Electronic Colloquium on Computational Complexity, Report No. 73 (2002)

Approximation Hardness for Small Occurrence Instances of NP-Hard Problems

Miroslav Chlebík
MPI for Mathematics in the Sciences, D-04103 Leipzig

Janka Chlebíková*
CAU, Institut für Informatik und Praktische Mathematik, D-24098 Kiel
jch@informatik.uni-kiel.de

Abstract

The paper contributes to the systematic study (started by Berman and Karpinski) of explicit approximability lower bounds for small occurrence optimization problems. We present parametrized reductions for some packing and covering problems, including 3-Dimensional Matching, and prove the best known inapproximability results even for highly restricted versions of them. For example, we show that it is NP-hard to approximate Max-3DM within $\frac{141}{140}$ even on instances with exactly two occurrences of each element.

Our reductions from Max-E3-Lin-2 depend on parameters of amplifiers that provably exist, we need not restrict ourselves to amplifiers that can be constructed efficiently. New structural results which improve the known bounds for 3-regular amplifiers and hence the inapproximability results for numerous small occurrence problems studied by Berman and Karpinski in the article "On some tighter inapproximability results" (ECC/65, 1998) are also presented.

Keywords: Approximation Algorithms, Approximation Hardness, NP-hard combinatorial optimization problems

1 Introduction

The research on the hardness of bounded occurrence (resp. bounded degree) optimization problems is focused on the case of very small value of the bound parameter. For example, considerable effort of Berman and Karpinski (see [2] and references therein) has gone into the developing of a new method of reductions for determining the inapproximability of Maximum Independent Set and Minimum Node Cover in graphs of maximum degree 3 or 4. The study of problems with small value of a bound parameter is very well motivated; they are useful as intermediate steps in reductions to many important problems.

Related Work. This work is a contribution to a systematic study of bounded occurrence optimization problems with applications to other optimization problems.

Tight hardness results for optimization problems usually build on the PCP characterization of NP. But for many small parameter problems they can be hardly achieved

*The second author has been supported by the EU-Project ARACNE, Approximation and Randomized Algorithms in Communication Networks, HPRN-CT-199-00112.

ISSN 1433-8092

directly. Rather, one has to use an expander/amplifier method. A restricted version of MAXIMUM LINEAR EQUATIONS over $\mathbb{Z}_2$ with exactly 3 variables per equation seems to be very efficient as the canonical problem for starting gap-preserving reductions. The problem has a simple 2-approximation algorithm. But Håstad ([4]), based on his verifier, proved hard gap result implying inapproximability within $2-\varepsilon$ even when restricted to instances where each variable appears at most $2^{\varepsilon^{-d}}$ times, for some explicit absolute constant d and $\varepsilon \in (0, \frac{1}{2})$. Thus, the optimal hard gap result extends, with negligible loss, to the case where each variable occurs bounded (even constant) number of times. It can be used to prove a hard gap result, say, for 3-occurrence case using amplifiers. But the size of an amplifier is related to the number of occurrences of each variable in Håstad's result. Since we reduce from the problem in which this is a constant independent on an input, an amplifier is a constant element in our reduction. Therefore we need not restrict ourselves to amplifiers that can be constructed in polynomial time, to prove NP-hard gap result. Any (even nonconstructive) proof of existence of amplifiers (or expanders) with better parameters than those currently known implies the *existence* of (deterministic, polynomial) gap-preserving reductions leading to better inapproximability result.

This is our paradigm towards tighter inapproximability results inspired by the paper of Papadimitriou and Vempala on Traveling Salesman problem ([10]), that we have already used for Steiner Tree problem in [6].

Our results. As a starting point to our gap preserving reductions we state in Theorems 2 and 3 the versions of NP-hard gap results on bounded (constant) occurrence MAX-E3-LIN-2. Their weaker forms are known to experts and have been already used ([10], [3], [11], [6]).

We prove structural results about 3-regular amplifiers which play a crucial role in proving explicit inapproximability results for bounded occurrence optimization problems. A $(2,3)$-graph $G=(V,E)$, i.e. a graph with nodes only of degree 2 (*Contacts*) and 3 (*Checkers*), is an *amplifier* (more precisely, a 3-regular amplifier for its checker nodes) if for very $A \subseteq V$ either $|\mathrm{Cut}\, A| \geq |Contacts \cap A|$, or $|\mathrm{Cut}\, A| \geq |Contacts \setminus A|$. The parameter $\tau(G) := \frac{|V|}{|Contacts|}$ measures the quality of an amplifier; the smaller, the better. We are able to prove for many bounded occurrence problems a tight correspondence between $\tau_* := \inf\{\gamma : \tau(G) < \gamma$ for infinity many amplifiers $G\}$ and inapproximability results. There is a substantial gap between the best upper and lower bounds on parameters of amplifiers and expanders. Berman and Karpinski have proved that $\tau_* \leq 7$. In this paper we show slight improvement, $\tau_* \leq 6.9$, based on the results from [2] and our structural amplifier analysis. But further improvements of estimates on amplifier parameters of randomly generated graphs, pushing the method to its limits, is in progress.

Consider a probabilistic model of generating $(2,3)$-graphs (with sets of *Contacts* and *Checkers* fixed and large enough). In such situation we need to estimate the probability that the random $(2,3)$-graph is an amplifier. It fails to be an amplifier if there are "bad sets" which violate the condition from the definition. For a single bad set it is simple to estimate the probability that this bad set doesn't occur. At the heart of the problem lies the question of how to estimate the union bound over all bad sets in better way, than adding all single probabilities. Our structural analysis of amplifiers allowed to produce significantly smaller list of bad sets which are sufficient to test for the presence. That leads to more efficient counting methods showing that most of random $(2,3)$-graphs with $\tau(G) \geq 6.9$ are amplifiers. This result combined with reductions used by Berman and Karpinski immediately improve (slightly) lower bounds for numerous bounded occurrence optimization problems.

We developed our method of parametrized reductions (a parameter is a fixed amplifier) to prove inapproximability results for E3-Occ-Max-3-Lin-2 problem, and problems Max Independent Set and Min Node Cover on 3-regular graphs (Section 4). The similar method can be applied to all problems studied in [2] (with modification of amplifiers to bipartite-like for Max Cut) to improve the lower bound on approximability. Similarly, for the problem TSP with distances 1 and 2 ([3]).

We include also reductions to some packing and covering problems to state the best known inapproximability results on (even highly restricted) version of Triangle Packing, 3-Set Packing, and 3-Set Covering problems (Section 4). These reductions are quite straightforward from 3-MIS, resp. 3-NC and they are included as inspiration to the new reduction for 3-Dimensional Matching problem (Max-3DM) (Section 5). APX-completeness of the problem has been well known even on instances with at most 3 occurrences of any element, but our lower bound applies to the instances with exactly 2 occurrences. We do not know about any previous hardness result on the problem with the bound 2 on the number of occurrences of elements. The best to our knowledge upper and lower approximation bounds for theses packing and covering problems are summarized in the following table. The upper bounds are from [5].

Problem	Approx. upper bound	Param. lower bound	Approx. lower bound ($\tau_* = 6.9$)
Max Triangle Packing	$1.5 + \varepsilon$	$1 + \frac{1}{18\tau_* + 13}$	1.0072
3-Set Packing	$1.5 + \varepsilon$	$1 + \frac{1}{18\tau_* + 13}$	1.0072
3-Set Covering	$1.4 + \varepsilon$	$1 + \frac{1}{18\tau_* + 18}$	1.0070
Max 3-DM	$1.5 + \varepsilon$	$1 + \frac{1}{18\tau_* + 15}$	1.0071

Our inapproximability result on Max-3DM can be applied to obtain explicit lower bounds for several problems of practical interest, e.g. scheduling problems, some (even highly restricted) cases of Generalized Assignment problem, or the other more general packing problems.

Gap Problems. To achieve strong (or even tight) inapproximability results for an optimization problem P, it is useful to study various partial decision subproblems of P dealing with instances for which the optimum is promised to be either "very high" or "very low". We associate to P a *promise problem*, depending on a pair of polynomial time computable functions l, h of the size of an input instance and satisfying $0 < l(\cdot) < h(\cdot)$.

Definition 1 *Let P be an optimization problem, and let $0 < l(\cdot) < h(\cdot)$ be polynomial time computable functions of the size of an input instance. Define H ("high" instances) and L ("low" instances), the subsets of the set of instances of the problem P as follows: $H = \{I : \mathrm{OPT}(I) > h(|I|)\}$, $L = \{I : \mathrm{OPT}(I) < l(|I|)\}$. The $\langle l, h\rangle$-gap version of P is a decision problem whose task is to decide whether a given instance x is an H or an L instance.*

It is rather obvious that showing NP hardness of the $\langle l, h\rangle$-gap version of P implies that it is NP-hard to approximate P within h/l.

2 Inapproximability of subproblems of Max-E3-Lin-2

In proving inapproximability results we produce new "hard gaps" from those already known using gap-preserving reductions and their compositions. We start with a restricted version of MAX-LINEAR equations over $\mathbb{Z}_2$, namely MAX-E3-LIN-2.

Definition 2 MAX-E3-LIN-2 *is the following optimization problem: Given a system I of linear equation over $\mathbb{Z}_2$, with exactly 3 (distinct) variables in each equation. The goal is to maximize, over all assignments ψ to the variables, the ratio $\frac{N(I,\psi)}{|I|}$, where $N(I,\psi)$ is the number of equations of I satisfied by ψ.*

Notation. We use the notation Ek-OCC-MAX-Ed-LIN-2 for the same maximization problem, where each equation has exactly d variables (hence Ed) and each variable occurs exactly k times (hence Ek). If we drop an "E" than we have "at most d variables" and/or "at most k occurrences".

Denote $Q(\varepsilon, k)$ the following restricted version of MAX-E3-LIN-2: *Given an instance of* Ek-OCC-MAX-E3-LIN-2. *The problem is to decide if the fraction of more than $(1-\varepsilon)$ or less than $(\frac{1}{2}+\varepsilon)$ of all equations is satisfied by the optimal (i.e. maximizing) assignment.*

MAX-E3-LIN-2 problem plays a role of the canonical problem in our reductions due to its simple structure and efficiency. In many gap-preserving reductions it is easier to start with instances of MAX-E3-LIN-2 where each variable appears bounded (or even constant) number of times. The corresponding NP-hard gap results even for such restricted versions of the problem are due to Håstad [4]. They imply the Theorem 1 below.

Theorem 1 *For every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is an integer $k_0(\varepsilon)$ such that the partial decision subproblem $Q(\varepsilon, k_0(\varepsilon))$ of* MAX-E3-LIN-2 *is* NP*-hard.*

Proof. Strictly speaking, Håstad's proof of Theorem 5.4 in [4] shows that for every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is an integer k ($k \leq 2^{\varepsilon^{-d}}$ for some explicit absolute constant d) such that $\langle \frac{1}{2}+\varepsilon, 1-\varepsilon \rangle$-gap version of MAX-E3-LIN-2 is NP-hard even if restricted to instances with occurrences of each variable bounded by k. But, keeping $k = k(\varepsilon)$ as above fixed, there is a polynomial time reduction from the version with occurrences of variables bounded by k, to the version of MAX-E3-LIN-2 with exactly $k_0(\varepsilon) := (k!)^3$ occurrences of each variable, such that it preserves optimal value, as follows:

Given an instance I with the set of variables $\mathcal{V}(I)$, where each variable $x \in \mathcal{V}(I)$ occurs $s(x)$ times, $1 \leq s(x) \leq k$. We transform I to a new instance I' in which each variable of $\mathcal{V}(I')$ occurs exactly $(k!)^3$ times, and with $\mathrm{OPT}(I') = \mathrm{OPT}(I)$. For each variable $x \in \mathcal{V}(I)$ we put into $\mathcal{V}(I')$ exactly $s(x)$ variants of x, namely x_i, $i = 1, 2, \ldots, s(x)$.

For each equation

$$x + y + z = a, \quad a \in \{0, 1\} \tag{E}$$

of I, we put into I' a multiset (E') containing exactly $\frac{(k!)^3}{s(x)s(y)s(z)}$ copies of an equation

$$x_i + y_j + z_l = a,$$

for each triple (i,j,l) with $1 \le i \le s(x)$, $1 \le j \le s(y)$ and $1 \le l \le s(z)$. If we have done this for all equations (E) of I, the result is the new instance I' with exactly $(k!)^3$ occurrences of each variable of $\mathcal{V}(I')$.

Clearly $\mathrm{OPT}(I') \ge \mathrm{OPT}(I)$, as we can assign to all variants x_i $(1 \le i \le s(x))$ for each variable $x \in \mathcal{V}(I)$ the same value as to x. The key point to prove $\mathrm{OPT}(I') = \mathrm{OPT}(I)$ is to show that $\mathrm{OPT}(I')$ is achieved on assignments to variables $\mathcal{V}(I')$ with the property that for each variable $x \in \mathcal{V}(I)$ all the variants x_i $(1 \le i \le s(x))$ take the same value. To show this, it is sufficient to make the following observation: Given an assignment ϕ to variables $\mathcal{V}(I')$ and one fixed variable of $\mathcal{V}(I)$, say x. One can modify ϕ on variants of x only, making all variants of x take the same value, without decreasing the number of satisfied equations by ϕ.

This follows since keeping assignment to the variants of the other variables fixed, we can change the status of equations of I' containing variants of x. They are now of the form $x_i = a$ $(a \in \{0,1\}, 1 \le i \le s(x))$.

But our construction ensures that all variants of x appear in exactly the same contexts, and in this context one of the two values (0 or 1) of the variable is at least as good as the other. Hence we can make all variants of x take this value.

It follows that optimal values and hard gaps are preserved by the above transformation $I \mapsto I'$. □

For our applications the strengthening contained in Theorems 2 and 3 are more convenient.

Theorem 2 *For every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is a constant $k(\varepsilon)$ such that for every integer $k \ge k(\varepsilon)$ the partial decision subproblem $Q(\varepsilon, k)$ of* MAX-E3-LIN-2 *is NP-hard.*

Proof. Theorem 1 says that for every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is an integer $k_0(\varepsilon)$ such that the gap version $Q(\varepsilon, k_0(\varepsilon))$ of MAX-E3-LIN-2 is NP-hard. Repeating each equation of an instance of $Q(\varepsilon, k_0(\varepsilon))$ exactly r times (r being a fixed positive integer) shows that $Q(\varepsilon, k)$ with k being a multiple of $k_0(\varepsilon)$, is NP-hard as well. To prove NP-hardness for all sufficiently large k we have to proceed more carefully.

Given $\varepsilon \in \left(0, \frac{1}{4}\right)$, put $r_0 = \lceil \frac{4}{\varepsilon} \rceil$ and fix an integer $k \ge k(\varepsilon) := r_0 k_0(\frac{\varepsilon}{2})$. To prove NP-hardness of $Q(\varepsilon, k)$, we will provide the following reduction from NP-hard problem $Q\left(\frac{\varepsilon}{2}, k_0(\frac{\varepsilon}{2})\right)$ to $Q(\varepsilon, k)$.

Take an instance I of $Q\left(\frac{\varepsilon}{2}, k_0(\frac{\varepsilon}{2})\right)$, let $\mathcal{V}(I)$ stand for the set of variables in I and $m := |\mathcal{V}(I)|$. Clearly, $3 | m k_0(\frac{\varepsilon}{2})$ and I has exactly $\frac{1}{3} m k_0(\frac{\varepsilon}{2})$ equations. Omitting just finitely many instances, we can assume that m is sufficiently large, say $m \ge \frac{25}{\varepsilon}$. Since $k \ge r_0 k_0(\frac{\varepsilon}{2})$, we can fix an integer $r \ge r_0$ such that

$$r k_0(\frac{\varepsilon}{2}) \le k < (r+1) k_0(\frac{\varepsilon}{2}).$$

Now we create an instance I' in which each variable occurs exactly k times. If $k = r k_0(\frac{\varepsilon}{2})$, I' consists of r copies of I. If $\widetilde{k} := k - r k_0(\frac{\varepsilon}{2}) > 0$ (clearly $\widetilde{k} < k_0(\frac{\varepsilon}{2})$), we have to add some equations to r copies of I, to ensure the additional $\widetilde{k}$ occurrences of each variable.

a) If $3 \,|\, m\widetilde{k}$ one can easily create additional $\frac{m\widetilde{k}}{3}$ equations over $\mathcal{V}(I)$ in which each variable occurs exactly $\widetilde{k}$ times (any bipartite graph with bipartition (A, B), $|A| = m$, $|B| = \frac{m\widetilde{k}}{3}$, $v \in A \implies \deg(v) = \widetilde{k}$, $v \in B \implies \deg(v) = 3$, can be taken as a pattern for $A = \{\text{variables}\}$, $B = \{\text{equations}\}$). It is easy to see that such graph exists.

b) If $3 \not| \, m\widetilde{k}$, take $s \in \{4, 5\}$ such that $3 | (m\widetilde{k} + sk)$, and let Y be the set of new variables not occurring in I. We can create additional $\frac{1}{3}(m\widetilde{k} + sk)$ equations in which each variable of I occurs exactly $\widetilde{k}$ times and each new variable exactly k times (we can use any bipartite graph with bipartition $(A \cup Y, B)$, $|A| = m$, $|Y| = s$, $|B| = \frac{1}{3}(m\widetilde{k} + sk)$, $v \in A \implies \deg(v) = \widetilde{k}$, $v \in Y \implies \deg(v) = k$, $v \in B \implies \deg(v) = 3$, as a pattern). Analogously as in the previous case, it is easy to see that such graph exists.

The reduction $f : I \to I'$ just described is clearly of polynomial time. Now we observe that it is in fact a gap-preserving reduction from $Q\left(\frac{\varepsilon}{2}, k_0(\frac{\varepsilon}{2})\right)$ (with finitely many "small" instances omitted) to $Q(\varepsilon, k)$. If $I \in Q\left(\frac{\varepsilon}{2}, k_0(\frac{\varepsilon}{2})\right)$ satisfies $\mathrm{OPT}(I) > \left(1 - \frac{\varepsilon}{2}\right)$, then an optimal assignment satisfies more than $\left(1 - \frac{\varepsilon}{2}\right)|I|$ equations in I. The same assignment to variables in I' (extended arbitrarily to new variables) satisfies at least all copies of equations satisfied in I. Hence

$$\mathrm{OPT}(I') > \frac{\left(1 - \frac{\varepsilon}{2}\right)|I|r}{|I'|} > (1 - \varepsilon).$$

(The last inequality can be easily checked from our choice of parameters.) On the other hand, if $I \in Q\left(\frac{\varepsilon}{2}, k_0(\frac{\varepsilon}{2})\right)$ satisfies $\mathrm{OPT}(I) < \frac{1}{2} + \frac{\varepsilon}{2}$, then any assignment to variables in I satisfies less than $\left(\frac{1}{2} + \frac{\varepsilon}{2}\right)|I|$ equations in I. Consequently, any assignment to variables in I' satisfies less than $\left(\frac{1}{2} + \frac{\varepsilon}{2}\right)|I|r$ of copied equations and at most all $|I'| - |I|r$ additional equations. Hence

$$\mathrm{OPT}(I') < \frac{\left(\frac{1}{2} + \frac{\varepsilon}{2}\right)|I|r + (|I'| - |I|r)}{|I'|} < \frac{1}{2} + \varepsilon.$$

Now NP hardness of $Q(\varepsilon, k)$ follows from NP hardness of $Q\left(\frac{\varepsilon}{2}, k_0(\frac{\varepsilon}{2})\right)$. $\square$

To prove hard gap results for some problems using reduction from MAX-E3-LIN-2 it is sometimes useful, if all equations have the same right hand side. This can be easily enforced if we allow flipping some variables. The canonical gap versions $Q_i(\varepsilon, 2k)$ of MAX-E3-LIN-2 of this kind are as follows: *Given an instance of* MAX-E3-LIN-2 *such that all equations are of the form $x + y + z = i$ and each variable appears exactly k times negated and k times unnegated. The task is to decide if the fraction of more than $(1 - \varepsilon)$ or less than $(\frac{1}{2} + \varepsilon)$ of all equations is satisfied by the optimal (i.e. maximizing) assignment.*

The corresponding hard-gap result for this restricted version reads as follows.

Theorem 3 *For every $\varepsilon \in \left(0, \frac{1}{4}\right)$ there is a constant $k(\varepsilon)$ such that for every integer $k \geq k(\varepsilon)$ the partial decision subproblems $Q_0(\varepsilon, 2k)$ and $Q_1(\varepsilon, 2k)$ of* MAX-E3-LIN-2 *are NP-hard.*

Proof. Fix $\varepsilon \in \left(0, \frac{1}{4}\right)$ and $k(\varepsilon)$ from Theorem 2 such that for every $k \geq k(\varepsilon)$ the problem $Q(\varepsilon, k)$ is NP-hard. We demonstrate how to prove, for example, NP-hardness of $Q_0(\varepsilon, 2k)$ for k sufficiently large.

Fix $k \geq k(\varepsilon)$ and take any instance I of $Q(\varepsilon, k)$. We can simply create an instance I' of $Q_0(\varepsilon, 4k)$ as follows: For each equation $x + y + z = 0$ in I we put 4 equations, $x + y + z = 0$, $\overline{x} + \overline{y} + z = 0$, $\overline{x} + y + \overline{z} = 0$, and $x + \overline{y} + \overline{z} = 0$ (all equivalent to the original one) into I', and for each equation $x + y + z = 1$ in I we put equations $\overline{x} + y + z = 0$, $x + \overline{y} + z = 0$, $x + y + \overline{z} = 0$ and $\overline{x} + \overline{y} + \overline{z} = 0$ into I'.

This is a rather trivial gap-preserving reduction proving NP-hardness of $Q_0(\varepsilon, 2k)$ for all *even* $k \geq 2k(\varepsilon)$. To have the proof even for *odd* (and sufficiently large) k, we can use the method of the proof of Theorem 2. Namely, for k large, we reduce from $Q\left(\frac{\varepsilon}{2}, k(\frac{\varepsilon}{2})\right)$ first to $Q_0\left(\frac{\varepsilon}{2}, 4k(\frac{\varepsilon}{2})\right)$ as above, then take r copies off each equation (with r such that $0 \leq 2k - 4k(\frac{\varepsilon}{2})r < 4k(\frac{\varepsilon}{2})$), and add a few more equations to have an instance of $Q_0(\varepsilon, 2k)$. We omit the details, that are similar to those of the proof of Theorem 2. $\square$

3 Amplifiers

In this section we describe our results about the structure and parameters of 3-regular amplifiers, that we use in our reductions.

Definition 3 *A graph $G = (V, E)$ is a $(2,3)$-graph if G contains only the nodes of degree 2 (contacts) and 3 (checkers). We denote Contacts $= \{v \in V : \deg_G(v) = 2\}$, and Checkers $= \{v \in V : \deg_G(v) = 3\}$. Furthermore, a $(2,3)$-graph G is an* amplifier *(more precisely, it is a 3-regular amplifier for its contact nodes) if for every $A \subseteq V$: $|\mathrm{Cut}\, A| \geq |Contacts \cap A|$, or $|\mathrm{Cut}\, A| \geq |Contacts \setminus A|$, where $\mathrm{Cut}\, A = \{v \in A$, such that there exists $u \in V \setminus A,\ (u, v) \in E\}$.*

Notation. An amplifier G is called a (k, τ)-amplifier if $|Contacts| = k$ and $|V| = \tau k$. We introduce the notation $\tau(G) := \frac{|V|}{|Contacts|}$ for an amplifier G. Let us denote $\tau_* = \inf\{\gamma : \tau(G) < \gamma$ for infinitely many amplifiers $G\}$.

We have studied several probabilistic models of generating $(2,3)$-graphs randomly. In such situation we need to estimate the probability that the random $(2,3)$-graph G is an amplifier. It fails to be an amplifier iff the system of so-called *bad* sets $\mathcal{B} := \big\{A \subseteq V : |\mathrm{Cut}\, A| < \min\{|Contacts \cap A|, |Contacts \setminus A|\}\big\}$ is nonempty. For a fixed bad set it is quite simple to estimate the probability that this candidate for a bad set doesn't occur. But the question is how to estimate the union bound over all bad sets in better way, than by adding all single probabilities. It is useful to look for a small list $\mathcal{B}_* \subseteq \mathcal{B}$, such that if $\mathcal{B} \neq \emptyset$ then $\mathcal{B}_* \neq \emptyset$ as well. In [2] the role of $\mathcal{B}_*$ play elements of $\mathcal{B}$ of the minimum size. Our analysis shows that one can produce the significantly smaller list of bad sets which is sufficient to exclude to be sure that a graph is an amplifier. We elaborate in details on our general results in the concrete model of randomly generated (k, τ)-wheels, which generalizes slightly the notion of a wheel-amplifier used by Berman and Karpinski ([2]).

A (k, τ)-wheel is a $(2,3)$-graph $G = (V, E)$ with $|V| = \tau k$ and $|Contacts| = k$, and with the edge set E splited into two parts E_C and E_M. E_C is an edge set of several disjoint cycles in G collectively covering V. In each cycle consecutive contacts of G are separated by a chain of several (at least 2) checkers. E_M is a perfect matching for the set of checkers. Here τ is a rational number, $(\tau - 1)k$ is an even integer.

For purpose of this paper we can confine ourselves to the model with E_C consisting of 2 cycles. One consists of $(1 - \theta)k$ ($\theta \in (0, 1)$) contacts, separated by chains of checkers of length 6, and in the second one θk contacts are separated by chains of checkers of length 5. For fixed parameters θ and k consider two cycles with contacts and checkers as above and take a random perfect matching for the set of checker nodes. Then, with the high probability, the produced $(k, 7 - \theta)$-wheel will be an amplifier. More precisely, for an explicit constant $\theta_0 \in (0, 1)$, for any rational $\theta \in (0, \theta_0)$, and any sufficiently large positive integer k for which θk is an even integer, $(k, 7 - \theta)$-amplifiers exist.

Theorem 4 *$\tau_* \leq 6.9$.*

The proof of this theorem is quite technical and based on the series of lemmas. For a $(2,3)$-graph $G = (V, E)$ we define the relation $\preceq$ on the set $\mathcal{P}(V)$ of all subsets V

$$A \preceq B \quad \text{iff} \quad |\mathrm{Cut}\, A| \leq |\mathrm{Cut}\, B| - |(A \triangle B) \cap Contacts|$$

whenever $A, B \subseteq V$ and $A \triangle B$ stands for $(A \setminus B) \cup (B \setminus A)$.

Clearly, the relation $\preceq$ is reflexive and transitive. So, $\preceq$ induces a partial order on the equivalence classes $\mathcal{P}(V)\,/\approx$. The equivalence relation $\approx$ can be more simply characterized by $A \approx B$ iff $A \cap \mathit{Contacts} = B \cap \mathit{Contacts}$ & $|\mathrm{Cut}\, A| = |\mathrm{Cut}\, B|$, for $A, B \subseteq V$. Moreover, for every $A \subseteq V$, $A \preceq B$ iff $V \setminus A \preceq V \setminus B$.

Using this relation one can describe the set $\mathcal{B}$ of bad sets, as

$$\mathcal{B} := \mathcal{B}(G) = \{B \subseteq V: \text{ neither } \emptyset \preceq B, \text{ nor } V \preceq B\}. \tag{1}$$

Clearly, for every $A, B \subseteq V$, $B \in \mathcal{B}$ & $A \preceq B$ imply $A \in \mathcal{B}$. The minimal elements of the partial order $(\mathcal{P}(V), \preceq)$ play an important role in what follows. We denote

$$\mathcal{B}_0 := \mathcal{B}_0(G) = \{B \subseteq V : B \text{ is a minimal element of } (\mathcal{B}, \preceq)\}.$$

Clearly, a set $\mathcal{B}_0$ is closed on the complementation operation $A \mapsto V \setminus A$ for any subset $A \subseteq V$.

Lemma 1 *Let G be a $(2,3)$-graph and $B \in \mathcal{B}_0(G)$ be given. Then for every set $Z \subseteq B$ the inequality $2 \cdot |\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| \leq |\mathrm{Cut}\, Z| + |Z \cap \mathit{Contacts}|$ holds with the equality iff $B \setminus Z \approx B$. In particular, if $Z \cap \mathit{Contacts} \neq \emptyset$ the inequality is strict.*

Proof. Clearly,

$$2 \cdot |\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| - |\mathrm{Cut}\, Z| = |\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| - |\mathrm{Cut}\, Z \setminus \mathrm{Cut}\, B|. \tag{2}$$

Using notation $(X, Y) := \{(x, y) \in E : x \in X, y \in Y\}$ for any disjoint sets $X, Y \subseteq V$, we have (assuming $Z \subseteq B$)

$$\begin{aligned}
\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B &= (Z, V \setminus B), \\
\mathrm{Cut}\, Z \setminus \mathrm{Cut}\, B &= (Z, B \setminus Z), \\
\mathrm{Cut}\,(B \setminus Z) &= (Z, B \setminus Z) \cup (B \setminus Z, V \setminus B) \qquad \text{(a partition)}, \\
\mathrm{Cut}\, B &= (Z, V \setminus B) \cup (B \setminus Z, V \setminus B) \qquad \text{(a partition)}.
\end{aligned}$$

Hence we have

$$|\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| - |\mathrm{Cut}\, Z \setminus \mathrm{Cut}\, B| = |\mathrm{Cut}\, B| - |\mathrm{Cut}\,(B \setminus Z)|. \tag{3}$$

If $B \setminus Z \approx B$ then $|\mathrm{Cut}\, B| = |\mathrm{Cut}\,(B \setminus Z)|$ and $Z \cap \mathit{Contacts} = \emptyset$. If $B \setminus Z \not\approx B$ then we get $|\mathrm{Cut}\, B| - |\mathrm{Cut}\,(B \setminus Z)| < |Z \cap \mathit{Contacts}|$. Hence

$$|\mathrm{Cut}\, B| - |\mathrm{Cut}\,(B \setminus Z)| \leq |Z \cap \mathit{Contacts}| \tag{4}$$

with the strict inequality exactly if $B \setminus Z \not\approx B$, in particular if $Z \cap \mathit{Contacts} \neq \emptyset$.

Putting (2), (3) and (4) together completes the proof. □

Lemma 2 *Let G be a $(2,3)$-graph. Then for every $B \in \mathcal{B}_0(G)$ the set $\mathrm{Cut}\, B$ is a matching in the graph G.*

Proof. Let $v \in V$ be fixed. Put $Z := \{v\}$ and $c(v) := |\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B|$. Obviously $c(v)$ is the number of edges of $\mathrm{Cut}\, B$ adjacent to v. Applying Lemma 1 (if $v \notin B$ we apply it with $V \setminus B$ instead of B) we get $2c(v) \leq 3$. Hence $c(v) \in \{0, 1\}$ for every $v \in V$ and $\mathrm{Cut}\, B$ is a matching. □

For a $(2,3)$-graph $G = (V, E)$ let $Z \subseteq V$ be given. Let $G_Z = (Z, E_Z)$ stand for the subgraph of G induced by the node set Z. To see that $|\mathrm{Cut}\, Z| + |Z \cap \mathit{Contacts}| = 3|Z| - 2|E_Z|$, we can argue as follows:

$$|\mathrm{Cut}\, Z| = \sum_{v \in Z \cap \mathit{Checkers}} (3 - \deg_{G_Z}(v)) + \sum_{v \in Z \cap \mathit{Contacts}} (2 - \deg_{G_Z}(v))$$
$$= \sum_{v \in Z} (3 - \deg_{G_Z}(v)) - |Z \cap \mathit{Contacts}| = 3|Z| - 2|E_Z| - |Z \cap \mathit{Contacts}|.$$

Given $B \in \mathcal{B}_0(G)$, $\mathrm{Cut}\, B$ is a matching in G as follows from Lemma 2. Let $\mathit{Cutters}(B)$ stand for the set of nodes in B adjacent to $\mathrm{Cut}\, B$. Clearly for any $Z \subseteq B$, an edge of $\mathrm{Cut}\, B$ adjacent to $v \in \mathit{Cutters}(B)$ belongs to $\mathrm{Cut}\, Z$ iff $v \in Z$. Therefore $|\mathrm{Cut}\, Z \cap \mathrm{Cut}\, B| = |Z \cap \mathit{Cutters}(B)|$. Hence we can reformulate Lemma 1 as follows:

Lemma 3 *Let G be a $(2,3)$-graph and $B \in \mathcal{B}_0(G)$ be given. Then for every set $Z \subseteq B$ the inequality $|Z \cap \mathit{Cutters}(B)| \leq \frac{3}{2}|Z| - |E_Z|$ holds with the equality iff $B \setminus Z \approx B$. In particular, if $Z \cap \mathit{Contacts} \neq \emptyset$ the inequality is strict.*

The purpose of this lemma is to derive some restrictions on local patterns of $\mathrm{Cut}\, B$ for a general set $B \in \mathcal{B}_0(G)$. Given $B \in \mathcal{B}_0(G)$, we can test it with many various $Z \subseteq B$ (typically with G_Z being a small connected graph) to obtain restrictions on possible patterns of $\mathit{Cutters}(B)$ in B. Some of basic results of this kind are stated in the following lemma.

Lemma 4 *Let G be a $(2,3)$ graph, $B \in \mathcal{B}_0(G)$ and $Z \subseteq B$ be given.*

(i) If G_Z is a tree and $|Z| = 2k - 1$ $(k = 1, 2, \dots)$ then $|Z \cap \mathit{Cutters}(B)| \leq k$.

(ii) If G_Z is a tree and $|Z| = 2k$ then $|Z \cap \mathit{Cutters}(B)| \leq k + 1$. Moreover, this inequality is strict if $Z \cap \mathit{Contacts} \neq \emptyset$.

(iii) If G_Z is a $(2k+1)$-cycle then $|Z \cap \mathit{Cutters}(B)| \leq k$.

(iv) If G_Z is a $2k$-cycle then $|Z \cap \mathit{Cutters}(B)| \leq k$. Moreover, this inequality is strict if $Z \cap \mathit{Contacts} \neq \emptyset$.

Lemma 5 *Let $G = (V, E)$ be a $(2,3)$-graph and $B \in \mathcal{B}_0(G)$ be given.*

(i) If $a, b \in \mathit{Cutters}(B)$ and $(a, b) \in E$, then $a, b \in \mathit{Checkers}$ and there are 2 distinct nodes a', $b' \in B \setminus \mathit{Cutters}(B)$ such that $(a, a') \in E$ and $(b, b') \in E$.

(ii) If $a, c \in \mathit{Cutters}(B)$, $b \in B$, $(a, b) \in E$, $(b, c) \in E$, and if exactly one of nodes a, b and c belongs to Contacts, then there are 2 distinct nodes d, $e \in B \setminus \mathit{Cutters}(B)$, each adjacent to one of two nodes in $\{a, b, c\} \cap \mathit{Checkers}$.

Proof. (i) If at least one of a, b belongs to *Contacts*, applying Lemma 4(ii), with $Z := \{a, b\}$, we obtain a contradiction. Hence $a, b \in \mathit{Checkers}$ and knowing that $\mathrm{Cut}\, B$ is a matching, there are a', $b' \in B$ such that (a, a') and (b, b') belong to E. If $a' = b'$, we get a contradiction using Lemma 4(iii) with $Z := \{a, b, a'\}$. Hence $a' \neq b'$. To complete the proof one can conclude that $a' \in B \setminus \mathit{Cutters}(B)$ (respectively, $b' \in B \setminus \mathit{Cutters}(B)$) applying the Lemma 4(i) with $Z := \{a, b, a'\}$ (respectively, with $Z := \{a, b, b'\}$).

(ii) Clearly there are $d, e \in B$ distinct from any of a, b, and c, such that the one of two nodes in $\{a, b, c\} \cap \mathit{Checkers}$ is adjacent to d, and the other one to c. Applying Lemma 3 with $Z := \{a, b, c, d\}$ (respectively with $Z := \{a, b, c, e\}$) we easily get that $d \neq e$ and both, $d \in B \setminus \mathit{Cutters}(B)$ and $e \in B \setminus \mathit{Cutters}(B)$. □

It is useful to work with the smaller list of bad sets than $\mathcal{B}_0(G)$. For the purpose to provide such more restricted list we make our partial order $\preceq$ finer inside the equivalence classes $\mathcal{P}(V)\,/\approx$.

For a given $(2,3)$-graph $G=(V,E)$ let a subset F of E of "distinguished edges" be fixed. We define the following relations on the set $\mathcal{P}(V)$ of all subset V, whenever $A, B \subseteq V$:

$$A \overset{F}{\preceq} B \quad \text{iff} \quad \text{either } A \preceq B \;\&\; A \not\approx B \text{ or } A \approx B \;\&\; |F \cap \operatorname{Cut} A| \leq |F \cap \operatorname{Cut} B|$$

$$A \overset{F}{\preceq}_* B \quad \text{iff} \quad \text{either } A \overset{F}{\preceq} B \;\&\; A \overset{F}{\not\approx} B, \text{ or } A \overset{F}{\approx} B \;\&\; \min\{|A \cap \mathit{Checkers}|, |\mathit{Checkers} \setminus A|\} \\ \leq \min\{|B \cap \mathit{Checkers}|, |\mathit{Checkers} \setminus B|\}.$$

Denote

$$\mathcal{B}_F(G) := \{B \subseteq V : B \text{ is a minimal element of } (\mathcal{B}(G), \overset{F}{\preceq})\},$$
$$\mathcal{B}^*_F(G) := \{B \subseteq V : B \text{ is a minimal element of } (\mathcal{B}(G), \overset{F}{\preceq}_*)\}.$$

The equivalence relation $\overset{F}{\approx}$ is defined by: $A \overset{F}{\approx} B$ iff $A \overset{F}{\preceq} B$ & $B \overset{F}{\preceq} A$. Clearly $\overset{F}{\preceq}$ is a partial order on equivalence classes $\mathcal{P}(V)\,/\overset{F}{\approx}$. The equivalence relation $\overset{F}{\approx}$ can be also characterized by

$$A \overset{F}{\approx} B \quad \text{iff} \quad A \cap \mathit{Contacts} = B \cap \mathit{Contacts} \;\&\; |\operatorname{Cut} A| = |\operatorname{Cut} B| \;\&\; |F \cap \operatorname{Cut} A| = |F \cap \operatorname{Cut} B|.$$

Clearly $\mathcal{B}^*_F(G) \subseteq \mathcal{B}_F(G) \subseteq \mathcal{B}_0(G)$, and

$$B \in \mathcal{B}_F(G) \quad \text{iff} \quad B \in \mathcal{B}_0(G) \;\&\; (A \approx B \text{ implies } |F \cap \operatorname{Cut} B| \leq |F \cap \operatorname{Cut} A|).$$

$$B \in \mathcal{B}^*_F(G) \quad \text{iff} \quad B \in \mathcal{B}_F(G) \;\&\; A \overset{F}{\approx} B \text{ implies } \min\{|B \cap \mathit{Checkers}|, |\mathit{Checkers} \setminus B|\} \\ \leq \min\{|A \cap \mathit{Checkers}|, |\mathit{Checkers} \setminus A|\}, \tag{5}$$

Lemma 6 *Let G be a $(2,3)$-graph and $B \in \mathcal{B}_F(G)$ be given. Then for every set $Z \subseteq B$ such that $B \setminus Z \approx B$ (equivalently, $Z \subseteq B \cap \mathit{Checkers}$ and $|Z \cap \mathit{Cutters}(B)| = \frac{3}{2}|Z| - |E_Z|$) $|F \cap \operatorname{Cut} Z \cap \operatorname{Cut} B| \leq \frac{1}{2}|F \cap \operatorname{Cut} Z|$ holds, with the equality iff $B \setminus Z \overset{F}{\approx} B$.*

Proof. Similarly as in the proof of Lemma 1 we get

$$2 \cdot |F \cap \operatorname{Cut} Z \cap \operatorname{Cut} B| - |F \cap \operatorname{Cut} Z| = |F \cap \operatorname{Cut} B| - |F \cap \operatorname{Cut}(B \setminus Z)| \leq 0.$$

Assuming $B \setminus Z \approx B$, the equality means exactly $B \setminus Z \overset{F}{\approx} B$. □

Lemma 7 *Let G be a $(2,3)$-graph, $B \in \mathcal{B}^*_F(G)$ and $\emptyset \neq Z \subseteq B$ such that $B \setminus Z \approx B$ and $2 \cdot |B \cap \mathit{Checkers}| < |\mathit{Checkers}| + |Z|$. Then $|F \cap \operatorname{Cut} Z \cap \operatorname{Cut} B| < \frac{1}{2} \cdot |F \cap \operatorname{Cut} Z|$.*

Proof. Put $A := B \setminus Z$. From our assumption $2|B \cap \mathit{Checkers}| < |\mathit{Checkers}| + |Z|$ follows $|A \cap \mathit{Checkers}| \leq |\mathit{Checkers} \setminus A|$ and $|B \cap \mathit{Checkers}| - |Z| < |\mathit{Checkers} \setminus B|$. As $B \setminus Z \approx B$ we obtain $Z \subseteq B \cap \mathit{Checkers}$. Hence,

$$\min\{|A \cap \mathit{Checkers}|, |\mathit{Checkers} \setminus A|\} = |B \cap \mathit{Checkers}| - |Z| \\ < \min\{|B \cap \mathit{Checkers}|, |\mathit{Checkers} \setminus B|\}.$$

Due to (5) we see that $B \setminus Z \overset{F}{\not\approx} B$ and so in the inequality guaranteed by Lemma 6 we have in fact the strict inequality. □

Let us consider a $(2,3)$-graph $G = (V, E)$. For $B \subseteq V$, we denote $B_{\text{red}} := B \cap$ *Checkers*. Assume further that no pair of nodes in *Contacts* is adjacent by an edge. We convert G to a 3-regular (multi-)graph G_{red} with a node set V_{red} equals to *Checkers*. Each node $v \in$ *Contacts* and two edges adjacent to v in G are replaced with an edge $e(v)$ (later called a *contact* edge) that connects the pair of nodes that were adjacent to v in G. For any $A \subseteq V_{\text{red}}$ let $\text{Cut}_{\text{red}} A$ stand for a cut of A in G_{red}, and $\mathit{Cutters}_{\text{red}}(A)$ stand for the set of nodes of A adjacent in G_{red} to an edge of $\text{Cut}_{\text{red}} A$.

Lemma 8 *Let G be a $(2,3)$-graph with no edge between contact nodes, and let $B \in \mathcal{B}_0(G)$. Then $|\text{Cut}\, B| = |\text{Cut}_{red}(B_{red})|$, and if any pair of nodes in Contacts is at least at distance 3 apart, $\text{Cut}_{red}(B)$ is a matching in G_{red}.*

Proof. It easily follows from the fact that $\text{Cut}\, B$ is a matching in G (Lemma 2) and from Lemma 5(i). □

We can summarize the above results for the concrete model of (k,τ)-wheels. Given a (k,τ)-wheel G with the edge set consisting of E_C (which is union of two disjoint cycles C_1 and C_2) and E_M (which is a perfect matching for the checkers of G). We consider here the choice $F := E_C$ for the special subset of "distinguished edges" in our amplifier analysis.

Given a bad set B, we will refer to fragments of B, the connected components of B within cycles C_1 and C_2, and to reduced fragments of B_{red}, the connected components of B_{red} within corresponding reduced cycles.

Proposition 1 *Let G be a (k,τ)-wheel. Then every set $B \in \mathcal{B}_0(G)$ has the following properties:*

(i) *B is a bad set, i.e. $|\text{Cut}\, B| < \min\{|\mathit{Contacts} \cap B|, |\mathit{Contacts} \setminus B|\}$.*

(ii) *$\text{Cut}\, B$ is a matching in G.*

(iii) *$|\text{Cut}_{red}(B_{red})| = |\text{Cut}\, B|$, and $\text{Cut}_{red}(B_{red})$ is a matching in G_{red}.*

(iv) *Any fragment of B contains at least 2 checkers.*

(v) *End nodes of any reduced fragment of B_{red} are not incident to $E_M \cap \text{Cut}\, B$.*

(vi) *Any fragment of B consisting of 3 checkers has none of its nodes incident to $E_M \cap \text{Cut}\, B$.*

(vii) *Any fragment of B consisting of 2 checkers and 1 contact has both its checkers matched with $B \setminus \mathit{Cutters}(B)$ nodes.*

(viii) *Any fragment of B consisting of 2 checkers has both its nodes matched with $B \setminus \mathit{Cutters}(B)$ nodes.*

Every set $B \in \mathcal{B}_F(G)$ additionally has the following properties:

(ix) *Any fragment of B contains at least 3 nodes.*

(x) *Any fragment of B consisting of 3 checkers has all its nodes matched with $B \setminus \mathit{Cutters}(B)$.*

(xi) *Any fragment of B consisting of 4 checkers has none of its nodes incident to $E_M \cap \operatorname{Cut} B$.*

All the above properties apply at the same time to B and $\tilde{B} :=$ Checkers $\setminus B$. The following is less symmetric, it says something more about the smaller of the sets B, $\tilde{B}$, if $B \in \mathcal{B}(G)^*_F$.

(xii) *If $B \in \mathcal{B}(G)^*_F$ with $|B \cap$ Checkers$| \leq \frac{1}{2}|$Checkers$|$, then no pair of checkers that are end nodes of (possibly distinct) fragments of B, are matched.*

Proof. (i) follows from the definition, (ii) is just Lemma 2, (iii)–(viii) can be easily proved using (ii), Lemma 2, 3 and 5, (ix)–(xi) using Lemma 6, and (xii) using Lemma 7. □

Proof of Theorem 4

We do not try to optimize the value of τ_* here, just to show that $\tau_* \leq 6.9$. Hence we can keep the proof close to the one of Berman and Karpinski ([2]). In situations, where counting method is similar, we refer to their proof for more details.

Given k, for simplicity even, $k = 2r$. Further given θ with θk even, $(1-\theta)k = 2r_1$, $\theta k = 2r_2$, $r_1 + r_2 = r$. Consider a cycle C_1 with $2r_1$ contacts (of G, to be produced) and each 2 consecutive contacts separated by chain of 6 checkers, and a cycle C_2 (disjoint from C_1) with $2r_2$ contacts with those chains of length 5. Take a random perfect matching for the set of checkers. Our aim is to prove that with probability tending to 1 (as $k \to \infty$) the random graph G generated in this way is an amplifier. (Strictly speaking, conditionally that G is a simple graph. But it is standard to prove that resulting multigraph is a simple graph with probability bounded away from 0.)

We want to estimate from above the probability that resulting graph G is *not* an amplifier, i.e. the probability of the event $\mathcal{B}^*_F(G) \neq \emptyset$. (Here $F := E_C$ is the set of edges of $C = C_1 \cup C_2$.) As in our random matchings checkers are relevant only, it is more convenient to look at the situation in reduced graph G_{red}. Put

$$\mathcal{A}(G) = \{B_{\text{red}} : B \in \mathcal{B}^*_F(G)\}.$$

From Proposition 1 the properties of any $A \in \mathcal{A}(G)$ can be derived, e.g.:

(i) $\operatorname{Cut}_{\text{red}} A$ is a matching in G_{red}.

(ii) Any reduced fragment of A contains at least 2 checkers, and if it is not incident to any contact edge it contains at least 3 checkers.

The same can be applied to $\tilde{A} :=$ *Checkers* $\setminus A$.

(iii) End nodes of any reduced fragment of A are matched with nodes of A, and end nodes of any reduced fragment of $\tilde{A}$ are matched with nodes of $\tilde{A}$.

For $A \in \mathcal{A}(G)$ and $j = 1, 2$ we use the following notation:

$$a^1_j = \text{the number of contact edges in } C_j \cap \operatorname{Cut}_{\text{red}}(A),$$
$$a^2_j = \text{the number of contact edges of } C_j \text{ inside fragments of } A,$$
$$\tilde{a}^2_j = \text{the number of contact edges of } C_j \text{ inside fragments of } \tilde{A}.$$

Hence

$$\begin{aligned}
a_j &:= a_j^1 + a_j^2 = \text{the number of contact edges of } C_j \text{ incident to } A,\\
\tilde{a}_j &:= a_j^1 + \tilde{a}_j^2 = \text{the number of contact edges of } C_j \text{ incident to } \tilde{A}.\\
a^1 &:= a_1^1 + a_2^1 = \text{the number of contact edges in } E_C \cap \mathrm{Cut}_{\mathrm{red}}(A),\\
a^2 &:= a_1^2 + a_2^2 = \text{the number of contact edges inside fragments of } A,\\
\tilde{a}^2 &:= \tilde{a}_1^2 + \tilde{a}_2^2 = \text{the number of contact edges inside fragments of } \tilde{A}.
\end{aligned}$$

Denote f_j = the number of fragments of A (and $\tilde{A}$ as well) in C_j $(j = 1, 2)$, and $f = f_1 + f_2$. Clearly, $2f_j$ is the size of $\mathrm{Cut}_{\mathrm{red}}(A)$ inside C_j. Put

$$\begin{aligned}
s_j &= \text{ the number of checkers in } A \cap C_j,\\
\tilde{s}_j &= 2(7-j)r_j - s_j = \text{ the number of checkers in } \tilde{A} \cap C_j,\\
s &= s_1 + s_2, \quad \tilde{s} = \tilde{s}_1 + \tilde{s}_2 = 2(6-\theta)r - s.
\end{aligned}$$

There are possible many $B \in \mathcal{B}_F^*(G)$ such that $A = B_{\mathrm{red}}$. But for any such B the important number $\min\{|B \cap \mathit{Contacts}|, |\mathit{Contacts} \setminus B|\}$ can be bounded from above by

$$a := \min\{r, a^1 + \min\{a^2, \tilde{a}^2\}\}.$$

Hence $2f < a$ is a necessary condition for A. For any $B \in \mathcal{B}_F^*(G)$ if $A = B_{\mathrm{red}}$ and $c := |E_M \cap \mathrm{Cut}\, B|$, necessarily $c + 2f < a$. Moreover $s - c$, $\tilde{s} - c$ have to be even, so for c we have as possible choices those integers from $\{0, 1, 2, \dots, a - 2f - 1\}$ for which $s - c$ is even. Now we will prove bounds on s_j in terms of a_j, f_j, a_j^1. Any fragment (in C_j) incident to i contact edges (suppose $i \geq 2$) contains at least $(7-j)(i-1)$ checkers, and at most $(7-j)(i+1) - 2$ checkers.

For $i = 0$ and $i = 1$ we obtain lower estimates using Proposition 1, the upper bounds are trivial. By adding sizes of all fragments we obtain

$$s_j \geq (7-j)(a_j - f_j) + (2f_j - a_j^1) + (8-j)f_j^0,$$

where f_j^0 means the number of fragments of A in C_j incident to no contact edge. Similarly,

$$s_j \leq (7-j)a_j + (5-j)f_j - (6-j)a_j^1, \quad j = 1, 2.$$

Analogous bounds we obtain for $\tilde{s}_j$, replacing a_j, f_j, f_j^0 by $\tilde{a}_j$, $\tilde{f}_j$, $\tilde{f}_j^0$. Adding, as $s = s_1 + s_2$,

$$s \geq 5a - 6f + a_1 + f_2 + (2f - a^1),$$

in particular $s > 2a$ follows.

Similarly, assuming $a < \frac{r}{2}$ and $s \leq \tilde{s}$, we get

$$s \leq 6a + 4f - f_1 - a_1 - 5a_1^1 - 4a_2^1,$$

and $s < 8a < 4r$.

Now let A satisfying properties (i)–(iii) with parameters a, f, s, be given. We can assume that $s \leq \tilde{s}$, otherwise we change the role of A and $\tilde{A}$. Keeping A fixed let us estimate the probability $p = p(a, f, s)$ that the random matching will produce G such that our fixed A belongs to $\mathcal{A}(G)$.

Fix $c \in \{0, 1, \dots, a - 2f - 1\}$ such that $s - c$ is even. Now we estimate the number of perfect matchings of checkers such that $A \in \mathcal{A}(G)$ for resulting graph G and $|E_M \cap \mathrm{Cut}\, A| = c$.

The end points of $E_M \cap \operatorname{Cut} A$ in A can be chosen in $\binom{s-2f}{c}$ ways ($2f$ end nodes of reduced fragments of A are excluded), similarly end points in $\tilde{A}$ in $\binom{\tilde{s}-2f}{c}$ ways, there are $c!$ possibilities for the matching with the end points chosen as above. Now, denoting by

$$\mu(m) := \frac{(2m)!}{m!2^m},$$

the number of perfect matchings in a clique with $2m$ nodes, we can simply estimate the number of perfect matchings of checkers leading to graph with $A \in \mathcal{A}(G)$ and $|E_M \cap \operatorname{Cut} A| = c$ by the number

$$g_{f,s}(c) := \mu\left(\frac{s-c}{2}\right)\mu\left(\frac{\tilde{s}-c}{2}\right)\binom{s-2f}{c}\binom{\tilde{s}-2f}{c}c!.$$

One can prove that $g_{f,s}(= g)$ is, in fact, increasing for $c \in \langle 0, a-2f\rangle$ in our setting, hence

$$g(c) \le g(a-2f), \quad \text{and} \quad \sum_c g(c) \le rg(a-2f).$$

Consequently, the probability $p(a, f, s)$ defined above can be estimated like

$$p(a,f,s) \le \frac{\mu(\frac{s-a-2f}{2})\mu(\frac{\tilde{s}-a+2f}{2})}{\mu\big((6-\theta)r\big)}\binom{s-2f}{a-2f}\binom{\tilde{s}-2f}{a-2f}(a-2f)!.$$

Now we have to estimate how many sets A satisfying (i)–(iii), and with fixed vector of parameters (a, f, s), there exist. (Of course, only triples satisfying all relations derived above for sets in $\mathcal{A}(G)$ are relevant.) Let us denote that number by $N(a, f, s)$.

We look at all vectors of parameters (c_1, f_1, s_1) that are compatible with fixed (a, f, s). Any such (a_1, f_1, s_1) determines compatible (a_2, f_2, s_2) and we estimate $N(a, f, s) \le O(r^6) \cdot \max\{N_1(a_1, f_1, s_1)N_2(a_2, f_2, s_2)\}$, where maximum is taken over all compatible triples (a_1, f_1, s_1), and $N_j(a_j, f_j, s_j)$ stands for the number of sets in C_j with corresponding parameters (a_j, f_j, s_j).

Now we have several possibilities how to estimate $N_j(a_j, f_j, s_j)$ from above.

Trivial estimate. If $2f_j$ reduced fragments, each of length at least 2, forms a partition of C_j, it can be generated by less than $\binom{2(7-j)r_j-2f_j}{2f_j}$ ways.

Estimate for s small. Let us assume the clockwise orientation on C_j. We select f_j "first ends" of the fragments; this can be done in at most $\binom{2(7-j)r_j}{f_j}$ ways. Next we distribute the sizes of the fragments; because the sum is s_j and each size is at least 2, it can be done in $\binom{s_j-f_j}{f_j}$ ways, hence we have estimate $\binom{2(7-j)r_j}{f_j} \cdot \binom{s_j-f_j}{f_j}$. This works well, e.g. for $a \le \frac{r}{10}$.

One can get in the elementary way estimate

$$p(a,f,s) \cdot N_1(a_1,f_1,s_1)N_2(a_2,f_2,s_2) = O(r^{-10})(\frac{9}{10})^a,$$

just comparing terms when expressed as fractions of factorial products. From this, probability that a set A with $a \le \frac{r}{10}$ will belong to $\mathcal{A}(G)$ for a produced graph G tends to 0 as r approaches infinity.

Alternative countings. If s_j is relatively low, in the sense that s_j is close to its lower estimate given earlier, then A in C_j is very close to its inner approximant whose end points are adjacent to contact edges. In fact, the difference $d_j := s_j - (7-j)(a_j - f_j)$ measures the deviation of A in C_j from this approximant. If d_j is relatively small, there are only a few candidates for a set $A \cap C_j$ with these parameters a_j, f_j, s_j. This leads to estimate

$$N_j(a_j,f_j,s_j) \le \binom{2r_j+2f_j}{2f_j}\binom{d_j+2f_j}{2f_j}.$$

Similarly, if s_j is close to its upper estimate, then A in C_j is very close to its outer approximant. Now, the deviation is computed differently: $d_j := (7-j)(a_j+f_j)-s_j$. The resulting formula estimating $N_j(a_j,f_j,s_j)$ can be taken the same with this choice. See [2] for more details on this particular way of computing.

Now, for $N_j(a_j,f_j,s_j)$ we take minimum from bounds computed above for bound on

$$p(a,f,s)\cdot N_1(a_1,f_1,s_1)\cdot N_2(a_2,f_2,s_2).$$

Using Stirling's formula and binary entropy function we estimate $\frac{1}{a}\log$ of above expressions over the range of parameters given by our earlier estimates and $a \geq \frac{r}{10}$. It stays negative and bounded away from zero for $\theta \in (0, 0.1)$. For such θ, the probability that a set A with $a \geq \frac{r}{10}$ will belong to $\mathcal{A}(G)$ for a produced graph G is bounded from above by $O(0.95^a)$.

Consequently, $\tau_* \leq 6.9$ easily follows.

4 Amplifier parametrized known reductions

We call HYBRID a system of linear equations over $\mathbb{Z}_2$, each equation either with 2 or with 3 variables. We are interested in hard gap results for instances of HYBRID with exact 3 occurrences of each variable (a subproblem of E3-Occ-Max-3-Lin-2). As suggested in [2], one can produce hard gaps for such restricted instances of HYBRID by gap-preserving reduction from Max-E3-Lin-2.

Our approach is simpler than in [2], since we start the reduction from the problem which is already of bounded (even constant, and possibly very large) occurrence. This is a crucial point, since the number of occurrences of variables is just the value that has to be amplified using the expander or amplifier method. In our reductions an amplifier plays a role of a constant. Therefore we need not restrict ourselves to amplifiers that can be effectively constructed; any proof of the existence of amplifiers with better parameters than those currently known, further improves all our inapproximability results.

Reduction from $Q(\varepsilon, k)$ to HYBRID(G)

Let $\varepsilon \in \left(0, \frac{1}{4}\right)$, and k be a positive integer such that $Q(\varepsilon, k)$ is NP-hard. Now we describe a gap-preserving reduction from $Q(\varepsilon, k)$ to the corresponding gap-version of HYBRID. Assume that $G=(V,E)$ is a fixed (k,τ)-amplifier with $|Contacts| = k$ and $|V| = \tau k$. Let an instance I of $Q(\varepsilon, k)$ be given, denote by $\mathcal{V}(I)$ the set of variables in I, $m := |\mathcal{V}(I)|$. Take m disjoint copies of G, one for each variable from $\mathcal{V}(I)$. Let G_x denote a copy of G that corresponds to a variable x. The contact nodes of G_x represent k occurrences of x in equations of I. Distinct occurrences of a variable x in I are now represented by distinct contact nodes of G_x. For each equation $x+y+z=i$ of I ($i \in \{0,1\}$) we create a hyperedge of size 3, labeled by i. A hyperedge connects a triple of contact nodes, one from each G_x, G_y and G_z. The edges inside each copy G_x are labeled by 0 and any such edge (u,v) represents the equivalence equation $u+v=0$.

The produced instance I' of HYBRID corresponds simultaneously to a system of equations and a labeled hypergraph. Clearly, nodes correspond to variables, and labeled (hyper-)edges to equations in an obvious way. The restriction of HYBRID to these instances will be called as HYBRID(G) in what follows. The most important property of a produced instance I' is that each variable occurs exactly 3 times in equations. In particular, each contact node occurs exactly in one hyperedge. If an instance I has m variables with $|I| = \frac{mk}{3}$ equations, then I' has $m\tau k$ variables,

$\frac{mk}{3}$ equations with 3 variables, and $\frac{mk}{2}(3\tau - 1)$ equations with 2 variables. Hence $|I'| = \frac{mk}{6}(9\tau - 1)$ equations in total.

Now we show that the above reduction from I to I' preserves the hard gap of $Q(\varepsilon, k)$. In fact, there is an affine dependence of $\mathrm{OPT}(I')$ on $\mathrm{OPT}(I)$. Clearly, any assignment to variables from $\mathcal{V}(I)$ generates an assignment to variables of I' in natural way; the value of a variable x is assigned to all variables of G_x. Such assignments to variables of I' are called *standard*. To show that the optimum $\mathrm{OPT}(I')$ is achieved on standard assignments is easy. In fact, any given assignment φ to variables of I' can be converted to a standard one in such way that the number of satisfied equations does not decrease, as follows: consider a variable x from $\mathcal{V}(I)$. Assign to all variables of G_x the same value as is assigned to the majority of contact nodes in G_x by the assignment φ. The fact that G_x is an amplifier ensures that the number of unsatisfied equations does not increase. Now, if we do the same for all variables from $\mathcal{V}(I)$ one after another, the result will be a standard assignment. Consequently, $\mathrm{OPT}(I')$ is achieved on standard assignments. But for a standard assignment the number of unsatisfied equations for I' is the same as for I. Consequently, $\mathrm{OPT}(I')$ depends affinely on $\mathrm{OPT}(I)$, namely $(1 - \mathrm{OPT}(I'))|I'| = (1 - \mathrm{OPT}(I))|I|$. Now we see that

$$\mathrm{OPT}(I) > 1 - \varepsilon \quad \text{implies} \quad \mathrm{OPT}(I') > 1 - \frac{2\varepsilon}{9\tau - 1}, \quad \text{and}$$
$$\mathrm{OPT}(I) < \frac{1}{2} + \varepsilon \quad \text{implies} \quad \mathrm{OPT}(I') < \frac{9\tau - 2}{9\tau - 1} + \frac{2\varepsilon}{9\tau - 1}.$$

This proves NP-hardness of $\langle \frac{9\tau-2}{9\tau-1} + \frac{2\varepsilon}{9\tau-1}, 1 - \frac{2\varepsilon}{9\tau-1} \rangle$-gap version of HYBRID($G$).

Hence, we have just proved the following:

Theorem 5 *Assume that $\varepsilon \in \left(0, \frac{1}{4}\right)$, let k be an integer such that $Q(\varepsilon, k)$ is* NP*-hard, and G be a (k, τ)-amplifier. Then $\langle \frac{9\tau-2}{9\tau-1} + \frac{2\varepsilon}{9\tau-1}, 1 - \frac{2\varepsilon}{9\tau-1} \rangle$-gap version of* HYBRID($G$) *is* NP*-hard.*

Corollary 1 *It is* NP*-hard to approximate the solution of* E3-Occ-Max-3-Lin-2 *within any constant smaller than $1 + \frac{1}{9\tau_* - 2}$.*

Proof. Inapproximability within $\frac{9\tau-1-2\varepsilon}{9\tau-2+2\varepsilon} = 1 + \frac{1-4\varepsilon}{9\tau-2+2\varepsilon}$ follows from the Theorem 5. But for $\varepsilon > 0$ arbitrarily small there exists (k, τ)-amplifier with τ arbitrarily close to τ_* and with k so large that $Q(\varepsilon, k)$ is NP-hard. If $r < 1 + \frac{1}{9\tau_* - 2}$ we can take those parameters such that $r < 1 + \frac{1-4\varepsilon}{9\tau-2+2\varepsilon}$, and inapproximability within r follows. □

Reductions from HYBRID(G) to other problems

We refer to [2] where Berman and Karpinski provide gadgets for reductions from HYBRID to small bounded instances of Maximum Independent Set and Minimum Node Cover. We can use exactly the same gadgets in our context, but instead of their wheel-amplifier we use a general (k, τ)-amplifier. The proofs from [2] apply in our context as well.

Theorem 6 *Let $\varepsilon \in \left(0, \frac{1}{4}\right)$, k be a positive integer such that $Q(\varepsilon, k)$ is* NP*-hard, and τ be such that a (k, τ)-amplifier exists. It is* NP*-hard to decide whether an instance of* Max-3-IS *with n nodes has the maximum size of an independent set above $\frac{18\tau+14-2\varepsilon}{4(9\tau+8)}n$, or below $\frac{18\tau+13+2\varepsilon}{4(9\tau+8)}n$. Consequently, it is NP-hard to approximate the solution of* Max-3-IS *within any constant smaller than $1 + \frac{1}{18\tau_* + 13}$.*

Similarly, it is NP-hard to decide whether an instance of MIN-3-NC *with* n *nodes has the minimum size of a node cover above* $\frac{18\tau+19-2\varepsilon}{4(9\tau+8)}n$*, or below* $\frac{18\tau+18+2\varepsilon}{4(9\tau+8)}n$*. Consequently, it is NP-hard to approximate the solution of* MIN-3-NC *within any constant smaller than* $1+\frac{1}{18\tau_*+18}$*. The same hard-gap and inapproximability results apply to* 3*-regular triangle-free graphs.*

In the following we present the inapproximability results for three similar APX-complete problems: MAXIMUM TRIANGLE PACKING, MAXIMUM 3-SET PACKING problem and MINIMUM 3-SET COVERING problem. From L-reductions used in the proofs of Max-SNP completeness (see [7] or [9]) some lower bounds can be counted but results will be worse as the lower bounds presented here.

Maximum Triangle Packing problem. A triangle packing for a graph $G=(V,E)$ is a collection $\{V_i\}$ of disjoint 3-sets of V, such that every V_i induces a 3-clique in G. The goal is to find cardinality of maximum triangle packing. The problem is APX-complete even for graphs with maximum degree 4 ([7]).

Maximum 3-Set Packing problem. Given a collection C of sets, the cardinality of each set in C is at most 3. A set packing is a collection of disjoint sets $C' \subseteq C$. The goal is to find cardinality of maximum set packing. If the number of occurrences of any element in C is bounded by a constant K, $K \geq 2$, the problem is still APX-complete ([1]).

Minimum 3-Set Covering Problem. Given a collection C of subsets of a finite set S, the cardinality of each set in C is at most 3. The goal is to find cardinality of minimum subset $C' \subseteq C$ such that every element in S belongs to at least one member of C'. If the number of occurrences of any element in sets of C is bounded by a constant $K \geq 2$, the problem is still APX-complete [9].

Theorem 7 *Assume that* $\varepsilon \in \left(0, \frac{1}{4}\right)$*,* k *a positive integer such that* $Q(\varepsilon, k)$ *is* NP*-hard, and* τ *such that there is a* (k,τ)*-amplifier.*

(i) It is NP*-hard to decide whether an instance of* TRIANGLE PACKING *with* n *nodes has the maximum size of a triangle packing above* $\frac{18\tau+14-2\varepsilon}{6(9\tau+8)}n$*, or below* $\frac{18\tau+13+2\varepsilon}{6(9\tau+8)}n$*. Consequently, it is NP-hard to approximate the solution of* MAXIMUM TRIANGLE PACKING *problem (even on* 4*-regular graphs) within any constant smaller than* $1+\frac{1}{18\tau_*+13}$*.*

(ii) It is NP*-hard to decide whether an instance of* 3-SET PACKING *with* n *triples and the occurrence of each element exactly in two triples has the maximum size of a packing above* $\frac{18\tau+14-2\varepsilon}{4(9\tau+8)}n$*, or below* $\frac{18\tau+13+2\varepsilon}{4(9\tau+8)}n$*. Consequently, it is* NP*-hard to approximate the solution of* 3-SET PACKING *with exactly two occurrences of each element within any constant smaller than* $1+\frac{1}{18\tau_*+13}$*.*

(iii) It is NP*-hard to decide whether an instance of* 3-SET COVERING *with* n *triples and the occurrence of each element exactly in two triples has the minimum size of a covering above* $\frac{18\tau+19-2\varepsilon}{4(9\tau+8)}n$*, or below* $\frac{18\tau+18+2\varepsilon}{4(9\tau+8)}n$*. Consequently, it is NP-hard to approximate the solution of* 3-SET COVERING *with exactly two occurrences of each element within any constant smaller than* $1+\frac{1}{18\tau_*+18}$*.*

Proof. Consider a 3-regular triangle-free graph G as an instance of MAX-3-IS from Theorem 6. (i) Take a line-graph $L(G)$ of G. Nodes of G are transformed to triangles in $L(G)$ and this is one-to-one correspondence, as G was triangle-free. Clearly, independent sets of nodes in G are in one-to-one correspondence with triangle packings in $L(G)$, so the conclusion easily follows from Theorem 6. (ii) Create an instance of 3-SET PACKING that uses for 3-sets exactly triples of edges of G adjacent to each node of G. Clearly, independent sets of nodes in G are in one-to-one correspondence with packings of triples in the corresponding instance. Now the conclusion easily follows from the hard-gap for MAX-3-IS problem. (iii) Now a graph G from Theorem 6 is viewed as an instance of MIN-3-NC. Using the same collection of 3-sets as in the part (ii) we see that node covers in G are in one-to-one correspondence with coverings by triples in the new instance. The conclusion follows from the hard-gap result for MIN-3-NC from Theorem 6. □

5 New reduction for 3-Dimensional Matching

Definition and known results. Given the disjoint sets A, B, and C and a set $T \subseteq A \times B \times C$. A matching for T is a subset $T' \subseteq T$ such that no elements in T' agree in any coordinate. The goal of the MAXIMUM 3-DIMENSIONAL MATCHING problem (shortly, MAX-3DM) is to find cardinality of a maximum matching.

The problem is APX-complete even in the case if the number of occurrences of any element in A, B or C is bounded by a constant K ($K \geq 3$) [7]. For 'planar' instances problem admits a PTAS ([8]).

Recall that usually the hardness of MAX-3DM is proved by reduction from bounded instances of MAX-3-SAT. This approach is used in [7] and from given L-reduction lower bound $1 + \varepsilon$ (for a very small ε) can be counted. In what follows we present the new transformation from HYBRID to edge 3-colored instances of MAX-3-IS to obtain better inapproximability result for MAX-3DM.

Idea: If we have hardness result for MAX-3-IS on 3-regular edge-3-colored graphs, it is at the same time the result for MAX-3DM due to the following natural transformation. Suppose that edges of graph $G = (V, E)$ are properly colored with three colors a, b, c. Now define the sets $A = \{$all edges of color $a\}$, $B = \{$all edges of color $b\}$, $C = \{$all edges of $c\}$ and a set $T \subseteq A \times B \times C$ as $T = \{(e_a(v), e_b(v), e_c(v))$, for all $v \in V\}$, where $e_i(v)$ denotes an edge of color i incident to the node v.

It is easy to see that independent sets of nodes in G are in one-to-one correspondence with matchings of an instance obtained by the reduction above. So, the hardness result for MAX-3DM will immediately follow from the hardness result for MAX-3-IS on edge-3-colored graphs. It is not clear from Berman-Karpinski reduction ([2]) if their hardness results applies to the instances of MAX-3-IS are (in polynomial time) edge-3-colorable. Therefore we modify their reduction to obtain hardness result even for instances of MAX-3-IS that are easily edge-3-colorable.

Theorem 8 *Given $\varepsilon \in \left(0, \frac{1}{4}\right)$ and k be an integer such that $Q(\varepsilon, k)$ is* NP*-hard. Assume τ is such that there is a (k, τ)-amplifier. Then it is* NP*-hard to decide whether a 3-regular edge-3-colored instance of* MAX-3-IS *with n nodes has the maximum size of an independent set above $\frac{18\tau+16-2\varepsilon}{36(\tau+1)}n$, or below $\frac{18\tau+15+2\varepsilon}{36(\tau+1)}n$. Consequently, it is* NP*-hard to decide whether an instance of* MAX-3DM *with n-triples, each element occurring in exactly two triples, has the maximum size of a matching above $\frac{18\tau+16-2\varepsilon}{36(\tau+1)}n$, or below $\frac{18\tau+15+2\varepsilon}{36(\tau+1)}n$. Hence it is* NP*-hard to approximate* MAX-3DM *within any constant smaller than $1 + \frac{1}{18\tau_*+15}$ even on instances with exactly two occurrences of each element.*

Proof. Let $\varepsilon \in \left(0, \frac{1}{4}\right)$, k be a positive integer such that $Q(\varepsilon, k)$ is NP-hard, and $G = (V, E)$ be a fixed (k, τ)-amplifier. We describe a reduction of an instance I of HYBRID(G) to an edge-3-colored instance G' of MAX-3-IS. Recall that I has $\frac{6\tau}{9\tau-1}|I|$ variables, $i_3 := \frac{2}{9\tau-1}|I|$ equations with 3 variables, and $i_2 := \frac{9\tau-3}{9\tau-1}|I|$ equations with 2

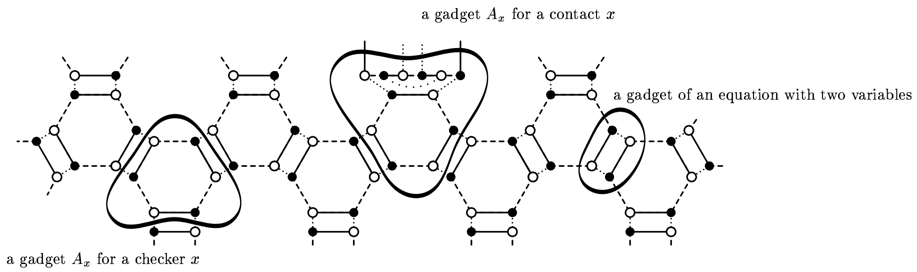

Figure 1: Example of gadgets for checkers, contacts and equations with two variables.

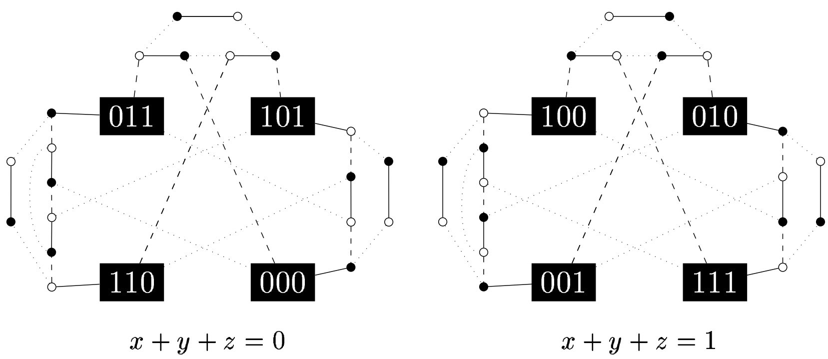

Figure 2: Equation gadgets with three variables.

variables. Each variable x of I is replaced with a gadget A_x. The gadget of a checker is a hexagon H_x in which nodes with labels 0 and 1 alternate. A gadget of a contact is a hexagon H_x augmented with a trapezoid T_x, a cycle of 6 (respectively, 8) nodes that shares one edge with a hexagon H_x. Furthermore, we add a chord for an 8-cycle (see Fig. 1). Again, labels 0 and 1 of nodes in those cycles alternate.

If two variables (i.e. nodes) x, y are connected by an equation (i.e. disjoint edge), $x = y$, we connect their hexagons with a pair of edges (so called "connections") to form a rectangle in which the in which nodes with label 0 and 1 alternate. The rectangle thus formed is a gadget of an equation with two variable (see Fig. 1).

If three variables are connected by an equation (i.e. an hyperedge), say, $x+y+z = 0$, the trapezoids T_x, T_y and T_z are coupled with the set S_{xyz} of four *special* nodes (see Fig. 2). The trapezoid of the first variable of the equation, T_x, is an 8-cycle with a chord, trapezoids T_y and T_z are 6-cycles.

The produced instance $G' = (V', E')$ has n nodes, where $n = 4i_2 + 24i_3 = \frac{36(\tau+1)}{9\tau-1}|I|$. Edges of G' are 3-colored as depicted on Figures 1–2 such that all connections are of the same color (dotted lines), which alternates on rectangles with the second color (full lines). On hexagons full line alternates with dashed.

Given an assignment $\varphi : \mathcal{V}(I) \to \{0, 1\}$, we can describe an IS in G' that corresponds to the assignment as follows: Take preliminary a set N without special nodes with the property that for each $x \in \mathcal{V}(I)$ $N \cap A_x$ consists of the set of all nodes in A_x labeled by $\varphi(x)$. Now we modify N:

- For each equation of the form $x = y$ which is not satisfied by φ we remove one (arbitrary) of two nodes of N from that rectangle, which corresponds to the gadget of the equation.

- For each equation of the form $x+y+z=0$ (respectively, $x+y+z=1$) satisfied by φ we add to N the special node from S_{xyz} labeled by triple $\varphi(x)\varphi(y)\varphi(z)$.

Any independent set N obtained from some assignment in this way is called *standard.* It is easy to count the cardinality of any standard IS, it is $2i_2 + 11i_3 - \text{UNSAT}(\varphi)$, where $\text{UNSAT}(\varphi)$ means the number of equations of I unsatisfied by φ. Taking φ optimal, i.e. such that $\text{UNSAT}(\varphi) = (1 - \text{OPT}(I)) \cdot |I|$, implies

$$\begin{aligned}\text{OPT}(G') &\geq \max\{|N| : N \text{ is a standard IS}\} = 2i_2 + 11i_3 - (1 - \text{OPT}(I))|I| \\ &= \frac{n}{36(\tau+1)}\left(18\tau + 16 - (1 - \text{OPT}(I))(9\tau - 1)\right).\end{aligned}$$

In what follows we show that, in fact, equality holds, showing that there is a standard IS in G' of cardinality $\text{OPT}(G')$. Having this simple functional dependence of $\text{OPT}(G')$ on $\text{OPT}(I)$ for granted we immediately conclude the hard gap result claimed in this theorem from that of $\text{HYBRID}(G)$.

Put $\mathscr{M} = \{M : M \text{ is an IS in } G' \text{ with } |M| = \text{OPT}(G')\}$. We prove that there is $M \in \mathscr{M}$ such that M is a standard IS showing that some extremal elements of $\mathscr{M}$ have to be standard. The definition of extremality we use looks artificial, but it fits well to our purpose. Let us keep $M \in \mathscr{M}$ fixed. If for a subset A of a node gadget A_v the set $A \cap M$ contains nodes of one type only (i.e. only 0-nodes, or only 1-nodes), the set A is termed *pure*; otherwise it is *dirty.* Let us introduce the following notation for M: $n_1(M)$ is the number of dirty hexagons, $n_2(M)$ is the number of special nodes in M, $n_3(M)$ is the number of nodes $x \in M$ such that for some contact u with H_u pure, $x \in T_u$ and of distinct type from $H_u \cap M$. For M_1, $M_2 \in \mathscr{M}$ we write $M_1 \prec M_2$ whenever $(n_1(M_1), n_2(M_1), n_3(M_1)) < (n_1(M_2), n_2(M_2), n_3(M_2))$ in a lexicographic order.

Now we prove in the series of claims that any minimal element M of $(\mathscr{M}, \prec)$ is a standard IS of G'.

Claim 1 *If u is a checker then A_u $(= H_u)$ is pure.*

Proof. If H_u is dirty, then $H_u \cap M$ consists of two opposite nodes a_0, a_1 of H_u. (Here a_i has a label i, the same holds in what follows for all indices.) One of two sides of H_u that are "parallel" to the main diagonal (a_0, a_1) of H_u, say (b_1, b_0), has both nodes in the same rectangle. Let c_0, c_1 stand for the other vertices of this rectangle. Since $(c_0, c_1) \in E'$ there is $j \in \{0,1\}$ such that $c_j \notin M$. But replacing $a_j \in M$ by b_{1-j} leads to $M' \in \mathscr{M}$ with $M' \prec M$, a contradiction. Hence H_u is pure. □

Claim 2 *If u is a contact with dirty hexagon H_u of its gadget A_u, then T_u is a 6-cycle and $A_u \cap M = \{a_0, a_1, c_0, c_1\}$, where $a_0, a_1 \in H_u \setminus T_u$ are vertices connected to T_u, and $c_0, c_1 \in T_u \setminus H_u$ are connected to H_u.*

Moreover, if T_u is a 6-cycle with vertices b_0, b_1, c_0, d_1, d_0, c_1 in this order, then from the 3-variable equation gadget corresponding to u, exactly two special vertices are in M, those adjacent to d_0 and d_1.

Proof. H_u is dirty, thus $H_u \cap M$ are opposite vertices a_0, a_1 of H_u. One of two sides of H_u "parallel" to (a_0, a_1), say (b_1, b_0), has both vertices in the same equation gadget. Let c_0, c_1 stand for the vertices of this gadget that are connected to b_1 and b_0. If there is $j \in \{0,1\}$ such that $c_j \notin M$ (which is always true if b_0, b_1, c_0 and c_1 forms a rectangle) we get the contradiction in the same way as in the proof of Claim 1; replacing a_j by b_{1-j} in M we get $M' \in \mathscr{M}$ with less dirty hexagons. Consequently, (b_1, b_0) is an edge of a trapezoid T_u, and $c_i \in M$ for both $i \in \{0,1\}$ (see Fig. 3).

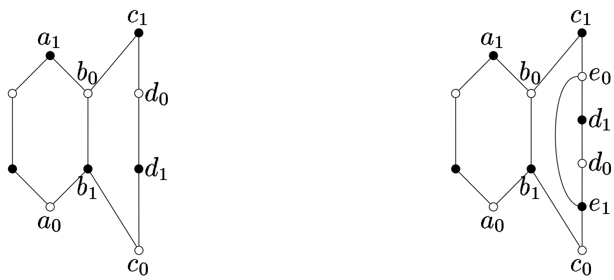

Figure 3:

Assume now, that T_u is a 8-cycle with a chord and with nodes b_0, b_1, c_0, e_1, d_0, d_1, e_0, c_1 in this order (see Fig. 3). As $(d_0, d_1) \in E'$, there is $j \in \{0, 1\}$ such that $d_j \notin M$. Now replacing $c_j \in M$ by e_{1-j} and $a_j \in M$ by b_{1-j} we get $M' \prec M$, a contradiction. Consequently, T_u is a 6-cycle, with $\{a_0, a_1, c_0, c_1\} \subseteq M$. Thus clearly $A_u \cap M = \{a_0, a_1, c_0, c_1\}$.

Obviously, special nodes adjacent to either c_0 or c_1 are not in M. To complete the proof, assume that for some $j \in \{0, 1\}$ the special node adjacent to d_j is not in M. But replacing $c_{1-j} \in M$ by d_j and a_{1-j} by b_j we obtain less dirty hexagons, a contradiction that complete the proof of claim. □

Claim 3 *Assume that u is a contact with pure H_u. Then $H_u \cap T_u \cap M$ contains a node, say b_j, ($j \in \{0, 1\}$ and the notation of nodes is the same as in the proof of Claim 2).*

(i) If T_u is pure, then it is also full, i.e. $T_u \cap M$ contains all nodes of T_u labeled by j.

(ii) If T_u is a dirty 6-cycle, then $T_u \cap M = \{b_j, d_{1-j}\}$ and both special nodes adjacent to c_j and d_j belong to M.

(iii) If T_u is a dirty 8-cycle with a chord, then $T_u \cap M = \{b_j, e_{1-j}, d_{1-j}\}$ and both special nodes adjacent to c_j and d_j belong to M.

Proof. Take $j \in \{0, 1\}$ that coincides with label of each node in $H_u \cap M$. We first prove that $b_j \in M$ (recall that $b_j \in H_u \cap T_u$). If $b_j \notin M$ and $M \in \mathscr{M}$, a neighbor of b_j, necessarily c_{1-j}, belongs to M. But replacing $c_{1-j} \in M$ by b_j leads to $M' \in \mathscr{M}$ with $M' \prec M$ a contradiction. Consequently, $b_j \in H_u \cap T_u \cap M$.

(i) Assume that T_u is pure, i.e. $T_u \cap M$ contains nodes labeled by j only. We prove that T_u is full. Fix a node $v \in T_u$ labeled by j. Assume, on the contrary, that $v \notin M$. As $M \in \mathscr{M}$, it is only possible if a neighbor of v (necessarily one of special nodes) belongs to M. Replacing this special node in M by v we obtain $M' \in \mathscr{M}$ with $M' \prec M$, a contradiction. Consequently, each $v \in T_u$ labeled by j belongs to M, hence T_u is pure and full.

(ii) If T_u is a dirty 6-cycle, as $b_j \in M$, necessarily $T_u \cap M = \{b_j, d_{1-j}\}$. If the special node adjacent to c_j (respectively, d_j) does not belong to M, replacing $d_{1-j} \in M$ by c_j (respectively, by d_j) leads to $M' \in \mathscr{M}$ with $M' \prec M$, a contradiction. Hence both those special nodes belong to M.

(iii) Assume T_u is a dirty 8-cycle with a chord. Clearly, at least one of d_{1-j} and e_{1-j} belongs to M. If only one of them belongs to M, then replacing it in M by e_j leads to $M' \prec M$, a contradiction. Now it is obvious that $T_u \cap M = \{b_j, e_{1-j}, d_{1-j}\}$. If the special vertex adjacent to c_j (respectively, d_j) does not belong to M, replacing $e_{1-j} \in M$ by c_j (respectively, d_j) and d_{1-j} by e_j leads to $M' \in \mathscr{M}$ with $M' \prec M$, a contradiction that completes the proof of claim. □

Let us summarize the properties of $M \in \mathscr{M}$:

(i) The gadgets of all checkers are pure.

(ii) If a gadget of a contact is pure, it has a full trapezoid.

The following are the only possibilities of a dirty gadget of a contact u:

(iii) T_u is an 8-cycle with a chord, H_u is pure, and for some $j \in \{0,1\}$ $T_u \cap M = \{b_j, e_{1-j}, d_{1-j}\}$ and both special nodes adjacent to c_j and d_j belong to M.

(iv) T_u is a 6-cycle, H_u is pure, and for some $j \in \{0,1\}$, $T_u \cap M = \{b_j, d_{1-j}\}$ and both special nodes adjacent to c_j and d_j belong to M.

(v) T_u is a 6-cycle, $A_u \cap M = \{a_0, a_1, c_0, c_1\}$, and from the 3-variable equation gadget corresponding to u exactly 2 special nodes belong to M, namely those adjacent to d_0 and d_1.

Our aim is to prove that, in fact, the gadget of no contact is dirty. As then all trapezoids are pure and full, it easily follows (as $M \in \mathscr{M}$) that M is a standard independent set, completing the proof.

To obtain the proof by contradiction, assume that a gadget of some contact is dirty. Consider the 3-variable equation gadget in which this dirty contact appears. We can assume that the corresponding equation is $x + y + z = 0$, the case of $x + y + z = 1$ being completely analogous. Recall that T_x is an 8-cycle with a chord, T_y and T_z are 6-cycles.

If T_x is a dirty set, then for some $j \in \{0,1\}$, $T_x \cap M = \{b_j, e_{1-j}, d_{1-j}\}$ and both special nodes adjacent to c_j and d_j belong to M. But in T_y (resp. T_z) these special nodes are connected with nodes of distinct type. It implies, T_y and T_z cannot be full trapezoids and due to (ii), both their gadgets are dirty. The case (v) does not occur for any of them, because both special nodes from M are not adjacent to d-nodes of T_y (resp. T_z). Hence the case (iv) occurs for both, which in turn implies that the special node adjacent to d_{1-j} (in T_x) belongs to M, a contradiction.

If T_x is a pure set, then it is full (according (ii)) and M contains at most 2 of special nodes. If we suppose T_y or T_z dirty, then M contains exactly 2 special nodes. But from the property of M, the nodes from $M \cap S_{xyz}$ are adjacent in T_y with two nodes of distinct type, but one is not d-node. It means (iv) and (v) does not occur and T_y is pure. The same arguments hold also for T_z. □

Conclusion. There is still substantial gap between the lower and upper approximation bounds for small occurrence combinatorial optimization problems. The method of parametrized amplifiers shows better the quality of used reductions and the possibilities for further improvement of lower bounds. But it is quite possible that the upper bounds can be improved more significantly.

References

[1] P. Berman and T. Fujito: *Approximating independent sets in degree* 3 *graphs*, Proc. 4th Workshop on Algorithms and Data Structures, Lecture Notes in Computer Science 955, Springer-Verlag, Berlin, 1995, 449–460.

[2] P. Berman and M. Karpinski: *On some tighter inapproximability results, further improvements*, Electronic Colloquium Complexity, Report No. **65**, 1998.

[3] L. Engebretsen and M. Karpinski: *Approximation hardness of TSP with bounded metrics*, Proceedings of 28th International Colloquium on Automata, Languages and Programming (ICALP), Crete, Greece, 2001, LNCS 2076, 201–212.

[4] J. Håstad: *Some optimal inapproximability results*, Journal of ACM **48** (2001), 798–859.

[5] C. A. J. Hurkens and A. Schrijver: *On the size of systems of sets every t of which have an SDR, with an application to the worst-case ratio of heuristics for packing problems*, SIAM J. Discrete Mathematics **2** (1989), 68–72.

[6] M. Chlebík and J. Chlebíková: *Approximation Hardness of the Steiner Tree Problem on Graphs*, Proceedings of the 8th Scandinavian Workshop on Algorithm Theory, SWAT 2002, Turku, Finland, 2002, Springer, LNCS **2368** (2002), 170–179.

[7] V. Kann: *Maximum bounded 3-dimensional matching is MAX SNP-complete*, Information Processing Letters **37** (1991), 27–35.

[8] T. Nishizeki and K. Kashiwagi: *On the 1.1 edge-coloring of multigraphs*, SIAM J. Discrete Mathematics **3** (1990), 391–410.

[9] C. H. Papadimitriou and M. Yannakakis: *Optimization, approximation, and complexity classes*, J. Computer and System Sciences **43** (1991), 425–440.

[10] C. H. Papadimitriou and S. Vempala: *On the Approximability of the Traveling Salesman Problem*, Proceedings of the 32nd ACM Symposium on the theory of computing, Portland, 2000.

[11] M. Thimm: *On the Approximability of the Steiner Tree Problem*, Proceedings of the 26th International Symposium, MFCS 2001 Mariánske Lázne, Czech Republic, August 27-31, 2001, Springer, Lecture Notes in Computer Science **2136** (2001), 678–689.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'