circle Border Equations: Let $\mathcal{C}_l$ and $\mathcal{C}_{l+1}$ be circles in $\mathcal{H}$. In addition, let $x^l_1 \oplus x^l_n = 0$ be the circle border equation of $\mathcal{C}_l$. Recall that x^l_n also occurs in an equation g^3_c with three variables in $\mathcal{H}$. Assuming $g^3_c \equiv x^l_n \oplus y \oplus z = 0$, we introduce the vertex b_l, b_{l+1} and the parity graph $P^l_{\{n,1\}}$. Then, we create $(b_l, v^{l1}_{\{n,1\}})$, $(v^{l0}_{\{n,1\}}, v^{l1}_n)$, (b_l, v^{l0}_1), (v^{l0}_1, b_{l+1}) and (v^{l1}_n, b_{l+1}). (In case of $g^3_c \equiv \bar{x}^l_n \oplus y \oplus z = 0$, we add (b_l, v^{l0}_1), (v^{l0}_1, b_{l+1}), (b_l, v^{l1}_n), $(v^{l1}_n, v^{l0}_{\{n,1\}})$ and $(v^{l1}_{\{n,1\}}, b_{l+1})$ instead.) Finally, we set $b_{m+1} = s_1$, where s_1 is the starting vertex of D^3_1.

Constructing a Tour from an Assignment

Let $\mathcal{H}$ be an instance of the Hybrid problem consisting of the circles $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$, 60ν equations with 2 variables and 2ν equations with three variables. Given an assignment ϕ to the variables of $\mathcal{H}$ leaving $\delta \cdot \nu$ equations unsatisfied for a constant $\delta \in (0,1)$, we are going to construct the associated Hamiltonian tour σ_ϕ in $D_\mathcal{H}$. According to Theorem 1, we may assume that all equations with 2 variables in $\mathcal{H}$ are satisfied by ϕ. Thus, all variables associated to a circle have the same value. Then, the Hamiltonian tour σ_ϕ in $D_\mathcal{H}$ starts at the vertex b_1. From a high-level view, σ_ϕ traverses all graphs corresponding to the equations associated with the circle $\mathcal{C}_1$ using the $\phi(x^1_1)$-traversal of all parity graphs corresponding to circle equations of $\mathcal{C}_1$ ending with the vertex b_2. Successively, it passes all graphs for each circle in $\mathcal{H}$ until it reaches the vertex $b_{m+1} = s_1$ as s_1 is the starting vertex of the graph D^3_1.

At this point, the tour begins to traverse the remaining graphs D^3_c, which are simulating the equations with three variables in $\mathcal{H}$. By now, some of the parity graphs appearing in graphs D^3_c already have been traversed in the *inner loop* of σ_ϕ. The *outer loop* checks whether for each graph D^3_c, an even number of parity graphs has been traversed in the inner loop. In every situation, in which ϕ does not satisfy the underlying equation, the tour needs to use a 2-arc.

Constructing an Assignment from a Tour

Let $\mathcal{H}$ be an instance of the Hybrid problem, $D_\mathcal{H} = (V_\mathcal{H}, A_\mathcal{H})$ the associated instance of the $(1,2)$-ATSP problem and σ a tour in $D_{\mathcal{H}}$. We are going to define the corresponding assignment ψ_σ to the variables in $\mathcal{H}$. In addition, we establish a connection between the length of σ and the number of satisfied equations by ψ_σ. First of all, we introduce the notion of consistent tours.

Definition 2 (Consistent Tour)**.** *Let $\mathcal{H}$ be an instance of the Hybrid problem and $D_{\mathcal{H}}$ the associated instance of the $(1,2)$–ATSP problem. A tour in $D_{\mathcal{H}}$ is called consistent if the tour uses only $0/1$-traversals of all in $D_{\mathcal{H}}$ contained parity graphs.*

Due to the following proposition, we may assume that the underlying tour is consistent.

Proposition 2. *Let $\mathcal{H}$ be an instance of the Hybrid problem and $D_{\mathcal{H}}$ the associated instance of the $(1,2)$–ATSP problem. Any tour σ in $D_{\mathcal{H}}$ can be transformed in polynomial time into a consistent tour with at most the same length as σ.*

Proof. For every parity graph contained in $D_{\mathcal{H}}$, it can be seen by considering all possibilities exhaustively that any tour in $D_{\mathcal{H}}$ that is not using the corresponding $0/1$-traversals can be modified into a tour with at most the same number of 2-arcs. The less obvious cases are shown in the full version [KS12]. □

Let us define the corresponding assignment ψ_σ given a tour σ in $D_{\mathcal{H}}$.

Definition 3 (Assignment ψ_σ)**.** *Let $\mathcal{H}$ be an instance of the Hybrid problem, $D_{\mathcal{H}} = (V_{\mathcal{H}}, A_{\mathcal{H}})$ the associated instance of the $(1,2)$-ATSP problem. Given a consistent tour σ in $D_{\mathcal{H}}$, the corresponding assignment ψ_σ is defined as $\psi_\sigma(x_i^l) = 1$ if σ uses a 1-traversal of P_i^l, and 0 otherwise.*

Let us start with the analysis. In the remainder, we assume that the underlying tour σ is consistent.

Matching Equations: Given the equations $x_i \oplus x_{i+1} = 0$, $x_i \oplus x_j = 0$, $x_j \oplus x_{j+1} = 0$ and a tour σ, we are going to analyze the relation between the length of the tour and the number of satisfied equations by ψ_σ.

1.Case ($\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 0$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 0$ & $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 0$): Given $\psi_\sigma(x_i) = \psi_\sigma(x_j) = \psi_\sigma(x_j) = \psi_\sigma(x_{j+1}) = 1$, the cost of a tour traversing this part of $D_{\mathcal{H}}$ can be bounded from below by 5. In this case, σ contains (v_i^{l1}, v_{i+1}^{l0}), (v_j^{l1}, v_e^{lj}), $(v_e^{lj}, v_e^{l\perp})$, $(v_e^{l\perp}, v_e^{l(i+1)})$ and $(v_e^{l(i+1)}, v_{j+1}^{l0})$. The case $\psi_\sigma(x_i)$=$\psi_\sigma(x_{i+1})$=$\psi_\sigma(x_j)$=$\psi_\sigma(x_{j+1})$=0 can be discussed analogously. In both cases, we obtain the local length 5 for this part of σ while ψ_σ satisfies all 3 equations.
2.Case ($\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 0$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 1$ & $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 0$): In both cases, we associate only the cost of one 2-arc yielding a lower bound of 6 on the local length, which corresponds to the fact that ψ_σ leaves the equation $x_i \oplus x_j = 0$ unsatisfied. Note that a similar situation holds in case of $\psi_\sigma(x_i) = \psi_\sigma(x_{i+1}) = 0$ and $\psi_\sigma(x_j) = \psi_\sigma(x_{j+1}) = 1$.

3. Case ($\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 0$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 0$ & $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 1$):

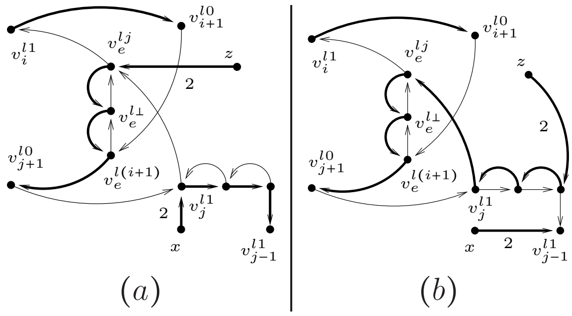

Figure 7: 5.Case with $\psi_\sigma(x_i) = \psi_\sigma(x_{i+1}) = 1$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 1$.

Given $\psi_\sigma(x_i) = \psi_\sigma(x_{i+1}) = 1$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 0$, we are forced to use two 2-arcs increasing the cost by 2. Thus, we obtain a lower bound of $4 + 2$. The case $\psi_\sigma(x_i) = \psi_\sigma(x_{i+1}) = 0$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 1$ can be analyzed analogously. A similar argumentation holds for $\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 1$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 0$ and $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 0$.

4.Case ($\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 1$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 0$ & $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 1$): Given $\psi_\sigma(x_i) \neq \psi_\sigma(x_{i+1}) = 0$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 0$, we are forced to use four 2-arcs in order to connect all vertices. Consequently, it yields the lower bound of 7. The case, in which $\psi_\sigma(x_i) \neq \psi_\sigma(x_{i+1}) = 0$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 0$ holds, can be discussed analogously.

5.Case ($\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 0$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 1$ & $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 1$): Let the tour σ be characterized by $\psi_\sigma(x_i) = \psi_\sigma(x_{i+1}) = 1$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 1$. Let us assume that σ uses the arc (v_i^{l1}, v_{i+1}^{l0}). The corresponding situation is illustrated in Figure 7(a). We transform σ such that it traverses the parity graph P_j^l in the other direction and obtain $\psi_\sigma(x_j) = 1$. This transformation induces a tour with at most the same cost. On the other hand, the corresponding assignment ψ_σ satisfies at least $2 - 1$ more equations since $x_j^l \oplus x_{j-1}^l = 0$ might get unsatisfied. In this case, we associate the local costs of 6 with σ. In the other cases, in which $\psi_\sigma(x_i) = \psi_\sigma(x_{i+1}) = 0$ & $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 0$ or $\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 1$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 1$ & $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 0$ holds, we may argue similarly.

6.Case ($\psi_\sigma(x_i) \oplus \psi_\sigma(x_{i+1}) = 1$, $\psi_\sigma(x_i) \oplus \psi_\sigma(x_j) = 1$ and $\psi_\sigma(x_j) \oplus \psi_\sigma(x_{j+1}) = 1$): Given a tour σ with $\psi_\sigma(x_i) \neq \psi_\sigma(x_{i+1}) = 1$ and $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 0$, we transform σ such that it traverses the parity graph P_j^l in the opposite direction meaning $\psi_\sigma(x_j) = 0$. This transformation enables us to use the arc (v_{j+1}^{l0}, v_j^{l1}). Furthermore, it yields at least one more satisfied equation in $\mathcal{H}$. In order to connect the remaining vertices, we are forced to use at least two 2-arcs. In summary, we associate the local length 7 with this situation in conformity with the at most 2 unsatisfied equations by ψ_σ. The case, in which $\psi_\sigma(x_i) \neq \psi_\sigma(x_{i+1}) = 0$ & $\psi_\sigma(x_j) \neq \psi_\sigma(x_{j+1}) = 1$ holds, can be discussed analogously.

In summary, we obtain the following statement.

Proposition 3. *Let* $E = \{x_i^l \oplus x_{i+1}^l = 0,\ x_i^l \oplus x_j^l = 0,\ x_j^l \oplus x_{j+1}^l = 0\}$ *be a subset of* $\mathcal{H}$ *with* $\{i,j\} \in M_l$*. Then, it is possible to transform in polynomial time a given tour* σ *passing through the graphs corresponding to* $g \in E$ *into a tour* π *that has local cost* $(5+\alpha)$ *and the number of unsatisfied equations in* E *by* ψ_π *is at most* α*.*

Equations with Three Variables: Let $g_c^3 \equiv x_i^l \oplus x_j^s \oplus x_k^r = 0$ be an equation with three variables in $\mathcal{H}$. Furthermore, let $\mathcal{C}_l$ be a circle in $\mathcal{H}$ and $x_i^l \oplus x_{i+1}^l = 0$ a circle equation. For notational simplicity, we set $e = \{i, i+1\}$. We are going to analyze the number of satisfied equations by ψ_σ in dependence to the local length of σ in the graphs P_i^l, P_{i+1}^l, P_e^l and D_c^3. First, we transform the tour traversing the graphs P_i^l, P_{i+1}^l and P_e^l such that it uses the $\psi_\sigma(x_i^l)$-traversal of P_e^l. Afterwards, due to the construction of D_c^3 and Proposition 1, the tour can be transformed such that it has local length of $3 \cdot 3 + 4$ if it passes an even number of parity graphs $P \in \{P_e^l, P_{\{j,j+1\}}^s, P_{\{k,k+1\}}^r\}$ while using a simple path through D_c^3. Otherwise, it yields a local length of $13 + 1$.

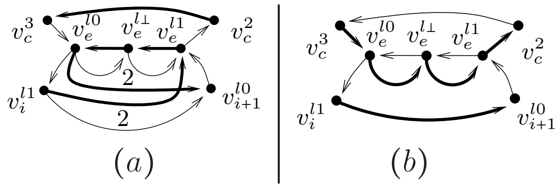

Figure 8: Case $\psi_\sigma(x_i^l) = 1$ and $\psi_\sigma(x_{i+1}^l) = 1$

Let us start to analyze the local cost of σ in the graph corresponding to $x_i^l \oplus x_{i+1}^l = 0$:

1. Case ($\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_{i+1}^l) = 0$**):** In both cases, we transform the tour such that it uses the $\psi_\sigma(x_i^l)$-traversal of P_e^l without increasing its length. Exemplary, we display such a scenario for the case ($\psi_\sigma(x_i^l) = 1$ & $\psi_\sigma(x_{i+1}^l) = 1$) in Figure 8$(a)$ and (b) (transformed tour in Figure 8(b)). For both cases, we associate a lower bound of 1 on the local cost.

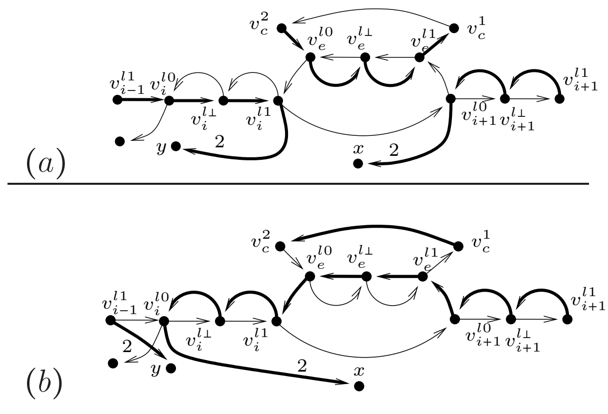

Figure 9: Case ($\psi_\sigma(x_i^l) = 1$, $\psi_\sigma(x_{i+1}^l) = 0$ and $\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_j^s) \oplus \psi_\sigma(x_k^r) = 1$).

2.Case ($\psi_\sigma(x_i^l) = 1$ **&** $\psi_\sigma(x_{i+1}^l) = 0$**):** Let us assume that $\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_j^s) \oplus \psi_\sigma(x_k^r) = 0$ holds. Due to Proposition 1, it is possible to transform the tour such that it uses the 0-traversal of the parity graph P_e^l without increasing the length. In the other case, i.e. $\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_j^s) \oplus \psi_\sigma(x_k^r) = 1$, we will change the value of $\psi_\sigma(x_i^l)$ achieving in this way at least $2-1$ more satisfied equation. Let us examine the scenario and the corresponding transformation in Figure 9(a) and (b), respectively. Accordingly, the tour uses the 0-traversal of the parity graph P_e^l, which enables σ to pass the parity check in D_c^3. In both cases, we obtain the local length of 2 in conformity with the at most one unsatisfied equation by ψ_σ.

3.Case ($\psi_\sigma(x_i^l) = 0$ & $\psi_\sigma(x_{i+1}^l) = 1$): Assuming $\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_j^s) \oplus \psi_\sigma(x_k^r) = 0$, the tour will be modified such that the parity graphs P_i^l and P_e^l are traversed in the same direction. Since we have $\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_j^s) \oplus \psi_\sigma(x_k^r) = 0$, we are able to uncouple the parity graph P_e^l from the tour σ through D_c^3 without increasing the length of σ. Assuming $\psi_\sigma(x_i^l) \oplus \psi_\sigma(x_i^l) \oplus \psi_\sigma(x_i^l) = 1$, we transform σ such that the parity graph P_e^l is traversed when σ is passing through D_c^3 meaning $v_c^3 \to v_e^{l0} \to v_e^{l\perp} \to v_e^{l1} \to v_c^2$ is a part of the tour. In addition, we change the value of $\psi_\sigma(x_i^l)$ yielding at least $2-1$ more satisfied equations. In both cases, we associate the local length of 2 with σ. On the other hand, ψ_σ leaves at most one equation unsatisfied.

The construction for $x_k^r \oplus x_{k+1}^r = 0$ and $x_j^s \oplus x_{j+1}^s = 0$ can be analyzed analogously yielding the following statement.

Proposition 4. *Let $E = \{x_i^l \oplus x_j^s \oplus x_k^r = 0, x_i^l \oplus x_{i+1}^l = 0, x_j^s \oplus x_{j+1}^s = 0, x_k^r \oplus x_{k+1}^r = 0\}$ be a subset of $\mathcal{H}$. Then, it is possible to transform in polynomial time a given tour σ passing through the graph corresponding to $g \in E$ into a tour π that has local length $(4+3\cdot 3+3+\alpha)$ and the number of unsatisfied equations in E by ψ_π is at most α.*

The construction for circle border equations can be analyzed similarly to the the construction for equations with three variables. We obtain the following statement.

Proposition 5. *Let $x_1^l \oplus x_n^l = 0$ be a circle border in $\mathcal{H}$. Then, it is possible to transform in polynomial time a given tour σ passing through the graph corresponding to $x_1^l \oplus x_n^l = 0$ into a tour π that has local length at least 2 if $x_1^l \oplus x_n^l = 0$ is satisfied by ψ_π, and at least 3 otherwise.*

Thus far, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. Let $\mathcal{H}$ be an instance of the Hybrid problem consisting of n circles, 60ν equations with two variables and 2ν equations with three variables. Then, we construct in polynomial time the corresponding instance $D_{\mathcal{H}}$ of the $(1,2)$-ATSP problem.

(i) Let ϕ be an assignment to the variables in $\mathcal{H}$ leaving $\delta\nu$ equations in $\mathcal{H}$ unsatisfied for a constant $\delta \in (0,1)$. Then, it is possible to construct in polynomial time a tour with length at most $3\cdot 60\nu + (4+3\cdot 3)\cdot 2\nu + n + 1 + \delta\nu$.

(ii) Let σ be a tour in $D_{\mathcal{H}}$ with length $206\nu + n + 1 + \delta\nu$. Due to Proposition 2 we may assume that σ uses only $0/1$-traversals of every parity graph included in $D_{\mathcal{H}}$.

According to Definition 3, we associate the corresponding assignment ψ_σ with the underlying tour σ. Recall from Proposition 3 – 5 that it is possible to convert σ in polynomial time into a tour π without increasing the length such that ψ_π leaves at most $\delta\nu$ equations in $\mathcal{H}$ unsatisfied. □

6 The $(1,2)$-TSP Problem

In order to prove Theorem 3.2, we apply the reduction method used in the previous section to the $(1,2)$–TSP problem. As for parity gadget, we use the graph displayed in Figure 10 with its corresponding traversals. The traversed edges are illustrated by thick lines.

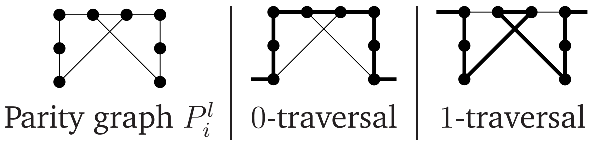

Figure 10: $0/1$-Traversals of the graph P_i^l.

Let $\mathcal{H}$ be an instance of the hybrid problem and $x_i^l \oplus x_j^l = 0$ a contained matching equation. Let $x_i^l \oplus x_{i+1}^l = 0$ and $x_j^l \oplus x_{j+1}^l = 0$ be the corresponding circle equations. Then, we connect the associated parity graphs P_i^l, P_{i+1}^l, $P_{\{i,j\}}^l$, P_j^l and P_{j+1}^l as displayed in Figure 11.

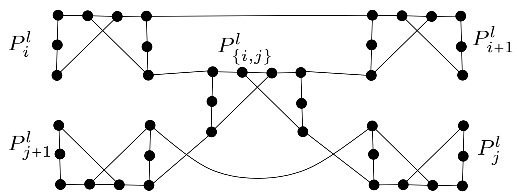

Figure 11: Graphs corresponding to equations $x_i^l \oplus x_j^l = 0$, $x_i^l \oplus x_{i+1}^l = 0$ and $x_j^l \oplus x_{j+1}^l = 0$.

For equations with three variables $g_c^3 \equiv x \oplus y \oplus z = 0$ in $\mathcal{H}$, we use the graph G_c^3 depicted in Figure 12. For this graph, Engebretsen and Karpinski [EK06] proved the following statement.

Proposition 6 ([EK06]). *There is a simple path from s_c to s_{c+1} in Figure 12 containing v_c^1 and v_c^2 if and only if an even number of parity graphs is traversed.*

Let $\mathcal{C}_l$ and $\mathcal{C}_{l+1}$ be circles in $\mathcal{H}$. Let $x_1^l, \ldots, x_m^l$ be the variables contained in $\mathcal{C}_l$. For the circle border equation of $\mathcal{C}_l$, we introduce the path $p_l = b_l^1 - b_l^2 - b_l^3$ and the parity graph $P_{\{1,m\}}^l$. In addition, we connect b_l^3 and b_{l+1}^1 to the parity graphs P_1^l, P_n^l and $P_{\{1,m\}}^l$ in a similar way as in the reduction from the Hybrid problem to the $(1,2)$–ATSP problem. Let $\mathcal{C}_n$ be the last circle in $\mathcal{H}$. Then, we introduce the path $p_{n+1} = b_{n+1}^1 - b_{n+1}^2 - s_1$, where s_1 is a vertex of the graph G_1^3 associated to the equation g_1^3 with three variables in $\mathcal{H}$. This is the whole description of the

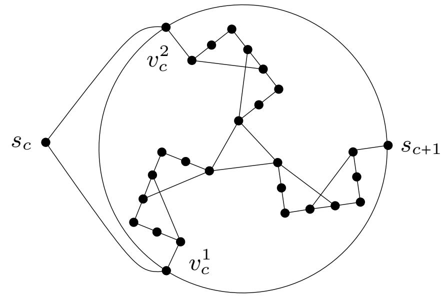

Figure 12: Graph G_c^3 corresponding to $x \oplus y \oplus z = 0$.

corresponding graph $G_{\mathcal{H}}$.

We are ready to give the proof of Theorem 3.2.

Proof of Theorem 3.2. Given $\mathcal{H}$ an instance of the Hybrid problem consisting of n circles, 60ν equations with two variables and 2ν equations with three variables, we construct in polynomial time the associated instance $G_{\mathcal{H}}$ of the $(1,2)$–TSP problem.

(i) Given an assignment ϕ to the variables of $\mathcal{H}$ leaving $\delta\nu$ equations unsatisfied in $\mathcal{H}$ for a constant $\delta \in (0,1)$, then, there is a tour in $G_{\mathcal{H}}$ with length at most $8 \cdot 60\nu + (3 \cdot 8 + 3) \cdot 2\nu + 3 \cdot (n+1) + 1 + \delta\nu$.

(ii) On the other hand, if we are given a tour σ in $G_{\mathcal{H}}$ with length $534\nu + 3(n + 1) + 1 + \delta\nu$, it is possible to transform σ in polynomial time into a tour σ' such that it uses only $0/1$-traversals of all contained parity graphs in $G_{\mathcal{H}}$ without increasing the length. Some cases are displayed in in the full version [KS12]. The remaining transformations described in the previous section can be straightforwardly adapted to the symmetric case since they only work with the connection edges of the parity graphs. Moreover, we are able to construct in polynomial time an assignment to the variables of $\mathcal{H}$, which leaves at most $\delta\nu$ equations in $\mathcal{H}$ unsatisfied. □

7 The Shortest Superstring Problem

In order to apply the arguments given in Section 5, we first describe a well-known reduction from the SSP problem to the ATSP problem. Let S be a collection of strings over Σ such that no string is a proper substring of another string in S. Then, we define an instance of the ASTP problem by (V_S, d_S), where $V_S = S \cup \{\Gamma\}$ with $\Gamma \notin \Sigma$ and $d_S(s_i, s_j) = |pref(s_i, s_j)|$ for all $s_i, s_j \in V_S$. Note that we can construct from a shortest tour in (V_S, d_S) of length $\ell + 1$ a shortest superstring for S of length ℓ.

We first give a high-level view of the reduction in order to build some intuition. Let $x_i \oplus x_{i+1} = 0$ be a circle equation of an instance $\mathcal{H}$ of the Hybrid problem such that x_i and x_{i+1} appears only in equations with two variables. The parity gadget of $x_i \oplus x_{i+1} = 0$ consists of two strings s_i^1 and s_i^2, which can be overlapped by two letters in two different ways. These two alignments, called $0/1$-alignments, define the assigned value to x_i. For any other string s in the corresponding instance $S_{\mathcal{H}}$, both s_i^1 and s_i^2 can be aligned with s by at most 1 letter. Then, a tour in $(V_{S_{\mathcal{H}}}, d_{S_{\mathcal{H}}})$ is called consistent with the parity gadget for $x_i \oplus x_{i+1} = 0$ if the tour contains the arc (s_i^1, s_i^2) or (s_i^2, s_i^1), i.e. a $0/1$-alignment of the strings s_i^1 and s_i^2. Moreover, it is not hard to see that a tour σ in $(V_{S_{\mathcal{H}}}, d_{S_{\mathcal{H}}})$ can be transformed into a tour π that is consistent with the parity gadget for $x_i \oplus x_{i+1} = 0$ without increasing the length.

Let us start with the description of $S_{\mathcal{H}}$. For every equation $g \in \mathcal{H}$, we define a set $S(g)$ of corresponding strings.

Strings for Circle Border Equations: Given a circle $\mathcal{C}_x$ and its border equation $x_1 \oplus x_n = 0$, we introduce six associated strings. Recall that x_n appears in an equation g_j^3 with three variables. The strings differ by the type of equation $x_n \oplus y \oplus z = \{0,1\}$. We begin with the case $x_n \oplus y \oplus z = 0$: The string $L_x C_x^l$ is used as the initial part of the superstring corresponding to this circle, whereas $C_x^r R_x$ is used as the end part. Furthermore, we introduce strings that represent an assignment that sets either the variable x_1 to 0 or the variable x_n to 1. The corresponding two strings are $C_x^l x_1^{m0} x_n^{l1} C_x^r$ and $x_n^{l1} C_x^r C_x^l x_1^{m0}$. Finally, we introduce $C_x^l x_1^{r1} x_n^{m0} C_x^r$ and $x_n^{m0} C_x^r C_x^l x_1^{r1}$ having a similar interpretation. The following two alignments are called the 0-*alignment* of the four strings. $\overbrace{C_x^l x_1^{m0} x_n^{l1} C_x^r} C_x^l x_1^{m0}$ and $\overbrace{x_n^{m0} C_x^r C_x^l x_1^{r1}} x_n^{m0} C_x^r$. On the other hand, we the define the 1-*alignment* as $x_n^{l1} C_x^r C_x^l x_1^{m0} x_n^{l1} C_x^r$ and $C_x^l x_1^{r1} x_n^{m0} C_x^r C_x^l x_1^{r1}$. For equations of the form $g_j^3 \equiv x_n \oplus y \oplus z = 1$, we use $L_x C_x^l$, $C_x^r R_x$, $C_x^l x_1^{m0} x_n^{m1} C_x^r$, $x_n^{m1} C_x^r C_x^l x_1^{m0}$ $C_x^l x_1^{r1} x_n^{l0} C_x^r$ and $x_n^{l0} C_x^r C_x^l x_1^{r1}$ (The bars above indicate original strings).

Strings Corresponding to Matching Equations: Let $x_i \oplus x_j = 0$ be a matching equation in $\mathcal{H}$ with $i < j$. Then, we introduce $x_j^{r0} x_j^{l0} x_i^{r1} x_i^{l1}$ and $x_i^{r1} x_i^{l1} x_j^{r0} x_j^{l0}$. We define the 0-*alignment* and 1-*alignment* as $x_j^{r0} x_j^{l0} x_i^{r1} x_i^{l1} x_j^{r0} x_j^{l0}$ and $x_i^{r1} x_i^{l1} x_j^{r0} x_j^{l0} x_i^{r1} x_i^{l1}$, respectively.

Strings for Equations with Three Variables: Let g_j^3 be an equation with three variables in $\mathcal{H}$. For every equation g_j^3, we define two corresponding sets $S^\alpha(g_j^3)$ and $S^\beta(g_j^3)$, both containing three strings. Finally, the set $S(g_j^3)$ is defined as the union $S^A(g_j^3) \cup S^B(g_j^3)$. An equation of the form $x \oplus y \oplus z = 0$ is represented by $S^\alpha(g_j^3)$ containing the strings $x^{r1\alpha} x^{l1} y^{r1} y^{l1}$, $y^{r1} y^{l1} x^{m0} C_j$, $x^{m0} C_j x^{r1\alpha} x^{l1}$.
The strings included in $S^\beta(g_j^3)$ are $x^{r1\beta} x^{l1} z^{r1} z^{l1}$, $z^{r1} z^{l1} C_j x^{m0}$, $C_j x^{m0} x^{r1\beta} x^{l1}$. The strings in $S^\alpha(g_j^3)$ can be overlapped by two letters in a cyclic fashion to obtain three different constellations. A suitable constellation can be used to connect with $0/1$-alignments corresponding to circle equations. The string $x^{r1\alpha} x^{l1} y^{r1} y^{l1} x^{m0} C_j x^{r1\alpha} x^{l1}$ represents the assignment $x = 1$, whereas the constellation $y^{r1} y^{l1} x^{m0} C_j x^{r1\alpha} x^{l1} y^{r1} y^{l1}$ is representing $y = 1$. Finally, the string $x^{m0} C_j x^{r1\alpha} x^{l1} y^{r1} y^{l1} x^{m0} C_j$ can be used to

overlap with $C_j x^{m0} x^{r1\beta} x^{l1} z^{r1} z^{l1} C_j x^{m0}$ consisting of the strings in $S^{\beta}(g_j^3)$ in the case ($x = 0$, $y = 0$, and $z = 0$). $z^{r1} z^{l1} C_j x^{m0} x^{r1\beta} x^{l1} z^{r1} z^{l1}$ is used in the case $z = 1$. The sets $S^{\alpha}(g_j^3)$ and $S^{\beta}(g_j^3)$ representing equations of the form $g_j^3 \equiv x \oplus y \oplus z = 1$ can be constructed analogously.

Strings for Circle Equations: Let $\mathcal{C}_x$ be a circle in $\mathcal{H}$ and M_x its associated matching. Furthermore, let $\{i,j\}$ and $\{i+1,j'\}$ be both contained in M_x. We assume that $i < j$. Then, we introduce the corresponding strings for $x_i \oplus x_{i+1} = 0$. If $i+1 < j'$, we have $x_i^{m0} x_{i+1}^{m0} x_i^{l1} x_{i+1}^{r1}$ and $x_i^{l1} x_{i+1}^{r1} x_i^{m0} x_{i+1}^{m0}$. We define the 0-*alignment* and 1-*alignment* as $x_i^{m0} x_{i+1}^{m0} x_i^{l1} x_{i+1}^{r1} x_i^{m0} x_{i+1}^{m0}$ and $x_i^{l1} x_{i+1}^{r1} x_i^{m0} x_{i+1}^{m0} x_i^{l1} x_{i+1}^{r1}$, respectively. In the case ($i+1 > j'$), we use $x_i^{m0} x_{i+1}^{r0} x_i^{l1} x_{i+1}^{m1}$ and $x_i^{l1} x_{i+1}^{m1} x_i^{m0} x_{i+1}^{r0}$. The strings for the remaining cases can be defined analogously.
If the variable x_i is contained in an equation $x_i \oplus y \oplus z = 0$, we introduce three strings for the equation $x_{i-1} \oplus x_i = 0$: $x_{i-1}^{l1} x_i^{r1\beta} x_{i-1}^{l1} x_i^{r1\alpha}$, $x_{i-1}^{l1} x_i^{r1\alpha} x_{i-1}^{m0} x_i^{m0}$ and $x_{i-1}^{m0} x_i^{m0} x_{i-1}^{l1} x_i^{r1\beta}$. The strings for the case $g_j^3 \equiv x \oplus y \oplus z = 1$ can be constructed analogously.

We are ready to give the proof of Theorem 3.3 and 3.4.

Proof of Theorem 3.3 and 3.4. Given $\mathcal{H}$ an instance of the Hybrid problem consisting of n circles, 60ν equations with two variables and 2ν equations with three variables, we construct in polynomial time the associated instance $S_{\mathcal{H}}$.

(i) Given an assignment ϕ to the variables of $\mathcal{H}$ leaving $\delta\nu$ equations with three variables unsatisfied for a constant $\delta \in (0,1)$, we are going to construct a superstring for $\mathcal{S}_{\mathcal{H}}$. Since we may assume that ϕ assigns to every variable x_i^l associated to a circle $\mathcal{C}_l$ the same value, we use the $\phi(x_1^l)$-alignment of the strings corresponding to equations contained in $\mathcal{C}_l$. These fragments can be overlapped by one letter from both sides. For equations with three variables, we use the appropriate constellations. It yields an overlap of 5 character if the underlying equation is satisfied, and 4 otherwise. Therefore, the resulting superstring has a length at most $60\nu \cdot 5 + 7 \cdot n + 16 \cdot 2\nu + \delta\nu$ and a compression at least $60\nu \cdot 8 + 12 \cdot n + 28 \cdot 2\nu - (7n + 332\nu + \delta\nu) = 5n + 60\nu \cdot 3 + 12 \cdot 2\nu - \delta\nu$.

(ii) Let s be a superstring for $\mathcal{S}_{\mathcal{H}}$ having length $7 \cdot n + 332\nu + \delta\nu$ or compression $5n + 60\nu \cdot 3 + 12 \cdot 2\nu - \delta\nu$. Recall that s can be transformed into a superstring for $S_{\mathcal{H}}$ using 0/1-alignments without increasing its length. The argumentation given in Section 5 for the $(1,2)$–ATSP problem can be adapted to analyze these fragments (0/1-alignments) and the corresponding instance $(V_{S_{\mathcal{H}}}, d_{S_{\mathcal{H}}})$ of the ATSP problem. Therefore, we define an assignment to the variables in $\mathcal{H}$ according to the 0/1-alignments used in s leaving at most $\delta\nu$ equations in $\mathcal{H}$ unsatisfied. □

References

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy, *Proof Verification and the Hardness of Approximation Problems*, J. ACM 45 (1998), pp. 501–555.

[AGM+10] A. Asadpour, M. Goemans, A. Madry, S. Gharan and A. Saberi, *An $O(\log n/\log\log n)$-Approximation Algorithm for the Asymmetric Traveling Salesman Problem*, Proc. 21st SODA (2010), pp. 379–389.

[BGS02] A. Barvinok, E. Gimadi and A. Serdyukov, *The Maximum Traveling Salesman Problem*, in: *The Traveling Salesman Problem and Its Variations*, pp. 585–607, G. Gutin and A. Punnen, eds., Kluwer, 2002.

[BK99] P. Berman and M. Karpinski, *On Some Tighter Inapproximability Results*, Proc. 26th ICALP (1999), LNCS 1644, 1999, pp. 200–209.

[BK03] P. Berman and M. Karpinski, *Improved Approximation Lower Bounds on Small Occurrence Optimization*, ECCC TR03-008, 2003.

[BK06] P. Berman and M. Karpinski, *8/7-approximation algorithm for $(1,2)$-TSP*, 17th SODA (2006), pp. 641–648.

[B04] M. Bläser, *A 3/4-Approximation Algorithm for Maximum ATSP with Weights Zero and One*, Proc. 8th APPROX-RANDOM (2004), pp. 61–71.

[C76] N. Christofides, *Worst-Case Analysis of a New Heuristic for the Traveling Salesman Problem*, Technical Report CS-93-13, Carnegie Mellon University, Pittsburgh, 1976.

[EPZ12] K. Elbassioni, K. Paluch and A. v. Zuylen, *Simpler Approximation of the Maximum Asymmetric Traveling Salesman Problem*, Proc. 29th STACS 2012, pp. 501–506.

[EK06] L. Engebretsen and M. Karpinski, *TSP with Bounded Metrics*, J. Comput. Syst. Sci. 72 (2006), pp. 509–546.

[H01] J. Håstad, *Some Optimal Inapproximability Results*, J. ACM 48 (2001), pp. 798–859.

[KLS+05] H. Kaplan, M. Lewenstein, N. Shafrir and M. Sviridenko, *Approximation Algorithms for Asymmetric TSP by Decomposing Directed Regular Multigraphs*, J. ACM 52 (2005), pp. 602–626.

[K72] R. Karp, *Reducibility Among Combinatorial Problems*, Compl. of Computer Computations, pp. 85–103, 1972.

[K01] M. Karpinski, *Approximating Bounded Degree Instances of NP-hard Problems*, Proc. 13th Symp. on Fundamentals of Comp. Theory, LNCS 2138, Springer, pp. 24–34, 2001.

[KS11] M. Karpinski and R. Schmied, *Improved Lower Bounds for the Shortest Superstring and Related Problems* CoRR abs/1111.5442, 2011.

[KS12] M. Karpinski and R. Schmied, *On Approximation Lower Bounds for TSP with Bounded Metrics*, CoRR abs/1201.5821, 2012.

[L12] M. Lampis, *Improved Inapproximability for TSP*, CoRR 1206.2497, 2012.

[L90] M. Li, *Towards a DNA Sequencing Theory*, Proc. 31th FOCS (1990), pp. 125–134.

[M12] M. Mucha, *Lyndon Words and Short Superstrings*, CoRR abs/1205.6787, 2012.

[PV06] C. Papadimitriou and S. Vempala, *On the Approximability of the Traveling Salesman Problem*, Combinatorica 26 (2006), pp. 101–120.

[PY93] C. Papadimitriou and M. Yannakakis, *The traveling Salesman Problem with Distances One and Two*, Math. Oper. Res. 18 (1993), pp. 1–11.

[S88] J. A. Storer, *Data Compression: Methods and Theory*, Computer Science Press, 1988.

[T00] L. Trevisan, *When Hamming Meets Euclid: The Approximability of Geometric TSP and Steiner Tree*, SIAM J. Comput. 30 (2000), pp. 475–485.

[V05] V. Vassilevska, *Explicit Inapproximability Bounds for the Shortest Superstring Problem*, Proc. 30th MFCS (2005), LNCS 3618, 2005, pp. 793–800.

[V92] S. Vishwanathan, *An Approximation Algorithm for the Asymmetric Travelling Salesman Problem with Distances One and Two*, Inf. Process. Lett. 44 (1992), pp. 297–302.