RC24848 (W0908-125) August 24, 2009
Computer Science

IBM Research Report

Optimal Long Code Test with One Free Bit

Nikhil Bansal
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Subhash Khot
NYU

Optimal Long Code Test with One Free Bit

Nikhil Bansal Subhash Khot

March 31, 2009

Abstract

For arbitrarily small constants $\varepsilon, \delta > 0$, we present a long code test with one free bit, completeness $1 - \varepsilon$ and soundness δ. Using the test, we prove the following two inapproximability results:

1. Assuming the Unique Games Conjecture of Khot [17], given an n-vertex graph that has two disjoint independent sets of size $(\frac{1}{2} - \varepsilon)n$ each, it is NP-hard to find an independent set of size δn.
2. Assuming a (new) stronger version of the Unique Games Conjecture, the scheduling problem of *minimizing weighted completion time with precedence constraints* is inapproximable within factor $2 - \varepsilon$.

1 Introduction

The celebrated PCP Theorem [12, 4, 3] states that every NP-statement has a *probabilistically checkable proof* where the verifier runs in polynomial time, reads only a constant number of bits from the proof and uses only a logarithmic amount of randomness. The verifier has the completeness property that every correct statement has a proof that is accepted with probability 1 and the soundness property that every proof of an incorrect statement is accepted with only a small probability, say 0.1. In general, one may have parameters $1 \geq c > s > 0$ that specify the probabilities in the completeness and soundness case respectively.

The PCP Theorem is intimately related to *inapproximability* results for NP-complete problems, i.e. results showing that for many NP-complete problems, even computing approximate solutions remains NP-hard. This connection has been used to design customized PCPs and prove strong (and in many cases optimal) inapproximability results for several NP-complete problems. One of the most notable results is by Håstad [16] showing that given a satisfiable 3SAT formula, each clause containing three distinct literals, it is NP-hard to find an assignment that satisfies a fraction c of its clauses for any constant $c > \frac{7}{8}$. Since a random assignment satisfies $\frac{7}{8}$ fraction, $\frac{7}{8}$ is a sharp approximability threshold for 3SAT. However, for many fundamental problems, their approximability threshold remains unknown. In this paper, we address two such problems, namely Vertex Cover and a scheduling problem denoted by $1|prec|\sum w_j C_j$, i.e. minimizing weighted completion time on a single machine with precedence constraints.

1.1 Vertex Cover

Given a graph $G(V, E)$, a vertex cover is a subset of vertices $V' \subseteq V$ such that every edge has at least one endpoint in V'. The complement $V \setminus V'$ of a vertex cover is an independent set. It is well-known that the minimum vertex cover can be approximated in polynomial time within factor 2. The best known inapproximability result is 1.36 NP-hardness due to Dinur and Safra [10]. Assuming the Unique Games conjecture, Khot and Regev [19] showed that it is NP-hard to approximate Vertex Cover within any factor better than 2. As observed by Bellare, Goldreich, and Sudan [6], inapproximability of the independent set problem (hence of vertex cover) is equivalent to constructing a PCP with *zero free bits*.

A PCP verifier is said to make f *free queries* if there are at most 2^f settings of query bits for which the verifier accepts. Note that there is no restriction on the number of queries itself. For example, suppose that the verifier reads three bits x, y, and z, accepting if and only if $x \oplus y \oplus z = 0$. Then the number of free queries is two, since there are $2^2 = 4$ accepting answers, namely $(0,0,0), (0,1,1), (1,0,1)$, and $(1,1,0)$. Let us denote by $\mathrm{fPCP}_{c,s}[r(n), f(n)]$ the class of languages that have a proof that can be probabilistically checked by a polynomial time verifier that uses at most $r(n)$ random bits, $f(n)$ free queries, and has completeness c and soundness s. Bellare et al [6] showed that:

Theorem 1.1 ([6]) *The following are equivalent:*

1. $\mathrm{NP} \subseteq \mathrm{fPCP}_{c,s}[O(\log n),\ \mathrm{zero}]$.
2. *There is a polynomial time reduction mapping a SAT formula ϕ to an n-vertex graph G such that if ϕ is satisfiable then G has an independent set of size cn and if ϕ is unsatisfiable, then G has no independent set of size sn.*

Note that in a PCP with zero free bits, there is only one accepting answer, and therefore the verifier already *knows* what answer to expect from the proof. Nevertheless, such PCPs are powerful enough to capture NP by the virtue of *imperfect* completeness. Note also that a reduction as in this theorem would imply that it is NP-hard to approximate Vertex Cover better than factor $\frac{1-s}{1-c}$. Using this equivalence between PCPs and the independence problem, Khot and Regev's $2-\varepsilon$ inapproximability result for vertex cover can be stated as:

Theorem 1.2 ([19]) *Assuming the Unique Games Conjecture, for arbitrarily small constants $\varepsilon, \delta > 0$,*

$$\mathrm{NP} \subseteq \mathrm{fPCP}_{\frac{1}{2}-\varepsilon,\delta}[O(\log n),\ \mathrm{zero}].$$

In this paper, we prove a stronger result:

Theorem 1.3 *Assuming the Unique Games Conjecture, for arbitrarily small constants $\varepsilon, \delta > 0$,*

$$\mathrm{NP} \subseteq \mathrm{fPCP}_{1-\varepsilon,\delta}[O(\log n),\ \mathrm{one}].$$

In words, assuming the UGC, we construct a PCP with one free bit (i.e. two accepting answers) that has near-perfect completeness and arbitrarily low soundness. Since a verifier may further decide, at random, to pick one of the two answers as an accepting answer, our construction implies a PCP with zero free bits (i.e. one accepting answer), completeness $\frac{1}{2}-\varepsilon$, and arbitrarily low soundness. Thus Theorem 1.3 implies Theorem 1.2. If one phrases these theorems in terms of independent set problem, we prove (assuming UGC) that given an n-vertex graph that has two disjoint independent sets of size $(\frac{1}{2}-\varepsilon)n$ each (i.e. the graph is almost 2-colorable), it is NP-hard to find an independent set of size δn. On the other hand, Khot and Regev prove that (assuming UGC) given an n-vertex graph that has an independent set of size $(\frac{1}{2}-\varepsilon)n$ each, it is NP-hard to find an independent set of size δn.

In addition to being stronger, our result has the following interesting features: (1) Both Khot-Regev [19] and its precursor Dinur-Safra [10] results, are more naturally viewed as combinatorial reductions whereas our result is more naturally viewed as a PCP construction[1]. (2) Unlike [19, 10], we do not need to use *biased* long codes (see below). (3) In [19], one first needs to transform a given Unique Games instance into another instance with some special properties; we do not need this transformation.

[1]Note however that the combinatorial and PCP views are equivalent by Theorem 1.1.

1.2 The Long Code Test

Most recent PCPs have, as a central building block, a probabilistic procedure to check a *long code*. A boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ is called a dictatorship if it depends on only co-ordinate, i.e. for some $1 \leq i \leq n$, $f(x_1, \ldots, x_n) = x_i$. A long code of an index $i \in \{1, \ldots, n\}$ is simply the truth-table of the dictatorship function of co-ordinate i. Thus we desire a probabilistic procedure that has access to the truth-table of a function f, accepts with probability at least c if f is a dictatorship, and accepts with probability at most s if f is *far from* a dictatorship. To formalize what we mean by *far from a dictatorship*, let influence of the co-ordinate j be defined as $\Pr_x[f(x) \neq f(x \oplus e_j)]$ where e_j is the input with the j^{th} co-ordinate equal to 1 and the rest of the co-ordinates 0. Note that if f is a dictatorship of co-ordinate i, then the influence of that co-ordinate equals 1 and all other influences are zero. Thus it makes sense, and turns out useful, to call a function as far from a dictatorship if all its *influences* are small[2].

Naturally, the properties that we desire of the PCP construction, namely the number of free queries, completeness probability c and the soundness probability s, are inherited from the respective properties of the long code test. We propose the following new test that is at the heart of our PCP constructions:

Theorem 1.4 *Given a function $f : \{0,1\}^n \mapsto \{0,1\}$ that is balanced, i.e. (say) $\frac{1}{3} \leq \mathbb{E}[f] \leq \frac{2}{3}$, and constants $\varepsilon, \delta > 0$, there is a constant $\eta = \eta(\varepsilon, \delta) > 0$ and a probabilistic test such that:*

1. *The test makes one free query.*
2. *(Completeness) If f is a dictatorship, the test accepts with probability at least $1 - \varepsilon$.*
3. *(Soundness) If all influences of f are smaller than η, then the test accepts with probability at most δ.*

It is quite easy to describe our test. It picks a random subset $S \subseteq \{1, \ldots, n\}$ of co-ordinates of size $|S| = \varepsilon n$ and a random input $x \in \{0,1\}^n$. Let $C_{x,S}$ denote the sub-cube of the hypercube defined as $C_{x,S} := \{z \in \{0,1\}^n \mid \forall j \notin S,\ z_j = x_j\}$. Note that the sub-cube contains $2^{|S|} = 2^{\varepsilon n}$ points. The test accepts if and only if the function f is constant on the sub-cube $C_{x,S}$. Since there are only two accepting answers, namely the constant 0 or constant 1 on the sub-cube, the test makes only one *free* query (the number of queries however is huge, namely $2^{|S|}$). If f is the dictatorship of co-ordinate i, then with probability $1-\varepsilon$, $i \notin S$, and in that case f is constant on the sub-cube $C_{x,S}$. Thus a dictatorship is accepted with probability at least $1 - \varepsilon$. Finally, we give a fairly short proof (see Section 3.1) of the soundness property, though it also follows directly from the *It Ain't Over till It's Over* Theorem, first conjectured by Friedgut and Kalai, and proved by Mossel, O'Donnell, and Oleszkiewicz [24]. In fact, Mossel et al's theorem is stronger in the following respect: they show that for a (balanced) function with all small influences, w.h.p. a random sub-cube of dimension εn has a non-negligible *fraction* of both 0s and 1s. On the other hand, we only show that w.h.p., a random sub-cube of dimension εn is non-constant (and this is all we need for our purpose). Nevertheless, our proof is quite different, without relying on the use of invariance principle. We believe that our test and the soundness analysis could be useful towards proving inapproximability of graph coloring problems.

We also use our long code test to prove a new inapproximability result for a scheduling problem that we describe next. In fact, the scheduling problem was our original motivation that led us to the new long code test and the application to the vertex cover problem.

[2]We actually need a more refined notion in terms of *low degree* influences.

1.3 Weighted Completion Time with Precedence Constraints

We consider the following scheduling problem known as minimum total weighted completion time with precedence constraints, denoted by $1|prec|\sum w_j C_j$ in the standard scheduling notation. There is a set $J = \{1, \ldots, n\}$ of jobs. Job i has size p_i and weight w_i. In addition there is a collection of precedence constraints specified by a directed acyclic graph (i.e. a partial order) $G = (J, P)$ where $(i, j) \in P$ implies that job i must be completed before job j is started. The goal is to find an ordering of jobs that minimizes the total weighted completion time $\sum_j w_j C_j$, where the completion time C_j of job j is the time at which it is completed in the schedule.

This is a classic scheduling problem and has been extensively studied since the 70s [29, 26, 21, 11]. Since then, many approximation algorithms that achieve a guarantee of 2 have been designed for it using several different techniques and various different linear programming relaxations [27, 15, 7, 14, 23, 25, 1]. However, despite much interest, there has been a relatively large gap in our understanding of the approximability of this problem. On one hand no approximation better than 2 is known, and yet until recently only NP-Hardness was known for it [21, 22]. The first inapproximability result of any kind was shown only recently by Ambuhl, Mastrolli and Svensson [2] who ruled out the possibility of a polynomial time approximation scheme. Specifically, they use the Quasi-random PCP due to Khot [18], which is based on the randomized subexponential time hypothesis, to show their result. Narrowing the approximability gap for $1|prec|\sum w_j C_j$ is considered one of the most important open problems in scheduling, see for example [28].

Understanding the approximability of $1|prec|\sum w_j C_j$ is also interesting because of its relation to Vertex Cover. While a connection to vertex cover was long suspected (see for example [28]), it has been established relatively recently in a series of papers [8, 9, 1]. The connection is quite non-trivial and we do not describe the details here, but at a high-level, the objective function of the problem can be divided into a fixed part and a variable part, where the variable part can be viewed as a vertex cover problem with a special structure. This structure has been used to obtain better than 2 approximation algorithms for various special cases of $1|prec|\sum w_j C_j$. Recently, [2] showed that the variable part in fact is as general as vertex cover, and gave a reduction where an arbitrary vertex cover instance can be converted into an instance of $1|prec|\sum w_j C_j$. However, even if we assume that vertex cover is $2-\varepsilon$ hard to approximate, their result does not imply even APX Hardness for $1|prec|\sum w_j C_j$ as the cost of fixed part is much larger than the variable part. In fact reducing the fixed part cost is quite challenging as it is known that if the fixed part is removed, then the problem becomes trivial and polynomially time solvable. One may speculate that perhaps the fixed part may always have substantial contribution in hard instances of the problem, and hence a better than 2 approximation may indeed be possible for $1|prec|\sum w_j C_j$. Contrary to such speculation, we prove:

Theorem 1.5 *Assuming a new variant of the Unique Games conjecture (specifically Hypothesis 5.1), it is* NP*-hard to approximate the scheduling problem* $1|prec|\sum w_j C_j$ *within any factor strictly less than* 2*.*

2 Preliminaries

This section describes some of the technical tools needed in the paper.

2.1 Influences and Friedgut's Theorem

Given a boolean function $f : \{0,1\}^n \mapsto \{0,1\}$, the influence of its i-th co-ordinate is defined as

$$\mathsf{Infl}_i(f) := \Pr_x[f(x) \neq f(x \oplus e_i)].$$

The i^{th} influence can also be expressed as

$$\mathsf{Infl}_i(f) = \sum_{S:i\in S} \widehat{f}(S)^2.$$

The total influence of a function is the sum of all influences $\mathsf{Infl}(f) := \sum_{i=1}^n \mathsf{Infl}_i(f) = \sum_S |S|\widehat{f}(S)^2$. The sum of influences is also referred to as the average sensitivity. The degree k-influence of i^{th} co-ordinate $\mathsf{Infl}_i^k(f)$ is defined as

$$\mathsf{Infl}_i^k(f) = \sum_{S:i\in S,|S|\le k} \widehat{f}(S)^2.$$

Since $\sum_S \widehat{f}(S)^2 \le 1$, the sum of all degree k-influences is at most k. A function $h : \{0,1\}^n \mapsto \{0,1\}$ is a k-junta if it only depends on k co-ordinates. Two functions f and g are γ-close if $\Pr_x[f(x) \neq g(x)] \le \gamma$.

Theorem 2.1 (Freidgut [13]) *Let f be a boolean function on $\{0,1\}^n$ with average sensitivity at most s and $\gamma > 0$. Then there exists a k-junta h that is γ-close to f, where $k = \lfloor e^{3s/\gamma} \rfloor$.*

In contra-positive, Freidgut's theorem says that a boolean function that is not close to any junta must have a large average sensitivity.

Lemma 2.2 *If $f : \{0,1\}^n \mapsto \{0,1\}$ is such that:*

1. $\mathbb{E}[f] \ge \delta$.
2. *There exists a set $B \subseteq [n], |B| = k$ and a string $s \in \{0,1\}^k$ such that*

$$\Pr_x\left[f(x) = 0 \middle| x|_B = s\right] \ge 1 - \delta/2.$$

Then there is a variable $i \in B$ such that $\mathsf{Infl}_i^k(f) \ge \delta^2/2^{2k+4}$.

Proof: Assume w.l.o.g. that $B = \{1, \dots, k\}$. We know that

$$\Pr\left[f(s_1, \dots, s_k, x_{k+1}, \dots, x_n) = 0\right] \ge 1 - \delta/2.$$

This implies that

$$\mathbb{E}_x\left[f \cdot \left(\prod_{i=1}^k \frac{1 + (-1)^{x_i \oplus s_i}}{2}\right)\right] \le \frac{\delta/2}{2^k}. \tag{1}$$

Since $\mathbb{E}[f] \ge \delta$, there exists input $t \in \{0,1\}^k,\ t \neq s$ such that

$$\Pr\left[f(t_1, \dots, t_k, x_{k+1}, \dots, x_n) = 1\right] \ge \delta.$$

This implies that

$$\mathbb{E}_x\left[f \cdot \left(\prod_{i=1}^k \frac{1 + (-1)^{x_i \oplus t_i}}{2}\right)\right] \ge \frac{\delta}{2^k}. \tag{2}$$

Subtracting Equation (1) from Equation (2), we get

$$\frac{1}{2^k} \sum_{A\subseteq[k],\ A\neq\emptyset} \mathbb{E}_x\left[f \cdot \left(\prod_{i\in A} (-1)^{x_i}\right)\right] \cdot \left(\prod_{i\in A} (-1)^{t_i} - \prod_{i\in A} (-1)^{s_i}\right) \ge \frac{\delta - \delta/2}{2^k} = \delta/2^{k+1},$$

which implies that

$$\frac{1}{2^k} \sum_{A \subseteq [k],\ A \neq \emptyset} |\widehat{f}(A)| \cdot 2 \ \ \geq \delta/2^{k+1}.$$

This implies that there exists $A \subseteq [k], A \neq \emptyset$ and $|\widehat{f}(A)| \geq \delta/2^{k+2}$. Any variable in A satisfies the requirement of the lemma. ■

Lemma 2.3 *Suppose $f, g : \{0,1\}^n \mapsto \{0,1\}$ are two functions such that*

$$f \leq g \quad \textit{and} \quad \delta \leq \mathbb{E}[f] \leq \mathbb{E}[g] \leq 1 - \delta.$$

Further assume that g is γ-close to a k-junta for some $\gamma \leq \delta^2/4$. Then $\exists\, i$ such that $\mathsf{Infl}_i^k(f) \geq \delta^2/2^{2k+4}$.

Proof: Let $h : \{0,1\}^k \mapsto \{0,1\}$ be a function such that (after re-ordering indices),

$$\Pr\left[g(x) = h(x_1, x_2, \ldots, x_k)\right] \geq 1 - \gamma.$$

Since $\mathbb{E}[g] \leq 1 - \delta$, we have $\mathbb{E}[h] \leq 1 - \delta + \gamma \leq 1 - \delta/2$. Also there exists input $s \in \{0,1\}^k$ such that $h(s) = 0$ and

$$\Pr\left[g(s_1, \ldots, s_k, x_{k+1}, \ldots, x_n) = h(s) = 0\right] \geq 1 - 2\gamma/\delta \geq 1 - \delta/2.$$

Since $f \leq g$, we have that

$$\Pr\left[f(s_1, \ldots, s_k, x_{k+1}, \ldots, x_n) = 0\right] \geq 1 - \delta/2.$$

Using this condition and the fact that $\mathbb{E}[f] \geq \delta$, we can apply Lemma 2.2 and conclude the existence of a variable $i \in \{1, \ldots, k\}$ such that $\mathsf{Infl}_i^k(f) \geq \delta^2/2^{2k+4}$. ■

2.2 Sunflower Lemma

A sunflower with k petals and a core Y is a collection of distinct sets $S_1 \ldots S_k$ such that $S_i \cap S_j = Y$ for all $i \neq j$. The sets $S_i \setminus Y$ are the petals. The following lemma is well-known.

Theorem 2.4 (Sunflower lemma) *Let $\mathcal{F}$ be family of sets each of cardinality s. If $|\mathcal{F}| > s!(k-1)^s$ then $\mathcal{F}$ contains a sunflower with k-petals.*

2.3 A Probability Estimate

Lemma 2.5 *Suppose $X_1, \ldots, X_m$ are $[0,1]$-valued random variables defined on the space $\Omega = [n]^m$, such that $\forall j \in \{1, \ldots, m\}$, X_j depends only on the first j co-ordinates $i_1, \ldots, i_j$ of $\omega = (i_1, \ldots, i_m) \in \Omega$ and for any values of $i_1, \ldots, i_{j-1}$, we have that $\mathbb{E}[X_j | i_1, \ldots, i_{j-1}] \geq \beta$. Then*

$$\mathbb{E}\left[e^{-\sum_{j=1}^m X_j}\right] \leq e^{-\beta m/2}.$$

Proof: We first note that if a random variable Y takes values in $[0,1]$ and $\mathbb{E}[Y] \geq \beta$, then by convexity, $\mathbb{E}[e^{-Y}] \leq (1-\beta)e^{-0} + \beta e^{-1} = 1-(1-1/e)\beta \leq e^{-\beta/2}$.

We prove the lemma by induction on m. By the observation above, the lemma clearly holds for $m=1$. Assume that it holds for m, and let $X_1,\ldots,X_{m+1}$ be random variables on $[n]^{m+1}$ satisfying the conditions of the lemma. Note that for every value $i_1 = t$, the variables $X_2|_{i_1=t},\ldots,X_{m+1}|_{i_1=t}$ are random variables on $[n]^m$ and satisfy the conditions of the lemma. Thus by induction hypothesis

$$\mathbb{E}\left[e^{-\sum_{j=2}^{m+1} X_j|_{i_1=t}}\right] \leq e^{-\beta m/2}.$$

Therefore

$$\begin{aligned}
\mathbb{E}\left[e^{-\sum_{j=1}^{m+1} X_j}\right] &= \sum_{t=1}^{n} \Pr[i_1 = t]\cdot e^{-X_1(t)} \cdot \mathbb{E}\left[e^{-\sum_{j=2}^{m+1} X_j|_{i_1=t}}\right] \\
&\leq \sum_{t=1}^{n} \Pr[i_1 = t]\cdot e^{-X_1(t)} \cdot e^{-\beta m/2} \\
&= \mathbb{E}[e^{-X_1}]\cdot e^{-\beta m/2} \leq e^{-\beta(m+1)/2},
\end{aligned}$$

completing the inductive step. ∎

2.4 The Unique Games Conjecture

Definition 2.6 *An instance $\mathcal{L}(G(V,W,E),[n],\{\pi_{v,w}\}_{(v,w)\in E})$ of* Unique Games *consists of a regular bipartite graph $G(V,W,E)$ and a set $[n]$ of* labels. *For each edge $(v,w)\in E$ there is a constraint specified by a permutation $\pi_{v,w} : [n] \mapsto [n]$. The goal is to find a* labelling *$\ell : V\cup W \to [n]$ of the vertices such that as many edges as possible are satisfied, where an edge $e=(v,w)$ is said to be satisfied if $\ell(v) = \pi_{v,w}(\ell(w))$.*

Definition 2.7 *Given a Unique Games instance $\mathcal{L}(G(V,W,E),[n],\{\pi_{v,w}\}_{(v,w)\in E})$ let $\mathrm{OPT}(\mathcal{L})$ denote the maximum fraction of simultaneously satisfied edges of $\mathcal{L}$ by any labeling, i.e.*

$$\mathrm{OPT}(\mathcal{L}) := \frac{1}{|E|}\max_{\ell:V\cup W\to[n]} |\{\, e\in E \,:\, \ell \textit{ satisfies } e \,\}|.$$

A formal statement of the Unique Games Conjecture appears below; the conjecture has been widely employed to prove strong inapproximability results for several problems.

Conjecture 2.8 ([17]) *For any constants $\zeta,\gamma > 0$, there is a sufficiently large constant $n = n(\zeta,\gamma)$ such that, for Unique Games instances $\mathcal{L}$ with label set $[n]$ it is* NP*-hard to distinguish between*

- $\mathrm{OPT}(\mathcal{L}) \geq 1-\zeta$,
- $\mathrm{OPT}(\mathcal{L}) \leq \gamma$.

3 Long Code Tests

A long code test, for our purpose, is a probabilistic test on a given boolean function f such that: (Completeness) If f is a dictatorship, then it passes with high probability. (Soundness) If f has no influential variable,[3] the test passes with low probability.

In the next four subsections, we present four long code tests denoted $T_\varepsilon, F_\varepsilon, T_{\varepsilon,t}, F_{\varepsilon,t}$ respectively, where ε and t are parameters. The test T_ε is the most basic test that we describe first. It is a zero free bit test with completeness close to $\frac{1}{2}$ and its soundness property works only when the given boolean function equals 1 on a constant fraction of inputs. The test is useful towards proving inapproximability of the scheduling problem. The test F_ε is a minor variant of the test T_ε. This is an optimal long code test with one free bit and useful towards proving inapproximability of vertex cover. The analysis of test F_ε follows easily from the analysis of test T_ε. For the actual PCP or inapproximability applications, we actually need to test that given a collection of boolean functions, all of them are indeed long codes and are *identical*. We modify the tests T_ε and F_ε towards this purpose and denote the tests by $T_{\varepsilon,t}$ and $F_{\varepsilon,t}$ respectively. Analysis of the test $T_{\varepsilon,t}$ is along the lines of that of the test T_ε, but more involved. Analysis of the test $F_{\varepsilon,t}$ follows easily from that of the test $T_{\varepsilon,t}$.

3.1 The Long Code Test T_ε

In this section, we propose and analyze the following test:

Test T_ε **Given** $f : \{0,1\}^n \mapsto \{0,1\}$ **and** $\varepsilon > 0$**.**

- Pick $x \in \{0,1\}^n$ at random.
- Let $m = \varepsilon n$. Pick indices $i_1, \ldots, i_m$ randomly and independently from $\{1, \ldots, n\}$ and let $S = (i_1, \ldots, i_m)$ be the sequence of these indices.[4]
- Define the sub-cube:
$$C_{x,S} := \{z \mid z_j = x_j \ \forall j \notin S\}.$$
- Accept if and only if f is identically zero on $C_{x,S}$.

Note that $C_{x,S}$ is a sub-cube determined by fixing the co-ordinates outside S and letting the co-ordinates in S take any setting of bits. The number of points in the sub-cube is $2^{m'}$ where m' is the number of distinct co-ordinates in the sequence S.

3.1.1 Completeness

Suppose f is the dictatorship of co-ordinate i. Then the test fails if either $i \in S$ or if $i \notin S$ and $f(x) = 1$. Since $|S| = \varepsilon n$, the dictatorship passes the test with probability at least $\frac{1}{2} - \varepsilon$.

[3] We in fact require the soundness property to hold even when f does not have a *low-degree* influential variable.

[4] We could instead pick S to be a random subset of size εn, but thinking of S as a sequence of independently picked indices makes the analysis easier.

3.1.2 Soundness

The soundness of the test is given by:

Theorem 3.1 *For every $\varepsilon, \delta > 0$, there exists $\eta > 0$ and integer k such that any $f : \{0,1\}^n \mapsto \{0,1\}$ that satisfies*

$$\mathbb{E}[f] \geq \delta \quad \textit{and} \quad \forall i \in [n], \mathsf{Infl}_i^k(f) \leq \eta,$$

passes the test T_ε with probability at most 2δ.

Intuitively, the test works for the following reason: suppose that f is a balanced function. Since all its influences are small, the function cannot be (even close to) a junta, and by Friedgut's theorem, the sum of influences must be large. This means that a relatively large fraction (β/n for a suitably large constant β) of the hypercube edges are non-monochromatic, i.e. value of f on its end-points is different. Hence a random sub-cube is likely to contain such an edge and the test rejects. In our argument the sub-cube has dimension εn and we need $\beta \gg 1/\varepsilon$ and the individual influences to be small enough to ensure that the sum of influences is at least β/n. We proceed with a formal proof.

Proof: Observe first that if $\mathbb{E}[f] \geq 1 - \delta$, we are done. This is because x chosen in the test T_ε satisfies $f(x) = 1$ with probability at least $1 - \delta$ and since $x \in C_{x,S}$, the test rejects for every choice of S. So henceforth we assume that $\delta \leq \mathbb{E}[f] \leq 1 - \delta$.

It is more convenient to view the test as first choosing the sequence $S = (i_1, \ldots, i_m)$ by picking m co-ordinates one at a time and then choosing x, specifying the sub-cube. As we pick indices in S, we will construct a sequence of functions $f = f_0 \leq f_1 \leq \ldots \leq f_m$ (these functions depend on S) such that $\forall j \in \{1, \ldots, m\}$:

1. f_j does not depend on co-ordinates $i_1, i_2, \ldots, i_j$.
2. For a fixed S and an arbitrary input x, the test T_ε accepts f_j if and only if it accepts $f = f_0$.

From the second property, we know that

$$\Pr_x \left[\text{Test } T_\varepsilon \text{ accepts } f \middle| S \right] = \Pr_x \left[\text{Test } T_\varepsilon \text{ accepts } f_m \middle| S \right].$$

We will show that with probability at least $1 - \delta$ over the choice of S, we have $\mathbb{E}[f_m] \geq 1 - \delta$. When this happens, we saw that the test T_ε accepts f_m (and hence f) with probability at most δ. Thus the test overall accepts f with probability at most $\delta + \delta = 2\delta$, completing the proof.

Now we define the functions f_j and prove that they satisfy properties (1) and (2) above. We let:

$$f_{j+1}(x) := \left\{ \begin{array}{lll} f_j(x) & \text{if} & f_j(x) = f_j(x \oplus e_{i_{j+1}}) \\ 1 & \text{otherwise} & \end{array} \right\} \tag{3}$$

Thus f_{j+1} is obtained by *symmetrizing* f_j over the co-ordinate i_{j+1} in a specific way. It is not difficult to see that an equivalent way to define f_{j+1} is the following: $f_{j+1}(x) = 0$ if and only if f is identically zero on $C_{x,(i_1,\ldots,i_{j+1})}$, the sub-cube defined by x and the prefix of indices $(i_1, \ldots, i_{j+1})$. Clearly, f_{j+1} is independent of co-ordinates $i_1, \ldots, i_{j+1}$. Moreover, for fixed S and an arbitrary x, f_{j+1} is identically zero on the sub-cube $C_{x,S}$ if and only if f is identically zero on $C_{x,S}$. This proves both property (1) and (2).

Now we prove the assertion that with probability at least $1-\delta$ over the choice of S, we have $\mathbb{E}[f_m] \geq 1-\delta$. From equation (3), it is clear that $f_j \leq f_{j+1}$ and

$$\mathbb{E}[f_{j+1}] = \mathbb{E}[f_j] + \mathsf{Infl}_{i_{j+1}}(f_j). \tag{4}$$

Since the sequence of functions $\{f_j\}_{j=1}^m$ is increasing, it suffices to prove that with probability at least $1-\delta$ over the choice of S, there exists $j^* \in \{1,\ldots,m\}$ such that $\mathbb{E}[f_{j^*}] \geq 1-\delta$ (the least index j^* for which this happens could depend on the choice of S).

Let $\gamma := \delta^2/4, k := 2^{16\log(1/\delta)/(\varepsilon\gamma)}, \eta := \delta^2/2^{2k+4}$. Assume on the contrary that for every $j \in \{1,\ldots,m\}, \mathbb{E}[f_j] \leq 1-\delta$. Thus,

$$f \leq f_j \quad \text{and} \quad \delta \leq \mathbb{E}[f] \leq \mathbb{E}[f_j] \leq 1-\delta,$$

and by the hypothesis of the theorem, $\forall i$, $\mathsf{Infl}_i^k(f) \leq \eta = \delta^2/2^{2k+4}$. Applying Lemma 2.3, we see that f_j is not γ-close to a k-junta. This implies, by Friedgut's Theorem, that the average sensitivity (i.e. sum of influences) of f_j is at least $4\log(1/\delta)/\varepsilon$. Thus when index i_{j+1} is chosen,

$$\mathbb{E}[\mathsf{Infl}_{i_{j+1}}(f_j)] \geq \frac{4\log(1/\delta)/\varepsilon}{n} := \beta.$$

Writing $X_{j+1} = \mathsf{Infl}_{i_{j+1}}(f_j)$,

$$\mathbb{E}[f_m] = \mathbb{E}[f] + \sum_{j=1}^{m} \mathsf{Infl}_{i_j}(f_{j-1}) = \mathbb{E}[f] + \sum_{j=1}^{m} X_j. \tag{5}$$

X_{j+1} depends on choice of $i_1,\ldots,i_{j+1}$, and we already noted that $\mathbb{E}[X_{j+1}|i_1,\ldots,i_j] \geq \beta$. Though $X_1,\ldots,X_m$ are not independent, we will prove that with probability at least $1-\delta$, $\sum_{j=1}^m X_j \geq 2$. This would be a contradiction since $\sum_{j=1}^m X_j \leq \mathbb{E}[f_m] \leq 1$.

Towards this end, we use Lemma (2.5) and conclude that $\mathbb{E}[e^{-\sum_{j=1}^m X_j}] \leq e^{-\beta m/2}$. Hence,

$$\Pr\left[X = \sum_{j=1}^m X_j \leq 2\right] = \Pr\left[e^{-X} \geq e^{-2}\right] \leq \frac{\mathbb{E}[e^{-X}]}{e^{-2}} \leq e^2 \cdot e^{-\beta m/2} = e^2 \cdot e^{-2\frac{\log(1/\delta)/\varepsilon}{n}\cdot \varepsilon n} \leq \delta.$$

This completes the proof. ∎

3.2 The One Free Bit Long Code Test F_ε

We now develop a long code test with one free bit, completeness close to 1 and soundness close to 0. The test is a minor variation of the test T_ε.

Test F_ε Given $f : \{0,1\}^n \mapsto \{0,1\}$

- Pick $x \in \{0,1\}^n$ at random. Let $\overline{x}$ denote the input obtained by flipping every bit of x.
- Let $m = \varepsilon n$. Pick indices $i_1,\ldots,i_m$ randomly and independently from $\{1,\ldots,n\}$ and let $S = (i_1,\ldots,i_m)$ be the sequence of these indices.

- Define the sub-cubes:
$$C_{x,S} := \{z \mid z_j = x_j \ \forall j \notin S\}.$$
$$C_{\overline{x},S} := \{z \mid z_j = \overline{x}_j \ \forall j \notin S\}.$$

- Accept if and only if for some bit $b \in \{0,1\}$, f is identically b on $C_{x,S}$ and identically $b \oplus 1$ on $C_{\overline{x},S}$.

Remark 3.2 *Probability that the functions f and $1-f$ pass the test F_ε is the same.*

3.2.1 Completeness

A dictatorship of co-ordinate i passes the test F_ε whenever $i \notin S$, i.e. with probability at least $1-\varepsilon$.

3.2.2 Soundness

Theorem 3.3 *For every $\varepsilon, \delta > 0$, there exists $\eta > 0$ and integer k such that any $f : \{0,1\}^n \mapsto \{0,1\}$ that satisfies*
$$\forall i \in [n], \mathsf{Infl}_i^k(f) \le \eta,$$
passes the test F_ε with probability at most 4δ.

Proof: Let η be as in Theorem 3.1. Assume that $\forall i \in [n], \mathsf{Infl}_i^k(f) \le \eta$. Since f and $1-f$ pass the test with the same probability, we may assume that $\mathbb{E}[f] \ge \frac{1}{2}$. Then

$$\begin{aligned}
\Pr[f \text{ passes } F_\varepsilon] &= \Pr_{x,S}[f|_{C_{x,S}} \equiv 0 \wedge f|_{C_{\overline{x},S}} \equiv 1] + \Pr_{x,S}[f|_{C_{x,S}} \equiv 1 \wedge f|_{C_{\overline{x},S}} \equiv 0] \\
&\le \Pr_{x,S}[f|_{C_{x,S}} \equiv 0] + \Pr_{x,S}[f|_{C_{\overline{x},S}} \equiv 0] \\
&= 2 \cdot \Pr_{x,S}[f|_{C_{x,S}} \equiv 0] = 2 \cdot \Pr[f \text{ passes } T_\varepsilon] \le 4\delta,
\end{aligned}$$

using Theorem 3.1. ∎

3.3 The Test $T_{\varepsilon,t}$ for Multiple Long Codes

For PCP applications, we need to test that given a collection of several long codes, they are (indeed long codes and are) identical. We modify the tests T_ε and F_ε for this purpose.

Test $T_{\varepsilon,t}$ Given $f^{(1)}, \ldots, f^{(t)} : \{0,1\}^n \mapsto \{0,1\}$

- Pick $x \in \{0,1\}^n$ at random.
- Let $m = \varepsilon n$. Pick indices $i_1, \ldots, i_m$ randomly and independently from $\{1, \ldots, n\}$ and let $S = (i_1, \ldots, i_m)$ be the sequence of these indices.
- Define the sub-cube:
$$C_{x,S} := \{z \in \{0,1\}^n \,|z_j = x_j \ \forall\ j \notin S\}.$$
- Accept if and only if every $f^{(\ell)}$, $1 \le \ell \le t$, is identically zero on $C_{x,S}$.

3.3.1 Completeness

Clearly, if for some $i \in [n]$, $f^{(1)}, \ldots, f^{(t)}$ are all dictatorship of the (same) i^{th} co-ordinate, then the test $T_{\varepsilon,t}$ passes with probability $\frac{1}{2} - \varepsilon$.

3.3.2 Soundness

Theorem 3.4 *For every $\varepsilon, \delta > 0$, there exists $\eta > 0$ and integers t, k such that any collection of functions $f^{(1)}, \ldots, f^{(t)} : \{0,1\}^n \mapsto \{0,1\}$ that satisfy:*

$$\forall j, \mathbb{E}[f_j] \geq \delta \quad \textit{and} \quad \forall i \in [n],\ \forall\, 1 \leq \ell_1 \neq \ell_2 \leq t,\ \ \min\{\ \mathsf{Infl}_i^k(f^{(\ell_1)}),\ \mathsf{Infl}_i^k(f^{(\ell_2)})\} \leq \eta,$$

pass the test $T_{\varepsilon,t}$ with probability at most 3δ.

Remark 3.5 *The influence condition says that in the collection of functions $f^{(1)}, \ldots, f^{(t)}$, no two functions share an influential variable (where the notion of influence is the degree-k influence).*

Proof: The proof is along the lines of proof of Theorem 3.1, but is more involved. The trouble is that we can only assume that the t functions do not *share* an influential variable, whereas proof of Theorem 3.1 assumes that the (single) function itself has no influential variable.

We start by noting that the test $T_{\varepsilon,t}$ includes the test T_ε on every $f^{(\ell)}$, $1 \leq \ell \leq t$. We attempt to carry out the proof of Theorem 3.1 for every $f^{(\ell)}$. Thus, we define the increasing sequence of functions $f^{(\ell)} = f_0^{(\ell)}, f_1^{(\ell)}, \ldots, f_m^{(\ell)}$ as in Equation (3). We assume that $\mathbb{E}[f_m^{(\ell)}] \leq 1 - \delta$ because otherwise the test T_ε on $f_m^{(\ell)}$ would reject with probability $1 - \delta$ and we are done.

Define $X_{j+1}^{(\ell)} = \mathsf{Infl}_{i_{j+1}}(f_j^{(\ell)})$ so that as in Equation (5),

$$\mathbb{E}[f_m^{(\ell)}] = \mathbb{E}[f^{(\ell)}] + \sum_{j=0}^{m-1} \mathsf{Infl}_{i_{j+1}}(f_j^{(\ell)}) = \mathbb{E}[f^{(\ell)}] + \sum_{j=0}^{m-1} X_{j+1}^{(\ell)}. \tag{6}$$

If $f^{(\ell)}$ did not have any influential variable, then as in the proof of Theorem 3.1, we could conclude that the sum of influences of $f_j^{(\ell)}$ is large, and thus $\mathbb{E}[X_{j+1}^{(\ell)} | i_1, \ldots, i_j] \geq \beta$ for some appropriate β. However, we can only assume that the t functions do not share an influential variable. So we instead sum-up Equation (6) over all ℓ and look at:

$$\frac{1}{t} \sum_{\ell=1}^{t} \mathbb{E}[f_m^{(\ell)}] = \left(\frac{1}{t} \sum_{\ell=1}^{t} \mathbb{E}[f^{(\ell)}] \right) + \sum_{j=0}^{m-1} \left(\frac{1}{t} \sum_{\ell=1}^{t} X_{j+1}^{(\ell)} \right).$$

Define $X_{j+1} = \frac{1}{t} \sum_{\ell=1}^{t} X_{j+1}^{(\ell)}$. We will prove that $\mathbb{E}[X_{j+1} | i_1, \ldots, i_j] \geq \beta$ where $\beta := \frac{4 \log(1/\delta)/\varepsilon}{n}$. After this, as in the proof of Theorem 3.1, we conclude that $\mathbb{E}[e^{-\sum_{j=0}^{m-1} X_{j+1}}] \leq e^{-\beta m/2}$, and hence $\Pr[\sum_{j=0}^{m-1} X_{j+1} \leq 2] \leq e^{-2} \cdot e^{-\beta m/2} \leq \delta$ by the choice of β. This would be a contradiction since $\sum_{j=0}^{m-1} X_{j+1} \leq 1$.

Now we will prove that $\mathbb{E}[X_{j+1} | i_1, \ldots, i_j] \geq \beta$, i.e.

$$\frac{1}{t} \sum_{\ell=1}^{t} \mathbb{E}_{i_{j+1}}[\mathsf{Infl}_{i_{j+1}}(f_j^{(\ell)})] \ \ \geq \ \ \beta.$$

If for $\frac{t}{2}$ out of t values of ℓ, the sum of influences of $f_j^{(\ell)}$ is at least $2\beta n$, we are done. Therefore we will assume that for at least $\frac{t}{2}$ values of ℓ, say w.l.o.g. $1 \le \ell \le \frac{t}{2}$, the sum of influences of $f_j^{(\ell)}$ is at most $2\beta n = 8\log(1/\delta)/\varepsilon$ and derive a contradiction. Henceforth we only restrict to indices $1 \le \ell \le \frac{t}{2}$.

Let $\gamma := \delta^4$, $k := \exp(24\log(1/\delta)/(\varepsilon\gamma))$, $r := 8\log(1/\delta)/\delta, t := 4k!(k+2r-1)^k$, and $\eta := \delta^2/(2^{2k+4})$. Applying Friedgut's Theorem, every $f_j^{(\ell)}$ for $1 \le \ell \le \frac{t}{2}$, is γ-close to a function $h^{(\ell)}$ that depends only on co-ordinates in a set A_ℓ with $|A_\ell| = k$. Applying Sunflower Lemma 2.4 to the collection of sets $\{A_\ell\}_{\ell=1}^{t/2}$, since $t/2 > k!(k+2r-1)^k$, there is a sunflower with *core* B and $k+2r$ petals. Assume w.l.o.g. that the sets in the sunflower are $A_1, \ldots, A_{k+2r}$.

Suppose on the contrary that for at least $k+1$ values of $\ell \in \{1, \ldots, k+2r\}$, $f^{(\ell)}$ has some variable $j_\ell \in B$ such that $\mathsf{Infl}_{j_\ell}^k(f^{(\ell)}) \ge \eta$. By pigeon-hole principle, there will be some $i \in B$ (recall $|B| \le k$) and indices $1 \le \ell_1 \ne \ell_2 \le k+2r$ for which

$$\mathsf{Infl}_i^k(f^{(\ell_1)}) \ge \eta, \quad \mathsf{Infl}_i^k(f^{(\ell_2)}) \ge \eta$$

contradicting the hypothesis of the theorem.

Therefore assume, w.l.o.g. that for $1 \le \ell \le 2r$, $f^{(\ell)}$ has no variable $i \in B$ such that $\mathsf{Infl}_i^k(f^{(\ell)}) \ge \eta$. A consequence of this, by Lemma 2.2, is that for every setting $s \in \{0,1\}^{|B|}$, we have

$$\Pr\Big[f^{(\ell)} = 0 \Big| x|_B = s\Big] \le 1 - \delta/2.$$

Since $f^{(\ell)} \le f_j^{(\ell)}$, we have

$$\Pr\Big[f_j^{(\ell)} = 0 \Big| x|_B = s\Big] \le 1 - \delta/2.$$

Now we prove that for a randomly chosen $x \in \{0,1\}^n$, with probability at least $1 - 3\delta$, at least one of $\{f_j^{(\ell)}(x)\}_{\ell=1}^{2r}$ equals 1. Whenever this happens, the test $T_{\varepsilon,t}$ rejects on input x. This would complete the proof.

For $1 \le \ell \le 2r$ and $s \in \{0,1\}^{|B|}$, call the pair (ℓ, s) bad if $\Pr[f_j^{(\ell)} \ne h^{(\ell)} | x|_B = s] \ge \gamma/\delta$. Since $f_j^{(\ell)}$ is γ-close to $h^{(\ell)}$, for every ℓ, the fraction of s such that the pair (ℓ, s) is bad is at most δ. This implies that for $1 - 2\delta$ fraction of s (call such s special), the number of $\ell \in \{1, \ldots, 2r\}$ for which (ℓ, s) is bad is at most r. Fix any such special s. Assume w.l.o.g. that for indices $\ell = 1, \ldots, r$ the pair (ℓ, s) is not bad. This means that for this fixed special s and $1 \le \ell \le r$,

$$\mathbb{E}[h^{(\ell)} | x|_B = s] \ge \mathbb{E}[f_j^{(\ell)} | x|_B = s] - \Pr[f_j^{(\ell)} \ne h^{(\ell)} | x|_B = s] \ge \delta/2 - \gamma/\delta \ge \delta/4.$$

It follows that for any fixed special s,

$$\begin{aligned}
\Pr[\vee_{\ell=1}^r f_j^{(\ell)} = 1 | x|_B = s] &\ge \Pr[\vee_{\ell=1}^r h^{(\ell)} = 1 | x|_B = s] - \sum_{\ell=1}^r \Pr[f_j^{(\ell)} \ne h^{(\ell)} | x|_B = s] \\
&= 1 - \prod_{\ell=1}^r (1 - \mathbb{E}[h^{(\ell)} | x|_B = s]) - \sum_{\ell=1}^r \Pr[f_j^{(\ell)} \ne h^{(\ell)} | x|_B = s] \\
&\ge 1 - (1 - \delta/4)^r - r\gamma/\delta \\
&\ge 1 - \delta.
\end{aligned}$$

The second step follows as the functions $\{h^\ell \mid x|_B = s\}$ do not share any variable (as the petals of a sunflower are disjoint). Since $1 - 2\delta$ fraction of s are special, it follows that $\Pr[\vee_{\ell=1}^r f_j^{(\ell)} = 1]$ is at least $(1 - 2\delta)(1 - \delta) \ge 1 - 3\delta$. ∎

3.4 The One Free Bit Test $F_{\varepsilon,t}$ for Multiple Long Codes

Now we describe the test that we will use to prove hardness of vertex cover.

Test $F_{\varepsilon,t}$ **Given** $f^{(1)},\ldots,f^{(t)} : \{0,1\}^n \mapsto \{0,1\}$

- Pick $x \in \{0,1\}^n$ at random. Let $\overline{x}$ denote the input obtained by flipping every bit of x.
- Let $m = \varepsilon n$. Pick indices $i_1,\ldots,i_m$ randomly and independently from $\{1,\ldots,n\}$ and let $S = (i_1,\ldots,i_m)$ be the sequence of these indices.
- Define the sub-cubes:
$$C_{x,S} := \{z \in \{0,1\}^n \mid z_j = x_j \;\forall\, j \notin S\}.$$
$$C_{\overline{x},S} := \{z \in \{0,1\}^n \mid z_j = \overline{x}_j \;\forall\, j \notin S\}.$$
- Accept if and only if for some bit $b \in \{0,1\}$, every $f^{(\ell)}, 1 \le \ell \le t$ is identically b on $C_{x,S}$ and identically $b \oplus 1$ on $C_{\overline{x},S}$.

3.4.1 Completeness

Clearly, if for some $i \in [n]$, $f^{(1)},\ldots,f^{(t)}$ are all dictatorship of the (same) i^{th} co-ordinate, then the test $F_{\varepsilon,t}$ passes with probability $1-\varepsilon$.

3.4.2 Soundness

Theorem 3.6 *For every $\varepsilon,\delta > 0$, there exists $\eta > 0$ and integers t,k such that any collection of functions $f^{(1)},\ldots,f^{(t)} : \{0,1\}^n \mapsto \{0,1\}$ that satisfy:*

$$\forall i \in [n],\; \forall\, 1 \le \ell_1 \ne \ell_2 \le t,\;\; \min\{\, \mathsf{Infl}_i^k(f^{(\ell_1)}),\; \mathsf{Infl}_i^k(f^{(\ell_2)})\} \le \eta,$$

passes the test $F_{\varepsilon,t}$ with probability at most 6δ.

Proof: For given ε,δ, let η,t',k be as given in Theorem 3.4 and let $t := 2t'$. Note that the collection of functions $\{1-f^{(\ell)}\}_{\ell=1}^t$ passes the test $F_{\varepsilon,t}$ with the same probability as the collection $\{f^{(\ell)}\}_{\ell=1}^t$. Therefore, complementing all functions if necessary, we can assume that for at least $t/2 = t'$ values of ℓ, say $\ell = 1,\ldots,t'$, $\mathbb{E}[f^{(\ell)}] \ge \frac{1}{2}$. Then,

$$\begin{aligned}
\Pr\left[\{f^{(\ell)}\}_{\ell=1}^t \text{ passes } F_{\varepsilon,t}\right] &\le \Pr\left[\{f^{(\ell)}\}_{\ell=1}^{t'} \text{ passes } F_{\varepsilon,t'}\right] \\
&= \Pr_{x,S}\left[\forall \ell \in [t'], f^{(\ell)}|_{C_{x,S}} \equiv 0 \wedge f^{(\ell)}|_{C_{\overline{x},S}} \equiv 1\right] \\
&\quad + \Pr_{x,S}\left[\forall \ell \in [t'], f^{(\ell)}|_{C_{x,S}} \equiv 1 \wedge f^{(\ell)}|_{C_{\overline{x},S}} \equiv 0\right] \\
&\le \Pr_{x,S}\left[\forall \ell \in [t'], f^{(\ell)}|_{C_{x,S}} \equiv 0\right] + \Pr_{x,S}\left[\forall \ell \in [t'], f^{(\ell)}|_{C_{\overline{x},S}} \equiv 0\right] \\
&= 2\cdot \Pr_{x,S}\left[\forall \ell \in [t'], f^{(\ell)}|_{C_{x,S}} \equiv 0\right] \\
&= 2\cdot \Pr\left[\{f^{(\ell)}\}_{\ell=1}^{t'} \text{ passes } T_{\varepsilon,t'}\right] \le 6\delta
\end{aligned}$$

using Theorem 3.4. ∎

4 Inapproximability of Vertex Cover

We now present, assuming the Unique Games Conjecture, a PCP with one free bit, completeness $1-2\varepsilon$ and soundness 12δ where $\varepsilon, \delta > 0$ are arbitrarily small constants. This proves Theorem 1.3.

Definition 4.1 *For $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ and $\pi : [n] \mapsto [n]$,*

$$\pi(x) := (x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

One Free Bit PCP Given Unique Game Instance $\mathcal{L}(G(V, W, E), [n], \{\pi_{v,w}\}_{(v,w)\in E})$.

- Pick a random vertex $v \in V$.
- Pick vertices $w_1, \ldots, w_t$ randomly and independently from $N(v) := \{w \in W \mid (v, w) \in E\}$. Let $\pi_\ell := \pi_{v,w_\ell}$.
- Let $f^{(\ell)}$ be the long code for w_ℓ for $1 \le \ell \le t$. Define functions $f^{(\ell)} \circ \pi_\ell$ as:
$$f^{(\ell)} \circ \pi_\ell(x) = f^{(\ell)}(\pi_\ell(x)).$$
- Run the test $F_{\varepsilon,t}$ on the collection $f^{(\ell)} \circ \pi_\ell,\ 1 \le \ell \le t$.

4.1 Completeness

Assume that the Unique Games instance has a labeling $\rho : V \cup W \mapsto [n]$ that satisfies $1-\zeta$ fraction of its edges. For every $w \in W$, let the long code corresponding to w be the long code (i.e. dictatorship) of $\rho(w)$.

When $(v, w_1, \ldots, w_t)$ are picked by the PCP verifier, with probability at least $1-\zeta t$, all the edges (v, w_ℓ) are satisfied, i.e. $\forall \ell \in \{1, \ldots, t\}, \pi_\ell(\rho(w_\ell)) = \rho(v)$. Assume that this holds. We will show that the test $F_{\varepsilon,t}$ then accepts with probability at least $1-\varepsilon$ (and hence the overall test accepts with probability $1-\zeta t-\varepsilon \ge 1-2\varepsilon$ for ζ small enough). From the completeness property of test $F_{\varepsilon,t}$ (Section 3.4.1), it suffices to show that $f^{(\ell)} \circ \pi_\ell$, $1 \le \ell \le t$ are dictatorships of the same co-ordinate. We will show that each of them is a dictatorship of the co-ordinate $\rho(v)$. Indeed, for any input x,

$$f^{(\ell)} \circ \pi_\ell(x) = f^{(\ell)}(x_{\pi_\ell(1)}, \ldots, x_{\pi_\ell(n)}) = x_{\pi_\ell(\rho(w_\ell))} = x_{\rho(v)},$$

since $\forall \ell$, $f^{(\ell)}$ is a dictatorship of $\rho(w_\ell)$ and $\pi_\ell(\rho(w_\ell)) = \rho(v)$.

4.2 Soundness

Suppose on the contrary that the PCP accepts with probability 12δ. Let k, η, t be as in Theorem 3.6. We will derive a labeling $\rho : V \cup W \mapsto [n]$ to the Unique Games instance that satisfies at least $6\delta\eta^2/(k^2t^2)$ fraction of its edges. This would be a contradiction if the soundness of the Unique Games instance was chosen to be small enough.

First define a set $L[w]$ of *candidate labels* for every $w \in W$ as:

$$L[w] := \{i \in [n] \mid \mathsf{Infl}_i^k(f^w) \ge \eta\}.$$

Since the number of co-ordinates with degree k-influence at least η is bounded by k/η, we have $|L[w]| \le k/\eta$. Now, for every $w \in W$, define $\rho(w)$ to be a random label from $L[w]$ (define $\rho(w)$ arbitrarily if $L[w]$ is empty). For every $v \in V$, pick its random neighbor $w \in N(v)$ and define $\rho(v) = \pi_{v,w}(\rho(w))$.

Since the acceptance probability of the verifier is 12δ, with probability 6δ after picking the tuple $(v, w_1, \ldots, w_t)$, the verifier accepts with probability at least 6δ. Call such a tuple good. From Theorem 3.6, it must be the case that there exist $1 \le \ell_1 \ne \ell_2 \le t$ such that $\exists i \in [n]$ and $\mathsf{Infl}_i^k(f^{(\ell_1)} \circ \pi_{\ell_1}) \ge \eta$ and $\mathsf{Infl}_i^k(f^{(\ell_2)} \circ \pi_{\ell_2}) \ge \eta$. This is same as saying that there exist $j \in L[w_{\ell_1}], j' \in L[w_{\ell_2}]$ such that $\pi_{\ell_1}(j) = \pi_{\ell_2}(j')$. Overall if we pick the tuple $(v, w_1, \ldots, w_t)$ at random and then w, w' at random from the set $\{w_1, \ldots, w_t\}$, then with probability 6δ the tuple $(v, w_1, \ldots, w_t)$ is good, with probability $1/t^2$ we have $w = w_{\ell_1}, w' = w_{\ell_2}$, and with probability $1/(k^2/\eta^2)$, the labeling procedure defines $j = \rho(w), j' = \rho(w')$. Thus,

$$\Pr_{v,w,w'}[\pi_{v,w}(\rho(w)) = \pi_{v,w'}(\rho(w'))] \ge 6\delta \cdot \frac{1}{t^2} \frac{1}{k^2/\eta^2}.$$

Since the labeling procedure defines the label of v by picking a random neighbor $w' \in N(v)$ and then letting $\rho(v) = \pi_{v,w'}(\rho(w'))$, we have, expected over the randomness of the labeling procedure,

$$\Pr_{(v,w)\in E}[\rho(v) = \pi_{v,w}(\rho(w))] \ge 6 \frac{\delta\eta^2}{k^2 t^2}.$$

This shows that there exists a labeling to the Unique Games instance that satisfies $6\delta\eta^2/(k^2t^2)$ fraction of its edges, completing the proof.

5 Inapproximability of Scheduling Problem

We first give some intuition for the basic gadget that we use to prove inapproximability result for the scheduling problem. Let $C_1, \ldots, C_m$ be a collection of sets on a universe $U = \{1, \ldots, n\}$ of n elements. We associate a natural scheduling instance with this set system as follows. For each element j in U, there is a job j of size 1 and weight 0. Similarly, for each set C_i there is a job C_i of size 0 and weight 1. The precedence constraints require that job C_i can only be scheduled after all the elements $j \in C_i$ have been scheduled[5]. As only the jobs in U have size 1, minimizing the weighted completion time for this instance is equivalent to (up to a multiplicative factor) finding a permutation σ of $\{1, \ldots, n\}$, that minimizes

$$\frac{1}{|U|}\mathbb{E}_{i\in\{1,\ldots,m\}}\left[\max_{j\in C_i} \sigma(j)\right].$$

Suppose we choose U to be the 2^n elements of the hypercube $\{0,1\}^n$ and the sets C to be all possible sub-cubes of dimension εn. Let σ be an ordering of U. We show that: (Completeness:) A dictatorship function on U corresponds to a good ordering σ, and (Soundness:) Any good ordering corresponds to a boolean function with a high influence co-ordinate.

For the completeness part, observe that a dictatorship function induces a partition $U = U' \cup U''$ with $|U'| = |U''| = 2^{n-1}$. Consider an ordering that first orders elements of U' arbitrarily, followed by elements of U'' ordered arbitrarily. A random sub-cube C of dimension εn lies entirely inside U' with probability at least $\frac{1-\varepsilon}{2}$. Therefore the value of this ordering is

$$\frac{1}{|U|}\mathbb{E}_C\left[\max_{j\in C} \sigma(j)\right] \le \frac{1}{2^n}\left(\frac{1-\varepsilon}{2} \cdot |U'| + \frac{1+\varepsilon}{2} \cdot |U|\right) = \frac{3}{4} + \frac{\varepsilon}{4}.$$

For the soundness part, suppose σ is an ordering such that $\frac{1}{|U|}\mathbb{E}_C\left[\max_{j\in C}\sigma(j)\right] \le 1 - 2\delta$. Let $U' \subseteq U$ be the set of jobs that are the last $\delta|U|$ of the jobs in the ordering and let f_σ be the indicator function of f_σ.

[5]Such instances were originally considered by Woeginger [31], and he showed that they are as hard as the general problem.

The fact that the value of the ordering is at most $1-2\delta$ implies that for at least δ fraction of the sub-cubes C, the entire sub-cube is contained inside $U \setminus U'$. In other words, for at least δ fraction of sub-cubes C, the function f_σ is identically zero on S. This is same as saying that the function f_σ is accepted with probability at least δ by the test T_ε in Section 3.1. Applying Theorem 3.1, f_σ must have an influential variable.

Thus we achieve a gap of $\frac{3}{4}+o(1)$ versus $1-o(1)$ between the completeness and the soundness case. We push this to the optimal gap of $\frac{1}{2}+o(1)$ versus $1-o(1)$ by repeating the construction over the Q-ary hypercube $[Q]^n$ and using functions $f : [Q]^n \mapsto Q$. In fact, we let $Q = 2^q$, identify $[Q]^n$ with $(\{0,1\}^q)^n = \{0,1\}^{qn}$ and reduce the analysis essentially to the binary case. We formally describe the gadget next.

5.1 Gadget for the Scheduling Problem

Let $Q = 2^q$ be an integer. The gadget consists of a collection of jobs and precedence constraints among them. There are two sets of jobs J and C.

- The jobs in J are called element jobs and correspond to vertices in the Q-ary hypercube $[Q]^n$. Each of these jobs has size 1 and weight 0. We will identify the set of integers $0, 1, \ldots, Q-1$ with $\{0,1\}^q$ by associating an integer with its binary representation. Thus $[Q]^n$ can be viewed as the binary hypercube $\{0,1\}^{qn} = (\{0,1\}^q)^n$.
- The jobs in C are called sub-cube jobs and are defined as follows: For $x \in \{0,1\}^{qn}$ and every sequence of indices $S = (i_1, \ldots, i_m) \in [qn]^m$, where $m = \varepsilon qn$, we have a job $C_{x,S}$ corresponding to the sub-cube $C_{x,S} := \{z \in \{0,1\}^{qn} \mid \forall j \notin S,\ z_j = x_j\ \}$. Each of these jobs has weight 1 and size 0.

  $C_{x,S}$ cannot be scheduled before any job in the sub-cube corresponding to it, i.e. there is a precedence constraint $z < C_{x,S}$ for every $z \in C_{x,S}$. These are the only precedence constraints.

As discussed above, minimizing the weighted completion time for this instance is equivalent to ordering the jobs in J i.e, find a one-to-one map $\sigma : \{0,1\}^{qn} \mapsto \{1, \ldots, 2^{qn}\}$, so as to minimize

$$\frac{1}{2^{qn}} \mathbb{E}_{x,S}\left[\max_{z \in C_{x,S}} \sigma(z)\right].$$

5.1.1 Completeness

We show that for every index $s \in [n]$, there is an ordering of jobs for which the objective function is at most $\frac{1}{2} + \varepsilon q$. This ordering corresponds to a *dictatorship function* of index s.

Fix the index $s \in [n]$. Denote an input $x \in \{0,1\}^{qn}$ as $x = (x_1^*, \ldots, x_n^*)$ where $x_j^* \in \{0,1\}^q$ for $1 \le j \le n$. Consider the map $\sigma^* : \{0,1\}^{qn} \mapsto \{0,1\}^q$:

$$\forall x \in \{0,1\}^{qn}, \quad \sigma^*(x_1^*, \ldots, x_n^*) = x_s^*.$$

σ^* gives a partition of the set of jobs J into 2^q equal sized subsets. Identify $\{0,1\}^q$ with the set of integers $\{0, 1, \ldots, Q-1\}$ by associating an integer with its binary representation. Denote the subsets in the partition given by σ^* as $J_0, \ldots, J_{Q-1}$. Thus,

$$\forall \ell \in \{0, 1, \ldots, Q-1\} = \{0,1\}^q, \quad J_\ell := \{x \in \{0,1\}^{qn} \mid x_s^* = \ell\}.$$

Now, let σ be the following ordering of the set of jobs. The sets $J_0, J_1, \ldots, J_{Q-1}$ appear in the ordering in that sequence, and within every set J_ℓ, the jobs are ordered arbitrarily.

We show that the ordering σ achieves objective value of at most $\frac{1}{2} + \varepsilon q$. Note that when the set of indices $S = (i_1, \ldots, i_m)$ is chosen at random, with probability $1 - \varepsilon q$, none of the indices is in the range $[(s-1)q + 1, \ldots, sq]$. Call this a good event. In the good event, the entire sub-cube $C_{x,S}$ is inside the set $J_{x_s^*}$. Therefore $\max_{z \in C_{x,S}} \sigma(z)$ is at most the index of the job in $J_{x_s^*}$ that appears last in the ordering, which is $(2^{qn}/2^q) \cdot (x_s^* + 1)$. Also, for a random x, x_s^* is equally likely to take the values in $\{0, 1, \ldots, Q-1\}$. On the other hand, when the good event fails to happen, $\max_{z \in C_{x,S}} \sigma(z)$ is at most the index of the last job in ordering σ, i.e. 2^{qn}. Therefore,

$$\mathbb{E}_{x,S}\left[\max_{z \in C_{x,S}} \sigma(z)\right] \leq (1 - \varepsilon q) \cdot \left(\frac{1}{Q} \sum_{\ell=0}^{Q-1} 2^{(qn-q)} \cdot (\ell + 1)\right) + \varepsilon q \cdot 2^{qn} \leq 2^{qn} \cdot (\frac{1}{2} + \varepsilon q)$$

provided $\varepsilon q \geq 1/Q$.

5.1.2 Soundness

Now we show that any ordering σ of $\{0,1\}^{qn}$ that achieves an objective value at most $1 - 3\delta$, must correspond to an influential co-ordinate. This follows as an easy application of Theorem 3.1.

Let $A \subseteq \{0,1\}^{qn}$ be the last δ fraction of the jobs in the ordering σ. Let f_A be the indicator function of A. Let η, k be as in Theorem 3.1. We will show that if $\forall i \in [qn], \inf_i^k(f_A) \leq \eta$, then the objective value is at least $1 - 3\delta$.

A sub-cube $C_{x,S}$ does not intersect A if and only if f_A is identically zero on $C_{x,S}$, i.e. if and only if the test T_ε accepts on (x, S). By Theorem 3.1 this happens with probability at most 2δ. Thus, at least $1 - 2\delta$ fraction of the sub-cubes $C_{x,S}$ contain a job from A. For all such sub-cubes $\max_{z \in C_{x,S}} \sigma(z) \geq (1 - \delta) \cdot 2^{qn}$. It follows that the objective value is at least $(1 - \delta) \cdot (1 - 2\delta) \geq 1 - 3\delta$.

5.2 UGC-based Hardness

We now describe how to use the ideas above to obtain a $2 - \varepsilon$ hardness result for the scheduling problem. To do this, we need a stronger variant of the Unique Games Conjecture that we describe next.

5.2.1 UGC variant

Hypothesis 5.1 *For arbitrarily small constants $\zeta, \gamma, \delta > 0$, there exists an integer $n = n(\zeta, \gamma, \delta)$ such that for a Unique Games instance $\mathcal{L}(G(V, W, E), [n], \{\pi_{v,w}\}_{(v,w) \in E})$, it is* NP*-hard to distinguish between:*

- *(YES Case:) There are sets $V' \subseteq V$, $W' \subseteq W$ such that $|V'| \geq (1 - \zeta)|V|$ and $|W'| \geq (1 - \zeta)|W|$ and an assignment to $\mathcal{L}$ such that all the edges between the sets (V', W') are satisfied.*
- *(NO Case:) No assignment to $\mathcal{L}$ satisfies even a γ fraction of edges. Moreover, the instance satisfies the following* expansion *property. For every set $S \subseteq V$, $|S| = \delta|V|$, we have $|\Gamma(S)| \geq (1 - \delta)|W|$, where $\Gamma(S) := \{w \in W \mid \exists v \in S,\ (v, w) \in E\}$.*

This UGC variant differs from the standard form in two ways. In the NO case, we require that the instance satisfy certain (arguably weak) expansion property, namely that any two sets of relative size δ contain an edge between them. This expansion requirement is not ruled out by the known algorithmic results for unique games on expanders [5, 30]. We would like to remark that it is always possible to turn the Unique Games instance into a strong expander by super-imposing a dummy expander on top, only slightly reducing its completeness.

In the YES case, we require not only that there is an almost satisfying assignment, but a stronger condition that there are large subsets $V' \subseteq V$ and $W' \subseteq W$ such that *all* the edges between (V', W') are satisfied. In fact, the same property is needed in Khot and Regev's paper [20] and they show that any Unique Games instance can be transformed into one with this property preserving the low soundness.

Thus each of the two extra properties in Hypothesis 5.1 (one in the YES case and the other in the NO case) can be achieved on its own by a suitable transformation of the Unique Games instance in the standard form. However we do not know how to achieve both properties simultaneously. In other words, we do not know whether Hypothesis 5.1 is equivalent to (the standard form of) the UGC.

5.2.2 Reduction

We now describe a reduction to the scheduling problem. Let $\mathcal{L}(G(V, W, E), [n], \{\pi_{v,w}\}_{(v,w)\in E})$ be a unique games instance as in Hypothesis 5.1. In the reduction, we will replace each vertex in V and W by a Q-ary hypercube $[Q]^n$. As in the gadget above, we will identify $[Q]^n$ with $(\{0,1\}^q)^n$. Given a permutation $\pi : [n] \to [n]$, and $x \in [Q]^n$ where $x = (x_1^*, x_2^*, \ldots, x_n^*)$ where $x_i^* \in [Q] = \{0,1\}^q$, we will denote $\pi(x) = (x^*_{\pi(1)}, \ldots, x^*_{\pi(n)})$. Consider the following instance of the scheduling problem.

- We replace each vertex $v \in V$ by a Q-ary hypercube $[Q]^n$. Let H_v denote the hypercube corresponding to v, and we denote the job in H_v corresponding to $x \in [Q]^n$ by (v, x). Let J denote the set of all such jobs, i.e. $J := \cup_{v\in V} H_v$. Each job in J has size 1 and weight 0.

- We replace each vertex $w \in W$ by a Q-ary hypercube $[Q]^n$. Let H_w denote the hypercube corresponding to w, and let (w, x) denote the job corresponding to $x \in H_w$. Let J' denote the set of all such jobs, i.e. $J' := \cup_{w\in W} H_w$. Each such job has size 0 and weight 0.

- For each edge $(v, w) \in E$ in the Unique Games instance, we add a precedence constraint between jobs in H_v and the "corresponding" job in H_w where the correspondence is via the permutation $\pi_{v,w}$, i.e. for every x, we add the constraint $(v, x) < (w, \pi_{v,w}(x))$.

- Finally, we define jobs that correspond to subsets of J' as follows. For $(v, w) \in E, x \in \{0,1\}^{qn}$ and indices $S = (i_1, \ldots, i_m) \in [qn]^m$, where $m = \varepsilon qn$, let $C_{v,x,S,w}$ denote the sub-cube

  $$C_{v,x,S,w} := \{(w, z) | z \in \{0,1\}^{qn} \text{ and } \forall j \notin S \;\; z_{\pi_{v,w}(j)} = x_j\}.$$

  Note that $C_{v,x,S,w}$ is the image of the sub-cube $C_{v,x,S}$ via $\pi_{v,w}$.

  For each vertex $v \in V$ and t neighbors $w_1, \ldots, w_t \in N(v)$, we define a job $C_{v,x,S,w_1,\ldots,w_t}$ corresponding to the subset $\cup_{i=1}^t C_{v,x,S,w_i}$. Let C denote the set of all such jobs. Each job in C has size 0 and weight 1. A job j in C can only be scheduled after all the jobs in the subset corresponding it are scheduled. That is, there is a precedence constraint $j < C_{v,x,S,w_1,\ldots,w_t}$ for every $j \in \cup_{i=1}^t C_{v,x,S,w_i}$.

Note that only the jobs in J have a non-zero size and only the jobs $C_{v,x,S,w_1,\ldots,w_t} \in C$ have non-zero weight. Let σ be an ordering of $2^{qn}|V|$ jobs in J, i.e. $\sigma : J \mapsto \{1, \ldots, 2^{qn}|V|\}$. For any $j \in J'$, by abuse of notation, let $\sigma(j)$ denote the earliest time at which j can be scheduled (i.e. immediately after each of its predecessors in J are complete). Thus the problem is to find an ordering $\sigma : J \mapsto \{1, \ldots, 2^{qn}|V|\}$ that minimizes

$$\frac{1}{2^{qn}|V|}\mathbb{E}_{v,x,S,w_1,\ldots,w_t}\left[\max_{j\in C_{v,x,S,w_1,\ldots,w_t}}\sigma(j)\right].$$

5.2.3 Completeness

Consider a labeling ρ of the Unique Games instance in the YES case of hypothesis 5.1. In this case we will show that there is a solution to the scheduling problem that satisfies all the precedence constraints and has value arbitrarily close to $\frac{1}{2}$. We will describe an ordering of jobs in J. This will automatically determine the ordering of jobs in J' and C since these jobs have size 0 and can be placed at the earliest position when all their predecessors have been scheduled.

Let $V' \subseteq V$ and $W' \subseteq W$ be such that $|V'| \geq (1-\zeta)|V|, |W'| \geq (1-\zeta)|W|$ and all the edges between V' and W' are satisfied. We will define a partition of jobs in J. Consider the jobs H_v for $v \in V'$. Each such job has the form (v,x) for $x = (x_1^*, \ldots, x_n^*)$. We assign each job (v,x) the label $x^*_{\rho(v)}$. This partitions H_v into Q equal sized sets $H_{v,0}, \ldots, H_{v,Q-1}$. We partition J as $P \cup J_0 \cup \ldots \cup J_{Q-1}$, where $P := \{(v,x)|x \in [Q]^n, v \in V \setminus V'\}$ is set of all jobs corresponding to $v \in V \setminus V'$ and $J_i := \cup_{v\in V'} H_{v,i}$ for $i = 0, \ldots, Q-1$.

Consider the ordering where we first schedule the jobs in P followed by the sets $J_0, J_1, \ldots, J_{Q-1}$ in that order. The jobs can be ordered arbitrarily within a set J_i or within P. As there are $\zeta 2^{qn}|V|$ jobs in P, and all sets J_i have equal size, any job in J_i appears no later than

$$\zeta|V|2^{qn} + (1-\zeta)|V|2^{qn}\frac{i+1}{Q} \leq 2^{qn}|V|(\zeta + \frac{i+1}{Q}).$$

As mentioned above this completely determines the schedule for all other jobs in the instance. First we place each job in J' at the earliest position by which all its predecessors in J have been scheduled. Next, we place each job in C at the earliest position by which all its predecessors in J' have been scheduled. By construction, the ordering satisfies all the precedence constraints.

Now consider a vertex $w \in W'$. For each job $(w,z) \in H_w$, where $z = (z_1^*, \ldots, z_n^*)$, we assign this job the label $z^*_{\rho(w)}$. This partitions H_w into Q equal sized sets $H_{w,0}, \ldots, H_{w,Q-1}$. Consider an edge (v,w) in unique games instance that is satisfied by the labeling. An important observation is the following: for any job $(w,z) \in H_w$, let (v,x) denote its "corresponding" job in H_v, i.e. $z = \pi_{v,w}(x)$ and we defined the predecessor constraint $(v,x) < (w,z)$. We claim that $(v,x) \in H_{v,i}$ and $(w,z) \in H_{w,i}$ for the same index i. This follows because, by definition, $(w,z) \in H_{w,z^*_{\rho(w)}}$ and $(v,x) \in H_{v,x^*_{\rho(v)}}$. As the labeling satisfies the edge (v,w) and $\pi_{v,w}(x) = z$, we have

$$z^*_{\rho(w)} = (\pi_{v,w}(x))^*_{\rho(w)} = x^*_{\pi_{v,w}(\rho(w))} = x^*_{\rho(v)}$$

as desired. We claim that for any job (w,z), all its predecessors have been placed by position $2^{qn}|V|(\zeta + \frac{z^*_{\rho(w)}+1}{Q})$. Indeed, any predecessor of (w,z) is of the form (v,x) for some $(v,w) \in E$ and $\pi_{v,w}(x) = z$. If $v \in V \setminus V'$, then $(v,x) \in P$ and hence is scheduled no later than $\zeta 2^{qn}|V|$ and the condition is trivially satisfied. On the other hand if $v \in V'$, then we are guaranteed that labeling satisfies the edge (v,w) and hence by the above observation this predecessor belongs to the set $J_{x^*_{\rho(v)}} = J_{z^*_{\rho(w)}}$ which implies the claim.

Finally we consider the jobs in C. Recall that jobs in C correspond to subsets $C_{v,x,S,w_1,\ldots,w_t}$. Consider the event where we choose $v \in V$, $S \in [qn]^m$ and $w_1, \ldots, w_t \in N(v)$ at random. Call this event good if $v \in V'$, each $w_1, \ldots, w_t$ in W' and none of the indices in S lies in $\{(\rho(v)-1)q+1, \ldots, \rho(v)q\}$. The probability that an event is good is at least $1 - \zeta - t\zeta - \varepsilon q$.

In the good event, for any choice of x, the entire sub-cube $C_{v,x,S} := \{(v,y)|\forall j \notin S\ y_j = x_j\}$ lies inside the set $H_{v,x^*_{\rho(v)}}$. Moreover, since the labeling satisfies all the edges $(v, w_1), \dots, (v, w_t)$, by the claim above all the jobs in the sub-cubes C_{v,x,S,w_i} finish no later than the last job in $J_{x^*_{\rho(v)}}$. Thus in this case, the job $C_{v,x,S,w_1,\dots,w_t}$ can be scheduled no later than $2^{qn}|V|(\zeta + \frac{x^*_{\rho(v)}+1}{Q})$. Also, for a random x, $x^*_{\rho(v)}$ is equally likely to take the values in $\{0, 1, \dots, Q-1\}$. On the other hand, when the good event fails to happen, the completion is at most the length of the schedule i.e. $|V|2^{qn}$. Therefore,

$$
\begin{aligned}
& \mathbb{E}_{v,x,S,w_1,\dots,w_t}\left[\max_{j \in C_{v,x,S,w_1,\dots,w_t}} \rho(j)\right] \\
\leq\ & (1 - \varepsilon q - (t+1)\zeta) \cdot \left(\frac{1}{Q}\sum_{\ell=0}^{Q-1} 2^{qn}|V| \cdot (\zeta + \frac{\ell+1}{Q})\right) + ((t+1)\zeta + \varepsilon q) \cdot 2^{qn}|V| \\
\leq\ & 2^{qn}|V| \cdot \left(\frac{1}{2} + \frac{Q}{2}\zeta + \frac{1}{2Q} + (t+1)\zeta + \varepsilon q\right) \ \leq\ 2^{qn}|V|\left(\frac{1}{2} + o(1)\right).
\end{aligned}
$$

5.2.4 Soundness

Now we show that for any schedule that achieves an objective value at most $1 - (8+t)\delta$, there exists a labeling of the unique games instance that satisfies at least a $\frac{3\delta\eta^2}{k^2t^2}$ fraction of the edges.

Let $A \subseteq V \times \{0,1\}^{qn}$ be the last 2δ fraction of the jobs J in the ordering ρ (i.e. all jobs that are scheduled at time $(1-2\delta)|J|$ or later. Let f_A be the indicator function of A, i.e. $\forall (v,x) \in |V| \times \{0,1\}^{qn}$, $f_A(v,x) = 1$ if and only if $(v,x) \in A$. This induces a function $f_v(x) := f_A(v,x)$ for each vertex $v \in V$. Since $\mathbb{E}_{v,x}[f_A] \geq 2\delta$, we have that $\mathbb{E}_x[f_v] \geq \delta$ for at least δ fraction of the vertices $v \in V$. Call such vertices good and let $S \subseteq V$ be the set of good vertices, $|S| \geq \delta|V|$. By the expansion property of the unique games instance, $|\Gamma(S)| \geq (1-\delta)|W|$, i.e. all but δ fraction of vertices in W have a neighbor in S.

For $w \in \Gamma(S)$, let $f_w : H_w \mapsto \{0,1\}$ be a function such that $f_w(z) = 1$ if and only if the job (w,z) is scheduled at time $(1-2\delta)|J|$ or later. Consider the edge (v,w) for some $v \in S$. We have a precedence constraint between every job in H_v and its corresponding job in H_w. This implies that $\mathbb{E}_x[f_w] \geq \mathbb{E}_x[f_v] \geq \delta$. For the sake of convenience, define $f_w \equiv 0$ for $w \notin \Gamma(S)$.

Suppose the objective value is less than $1-(8+t)\delta$. Then at least $(6+t)\delta$ fraction of the jobs in C must be scheduled before time $(1-2\delta)|J|$. For any such job, say $C_{v,x,S,w_1,\dots,w_t}$, f_{w_i} is identically zero on the sub-cube C_{v,x,S,w_i}. This can be interpreted as if the test $T_{\varepsilon,t}$ from Section 3.3 accepts the collection of functions $f_{w,i} \circ \pi_{v,w_i}, 1 \leq i \leq t$ on pair (x,S). Thus overall at least $(6+t)\delta$ fraction of the tests pass. When we choose a vertex v and its t neighbors at random, call a tuple $(v, w_1, \dots, w_t)$ good if $\forall i \in [t], w_i \in \Gamma(S)$ (we do not care whether v is good or not). Since the unique games graph is regular and $|\Gamma(S)| \geq (1-\delta)|W|$, a tuple is good with probability at least $1 - t\delta$. Thus the test must pass with probability at least $(6+t-t)\delta = 6\delta$ over good tuples. Thus for at least 3δ fraction of good tuples the test passes with probability at least 3δ.

Let t, k, η be as in proof of Theorem 3.4. Applying the theorem, for every good tuple, it must be the case that there exist $1 \leq \ell_1 \neq \ell_2 \leq t$ such that $\exists i \in [n]$ and $\mathsf{Infl}_i^k(f_{w_{\ell_1}} \circ \pi_{v,w_{\ell_1}}) \geq \eta$ and $\mathsf{Infl}_i^k(f_{w_{\ell_2}} \circ \pi_{v,w_{\ell_2}}) \geq \eta$. The analysis now is identical to that for vertex cover. For each vertex w, we define a candidate set of labels $L[w]$ as those variables i for which $\mathsf{Infl}_i^k(f_w) \geq \eta$, and note that $|L[w]| \leq k/\eta$. The above condition says that there exist $j \in L[w_{\ell_1}], j' \in L[w_{\ell_2}]$ such that $\pi_{v,w_{\ell_1}}(j) = \pi_{v,w_{\ell_2}}(j')$. The labeling procedure defines the label of v by picking a neighbor $w \in N(v)$ at random and setting $\rho(v) = \pi_{v,w}(\rho(w))$, and as in the vertex cover analysis we obtain that

$$\Pr_{(v,w)\in E}[\rho(v) = \pi_{v,w}(\rho(w))] \geq \frac{3\delta\eta^2}{k^2t^2}.$$

References

[1] C. Ambuhl and M. Mastrolilli. Single machine precedence constrained scheduling is a vertex cover problem. In *ESA*, pages 28–39, 2006.

[2] Christoph Ambuhl, Monaldo Mastrolilli, and Ola Svensson. Inapproximability results for sparsest cut, optimal linear arrangement, and precedence constrained scheduling. In *FOCS*, pages 329–337, 2007.

[3] S. Arora, C. Lund, R. Motawani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[4] S. Arora and S. Safra. Probabilistic checking of proofs : A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[5] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi. Unique games on expanding constraint graphs are easy. In *STOC*, pages 21–28, 2008.

[6] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability. *Electronic Colloquium on Computational Complexity, Technical Report TR95-024*, 1995.

[7] C. Chekuri and R. Motwani. Precedence constrained scheduling to minimize sum of weighted completion times on a single machine. *Discrete Applied Mathematics*, 98:29–38, 1999.

[8] F. A. Chudak and D. S. Hochbaum. A half-integral linear programming relaxation for scheduling precedenceconstrained jobs on a single machine. *Operations Research Letters*, 25:199–204, 1999.

[9] J. R. Correa and A. S. Schulz. Single machine scheduling with precedence constraints. *Mathematics of Operations Research*, 30(4):1005–1021, 2005.

[10] I. Dinur and S. Safra. The importance of being biased. In *Proc. 34th Annual ACM Symposium on Theory of Computing*, 2002.

[11] M. E. Dyer and L. A. Wolsey. Formulating the single machine sequencing problem with release dates as a mixed integer program. *Discrete Appl. Math.*, 26:255–270, 1990.

[12] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.

[13] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[14] M. X. Goemans and D. P. Williamson. Two-dimensional gantt charts and a scheduling algorithm of lawler. *SIAM J. Discrete Math.*, 13:281–294, 2000.

[15] L. A. Hall, A. S. Schulz, D. B. Shmoys, and J. Wein. Scheduling to minimize average completion time: off-line and on-line algorithms. *Mathematics of Operations Research*, 22:513–544, 1997.

[16] J. Hastad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.

[17] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symposium on Theory of Computing*, 2002.

[18] S. Khot. Ruling out ptas for graph min-bisection, densest subgraph and bipartite clique. In *FOCS*, pages 136–145, 2004.

[19] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2 - \varepsilon$. In *Proc. 18th IEEE Conference on Computational Complexity*, 2003.

[20] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within 2-epsilon. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.

[21] E. L. Lawler. Sequencing jobs to minimize total weighted completion time subject to precedence constraints. *Ann. Discrete Math.*, 2:75–90, 1978.

[22] J. K. Lenstra and A. Rinnooy Kan. Complexity of scheduling under precedence constraints. *Oper. Res.*, 26:22–35, 1978.

[23] F. Margot, M. Queyranne, and Y. Wang. Decompositions, network flows and a precedence constrained single machine scheduling problem. *Operations Research*, 51(6):981–992, 2003.

[24] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th IEEE Symposium on Foundations of Computer Science*, pages 21–30, 2005.

[25] N. Pisaruk. A fully combinatorial 2-approximation algorithm for precedence-constrained scheduling a single machine to minimize average weighted completion time. *Discrete Applied Mathematics*, 131(3):655–663, 2003.

[26] C. N. Potts. An algorithm for the single machine sequencing problem with precedence constraints. *Math. Programming Stud.*, 13:78–87, 1980.

[27] A. S. Schulz. Scheduling to minimize total weighted completion time: Performance guarantees of lp-based heuristics and lower bounds. In *IPCO*, pages 301–315, 1996.

[28] P. Schuurman and G. J. Woeginger. Polynomial time approximation algorithms for machine scheduling: ten open problems. *Journal of Scheduling*, 2:203–213, 1999.

[29] J. B. Sidney. Decomposition algorithms for single machine scheduling with precedence relations and deferral costs. *Oper. Res.*, 23:283–298, 1975.

[30] Luca Trevisan. Approximation algorithms for unique games. In *FOCS*, pages 197–205, 2005.

[31] G. J. Woeginger. On the approximability of average completion time scheduling under precedence constraints. *Discrete Appl. Math.*, 131:237–252, 2003.