Electronic Colloquium on Computational Complexity, Report No. 147 (2017)

Hardness of Rainbow Coloring Hypergraphs

Venkatesan Guruswami* Rishi Saket†

Abstract

A hypergraph is *k-rainbow colorable* if there exists a vertex coloring using k colors such that each hyperedge has all the k colors. Unlike usual hypergraph coloring, rainbow coloring becomes *harder* as the number of colors increases. This work studies the rainbow colorability of hypergraphs which are guaranteed to be nearly balanced rainbow colorable. Specifically, we show that for any $Q, k \geq 2$ and $\ell \leq k/2$, given a Qk-uniform hypergraph which admits a k-rainbow coloring satisfying:

- in each hyperedge e, for some $\ell_e \leq \ell$ all but $2\ell_e$ colors occur exactly Q times and the rest $(Q \pm 1)$ times,

it is NP-hard to compute an independent set of $(1 - \frac{\ell+1}{k} + \varepsilon)$-fraction of vertices, for any constant $\varepsilon > 0$. In particular, this implies the hardness of even (k/ℓ)-rainbow coloring such hypergraphs.

The result is based on a novel long code PCP test that ensures the strong balancedness property desired of the k-rainbow coloring in the completeness case. The soundness analysis relies on a mixing bound based on uniform reverse hypercontractivity due to Mossel, Oleszkiewicz, and Sen, which was also used in earlier proofs of the hardness of $\omega(1)$-coloring 2-colorable 4-uniform hypergraphs due to Saket, and k-rainbow colorable $2k$-uniform hypergraphs due to Guruswami and Lee.

1 Introduction

A hypergraph is a collection of *vertices* and subsets of the set of vertices called *hyperedges*. It is q-uniform if each hyperedge has exactly q vertices, in particular a 2-uniform hypergraph is the usual graph. An independent set of a hypergraph is a subset of vertices that does not contain all the vertices of any hyperedge. A fundamental property of a hypergraph is its *colorability*: it is said to be k-colorable if the set of vertices can be colored using k colors so that no hyperedge is monochromatic. These color classes partition the vertices into k disjoint independent sets.

The computational aspects of coloring graphs and hypergraphs have been the focus of a substantial body of research. In brief, it is known that 3-colorable graphs can efficiently be colored using $n^{O(1)}$-colors [24] (see also [38, 6, 7, 23, 19, 1]). Similar $n^{O(1)}$-approximations are known for 2-colorable 3-uniform and 4-uniform hypergraphs [11, 25, 31]. From the intractability perspective, on graphs the best lower bound only rules out efficiently coloring a 3-colorable graph

*Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA. Email: venkatg@cs.cmu.edu.
Research supported in part by NSF grant CCF-1526092. This work was initiated during a visit by the author to Microsoft Research, Bangalore.
†IBM Research, Bangalore, India. Email: rissaket@in.ibm.com

ISSN 1433-8092

using 4-colors [26, 16]. For hypergraphs much better lower bounds rule out efficiently coloring 2-colorable 8-uniform hypergraphs using $\exp((\log n)^{\Omega(1)})$ colors [30, 22, 37] building upon the previous $\exp(\exp(\Omega(\sqrt{\log \log n}))$ lower bound [12, 14]. For 2-colorable 4-uniform hypergraphs a corresponding $(\log n)^c$ lower bound is known [36]. These intractability results proceed by ruling out computing independent sets of density (i.e., fraction of vertices) $\alpha(n)$ which implies the same for $(1/\alpha(n))$-coloring the hypergraph. For 2-colorable 3-uniform hypergraphs however, [13] directly showed the hardness of coloring using constantly many colors.

Another notion of coloring in hypergraphs is that of *rainbow colorability* – a hypergraph can be k-rainbow colored if there exists a coloring of the vertices using k colors ensuring that each hyperedge contains vertices of all the k colors. Any $(k-1)$ of the k color classes therefore constitute an independent set, and thus such hypergraphs have at least one independent set of density $(1-1/k)$. Note that unlike usual hypergraph coloring, rainbow coloring becomes more restrictive as the number of colors increases, and the problem is to determine the largest k for which there exists a k-rainbow coloring. This has been studied by Bollobás et. al. [8] who gave structural upper and lower bounds for several classes of hypergraphs. In some scenarios (such as modelling fair resource allocations) it is desirable that the coloring also be *balanced*, i.e., the colors should occur roughly the same number of times in any hyperedge. This is related to minimizing the *discrepancy* of hypergraph 2-colorings, for which notable recent works [5, 32] have given constructive algorithms. Subsequent works have shown tight hardness results for zero discrepancy case in hypergraphs with unbounded hyperedges [10] and for bounded hypergraphs with nearly zero discrepancy [4].

A perfectly balanced rainbow coloring is one in which every color appears exactly the same number of times in each hyperedge. Such hypergraphs are easy to efficiently 2-color using semi-definite programming (see [18]). In particular, a k-uniform k-rainbow colorable hypergraph (a.k.a., a k-uniform k-partite hypergraph) can efficiently be 2-colored. On the other hand, efficient 2-coloring, or even finding an independent set of density 1/2 does not seem possible if the guaranteed coloring deviates from being perfectly balanced. Indeed, Guruswami and Lee [18] proved the hardness of approximately coloring a class of such hypergraphs. They proved that it is NP-hard to compute an independent set of density ε in a Qk-uniform hypergraph $(Q, k \geq 2)$ which is guaranteed to be k-rainbow colorable such that each color appears at least $(Q-1)$ times in every hyperedge. This implies the hardness of coloring such hypergraphs using constantly many colors as well as that of non-trivially (i.e., using at least 2 colors) rainbow coloring them. The results of [18] do not, however, say anything about coloring k-rainbow colorable q-uniform hypergraphs for $q < 2k$. Brakensiek and Guruswami [9], under a conjectured intractability of a problem called "V Label Cover" that they formulate, proved hardness of finding an independent set of density ε in a $(k+1)$-uniform k-rainbow colorable hypergraph. This generalized the case of 2-colorable 3-uniform hypergraphs for which the same hardness was shown by Khot and Saket [29] under the d-to-1 Games Conjecture of Khot [28]. Unlike these conjecture based works however, our focus is on unconditional results.

While the structural guarantee considered in [18] captures balanced rainbow colorings, it also allows those in which a particular color may appear up to $(k+Q-1)$ times in a hyperedge. This is quite far off from being balanced in the regime where k is comparable or larger than Q, for example when $Q=2$ which corresponds to the smallest uniformity relative to k for which the hardness applies. The focus of this work is the case of rainbow colorable hypergraphs with a stronger balancedness condition on the coloring: in each hyperedge the occurrences of some of the colors is each off by *at most* 1 from Q, and the rest of the colors have precisely Q occurrences.

In our main result we show that it is hard to rainbow color such hypergraphs even with far fewer colors. Formally we show the following:

Theorem 1.1. *For any $Q, k \geq 2$, $\ell \leq k/2$ and an arbitrarily small constant $\varepsilon > 0$, given a Qk-uniform hypergraph of size n which is guaranteed to be k-rainbow colorable such that:*

- *in each hyperedge e, for some $\ell_e \leq \ell$, there are ℓ_e colors that occur exactly $(Q-1)$ times, ℓ_e colors that occur exactly $(Q+1)$ times and the rest of the colors occur exactly Q times,*

it is NP*-hard to compute an independent set of density $\left(1 - \frac{(\ell+1)}{k} + \varepsilon\right)$. This implies, in particular, that it is* NP*-hard to (k/ℓ)-rainbow color such hypergraphs. Under* $\mathrm{DTIME}(N^{O(\log\log N)})$ *reductions from* 3SAT *one can choose $\varepsilon = (\log n)^{-c}$ where c is some positive constant depending on Q, k and ℓ.*

Our result yields (with $Q = k = 2$ and $\ell = 1$) the result of Saket [36] who showed the currently best hardness of $(\log n)^c$-coloring 2-colorable 4-uniform hypergraphs, improving on the $\Omega\left(\frac{\log\log n}{\log\log\log n}\right)$ lower bound by Guruswami, Håstad, and Sudan [15], and Holmerin [21]. For $\ell = 1$ and $Q, k \geq 2$, Guruswami and Lee [18] studied this problem and claimed[1] a weaker hardness – in an auxiliary result of their initial work – ruling out independent sets of density $(1 - 1/k)$.

1.1 Our Techniques

The hardness reduction presented in this paper follows the template of *long code* based Probabilistically Checkable Proofs (PCPs) for the Label Cover problem. The long code encodings of the supposed labels of the Label Cover variables constitute the proof, whose locations are the vertices of the resulting hypergraph instance. The PCP verifier queries a few locations of the proof in each of its random tests defining the set of hyperedges. The test accepts if the $[k] := \{1, \ldots, k\}$-valued labels at the queried locations describe a rainbow coloring satisfying the desired balancedness criterion.

Let us now illustrate what we consider the novelty of our reduction: the PCP test for proving Theorem 1.1 with constants $Q, k \geq 2$ and $1 \leq \ell \leq k/2$. The test queries k locations from Q long codes corresponding to Q vertices of the Label Cover instance with constraints projecting on a common neighbor. Its main building block is a distribution $\overline{\mathcal{P}}$ over $\prod_{j=1}^{Q}([k]^d)^k$ which gives the query locations restricted to the Q pre-images of a label of the common neighbor. To construct $\overline{\mathcal{P}}$ we define, for $1 \leq t \leq k$, μ_t to be the uniform distribution over all $(\overline{x}^{(1)}, \ldots, \overline{x}^{(t)}) \in ([k]^d)^t$ such that for every $i \in [d]$, $\overline{x}_i^{(1)}, \ldots, \overline{x}_i^{(t)}$ are distinct. Figure 1a illustrates a sample from μ_k.

Corresponding to the jth long code ($j \in [Q]$), let $(\overline{x}^{1,j}, \ldots, \overline{x}^{k,j})$ be an i.i.d. sample from μ_k. Now, for any choice of labels (having the fixed common projection) given by $i_1, \ldots, i_Q \in [d]$,

$$(\overline{x}_{i_1}^{(1,1)}, \ldots, \overline{x}_{i_1}^{(k,1)}, \ldots, \overline{x}_{i_j}^{(1,j)}, \ldots, \overline{x}_{i_j}^{(k,j)}, \ldots, \overline{x}_{i_Q}^{(1,Q)}, \ldots, \overline{x}_{i_Q}^{(k,Q)}) \in [k]^{kQ}$$

is perfectly balanced i.e., each color in $[k]$ occurs exactly Q times. For the PCP test to work, it requires some perturbation while ensuring that the resultant coloring above remains nearly balanced.

[1] However, Guruswami and Lee have since withdrawn this claim (Theorem 1.4 and Appendix D in the ECCC version [17]) from later versions of their paper [18].

2	4	1	5	3
3	1	2	4	5
5	2	4	1	3
1	2	3	4	5

(a) A sample from μ_k

4	2	1	4	3
3	1	1	2	3
3	5	4	1	5
5	4	5	4	2

(b) A sample from $(\mu_\ell \otimes \mu_{k-\ell})$

Figure 1: Samples from μ_k and $(\mu_\ell \otimes \mu_{k-\ell})$. $k = 5$, $\ell = 2$ and $d = 4$.

For this purpose, we choose $j^* \in [Q]$ at random and sample $(\overline{x}^{1,j^*}, \dots, \overline{x}^{k,j^*})$ from $\mu_\ell \otimes \mu_{k-\ell}$ (illustrated in Figure 1b). Formally, we define:

$$\overline{\mathcal{P}} := \frac{1}{Q} \sum_{j^*=1}^{Q} \left[\underbrace{\mu_k \otimes \cdots \otimes \mu_k}_{j^*-1 \text{ times}} \otimes (\mu_\ell \otimes \mu_{k-\ell}) \otimes \overbrace{\mu_k \otimes \cdots \otimes \mu_k}^{Q-j^* \text{ times}} \right].$$

Notice that the marginal of $\overline{\mathcal{P}}$ for any $j \in [Q]$ is $\mathcal{P} := (1 - 1/Q)\mu_k + (1/Q)(\mu_\ell \otimes \mu_{k-\ell})$. As desired, for any choice of $i_1, \dots, i_Q \in [d]$, for any $\prod_{j=1}^{Q} (\overline{x}^{(1,j)}, \dots, \overline{x}^{(k,j)})$ in the support of $\overline{\mathcal{P}}$,

$$(\overline{x}_{i_1}^{(1,1)}, \dots, \overline{x}_{i_1}^{(k,1)}, \dots, \overline{x}_{i_j}^{(1,j)}, \dots, \overline{x}_{i_j}^{(k,j)}, \dots, \overline{x}_{i_Q}^{(1,Q)}, \dots, \overline{x}_{i_Q}^{(k,Q)})$$

is nearly balanced: for some $\ell' \leq \ell$, ℓ' of the colors in $[k]$ occur exactly $(Q-1)$ times, ℓ' of them occur exactly $(Q+1)$ times, and the rest of the colors exactly Q times.

To extend the test distribution to the pre-images of L labels on smaller side of the Label Cover, we sample

$$(x^{(1,1)}, \dots, x^{(k,1)}, \dots, x^{(1,j)}, \dots, x^{(k,j)}, \dots, x^{(1,Q)}, \dots, x^{(k,Q)}) \in \prod_{j=1}^{Q} ([k]^{Ld})^k$$

by sampling independently for each $i \in [L]$, the restriction of the above locations to the coordinates $\{d(i-1)+1, \dots, di\}$ from $\overline{\mathcal{P}}$.

Let $f_j : [k]^{Ld} \to \{0,1\}$ denote the restriction to the jth long code of a sufficiently dense independent set. Our soundness analysis shows that with significant probability over the choice of the Q long codes of the PCP test, two of the functions $\{f_j\}_{j \in [Q]}$ are intersecting juntas. Otherwise, the expectations inside each long code would be uncorrelated yielding a hyperedge inside the supposed independent set. These juntas are then decoded into a good labeling for the Label Cover. This motivates the first step of the analysis: in a single long code lower bounding the expectation $\mathbb{E}\left[\prod_{s=1}^{k} f_j(x^{(s,j)})\right]$ for some sufficiently heavy f_j. We show that when $\mathbb{E}[f_j] \geq (1 - \frac{\ell+1}{k} + \varepsilon)$,

$$\mathbb{E}\left[\prod_{s=1}^{k} f_j(x^{(s,j)})\right] \geq \Omega(\varepsilon^c)$$

for some $c > 0$ depending on Q, k and ℓ. The proof of this lower bound proceeds by representing the product inside the expection as $A(X)B(Y)$ where $(X,Y) = ((x^{(1,j)}, \dots, x^{(\ell,j)}), (x^{(\ell+1,j)}, \dots, x^{(k,j)}))$.

It can be seen that X and Y are at most $(1-1/Q)$-correlated[2]. Further, using the structure of the PCP test we show that $\mathbb{E}[A], \mathbb{E}[B] \geq \Omega(\varepsilon)$. Using this coupled with a lower bound on the mixing of Markov chains that was shown by Mossel, Oleszkiewicz, and Sen [34] based on their generalized *reverse hypercontractivity*, yields the desired lower bounds. A similar use (for a simpler PCP test) of this technique was made by Saket [36]. Subsequently Guruswami and Lee [18] used reverse hypercontractivity to analyze a PCP test which sampled uniformly from $([k]^d)^k$ instead of $\mu_\ell \otimes \mu_{k-\ell}$ for the j^*th long code, and used it to show their NP-hardness result for $O(1)$-coloring k-rainbow colorable hypergraphs mentioned earlier.

In their work, Guruswami and Lee [18] leveraged a *smoothness* property (first defined by Khot [27]) of the Label Cover instance for their analysis which used Gaussian invariance theorems and decoded a labeling using influential coordinates. Unfortunately, achieving smoothness generates a significant blowup in the size of the Label Cover which renders the reduction somewhat inefficient. In contrast, our analysis (as also in [36]) is based on standard Fourier analysis and uses a projection size preservation property of the vanilla Label Cover shown by Håstad [20]. This limits the size blowup enabling us to upper bound the "error" ε in the NO case to $1/\text{poly}(\log n)$ under quasi-polynomial time reductions.

2 Preliminaries

Consider for $i=1$ to n, the product space $(\Omega_i^{(1)} \times \Omega_i^{(2)}, \mu_i)$ where the marginals of μ_i are $\mu_i^{(1)}, \mu_i^{(2)}$. Let $(\Omega^{(s)}, \mu^{(s)}) = (\prod_{i=1}^n \Omega_i^{(s)}, \otimes_{i=1}^n \mu_i^{(s)})$, for $s=1,2$. We say that $(X,Y) \in \Omega^{(1)} \times \Omega^{(2)}$ is ρ-correlated[3] if independently for each $i \in [n]$, (X_i, Y_i) is sampled from μ_i with probability ρ, and from $\mu_i^{(1)} \otimes \mu_i^{(2)}$ with probability $(1-\rho)$.

The following theorem is a straightforward generalization of the special case of $\Omega^{(1)} = \Omega^{(2)}$ and $\mu_i = \text{id}$ proved in [34]. The derivation is provided in [18] and we incorporate the explicit parameters from [34].

Theorem 2.1. *In above setup, let $A \subseteq \Omega^{(1)}$ and $B \subseteq \Omega^{(2)}$ be two sets such that $\mu^{(1)}\{A\}, \mu^{(2)}\{B\} \geq \delta \geq 0$. Let $(X,Y) \in (\Omega^{(1)} \times \Omega^{(2)})$ be ρ-correlated. Then,* $\Pr[X \in A, Y \in B] \geq \delta^{\frac{2-\sqrt{\rho}}{1-\sqrt{\rho}}}$.

We shall also use the *Efron-Stein* decompositions of functions over product spaces (see [33] for reference).

Proposition 2.2. *Let $(\Omega = \prod_{i=1}^n \Omega_i, \mu = \otimes_{i=1}^n \mu_i)$ be a product space. Then, any $f \in L^2(\Omega, \mu)$ can be decomposed uniquely as:*

$$f(x) = \sum_{S \subseteq [n]} f_S(x),$$

where f_S depends only on the coordinates in S and for $S' \not\supseteq S$, $E[f_S | x_{S'}] = 0$. In particular $\{f_S\}_{S \subseteq [n]}$ are orthogonal i.e., $\mathbb{E}[f_S f_{S'}] = 0$ for $S \neq S'$.

The starting point of the reduction is the LabelCover problem which is defined as follows.

[2] This is analogous to the notion of ρ-correlation used in [34] and was also used in the reverse hypercontractivity based mixing bounds of [18].

[3] See footnote 2.

Definition 2.3. *An instance $\mathcal{L}$ of* LabelCover *consists of a bipartite graph $G_{\mathcal{L}}(U, V, E)$ along with label sets $[L]$ and $[M]$ where $M = dL$. For each edge e between $u \in U$ and $v \in V$, there is a projection $\pi_{vu} : [M] \mapsto [L]$, such that $|\pi_{vu}^{-1}(j)| = d$ for each $j \in [L]$. A labeling $l_u \in [L]$ to u and $l_v \in [M]$ to v satisfies the edge if $\pi_{vu}(l_v) = l_u$. The goal is to find a labeling of U and V to satisfy the maximum number of edges.*

The inapproximability of LabelCover stated below follows from the PCP Theorem [3, 2], Raz's Parallel Repetition Theorem [35]. We also leverage a structural property proved by Håstad [20] showing that for any vertex in V the image of a large subset of its labels remains large under most of the projections incident on v.

Theorem 2.4. *For every positive integer r, there is a deterministic $N^{O(r)}$ time reduction from a* 3SAT *instance of size N to an instance $\mathcal{L}(G_{\mathcal{L}}(U, V, E), \{\pi_{vu}\}_{\{v,u\}\in E}, [L], [M])$ of* LabelCover *with the following properties:*

a. $|U|, |V| \leq N^{O(r)}$. $L, M, d \leq 2^{3r}$. G is bi-regular with left and right degrees bounded by $2^{O(r)}$.

b. There is a universal constant $c_0 > 0$ such that for any $v \in V$ and $S \subseteq [M]$, taking an expectation over a random neighbor u of v, $\mathbb{E}\left[|\pi_{vu}(S)|^{-1}\right] \leq |S|^{-2c_0}$. This implies that over the choice of a random neighbor u of v,

$$\Pr\left[|\pi_{vu}(S)| < |S|^{c_0}\right] \leq |S|^{-c_0}.$$

c. There is a universal constant $\gamma_0 > 0$ such that,

YES Case: *If the* 3SAT *instance is satisfiable then there is a labeling to U and V that satisfies all edges of $\mathcal{L}$.*

NO Case: *If the* 3SAT *instance is unsatisfiable then any labeling to U and V satisfies at most $2^{-\gamma_0 r}$ fraction of the edges.*

3 Proof of Theorem 1.1

In this section we prove the following hardness reduction which implies Theorem 1.1.

Theorem 3.1. *For any constant integers $Q, k \geq 2$ and $k/2 \geq \ell \geq 1$, and constant $\varepsilon > 0$, there is a polynomial time reduction from* 3SAT *to a Qk-uniform hypergraph G of size n such that:*

YES Case. *If the* 3SAT *instance is satisfiable then there is a k-coloring of the vertices of G such that for in each hyperedge e for some $\ell_e \leq \ell$ there exactly ℓ_e colors that appear $(Q-1)$ times each and ℓ_e colors that appear $(Q+1)$ times each, and the other colors appear exactly Q times each. In particular, the hypergraph is k-rainbow colorable.*

NO Case. *If the* 3SAT *instance is not satisfiable then there is no independent set in G of size $1 - \frac{(\ell+1)}{k} + \varepsilon$ fraction of vertices, implying that G is not (k/ℓ)-rainbow colorable.*

In the above, ε can be chosen to be $(\log n)^{-c}$ for some c depending on Q, k and ℓ if $N^{O(\log\log N)}$-time reduction from 3SAT *is allowed.*

The input is an instance $\mathcal{L}$ of LabelCover from Theorem 2.4 consisting of a bipartite graph $G_{\mathcal{L}}(U, V, E)$, label sets $[M]$ and $[L]$ with $M = dL$, and projections $\{\pi_{vu} : [M] \mapsto [L] \mid \{u, v\} \in E, u \in U, v \in V\}$ such that $|\pi_{vu}^{-1}(j)| = d$ for any $j \in [L]$.

For each vertex $v \in V$, there is a uniformly weighted long code $\mathcal{H}^v$ which is a copy of $[k]^M$. The set of vertices in the output G is $\bigcup_v \mathcal{H}^v$ and the Qk-uniform hyperedges correspond to the PCP test that is described in the next few paragraphs. The parameters Q, k and ℓ in Theorem 3.1 are fixed in the rest of this section.

3.1 Distributions

First we define $\mathcal{D}(t)$ for $t \leq k$ to be the uniform distribution over the set

$$\Gamma(t) := \{z \in [k]^t \mid z_i \neq z_j, \ \forall\, 1 \leq i < j \leq t\}. \tag{1}$$

Given this, let μ_t be the distribution over $([k]^d)^t$ where $(\overline{x}^{(1)}, \ldots, \overline{x}^{(t)}) \in ([k]^d)^t$ is sampled by independently for each $i \in [d]$ sampling $(\overline{x}_i^{(1)}, \ldots, \overline{x}_i^{(t)})$ from $\mathcal{D}(t)$. Using this we define the distribution $\overline{\mathcal{P}}$ over $\prod_{j=1}^{Q}([k]^d)^k$ by the following sampling procedure.

1. Choose $j^* \in [Q]$ uniformly at random.
2. For each $j \in [Q] \setminus \{j^*\}$ sample $(\overline{x}^{(1,j)}, \ldots, \overline{x}^{(k,j)})$ from μ_k.
3. For j^* sample $(\overline{x}^{(1,j^*)}, \ldots, \overline{x}^{(k,j^*)})$ from $(\mu_\ell \otimes \mu_{k-\ell})$.
4. Ouput $(\overline{x}^{(1,1)}, \ldots, \overline{x}^{(k,1)}, \ldots, \overline{x}^{(1,j)}, \ldots, \overline{x}^{(k,j)}, \ldots, \overline{x}^{(1,Q)}, \ldots, \overline{x}^{(k,Q)})$.

The marginal of $\overline{\mathcal{P}}$ restricted to any $j \in [Q]$ is the same distribution $\mathcal{P}$ where

$$\mathcal{P} := \left(1 - \frac{1}{Q}\right)\mu_k + \left(\frac{1}{Q}\right)(\mu_\ell \otimes \mu_{k-\ell}). \tag{2}$$

With the above in place the PCP test of the verifier is given below.

Test of PCP Verifier

The verifier expects a coloring $C_v : \mathcal{H}^v \to [k]$ for all $v \in V$ and executes the following steps.

1. The verifier chooses a random vertex $u \in U$ and Q of its neighbors $v_1, \ldots, v_Q$ with projections $\pi_j := \pi_{v_j u}$ and long codes $\mathcal{H}^j := \mathcal{H}^{v_j}$. Let the $C_j := C_{v_j}$ be the corresponding colorings.
2. The verifier samples:

$$\left(x^{(1,1)}, \ldots, x^{(k,1)}, \ldots, x^{(1,j)}, \ldots, x^{(k,j)}, \ldots, x^{(1,Q)}, \ldots, x^{(k,Q)}\right) \in \prod_{j=1}^{Q}([k]^{Ld})^k$$

from the distribution $\overline{\mathcal{Q}}$ which is described by sampling independently for each $i \in [L]$,

$$\left(x^{(1,1)}|_{\pi_1^{-1}(i)}, \ldots, x^{(k,1)}|_{\pi_1^{-1}(i)}, \ldots, x^{(1,Q)}|_{\pi_Q^{-1}(i)}, \ldots, x^{(k,Q)}|_{\pi_Q^{-1}(i)}\right)$$

from the distribution $\overline{\mathcal{P}}$ defined above. Let the marginal of $\overline{\mathcal{Q}}$ restricted to $j \in [Q]$ be $\mathcal{Q}_j$ noting that $\mathcal{Q}_j$ is identical to $\mathcal{P}^L$ up to permutation of coordinates.

3. The verifier accepts if the coloring $\left(C_j(x^{(s,j)})\right)_{s\in[k], j\in[Q]}$ has for some $\ell' \leq \ell$, ℓ' colors that appear exactly $(Q-1)$ times each, ℓ' colors that appear exactly $(Q+1)$ times each and the rest of the colors appearing exactly Q times each.

The rest of this section is devoted to proving the YES and NO cases of Theorem 3.1.

3.2 YES Case

In this case there is a labeling l_v for $v \in V$ such that for any $u \in U$ and its neighbors v, w, $\pi_{vu}(l_u) = \pi_{wu}(l_w)$. Letting $C_v(x) = x_{l_v}$ for $x \in \mathcal{H}^v$ and $v \in V$ yields a coloring of G that makes the verifier accept with probability 1. In particular, this coloring satisfies the YES case of Theorem 3.1.

3.3 NO Case

Suppose that G contains an independent set $\mathcal{I}$ of $\left(1 - \frac{(\ell+1)}{k} + 4\varepsilon\right)$ fraction of vertices. By standard averaging and using the bi-regularity of the LabelCover instance $\mathcal{L}$ we obtain that for at least ε fraction of "good" vertices $u \in U$, at least ε fraction of its neighbors are "heavy" vertices $v \in V$ which satisfy,

$$\mathbb{E}_{x\in[k]^M}\left[f_v(x)\right] \geq \left(1 - \frac{\ell+1}{k} + \varepsilon\right), \tag{3}$$

where $f_v : [k]^M \to \{0,1\}$ is the indicator of $\mathcal{I} \cap \mathcal{H}^v$.

3.3.1 Lower bound in each Long Code

Fix a choice of a good u and its heavy neighbors $v_1, \ldots, v_Q$ in the verifiers test. For convenience we let $f_j := f_{v_j}$. Fix some $j \in [Q]$, and consider the expectation

$$\mathbb{E}_{(x^{(1,j)},\ldots,x^{(k,j)})\leftarrow\mathcal{Q}_j}\left[\prod_{s=1}^{k} f_j(x^{(s,j)})\right]. \tag{4}$$

By rearranging the coordinates and omitting the subscript j, the above is equivalent to the following expectation:

$$\mathbb{E}_{(x^{(1)},\ldots,x^{(k)})\leftarrow\mathcal{P}^L}\left[\prod_{s=1}^{k} f(x^{(s)})\right], \tag{5}$$

where

$$\mathbb{E}_{x\in([k]^d)^L\simeq[k]^M}\left[f(x)\right]\geq\left(1-\frac{\ell+1}{k}+\varepsilon\right). \tag{6}$$

Using the definition of $\Gamma(.)$ in (1) and by abusing notation slightly let

$$X=(x^1,\dots,x^\ell)\in(\Gamma(\ell))^M,\quad\text{and}\quad Y=(x^{\ell+1},\dots,x^k)\in(\Gamma(k-\ell))^M, \tag{7}$$

and define functions

$$A(X):=\prod_{s=1}^{\ell}f(x^s),\quad\text{and}\quad B(Y):=\prod_{s=\ell+1}^{k}f(x^s). \tag{8}$$

Thus, the expectation in (5) is equivalent to

$$\mathbb{E}_{X,Y}\left[A(X)B(Y)\right], \tag{9}$$

where (X,Y) is sampled from $\mathcal{P}^{\otimes L}\simeq[(1-1/Q)\mu_k+(1/Q)(\mu_\ell\otimes\mu_{k-\ell})]^{\otimes L}$. Note that, the marginal distribution of X is $\mu_\ell^{\otimes L}$ and that of Y is $\mu_{k-\ell}^{\otimes L}$. Thus, X and Y are $(1-1/Q)$-correlated. Applying Theorem 2.1 we obtain,

$$\mathbb{E}_{X,Y}\left[A(X)B(Y)\right]\geq\left(\min\{\mathbb{E}[A(X)],\mathbb{E}[B(Y)]\}\right)^{3Q}. \tag{10}$$

The following argument lower bounds the RHS of the above.

Lemma 3.2. *For A and B defined as above,* $\mathbb{E}[A],\ \mathbb{E}[B]\ \geq\ \delta_0\ :=\ \varepsilon\Big/\binom{k}{\ell}$.

Proof. Consider a k-uniform hypergraph H on vertex set $[k]^M$ and hyperedge set $\{(x^1,\dots,x^k)\mid(x_i^{(1)},\dots,x_i^{(k)})\in\Gamma(k),\ \forall i\in[M]\}$. Observe that H is a regular hypergraph i.e. each vertex appears in the same number of hyperedges. Using the bound in (6) along with an averaging, we obtain that at least ε fraction of the hyperedges $(x^{(1)},\dots,x^{(k)})$ are "dense" satisfying

$$\left|\{s\ \mid\ f(x^{(s)})=1\}\right|\geq k-\ell. \tag{11}$$

Consider a random choice of $Y=(x^{(\ell+1)},\dots,x^{(k)})$ sampled from $\mu_{k-\ell}^{\otimes L}\simeq\mathcal{D}(k-\ell)^{\otimes M}$. This is equivalent to a u.a.r choice of a hyperedge in H and a u.a.r subset of $(k-\ell)$ of its vertices. From (11) we obtain $\mathbb{E}_Y[B(Y)]=\mathbb{E}_{(x^{(\ell+1)},\dots,x^{(k)})\leftarrow\mathcal{D}(k-\ell)^{\otimes M}}\left[\prod_{s=\ell+1}^{k}f(x^{(s)})\right]\geq\delta_0$. Further, since $\ell\leq k/2$ it is easy to see that $\mathbb{E}_X[A(X)]\geq\mathbb{E}_Y[B(Y)]$. □

Using Lemma 3.2 along with (10) we obtain $\mathbb{E}_{X,Y}\left[A(X)B(Y)\right]\geq\delta_0^{3Q}$, which is rewritten as:

$$\mathbb{E}_{(X^{(j)},Y^{(j)})\leftarrow\mathcal{Q}_j}\left[A_j(X^{(j)})B_j(Y^{(j)})\right]\geq\delta_1:=\delta_0^{3Q}. \tag{12}$$

3.3.2 Analyzing over Q Long Codes

Let us fix a choice of u and Q of its neighbors $v_1, \dots, v_Q$. Using the notation introduced in Section 3.3.1: we have functions $A_j : \Gamma(\ell)^M \to \{0,1\}$ and $B_j : \Gamma(k-\ell)^M \to \{0,1\}$ for $j = 1, \dots, Q$. From Proposition 2.2, their Efron-Stein decomposition is given as follows.

$$A_j = \sum_{S \subseteq [M]} A_{j,S}, \quad \text{and} \quad B_j = \sum_{S \subseteq [M]} B_{j,S}. \tag{13}$$

Let R be a parameter to be decided later. Define the following subsets of $(2^{[M]})^Q$:

$$\mathcal{S}_0 := \left\{ (S_1, \dots, S_Q) \in (2^{[M]})^Q \mid \pi_j(S_j) \cap \pi_{j'}(S_{j'}) = \emptyset, \ 1 \le j < j' \le Q \right\}, \tag{14}$$

$$\mathcal{S}_1 := \left\{ (S_1, \dots, S_Q) \in (2^{[M]})^Q \setminus \mathcal{S}_0 \mid \max_j |S_j| \le R \right\}, \tag{15}$$

$$\mathcal{S}_2 := \left\{ (S_1, \dots, S_Q) \in (2^{[M]})^Q \mid \max_j |\pi_j(S_j)| > R^{c_0} \right\}, \tag{16}$$

$$\mathcal{S}_3 := \left\{ (S_1, \dots, S_Q) \in (2^{[M]})^Q \mid \exists j \text{ s.t. } |S_j| > R, |\pi_j(S_j)| \le R^{c_0} \right\}, \tag{17}$$

where $c_0 > 0$ is the constant from Theorem 2.4. Note that $\bigcup_{p=0}^{3} \mathcal{S}_p \supseteq (2^{[M]})^Q$. Let us define

$$\delta := \prod_{j=1}^{Q} \mathbb{E}_{(X^{(j)}, Y^{(j)}) \leftarrow \mathcal{Q}_j} \left[A_j(X^{(j)}) B_j(Y^{(j)}) \right]. \tag{18}$$

Since $\mathcal{I}$ is an independent set we also have,

$$\mathbb{E}_{((X^{(1)}, Y^{(1)}), \dots, (X^{(Q)}, Y^{(Q)})) \leftarrow \overline{\mathcal{Q}}} \left[\prod_{j=1}^{Q} A_j(X^{(j)}) B_j(Y^{(j)}) \right] = 0. \tag{19}$$

Subtracting (19) from (18), expanding the Efron-Stein decomposition and using standard Fourier analysis we obtain

$$\Delta_0 + \Delta_1 + \Delta_2 + \Delta_3 \ge \delta, \tag{20}$$

where,

$$\begin{aligned} \Delta_0 = \sum_{(S_1, \dots, S_Q) \in \mathcal{S}_0} \Bigg(& \prod_{j=1}^{Q} \mathbb{E}_{(X^{(j)}, Y^{(j)}) \leftarrow \mathcal{Q}_j} \left[A_{j,S_j}(X^{(j)}) B_{j,S_j}(Y^{(j)}) \right] \\ & - \mathbb{E}_{((X^{(1)}, Y^{(1)}), \dots, (X^{(Q)}, Y^{(Q)})) \leftarrow \overline{\mathcal{Q}}} \left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) B_{j,S_j}(Y^{(j)}) \right] \Bigg) \end{aligned} \tag{21}$$

and for $p = 1, 2, 3$,

$$\begin{aligned} \Delta_p = \sum_{(S_1, \dots, S_Q) \in \mathcal{S}_p} \Bigg(& \prod_{j=1}^{Q} \left| \mathbb{E}_{(X^{(j)}, Y^{(j)}) \leftarrow \mathcal{Q}_j} \left[A_{j,S_j}(X^{(j)}) B_{j,S_j}(Y^{(j)}) \right] \right| \\ & + \left| \mathbb{E}_{((X^{(1)}, Y^{(1)}), \dots, (X^{(Q)}, Y^{(Q)})) \leftarrow \overline{\mathcal{Q}}} \left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) B_{j,S_j}(Y^{(j)}) \right] \right| \Bigg). \end{aligned} \tag{22}$$

The definition of $\mathcal{S}_0$ and the properties of Efron-Stein decompositions in Proposition 2.2 imply that each term of the sum in the RHS of (21) is zero. Thus,

$$\Delta_0 = 0. \tag{23}$$

The goal of the rest of the analysis is to show that for an appropriate choice of r in Theorem 2.4 and the parameter R, the expectation over the choice of u and $v_1, \ldots, v_Q$ of each Δ_p ($p = 1, 2, 3$) is small. Specifically, we shall show that a large $\mathbb{E}[\Delta_1]$ would yield a good labeling to $\mathcal{L}$ contradicting its NO case. Further, Δ_2 is bounded by the dampening induced by the presence of subsets with large projections in its sum, and $\mathbb{E}[\Delta_3]$ is bounded by property (b) of Theorem 2.4. On the other hand, $\mathbb{E}[\delta]$ is significant due to (12) thereby yielding for us a contradiction in (20).

We begin with the following upper bound on Δ_1.

Lemma 3.3.

$$\Delta_1 \leq 2 \cdot \left(\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}}.$$

Proof. Using $\mathbb{E}[fg] \leq \|f\|_2\|g\|_2$ observe that

$$\begin{aligned}
\Delta_1 &\leq \sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2 \left\|B_{j,S_j}\right\|_2 + \sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \left\| \prod_{j=1}^{Q} A_{j,S_j} \right\|_2 \left\| \prod_{j=1}^{Q} B_{j,S_j} \right\|_2 \\
&\leq \left(\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}} \left(\sum_{(S_1,\ldots,S_Q)} \prod_{j=1}^{Q} \left\|B_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}} \\
&\quad + \left(\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \left\| \prod_{j=1}^{Q} A_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}} \left(\sum_{(S_1,\ldots,S_Q)} \left\| \prod_{j=1}^{Q} B_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}},
\end{aligned} \tag{24}$$

where the last inequality uses the standard Cauchy-Schwartz inequality. Observe that $\left\| \prod_{j=1}^{Q} A_{j,S_j} \right\|_2^2 = \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2^2$ since $\{X^{(j)}\}_{j=1}^{k}$ are independent. Similarly, $\left\| \prod_{j=1}^{Q} B_{j,S_j} \right\|_2^2 = \prod_{j=1}^{Q} \left\|B_{j,S_j}\right\|_2^2$. Thus, (24) boils down to,

$$\begin{aligned}
\Delta_1 &\leq 2 \cdot \left(\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}} \left(\sum_{(S_1,\ldots,S_Q)} \prod_{j=1}^{Q} \left\|B_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}} \\
&\leq 2 \cdot \left(\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}} \left(\prod_{j=1}^{Q} \left\|B_j\right\|_2^2 \right)^{\frac{1}{2}} \\
&\leq 2 \cdot \left(\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \left\|A_{j,S_j}\right\|_2^2 \right)^{\frac{1}{2}}.
\end{aligned} \tag{25}$$

□

For analyzing Δ_2 and Δ_3, we use the following lemma which is proved in Appendix A.

Lemma 3.4.

$$\left|\mathbb{E}_{((X^{(1)},Y^{(1)}),\ldots,(X^{(Q)},Y^{(Q)}))\leftarrow\overline{\mathcal{Q}}}\left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \leq \left\|\prod_{j=1}^{Q} A_{j,S_j}\right\|_2 \left\|\prod_{j=1}^{Q} B_{j,S_j}\right\|_2 \left(1-\frac{1}{Q}\right)^{\max_j |\pi_j(S_j)|}.$$

Corollary 3.5.

$$\left|\mathbb{E}_{(X^{(j)},Y^{(j)})\leftarrow\mathcal{Q}_j}\left[A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \leq \left\|A_{j,S_j}\right\|_2 \left\|B_{j,S_j}\right\|_2 \left(1-\frac{1}{Q}\right)^{|\pi_j(S_j)|}.$$

Proof. Use Lemma 3.4 with $S_{j'} = \emptyset$ for all $j' \neq j$. □

Using the above we have the following bounds for the two sums in Δ_2.

Claim 3.6.

$$\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_2}\left|\mathbb{E}_{\overline{\mathcal{Q}}}\left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \leq Q\left(1-\frac{1}{Q}\right)^{R^{c_0}}. \tag{26}$$

Claim 3.7.

$$\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_2}\prod_{j=1}^{Q}\left|\mathbb{E}_{\overline{\mathcal{Q}}}\left[A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \leq \left(1-\frac{1}{Q}\right)^{R^{c_0}}. \tag{27}$$

Claims 3.6 and 3.7 are proved in Appendix B. Using them along with $p = 2$ in (22) directly yields the following lemma upper bounding Δ_2.

Lemma 3.8.

$$\Delta_2 \leq (Q+1)\left(1-\frac{1}{Q}\right)^{R^{c_0}}.$$

Similarly, we have the following bounds for the two sums in Δ_3.

Claim 3.9.

$$\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_3}\left|\mathbb{E}_{\overline{\mathcal{Q}}}\left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \leq \sum_{j=1}^{Q}\left(\sum_{\substack{S_j:|S_j|>R\\|\pi_j(S_j)|<R^{c_0}}}\left\|A_{j,S_j}\right\|_2^2\right)^{\frac{1}{2}} \tag{28}$$

Claim 3.10.

$$\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_3}\prod_{j=1}^{Q}\left|\mathbb{E}_{\overline{\mathcal{Q}}}\left[A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \leq \sum_{j=1}^{Q}\left(\sum_{\substack{S_j:|S_j|>R\\|\pi_j(S_j)|<R^{c_0}}}\left\|A_{j,S_j}\right\|_2^2\right)^{\frac{1}{2}} \tag{29}$$

Claims 3.9 and 3.10 are proved in Appendix C. Again, with $p = 3$ in (22) Claims 3.9 and 3.10 directly imply the following lemma.

Lemma 3.11.

$$\Delta_3 \leq 2 \cdot \sum_{j=1}^{Q} \left(\sum_{\substack{S_j : |S_j| > R \\ |\pi_j(S_j)| < R^{c_0}}} \left\| A_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}} .$$

Plugging in Lemmas 3.3, 3.8 and 3.11 into (20) we obtain that for such a fixed choice of u and $v_1, \ldots, v_Q$

$$\begin{aligned} \left(\frac{\delta}{Q+1}\right) \quad \leq \quad & \left(\sum_{(S_1,\ldots,S_Q) \in \mathcal{S}_1} \prod_{j=1}^{Q} \left\| A_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}} + \sum_{j=1}^{Q} \left(\sum_{\substack{S_j : |S_j| > R \\ |\pi_j(S_j)| < R^{c_0}}} \left\| A_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}} \\ & + \left(1 - \frac{1}{Q}\right)^{R^{c_0}} . \end{aligned} \tag{30}$$

For a good choice of u, and Q of its heavy neighbors $v_1, \ldots, v_Q$, as defined in (18), $\delta \geq \delta_1^Q$ due to the lower bound in (12). Taking an expectation of (30) over the verifiers choices and noting that with probability at least ε^{Q+1} u is good and $v_1, \ldots, v_Q$ are heavy, we obtain,

$$\begin{aligned} \left(\frac{\delta_1^Q \varepsilon^{Q+1}}{Q+1}\right) \quad \leq \quad & \mathbb{E}_{u,\{v_j\}_{j=1}^Q} \left[\left(\sum_{(S_1,\ldots,S_Q) \in \mathcal{S}_1} \prod_{j=1}^{Q} \left\| A_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}} + \sum_{j=1}^{Q} \left(\sum_{\substack{S_j : |S_j| > R \\ |\pi_j(S_j)| < R^{c_0}}} \left\| A_{j,S_j} \right\|_2^2 \right)^{\frac{1}{2}} \right] \\ & + \left(1 - \frac{1}{Q}\right)^{R^{c_0}} \\ \leq \quad & \left(\mathbb{E}_{u,\{v_j\}_{j=1}^Q} \left[\sum_{(S_1,\ldots,S_Q) \in \mathcal{S}_1} \prod_{j=1}^{Q} \left\| A_{j,S_j} \right\|_2^2 \right] \right)^{\frac{1}{2}} \\ & + \sum_{j=1}^{Q} \left(\mathbb{E}_{v_j} \left[\sum_{S_j : |S_j| > R} \left\| A_{j,S_j} \right\|_2^2 \Pr_u \left[|\pi_j(S_j)| < R^{c_0} \right] \right] \right)^{\frac{1}{2}} + \left(1 - \frac{1}{Q}\right)^{R^{c_0}} \\ \leq \quad & \left(\mathbb{E}_{u,\{v_j\}_{j=1}^Q} \left[\sum_{(S_1,\ldots,S_Q) \in \mathcal{S}_1} \prod_{j=1}^{Q} \left\| A_{j,S_j} \right\|_2^2 \right] \right)^{\frac{1}{2}} + \frac{Q}{R^{\frac{c_0}{2}}} + \left(1 - \frac{1}{Q}\right)^{R^{c_0}} \end{aligned} \tag{31}$$

where the last inequality uses property (b) of Theorem 2.4. Consider a labeling to the LabelCover instance $\mathcal{L}$ given by assigning each vertex $v \in V$ label l_v randomly chosen from a subset $S \subseteq [M]$ sampled with probability $\|A_{v,S}\|_2^2$. A vertex $u \in U$ is labeled by uniformly at random choosing

$(Q-1)$ of its neighbors $v_2, \ldots, v_Q$, random subsets S_j with probability $\|A_{v_j,S_j}\|_2^2$ independently for $j = 2, \ldots, Q$, and assigning a random label from $\bigcup_{j=2}^{Q} \pi_j(S_j)$. From the definition of $\mathcal{S}_1$ in (15) the expected number of edges satisfied by this strategy is at least

$$\frac{1}{Q} \cdot \frac{1}{R} \cdot \frac{1}{RQ} \cdot \mathbb{E}_{u,\{v_j\}_{j=1}^{Q}} \left[\sum_{(S_1,\ldots,S_Q)\in\mathcal{S}_1} \prod_{j=1}^{Q} \|A_{j,S_j}\|_2^2 \right],$$

which by (31) is at least

$$\left(\frac{1}{RQ}\right)^2 \left[\left(\frac{\delta_1^Q \varepsilon^{Q+1}}{Q+1} \right) - \frac{Q}{R^{\frac{c_0}{2}}} - \left(1 - \frac{1}{Q}\right)^{R^{c_0}} \right]^2.$$

For any constant $\varepsilon > 0$, choosing the parameter $R = \text{poly}(1/\varepsilon)$ and $r = \Theta(\log(1/\varepsilon))$ in Theorem 2.4 yields a contradiction to the NO Case of Theorem 2.4.

Ruling out $\varepsilon = (\log n)^{-c}$. Choosing $r = (\log\log N)/4$ in Theorem 2.4 we get that the reduction is of size $n = N^{O(r)} 2^{2^{3r}} \leq N^{O(\log\log N)}$. The soundness of $\mathcal{L}$ is $2^{-\Omega(\log\log N)} = 2^{-\Omega(\log\log n)}$. Combining this with the above analysis in the NO Case, choosing $\varepsilon = (\log n)^{-c}$ and $R = \varepsilon^{-c'}$ for some positive constants $c, c' > 0$ (depending on c_0, Q, ℓ, k and γ_0) we obtain a contradiction to the NO Case of Theorem 2.4.

4 Conclusion

Our work shows that in Qk-uniform k-rainbow colorable hypergraphs $(Q, k \geq 2)$ such that in each hyperedge at most 2ℓ of the colors appear $Q \pm 1$ times and the rest exactly Q times, it is NP-hard to find independent sets of density $> (1 - (\ell+1)/k)$. It is an open (challenging) question to prove the NP-hardness of finding independent sets of arbitrarily small constant density in such hypergraphs. The question of computing independent sets of density > 0.5 in perfectly balanced rainbow colorable hypergraphs is also similarly open.

References

[1] S. Arora, E. Chlamtac, and M. Charikar. New approximation guarantee for chromatic number. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 215–224, 2006.

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

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

[4] P. Austrin, V. Guruswami, and J. Håstad. $(2+\varepsilon)$-sat is NP-hard. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, pages 1–10, 2014.

[5] N. Bansal. Constructive algorithms for discrepancy minimization. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, pages 3–10, 2010.

[6] A. Blum. New approximation algorithms for graph coloring. *Journal of the ACM*, 41(3):470–516, 1994.

[7] A. Blum and D. R. Karger. An $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs. *Information Processing Letters*, 61(1):49–53, 1997.

[8] B. Bollobás, D. Pritchard, T. Rothvoß, and A. D. Scott. Cover-decomposition and polychromatic numbers. *SIAM Journal on Discrete Mathematics*, 27(1):240–256, 2013.

[9] J. Brakensiek and V. Guruswami. The quest for strong inapproximability results with perfect completeness. In *Proc. APPROX-RANDOM*, pages 4:1–4:20, 2017. URL: https://eccc.weizmann.ac.il/report/2017/080.

[10] M. Charikar, A. Newman, and A. Nikolov. Tight hardness results for minimizing discrepancy. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1607–1614, 2011.

[11] H. Chen and A. M. Frieze. Coloring bipartite hypergraphs. In *Proc. IPCO*, pages 345–358, 1996.

[12] I. Dinur and V. Guruswami. PCPs via low-degree long code and hardness for constrained hypergraph coloring. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, 2013.

[13] I. Dinur, O. Regev, and C. D. Smyth. The hardness of 3-uniform hypergraph coloring. *Combinatorica*, 25(5):519–535, 2005.

[14] V. Guruswami, J. Håstad, P. Harsha, S. Srinivasan, and G. Varma. Super-polylogarithmic hypergraph coloring hardness via low-degree long codes. *SIAM Journal of Computing*, 46(1):132–159, 2017.

[15] V. Guruswami, J. Håstad, and M. Sudan. Hardness of approximate hypergraph coloring. *SIAM Journal of Computing*, 31(6):1663–1686, 2002.

[16] V Guruswami and S. Khanna. On the hardness of 4-coloring a 3-colorable graph. *SIAM Journal of Discrete Mathematics*, 18(1):30–40, 2004.

[17] V. Guruswami and E. Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:43, 2014. URL: http://eccc.hpi-web.de/report/2014/043.

[18] V. Guruswami and E. Lee. Strong inapproximability results on balanced rainbow-colorable hypergraphs. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 822–836, 2015.

[19] E. Halperin, R. Nathaniel, and U. Zwick. Coloring k-colorable graphs using relatively small palettes. *J. Algorithms*, 45(1):72–90, 2002.

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

[21] J. Holmerin. Vertex cover on 4-regular hyper-graphs is hard to approximate within 2 - ε. In *Proceedings of the Annual IEEE Conference on Computational Complexity*, 2002.

[22] S. Huang. $2^{(\log n)^{1/10-o(1)}}$ hardness for hypergraph coloring. *CoRR*, abs/1504.03923, 2015.

[23] D. R. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *Journal of the ACM*, 45(2):246–265, 1998.

[24] K. Kawarabayashi and M. Thorup. Combinatorial coloring of 3-colorable graphs. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, pages 68–75, 2012.

[25] P. Kelsen, S. Mahajan, and R. Hariharan. Approximate hypergraph coloring. In *Proc. SWAT*, pages 41–52, 1996.

[26] S. Khanna, N. Linial, and S. Safra. On the hardness of approximating the chromatic number. *Combinatorica*, 20(3):393–415, 2000.

[27] S. Khot. Hardness results for coloring 3-colorable 3-uniform hypergraphs. In *Proceedings of the Annual Symposium on Foundations of Computer Science*, pages 23–32, 2002.

[28] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the ACM Symposium on the Theory of Computing*, pages 767–775, 2002.

[29] S. Khot and R. Saket. Hardness of finding independent sets in 2-colorable and almost 2-colorable hypergraphs. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1607–1625, 2014.

[30] S. Khot and R. Saket. Hardness of coloring 2-colorable 12-uniform hypergraphs with $2^{(\log n)^{\Omega(1)}}$ colors. *SIAM Journal of Computing*, 46(1):235–271, 2017.

[31] M. Krivelevich, R. Nathaniel, and B. Sudakov. Approximating coloring and maximum independent sets in 3-uniform hypergraphs. *Journal of Algorithms*, 41(1):99–113, 2001.

[32] S. Lovett and R. Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal of Computing*, 44(5):1573–1582, 2015.

[33] E. Mossel. Gaussian bounds for noise correlation of functions. *GAFA*, 19:1713–1756, 2010.

[34] E. Mossel, K. Oleszkiewicz, and A. Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013.

[35] R. Raz. A parallel repetition theorem. *SIAM Journal of Computing*, 27(3):763–803, 1998.

[36] R. Saket. Hardness of finding independent sets in 2-colorable hypergraphs and of satisfiable CSPs. In *Proceedings of the Annual IEEE Conference on Computational Complexity*, pages 78–89, 2014.

[37] G. Varma. Reducing uniformity in Khot-Saket hypergraph coloring hardness reductions. *Chicago J. Theor. Comput. Sci.*, 2016, 2016.

[38] A. Wigderson. Improving the performance guarantee for approximate graph coloring. *Journal of the ACM*, 30(4):729–735, 1983.

A Proof of Lemma 3.4

For $r = 1, \ldots, L$, $j^* = 1, \ldots, Q$, let us define the operator $T^{(r)}_{j^*}$ on the space of functions $g(X^{(1)}, \ldots, X^{(Q)})$ where $T^{(r)}_{j^*} g$ is the expectation of g over the rerandomization of $X^{(j^*)}|_{\pi_j^{-1}(r)} \leftarrow \mu_\ell$. Further, let

$$T^{(r)} g := \mathbb{E}_{j^* \in [Q]} \left[T^{(r)}_{j^*} g \right]. \tag{32}$$

Clearly,

$$T^{(r)}_{j^*} \prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) = \left(\prod_{j \in [Q] \setminus \{j^*\}}^{Q} A_{j,S_j}(X^{(j)}) \right) T^{(r)}_{j^*} A_{j^*,S_{j^*}}(X^{(j^*)}). \tag{33}$$

Moreover, from Proposition 2.2,

$$T^{(r)}_{j^*} A_{j^*,S_{j^*}}(X^{(j^*)}) = \begin{cases} 0 & \text{if } r \in \pi_{j^*}(S_{j^*}), \\ A_{j^*,S_{j^*}}(X^{(j^*)}) & \text{otherwise.} \end{cases} \tag{34}$$

Given $\overline{S} = (S_1, \ldots, S_Q)$ and $r \in [L]$ let $q(\overline{S}, r) := |\{j \in [Q] \mid r \in \pi_j(S_j)\}|$. From (32), (33) and (34) we obtain that for a fixed $(X^{(1)}, \ldots, X^{(Q)})$,

$$T^{(r)} \prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) = \left(1 - \frac{q(\overline{S}, r)}{Q}\right) \prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}), \tag{35}$$

and therefore,

$$T^{(1)} \ldots T^{(L)} \prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) = \left(\prod_{r=1}^{L} \left(1 - \frac{q(\overline{S}, r)}{Q}\right) \right) \prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}). \tag{36}$$

Now,

$$\begin{aligned}
&\mathbb{E}_{((X^{(1)},Y^{(1)}),\ldots,(X^{(Q)},Y^{(Q)})) \leftarrow \overline{\mathcal{Q}}} \left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) B_{j,S_j}(Y^{(j)}) \right] \\
&= \mathbb{E}_{\{(X^{(j)},Y^{(j)}) \leftarrow \mu_k^{\otimes L}\}_{j=1}^{Q}} \left[\left(T^{(1)} \ldots T^{(L)} \prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) \right) \prod_{j=1}^{Q} B_{j,S_j}(Y^{(j)}) \right] \\
&= \mathbb{E}_{\{(X^{(j)},Y^{(j)}) \leftarrow \mu_k^{\otimes L}\}_{j=1}^{Q}} \left[\left(\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)}) \right) \prod_{j=1}^{Q} B_{j,S_j}(Y^{(j)}) \right] \left(\prod_{r=1}^{L} \left(1 - \frac{q(\overline{S}, r)}{Q}\right) \right) \text{ (using (36))} \\
&\le \left\| \prod_{j=1}^{Q} A_{j,S_j} \right\|_2 \left\| \prod_{j=1}^{Q} B_{j,S_j} \right\|_2 \left(\prod_{r=1}^{L} \left(1 - \frac{q(\overline{S}, r)}{Q}\right) \right) \text{ (using } \mathbb{E}[fg] \le \|f\|_2 \|g\|_2\text{)} \\
&\le \left\| \prod_{j=1}^{Q} A_{j,S_j} \right\|_2 \left\| \prod_{j=1}^{Q} B_{j,S_j} \right\|_2 \left(1 - \frac{1}{Q}\right)^{\max_j |\pi_j(S_j)|},
\end{aligned}$$

which completes the proof.

B Proofs of Claims 3.6 and 3.7

Proof of Claim 3.6. Consider the term inside the sum on the LHS of (26):

$$\left|\mathbb{E}\left[\prod_{j=1}^{Q} A_{j,S_j}(X^{(j)})B_{j,S_j}(Y^{(j)})\right]\right| \tag{37}$$

Fix S_1 is such that $|\pi_1(S_1)| > R^{c_0}$. The sum over all $(S_2,\ldots,S_Q) \in (2^{[M]})^{(Q-1)}$ of (37) is upper bounded (using Lemma 3.4) by,

$$\left(1-\frac{1}{Q}\right)^{R^{c_0}} \sum_{(S_2,\ldots,S_Q)} \left\|\prod_{j=1}^{Q} A_{j,S_j}\right\|_2 \left\|\prod_{j=1}^{Q} B_{j,S_j}\right\|_2 .$$

Now note that,

$$\begin{aligned}
&\sum_{(S_2,\ldots,S_Q)} \left\|\prod_{j=1}^{Q} A_{j,S_j}\right\|_2 \left\|\prod_{j=1}^{Q} B_{j,S_j}\right\|_2 \\
\leq\ & \left(\sum_{(S_2,\ldots,S_Q)} \left\|\prod_{j=1}^{Q} A_{j,S_j}\right\|_2^2\right)^{\frac{1}{2}} \left(\sum_{(S_2,\ldots,S_Q)} \left\|\prod_{j=1}^{Q} B_{j,S_j}\right\|_2^2\right)^{\frac{1}{2}} && (38)\\
=\ & \left(\sum_{(S_2,\ldots,S_Q)} \mathbb{E}\left[\left|\prod_{j=1}^{Q} A_{j,S_j}\right|^2\right]\right)^{\frac{1}{2}} \left(\sum_{(S_2,\ldots,S_Q)} \mathbb{E}\left[\left|\prod_{j=1}^{Q} B_{j,S_j}\right|^2\right]\right)^{\frac{1}{2}} \\
=\ & \left(\mathbb{E}\left[\sum_{(S_2,\ldots,S_Q)} \left|\prod_{j=1}^{Q} A_{j,S_j}\right|^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[\sum_{(S_2,\ldots,S_Q)} \left|\prod_{j=1}^{Q} B_{j,S_j}\right|^2\right]\right)^{\frac{1}{2}} && (39)\\
=\ & \left(\mathbb{E}\left[\left(A_{1,S_1} \sum_{(S_2,\ldots,S_Q)} \prod_{j=2}^{Q} A_{j,S_j}\right)^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[\left(B_{1,S_1} \sum_{(S_2,\ldots,S_Q)} \prod_{j=2}^{Q} B_{j,S_j}\right)^2\right]\right)^{\frac{1}{2}} && (40)\\
=\ & \left(\mathbb{E}\left[\left(A_{1,S_1} \prod_{j=2}^{Q} A_j\right)^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[\left(B_{1,S_1} \prod_{j=2}^{Q} B_j\right)^2\right]\right)^{\frac{1}{2}} && (41)\\
\leq\ & \left(\mathbb{E}\left[A_{1,S_1}^2\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[B_{1,S_1}^2\right]\right)^{\frac{1}{2}} = \|A_{1,S_1}\|_2 \|B_{1,S_1}\|_2 , && (42)
\end{aligned}$$

where we applied Cauchy-Schwartz in (38) and used that fact that A_j and B_j are $\{0,1\}$ valued functions in (42). To see how (39) equals (40) observe that,

$$\mathbb{E}\left[\left(A_{1,S_1}\sum_{(S_2,\ldots,S_Q)}\prod_{j=2}^{Q}A_{j,S_j}\right)^2\right]-\mathbb{E}\left[\sum_{(S_2,\ldots,S_Q)}\left|\prod_{j=1}^{Q}A_{j,S_j}\right|^2\right]$$
$$= \mathbb{E}\left[A_{1,S_1}^2\right]\sum_{\substack{(S_2,\ldots,S_Q)\\ \neq(S_2',\ldots,S_Q')}}\prod_{j=2}^{Q}\mathbb{E}\left[A_{j,S_j}A_{j,S_j'}\right] = 0,$$

using the independence of $\{X^{(j)}\}_{j=1}^{Q}$ and the orthogonality of $\{A_{j,S}\}_{S\subseteq[M]}$ for each $j\in[Q]$. The same analysis holds for the second product in (39). Thus, the sum of (37) over all $(S_1,\ldots,S_Q)$ such that $|\pi_1(S_1)|>R^{c_0}$ is at most

$$\left(1-\frac{1}{Q}\right)^{R^{c_0}}\sum_{S_1:|\pi_1(S_1)|>R^{c_0}}\|A_{1,S_1}\|_2\,\|B_{1,S_1}\|_2$$
$$\leq \left(1-\frac{1}{Q}\right)^{R^{c_0}}\left(\sum_{S_1:|\pi_1(S_1)|>R^{c_0}}\|A_{1,S_1}\|_2^2\right)^{\frac{1}{2}}\left(\sum_{S_1:|\pi_1(S_1)|>R^{c_0}}\|B_{1,S_1}\|_2^2\right)^{\frac{1}{2}} \tag{43}$$
$$\leq \left(1-\frac{1}{Q}\right)^{R^{c_0}} \tag{44}$$

Summing the above bound for all $j\in[Q]$ for which $\pi_j(S_j)>R^{c_0}$ yields the claim. □

Proof of Claim 3.7. From Corollary 3.5 the LHS of (27) is upper bounded by

$$\sum_{(S_1,\ldots,S_Q)}\prod_{j=1}^{Q}\|A_{j,S_j}\|_2\,\|B_{j,S_j}\|_2\left(1-\frac{1}{Q}\right)^{R^{c_0}}$$
$$\leq \left(\sum_{(S_1,\ldots,S_Q)}\prod_{j=1}^{Q}\|A_{j,S_j}\|_2^2\right)^{\frac{1}{2}}\left(\sum_{(S_1,\ldots,S_Q)}\prod_{j=1}^{Q}\|B_{j,S_j}\|_2^2\right)^{\frac{1}{2}}\left(1-\frac{1}{Q}\right)^{R^{c_0}}$$
$$= \left(\prod_{j=1}^{Q}\left[\sum_{S_j}\|A_{j,S_j}\|_2^2\right]\right)^{\frac{1}{2}}\left(\prod_{j=1}^{Q}\left[\sum_{S_j}\|B_{j,S_j}\|_2^2\right]\right)^{\frac{1}{2}}\left(1-\frac{1}{Q}\right)^{R^{c_0}}$$
$$= \left(\prod_{j=1}^{Q}\|A_j\|_2^2\right)^{\frac{1}{2}}\left(\prod_{j=1}^{Q}\|B_j\|_2^2\right)^{\frac{1}{2}}\left(1-\frac{1}{Q}\right)^{R^{c_0}} \leq \left(1-\frac{1}{Q}\right)^{R^{c_0}}. \tag{45}$$

□

C Proofs of Claims 3.9 and 3.10

Proof of Claim 3.9. The proof is a slight variation to that of Claim 3.6 in Appendix B and proceeds in a similar manner till the bound in (43) except we don't have the $(1-1/Q)^{R^{c_0}}$ factor outside and the inner sum is over S_1 s.t. $|S_1| > R$ and $|\pi_1(S_1)| < R^{c_0}$ yielding an upper bound of

$$\left(\sum_{\substack{S_1 : |S_1| > R \\ |\pi_1(S_1)| < R^{c_0}}} \|A_{1,S_1}\|_2^2 \right)^{\frac{1}{2}},$$

for $j = 1$. Summing this over all $j \in [Q]$ completes the proof of the claim. □

Proof of Claim 3.10. The proof is similar to that of Claim 3.7 in Appendix B except that that it is done separately for each S_1 s.t. $|S_1| > R$ and $|\pi_1(S_1)| < R^{c_0}$, and upon summing over all such S_1 this yields an upper bound of

$$\sum_{\substack{S_1 : |S_1| > R \\ |\pi_1(S_1)| < R^{c_0}}} \|A_{1,S_1}\|_2 \|B_{1,S_1}\|_2 \le \left(\sum_{\substack{S_1 : |S_1| > R \\ |\pi_1(S_1)| < R^{c_0}}} \|A_{1,S_1}\|_2^2 \right)^{\frac{1}{2}},$$

for $j = 1$. Again, summing this over all $j \in [Q]$ completes the proof of the claim. □

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il