

An Optimal Space Lower Bound for Approximating MAX-CUT

Michael Kapralov
EPFL
michael.kapralov@epfl.ch

Dmitry Krachun
University of Geneva
dima@krachun.ru

November 27, 2018

Abstract

We consider the problem of estimating the value of MAX-CUT in a graph in the streaming model of computation. At one extreme, there is a trivial 2-approximation for this problem that uses only $O(\log n)$ space, namely, count the number of edges and output half of this value as the estimate for the size of the MAX-CUT. On the other extreme, for any fixed $\varepsilon > 0$, if one allows $\tilde{O}(n)$ space, a $(1 + \varepsilon)$ -approximate solution to the MAX-CUT value can be obtained by storing an $\tilde{O}(n)$ -size sparsifier that essentially preserves MAX-CUT value.

Our main result is that any (randomized) single pass streaming algorithm that breaks the 2-approximation barrier requires $\Omega(n)$ -space, thus resolving the space complexity of any non-trivial approximations of the MAX-CUT value to within polylogarithmic factors in the single pass streaming model. We achieve the result by presenting a tight analysis of the Implicit Hidden Partition Problem introduced by Kapralov et al. [SODA'17] for an arbitrarily large number of players. In this problem a number of players receive random matchings of $\Omega(n)$ size together with random bits on the edges, and their task is to determine whether the bits correspond to parities of some hidden bipartition, or are just uniformly random.

Unlike all previous Fourier analytic communication lower bounds, our analysis does not directly use bounds on the ℓ_2 norm of Fourier coefficients of a typical message at any given weight level that follow from hypercontractivity. Instead, we use the fact that graphs received by players are sparse (matchings) to obtain strong upper bounds on the ℓ_1 norm of the Fourier coefficients of the messages of individual players using their special structure, and then argue, using the convolution theorem, that similar strong bounds on the ℓ_1 norm are essentially preserved (up to an exponential loss in the number of players) once messages of different players are combined. We feel that our main technique is likely of independent interest.

Contents

1	Introduction	3
2	Communication problem and hard distribution	4
3	Preliminaries	7
3.1	Fourier analysis on the boolean hypercube	7
3.2	Total variation distance	9
3.3	Basic combinatorics of random matchings	9
4	The basic setup, (C, s^*)-bounded sets and their properties	10
4.1	The basic setup	10
4.2	(C, s^*) -bounded sets and their properties	11
5	Technical overview of our analysis	13
5.1	A component growing protocol for DIHP	13
5.2	Combinatorial analysis of the component growing protocol	15
5.3	Overview of general analysis (proof of Theorem 2.1)	17
5.3.1	Second level Fourier spectrum of the component growing protocol vs combinatorial analysis	17
5.3.2	Evolution of Fourier coefficients	19
6	Proof of main theorem (Theorem 2.1)	26
6.1	Evolution of Fourier coefficients	27
6.2	Putting it together	29
7	Proof of main technical lemma (Lemma 6.1)	30
7.1	Bounding mass transfer to low weight Fourier coefficients	33
7.1.1	Bounding S_0, S_1, S_2, S_3 (technical lemmas)	37
7.2	Bounding mass transfer to intermediate weight Fourier coefficients	42
7.2.1	Bounding T_1 and T_2 (technical lemmas)	47
8	Acknowledgements	50
A	Proof of Lemma 4.7	52
B	Proof of Lemma 6.3	54
C	Details omitted from Section 5	57
D	Useful facts	58
D.1	Concentration inequalities	58
D.2	Combinatorics	60
D.3	Inequalities	62
E	Fourier spectrum of the component growing protocol from section 5.1	64

F	Proof of Lemmas 3.3 and 3.4	65
F.1	General bound for the L_1 mass of the Fourier transform	65
F.2	Structure of the Fourier transform of a single player's message	66

1 Introduction

In the MAX-CUT problem an undirected graph is given as input, and the goal is to find a bipartition of the vertices of this graph (or, equivalently, a *cut*) that maximizes the number of edges that cross the bipartition. The size of a MAX-CUT is the number of edges that cross the optimal bipartition. In this paper, we study the space complexity of approximating the MAX-CUT size in an undirected graph in the streaming model of computation. Our main result is a strong lower bound (optimal to within polylogarithmic factors) on the space required for a non-trivial approximation to MAX-CUT size.

Specifically we consider a space bounded algorithm that is presented with a stream of edges of a graph on known vertex set $[n] \triangleq \{1, \dots, n\}$ and is required to output an α -approximation to the size of the maximum cut in the graph. An algorithm that simply counts m , the number of edges in the graph, and reports $m/2$ requires space $O(\log n)$ and produces a 2-approximation to the size of a maximum cut since the MAX-CUT size is at most m and at least $m/2$ in any undirected graph. On the other extreme, for any $\epsilon > 0$, it is possible to maintain a cut-preserving sparsifier of the graph using $\tilde{O}(n)$ space that allows one to recover a $(1 + \epsilon)$ -approximation to the maximum cut value – in fact, one can recover the actual vertex partition as well in this case. Till recently it was open as to whether such good approximations could be obtained with polylogarithmic space. This question was resolved in the negative by [KKS15, KK14]. In particular, [KKS15] showed that any $(2 - \epsilon)$ -approximation algorithm requires $\tilde{\Omega}(\sqrt{n})$ space, and [KKS17] ruled out the possibility of an approximation scheme in $o(n)$ space. This however left open the possibility that a non-trivial approximation (i.e. better than the trivial 2-approximation described above) can be achieved in $o(n)$ space. In this work, we settle this problem by showing that no sublinear space algorithm can achieve a strictly better than 2-approximation to the size of the maximum cut:

Theorem 1.1 *For every $\epsilon \in (0, 99/100)$, every randomized single-pass streaming algorithm that yields a $(2 - \epsilon)$ -approximation to MAX-CUT size must use $n/(1/\epsilon)^{O(1/\epsilon^2)} = \Omega(n)$ space, where n denotes the number of vertices in the graph.*

Our main technical contribution is a nearly optimal lower bound on the communication complexity of the **Implicit Hidden Partition** problem introduced in [KKS17]. The implicit hidden partition problem is a multiple-player communication game where many players are given labellings of sparse subsets of edges and must determine if they are consistent with a bipartition of vertices. This setting is in contrast to the vanilla “hidden partition problem” used in several previous works on MAX-CUT lower bounds, where one player is given a cut in a complete graph on n vertices and the other player is given a labelling of a linear number of edges of the graph where the labelling supposedly indicates whether those edge cross the cut, and the two players have to decide if the labelling is consistent with the given cut. Gavinsky et al [GKK⁺08] gave a tight $\Omega(\sqrt{n})$ lower bound on the one-way communication complexity of this problem, and this was used by [KKS15] for instance to give a $\tilde{\Omega}(\sqrt{n})$ lower bound on the space needed to find a $(2 - \epsilon)$ -approximation of MAX-CUT. Since in the *implicit* hidden partition problem no player has an explicit knowledge of the bipartition, this problem plausibly has a linear ($\Omega(n)$) lower bound on the communication complexity and this is what we prove.

The main technical tool underlying our analysis is a novel and general way of using the **convolution theorem** in Fourier analysis to analyze information conveyed by the combined messages of multiple players (which corresponds to the intersection of their individual messages). While this idea has recently been used for proving lower bounds on the streaming complexity of MAX-CUT [KKS17] and the sketching complexity of subgraph counting [KKP18], in both of these works the convolution theorem is applied in a rather restricted setting. Specifically, the structure of the Fourier transforms of the messages of the players is such that convolution simply amounts to multiplication in Fourier domain, i.e. only a single nonzero contributes to the corresponding sum of Fourier coefficients. In our setting convolving the Fourier transforms

of the players’ messages leads to contributions across different levels of the weight spectrum, and analyzing such processes requires a new technique. Our main insight is the idea of controlling the ℓ_1 norm of the Fourier transform of the intersection of the players’ messages as opposed to the ℓ_2 norm, bounds on which follow more naturally as a consequence of hypercontractivity (note that ℓ_2 bounds on various levels of the Fourier spectrum that follow from the hypercontractive inequality underlie the analysis of the Boolean Hidden Matching problem of [GKK⁺08], as well as recent works on streaming and sketching lower bounds through Fourier analysis [KKS15, KK14, KKS17, KKP18]). Conceptually, the idea of controlling the ℓ_1 norm stems from the fact that since individual players receive parity information of some hidden vector X across edges of a sparse graph (a matching), strong upper bounds on the ℓ_1 norm of the Fourier transform of the corresponding player’s message follow (due to sparsity of the graph), and these ℓ_1 bounds remain approximately preserved when the players’ functions are multiplied (i.e. when the players’ messages are combined).

Related work: The streaming model of computation, formally introduced in the seminal work of [AMS96] and motivated by applications in processing massive datasets, is an extremely well-studied model for designing sublinear space algorithms. The past decade has seen an extensive body of work on understanding the space complexity of fundamental graph problems in the streaming model; see, for instance, the survey by McGregor [McG14]. It is now known that many fundamental problems admit streaming algorithms that only require $\tilde{O}(n)$ space (i.e. they do not need space to load the edge set of the graph into memory) – e.g. sparsifiers [AG09, KL11, AGM12b, KLM⁺14], spanning trees [AGM12a], matchings [AG11, AG13, GKK12, Kap13, GO12, HRVZ15, Kon15, AKLY15], spanners [AGM12b, KW14]. Very recently it has been shown that it is sometimes possible to approximate the *cost* of the solution without even having enough space to load the *vertex set* of the graph into memory (e.g. [KKS14, EHL⁺15, CCE⁺15]). Our work contributes to the study of streaming algorithms, by providing a tight impossibility result for non-trivially approximating MAX-CUT value in $o(n)$ space.

Organization In section 2 we introduce our communication problem (**DIHP**), give a reduction from **DIHP** to MAX-CUT, state our main technical theorem (Theorem 2.1) on the communication complexity of **DIHP** in this section and prove Theorem 1.1 assuming this result. The rest of the paper is devoted to proving Theorem 2.1. Section 3 presents preliminaries on Fourier analysis and basic combinatorics of random matchings. Section 4 presents our basic setup for proving Theorem 2.1, introduces a crucial definition of (C, s^*) -bounded sets, and proves some of their basic properties. Section 5 presents a technical overview of the analysis: we start with a simple component growing protocol for **DIHP** that serves as a good model for our lower bound proof, give a direct analysis for this protocol, state the main technical components of our general lower bound proof and use the simple protocol to illustrate some of the steps. The rest of the paper is devoted to formally proving Theorem 2.1. Section 6 proves Theorem 2.1 assuming a key technical lemma, whose proof is given in Section 7.

2 Communication problem and hard distribution

In this section we introduce a T -player “sequential” communication problem and state our lower bound for this problem. We first describe the general model in which this problem is presented.

We consider a communication problem where T players receive, sequentially, public inputs M_t and private inputs w_t , for $t \in [T]$. Their goal is to compute some joint function $F(M_1, \dots, M_T; w_1, \dots, w_T)$. At stage t , the t th player announces its message $S_t = r_t(M_1, \dots, M_t; S_1, \dots, S_{t-1}; w_t)$ for some function r_t . The message S_T is defined to be the output of the protocol. The complexity of the protocol is the maximum length of the message $\{S_t\}_{t \in [T]}$. We consider the setting where the inputs are drawn from some distribution

μ and the error of the protocol is the probability that its output does not equal $F(M_1, \dots, M_T, w_1, \dots, w_T)$. We now describe the specific communication problem we consider in this work. For a protocol Π we let $|\Pi|$ denote the maximum size of messages posted by players.

Implicit Hidden Partition (IHP) Problem of [KKS17]. We define a parametrized class of problems **IHP**(n, α, T) for positive integers n and T and real $\alpha \in (0, 1/2)$ as follows: **IHP**(n, α, T) is a T -player problem with public inputs $M_t \in \{0, 1\}^{\alpha n \times n}$ being incidence matrices of matchings (so their rows sum to 2 and columns sum to at most 1), and the private inputs are $w_t \in \{0, 1\}^{\alpha n}$. Their goal is to decide the Boolean function $F(M_1, \dots, M_T; w_1, \dots, w_T)$ which is **YES** if and only if there exists $x \in \{0, 1\}^n$ such that $w_t = M_t x \pmod{2}$ for all $t \in [T]$, and **NO** otherwise.

By associating $[n]$ with the vertices of a graph G , we may think of x as a partition (cut) of the graph G whereas the edges are the set $\cup_{t \in [T]} M_t$. The condition $w_t = M_t x$ enforces that an edge crosses the cut if and only if it is labelled 1 in w_t . Thus the communication problem corresponds to asking if there is a cut consistent with the labelling of edges which are partitioned into matchings and presented as such.

Distributional Implicit Hidden Partition (DIHP) Problem. We will work with the following distributional version of **IHP**. We define a distribution \mathcal{D}^Y supported on **YES** instances which is obtained by sampling $X^* \in \{0, 1\}^n$ uniformly, sampling M_t 's independently and uniformly from the set of αn sized matchings on $[n]$ and setting $w_t = M_t X^*$. The distribution \mathcal{D}^N supported mostly on **NO** instances is obtained by sampling M_t 's as above, and w_t 's independently and uniformly from $\{0, 1\}^{\alpha n}$. The final distribution \mathcal{D} is simply $\mathcal{D} = \frac{1}{2}(\mathcal{D}^Y + \mathcal{D}^N)$. We use **DIHP** to denote the distributional **IHP** problem where instances are drawn from \mathcal{D} .

The following theorem is the main technical contribution of the paper:

Theorem 2.1 *There exists a constant $C_0 > 0$ such that for every $\epsilon \in (0, 1)$ and every sufficiently large n every protocol Π for **DIHP**(n, α, T) with $\alpha = 10^{-11}$, $T = 512/(\alpha \epsilon^2)$ that succeeds with probability at least $2/3$ satisfies $|\Pi| \geq n/(C_0/\epsilon)^{C_0/\epsilon^2}$.*

Remark 2.2 *We note that there exists a protocol Π with $|\Pi| = n/c^T$ which solves **DIHP**(n, α, T) for constant α , where $c = c(\alpha) > 1$, which makes our lower bound close to tight (up to the dependence on $1/\epsilon$ in the base of the exponent, which we think can likely be removed by a more careful, but somewhat more complex, analysis). The protocol works as follows. The first player posts bits on the first $s/(1 + \alpha)^T$ edges of the matching M_1 . Then each player posts bits on all edges incident to at least one of the edges which have been revealed previously. In other words, the players keep growing connected components in the graph whose edges they reveal the bits on. It can be shown that with high probability each player will post at most s bits and a cycle will be found thus solving the problem. See Section 5.1 for more details.*

Reduction from DIHP to MAX-CUT. We now give a reduction from **DIHP** to MAX-CUT (see Theorem 2.3 below). Our main result, Theorem 1.1, then follows by putting together Theorem 2.3 and the yet to be proved Theorem 2.1. The rest of the paper is then devoted to proving Theorem 2.1.

We start with the reduction. We show that for every $\epsilon > 0$, if there exists a single-pass streaming algorithm **ALG** that uses space $s = s(n)$ and approximates the MAX-CUT value to within a factor of $(2 - \epsilon)$ with probability at least $9/10$, then instances of **DIHP**(n, α, T) with $\alpha = 10^{-11}$ and $T = 512/(\alpha \epsilon^2)$ can be distinguished with probability at least $2/3$ by a protocol that uses messages of size at most s . This reduction combined with Theorem 2.1, establishes our main result, namely, every single-pass streaming algorithm needs $\Omega(n)$ space to approximate MAX-CUT value to within a factor of $(2 - \epsilon)$ for every fixed $\epsilon > 0$.

The main idea of the reduction is to map instances of **DIHP** to instances of **MAX-CUT** by only considering those edges where the corresponding entry of the w_t vector is 1. This in turn induces a distribution over **YES** and **NO** instances for **MAX-CUT** as below:

YES Let $G'_t = (V, E'_t)$ be the (bipartite) graph obtained by including those edges in E_t that cross the chosen (hidden) bipartition. Let $E' := E'_1 \cup E'_2 \cup \dots \cup E'_T$, and let $G'(V, E')$ denote the final graph.

NO Let $G'_t = (V, E'_t)$ be the graph obtained by including each edge in E_t independently with probability $1/2$. Let $E' := E'_1 \cup E'_2 \cup \dots \cup E'_T$, and let $G'(V, E')$ denote the final graph.

We denote the input distribution defined above by \mathcal{D}_{graph}^Y (**YES** case) and \mathcal{D}_{graph}^N (**NO** case) respectively. Let $\mathcal{D}_{graph} = \frac{1}{2}\mathcal{D}_{graph}^Y + \frac{1}{2}\mathcal{D}_{graph}^N$. We note that the graphs generated by our distribution \mathcal{D}_{graph} are in general multigraphs. We note that the expected number of repeated edges is only $O(1/\epsilon^2)$, and edge multiplicities are bounded by 2 with high probability.

Using this distribution we get

Theorem 2.3 (Reduction from DIHP to MAX-CUT) *For every $\epsilon, \alpha \in (n^{-1/10}, 1)$ and $T = 512/(\alpha\epsilon^2)$, the following conditions hold for sufficiently large n . Let **ALG** denote a (possibly randomized) single-pass streaming algorithm for approximating **MAX-CUT** value in (multi)graphs to within a factor of $(2 - \epsilon)$ using space $s = s(n)$ on graphs on n nodes with probability at least $9/10$. Then **ALG** can be used to obtain a deterministic protocol Π for the **DIHP**(n, α, T) with $|\Pi| \leq s$ that succeeds with probability at least $2/3$ over the randomness of the input. This holds even if **ALG** is only required to work on multigraphs that contain at most $O(1/\epsilon^2)$ repeated edges, and edge multiplicity is bounded by 2.*

The proof relies on the following Lemma, which establishes that there is almost a factor of 2 gap between **MAX-CUT** value in \mathcal{D}_{graph}^Y and \mathcal{D}_{graph}^N :

Lemma 2.4 *For every $\epsilon, \alpha \in (n^{-1/10}, 1)$, $\alpha < 1/4$ and $T = 512/(\alpha\epsilon^2)$ if $G'_T = (V, E'_T)$, $|V| = n$, $|E'| = m$ be generated according to the process above, then for sufficiently large n there exists $m_0 = m_0(n, \alpha, T)$ such that in the **YES** case the **MAX-CUT** value is at least m_0 , and in the **NO** case the **MAX-CUT** value is at most $m_0/(2 - \epsilon)$, both with probability at least $1 - 1/\sqrt{n}$.*

The proof of the lemma uses the following version of Chernoff bounds.

Lemma 2.5 *Let $X = \sum_{i=1}^n X_i$, where X_i are Bernoulli 0/1 random variables satisfying, for every $k \in [n]$, $\mathbb{E}[X_k | X_1, \dots, X_{k-1}] \leq p$ for some $p \in (0, 1)$. Let $\mu = np$. Then for all $\Delta > 0$*

$$\Pr[X \geq \mu + \Delta] \leq \exp\left(-\frac{\Delta^2}{2\mu + 2\Delta}\right).$$

The (somewhat standard) proof is given in Appendix D. We also need

Lemma 2.6 *Let G be a multigraph with n vertices and m edges (counted with multiplicities) in which each edge has multiplicity at most k . Let $S \subset [n]$ be a uniformly random subset of vertices and X be the number of edges crossing (S, \bar{S}) . Then for any $\delta > 0$ we have*

$$\Pr[X < m/2 \cdot (1 - \delta)] \leq \frac{k}{\delta^2 m}.$$

The proof is a simple application of Chebyshev’s inequality, and is presented in Appendix D. We now give **Proof outline of Lemma 2.4** We let $m_0 = \frac{\alpha n T}{2} \cdot (1 - \delta)$ with $\delta = \epsilon/100$. In the **YES** case the graph is bipartite so the value of MAX-CUT is equal to the number of edges in the graph. Since in the **YES** case we only keep those edges of the matchings which cross a fixed random bipartition, and in the union of matchings every edge has multiplicity at most T , Lemma 2.6 ensures that the probability that the number of edges is smaller than m_0 is at most

$$\frac{T}{\delta^2 \alpha n T} = \frac{1}{\delta^2 \alpha n} \leq 1/\sqrt{n},$$

since $\epsilon, \alpha > n^{-1/10}$ by assumption of the lemma.

We now consider the **NO** case. Since every edge of the matchings is kept with probability $1/2$ independently of the others, by Lemma 2.5 with probability at least $1 - \exp\left(-\frac{\delta^2 \alpha n T}{4(1+\delta)}\right)$ the number m of edges in the graph will be at most $\frac{\alpha n T}{2} \cdot (1 + \delta)$. One then shows using Lemma 2.5 that for every cut (S, \bar{S}) , where $S \subseteq V$, the probability that significantly more than half of the edges of the graph cross the cut is smaller than 2^{-2n} . Taking a union bound over all 2^n possible cuts completes the proof. The detailed proof is provided in Appendix D. ■

Equipped with Lemma 2.4, we can now give a proof of the reduction:

Proof of Theorem 2.3: By Lemma 2.4 in the **YES** case the MAX-CUT value is at least m_0 , and in the **NO** case the MAX-CUT value is at most $m_0/(2 - \epsilon)$, both with probability at least $1 - 1/\sqrt{n}$.

Thus with high probability, the MAX-CUT value in a **YES** instance of **DIHP** is at least $(2 - \epsilon)$ times the MAX-CUT value in a **NO** instance of **DIHP**. To complete the reduction, it now suffices to show that the algorithm **ALG** can be simulated by a **DIHP** protocol with message complexity at most s . This simulation can be done as follows: player t upon receiving its input (M_t, w_t) , runs **ALG** using the state posted by player $(t - 1)$ on the set of edges in M_t where the w_t value is 1. The state of the algorithm **ALG** at the end is then posted by player t on the board. The last player then outputs **YES** if **ALG** outputs **YES** and **NO** otherwise. Note that since we are evaluating the resulting protocol with respect to an input distribution, it is possible to make the protocol deterministic by fixing the randomness of **ALG** appropriately. ■

Given Theorem 2.3 and assuming Theorem 2.1, our main theorem follows easily.

Proof of Theorem 1.1: The proof follows by putting together Theorem 2.3 and Theorem 2.1. ■

3 Preliminaries

In this section we first review Fourier analysis on the boolean hypercube and give a version of the KKL bound that will be important in our analysis(Section 3.1), give basic facts about the total variation distance between distributions (Section 3.2) and state basic bounds on uniformly random matchings that we will use (Section 3.3).

3.1 Fourier analysis on the boolean hypercube

Let $p : \{0, 1\}^n \rightarrow \mathbb{R}$ be a real valued function defined on the boolean hypercube. We use the following normalization in the Fourier transform:

$$\hat{p}(v) = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} p(x) \cdot (-1)^{x \cdot v}.$$

With this normalization the inverse transform is given by

$$p(x) = \sum_{v \in \{0, 1\}^n} \hat{p}(v) \cdot (-1)^{x \cdot v}.$$

For a pair of functions $p, q : \{0, 1\}^n \rightarrow \mathbb{R}$ the convolution of p and q , denoted by $p * q$ is defined as

$$(p * q)(v) = \sum_{x \in \{0, 1\}^n} p(x)q(x + v).$$

We will use the relation between multiplication of functions in time domain and convolution in frequency domain to analyze the Fourier spectrum of $h_t = f_1 \cdot f_2 \cdot \dots \cdot f_t$ (recall that h_t is the indicator of \mathbf{B}_t as per Definition 4.1). With our normalization of the Fourier transform the convolution identity is

$$\widehat{(p \cdot q)}(v) = \sum_{x \in \{0, 1\}^n} \widehat{p}(x)\widehat{q}(x + v). \quad (1)$$

Thus, for each $t = 1, \dots, T$ we have that

$$\widehat{h}_t = \widehat{f}_1 * \dots * \widehat{f}_t.$$

This identity will form the basis of our proof.

We will also need Parseval's equality, which with our normalization takes form

$$\|\widehat{p}\|^2 = \sum_{v \in \{0, 1\}^n} \widehat{p}(v)^2 = \left(\frac{1}{2^n} \sum_{x \in \{0, 1\}^n} p(x) \cdot (-1)^{x \cdot v} \right)^2 = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} p(x)^2 = \frac{1}{2^n} \|p\|^2 \quad (2)$$

Remark 3.1 If $f(x) : \{0, 1\}^n \rightarrow \{0, 1\}$ is an indicator of a set $\mathbf{A} \subseteq \{0, 1\}^n$, we have $\|f\|^2 = |\mathbf{A}|$, so that $\|\widehat{f}\|^2 = \frac{|\mathbf{A}|}{2^n}$.

Definition 3.2 For a function $h : \{0, 1\}^n \rightarrow \{0, 1\}$ that is the indicator of a set $\mathbf{A} \subseteq \{0, 1\}^n$ we write \widetilde{h} to denote the Fourier transform of h scaled by $2^n/|\mathbf{A}|$. Specifically, for every $v \in \{0, 1\}^n$ we have

$$\widetilde{h}(v) = \frac{2^n}{|\mathbf{A}|} \widehat{h}(v) = \frac{1}{|\mathbf{A}|} \sum_{x \in \{0, 1\}^n} h(x) \cdot (-1)^{x \cdot v} = \mathbb{E}_{x \in \mathbf{A}} [(-1)^{x \cdot v}].$$

The following two important Lemmas will be proved in Appendix F.

Lemma 3.3 Let $A \subset \{0, 1\}^m$ be a set of cardinality at least 2^{m-d} with indicator function f . Then for every $y \in \{0, 1\}^m$ and every $q \leq d$ one has

$$\sum_{\substack{x \in \{0, 1\}^m \\ |x \oplus y| = q}} |\widetilde{f}(x)| \leq \sqrt{\binom{m}{q} \left(\frac{4d}{q}\right)^q},$$

where for $x \in \{0, 1\}^m$ by $|x|$ we denote the number of ones in x .

Lemma 3.4 Let $M_t \in \{0, 1\}^{\alpha n \times n}$ be the incidence matrix of a matching M , where the rows correspond to edges e of M ($M_{eu} = 1$ if e is incident on u and 0 otherwise). Let $g : \{0, 1\}^{\alpha n} \rightarrow \{0, 1\}^s$ for some $s > 0$. Let $a \in \{0, 1\}^s$ and let $q : \{0, 1\}^{\alpha n} \rightarrow \{0, 1\}$ be the indicator of the set $\mathbf{A}_{\text{reduced}} := \{z \in \{0, 1\}^{\alpha n} : g(z) = a\}$. Further, let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ denote the indicator of the set

$$\mathbf{A}_{\text{full}} := \{x \in \{0, 1\}^n : g(Mx) = a\}.$$

Then for any $v \in \{0, 1\}^n$

$$\hat{f}(v) = \begin{cases} 0, & \text{if } v \text{ cannot be perfectly matched via edges of } M \\ \hat{q}(w), & w \text{ the perfect matching of } v \text{ using edges of } M \text{ o.w.} \end{cases} \quad (3)$$

Furthermore, the perfect matching of v , when it exists, is unique. The second condition above is equivalent to the existence of $w \in \{0, 1\}^{\alpha n} = \{0, 1\}^M$ such that $v = M^T w$. Thus, Fourier coefficients of f are indexed by sets of edges of M . Note that nonzero weight k coefficients of \hat{q} are in one to one correspondence with nonzero weight $2k$ coefficients of \hat{f} , i.e. the only nonzero Fourier coefficients of \hat{f} are of the form $\hat{f}(M^T w) = \hat{q}(w)$ for some $w \in \{0, 1\}^M$.

3.2 Total variation distance

We define the notion of total variation distance between two distributions and state some of its useful properties. For a random variable X taking values on a finite sample space Ω we let $p_X(\omega), \omega \in \Omega$ denote the pdf of X . For a subset $A \subseteq \Omega$ we use the notation $p_X(A) := \sum_{\omega \in A} p_X(\omega)$. We will use the total variation distance $\|\cdot\|_{tvd}$ between two distributions:

Definition 3.5 (Total variation distance) Let μ, ν be two probability measures on a finite space Ω . The total variation distance between μ and ν is given by $V(\mu, \nu) = \max_{\Omega' \subseteq \Omega} (\mu(\Omega') - \nu(\Omega')) = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|$.

Definition 3.6 Let X, Y be two random variables taking values on a finite domain Ω . We denote the pdfs of X and Y by p_X and p_Y respectively. The total variation distance between X and Y is defined to be the total variation distance between p_X and p_Y . We will write $\|X - Y\|_{tvd}$ to denote the total variation distance between X and Y .

We will need the following claim.

Claim 3.7 (Claim 6.5 of [KKS15]) Let X, Y be two random variables. Let W be independent of (X, Y) . Then for any function f one has $\|f(X, W) - f(Y, W)\|_{tvd} \leq \|X - Y\|_{tvd}$.

3.3 Basic combinatorics of random matchings

Here we define several quantities related to random matchings and state Lemmas concerning random matchings. Proofs are given in Appendix D.2.

Definition 3.8 For every $\alpha \in (0, 1)$, every integer ℓ between 0 and $n/2$ we let $p_\alpha(\ell, n)$ denote the probability, over the choice of a uniformly random matching M of size αn , that a fixed set of 2ℓ points is perfectly matched by M (i.e., M restricted to this set is a perfect matching).

Definition 3.9 Let A be a set of cardinality $2k$. Then $q(k, i, n)$ is the probability that a random matching of size αn matches exactly $2i$ out of $2k$ points of A and there are no edges of the matching connecting points from A to points from A^c .

Definition 3.10 Let A be a set of cardinality $2k$. Then $q(k, i, b, n)$ is the probability that a random matching of size αn matches exactly $2i$ out of $2k$ points of A and there are exactly b edges of the matching connecting points from A to points from A^c .

Definition 3.11 For a set A and a matching M we say that an edge e of M is inner if it connects points from A , boundary if it connects a point from A and a point from A^c , external if it connects points from A^c .

Lemma 3.12 For every integer n and every $0 \leq \ell \leq n/2$

$$p(\ell, n) = \binom{\alpha n}{\ell} \binom{n}{2\ell}^{-1}.$$

Lemma 3.13 For every integer n and every $0 \leq i \leq k \leq n/2$

$$q(k, i, n) = \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k - i)} \binom{n}{2k}^{-1}.$$

Lemma 3.14 For all non-negative integers n, i, k, b satisfying $k \leq n/2$ and $2i + b \leq 2k$

$$q(k, i, b, n) = \binom{\alpha n}{i} \binom{\alpha n - i}{b} 2^b \binom{n - 2\alpha n}{2(k - i) - b} \binom{n}{2k}^{-1}.$$

Lemma 3.15 Suppose we have $k < n/10$, $\alpha < 1/100$, and $2i + b \leq 2k$ then

$$q(k, i, b, n) \leq q(k, i, n) 20^{-b} 4^{k-i}.$$

4 The basic setup, (C, s^*) -bounded sets and their properties

In this section we introduce our basic setup for analyzing the **DIHP** problem, define the notion of (C, s^*) -bounded sets and introduce some of their basic properties.

4.1 The basic setup

For a random variable J_t we write $J_{1:t}$ to denote the tuple (J_1, \dots, J_t) . In this notation, the inputs to **IHP**(n, α, T) are denoted by $M_{1:T}$ and $w_{1:T}$. Recall also that $S_t = r_t(M_{1:t}; S_{1:t-1}; w_t)$ denotes the message posted by the t th player. We use s to denote an upper bound on the size of the messages. The goal of our analysis is to show that if $s \ll n$ then the total variation distance between the distribution of messages S_t and matchings M_t posted on the blackboard at time T in the **YES** case and in the **NO** case is small.

More specifically, let $S_{1:T}^Y$ denote the random variables corresponding to the messages posted by the players when the input $(M_{1:T}, w_{1:T})$ is drawn from \mathcal{D}^Y , and let $S_{1:T}^N$ denote the corresponding sequence when the input is drawn from \mathcal{D}^N . Our goal is to show that the total variation distance between $(M_{1:T}, S_{1:T}^Y)$ and $(M_{1:T}, S_{1:T}^N)$, is vanishingly small. As we show in section 6 it suffices to consider the **YES** case only. In Lemma 6.3 we show that it suffices to show that with high probability for each $t = 1, \dots, T$ the distribution of $M_t X^*$ is close to uniform in $\{0, 1\}^{M_t}$. We now outline the techniques that we develop to prove this claim.

Our analysis relies on Fourier analytic techniques for reasoning about the distribution of $M_t X^*$. Conditioning on messages posted up to time t makes X^* uniformly random over a certain subset of the binary cube. We will analyze this subset of the hypercube, or, rather, the Fourier transform of its indicator function, and show that if communication is small, the distribution of X^* conditional on typical history is such that $M_t X^*$ is close to uniformly random in total variation distance.

We first define notation that lets us reason about the distribution of X^* . The knowledge that the players acquire about X^* is represented by some set $\mathbf{B}_t \subseteq \{0, 1\}^n$ with X^* being distributed uniformly on \mathbf{B}_t . These sets \mathbf{B}_t are constructed iteratively, by accumulating the information about X^* that each player's message conveys - specifically there exist sets \mathbf{A}_t (defined below) such that $\mathbf{B}_t = \mathbf{A}_1 \cap \mathbf{A}_2 \cap \dots \cap \mathbf{A}_t$. Recall that we only consider the **YES** case, so the state variables are superscripted accordingly.

Definition 4.1 (Sets $\mathbf{A}_t, \mathbf{B}_t$ and their indicator functions f_t, h_t) Fix $\alpha \in (0, 1)$ and integers $n, T \geq 1$ and $t \in [T]$. Consider a **YES** instance $(M_{1:T}, w_{1:T})$ of **DIHP** (n, α, T) with X^* being the (random) hidden partition (so $w_t = M_t X^*$). Recall $S_t^Y = r_t(M_{1:t}; S_{1:t-1}, M_t X^*)$. We define $\mathbf{A}_{\text{reduced}, t} \subseteq \{0, 1\}^{M_t}$ be the set of possible values of $w_t = M_t X^*$ that lead to message S_t^Y and \mathbf{A}_t to denote the values of $X^* \in \{0, 1\}^n$ that correspond to $\mathbf{A}_{\text{reduced}, t}$. Formally, letting $g_t(\cdot) := r_t(M_{1:t}, S_{1:t-1}^Y, \cdot) : \{0, 1\}^{M_t} \rightarrow \{0, 1\}^s$, we have

$$\mathbf{A}_{\text{reduced}, t} = g_t^{-1}(S_t^Y) \subseteq \{0, 1\}^{M_t},$$

$$\text{and } \mathbf{A}_t = \{x \in \{0, 1\}^n : M_t x \in \mathbf{A}_{\text{reduced}, t}\}. \quad (4)$$

Let $f_t : \{0, 1\}^n \rightarrow \{0, 1\}$ denote indicator functions of the sets $\mathbf{A}_t, t = 1, \dots, T$. For each t let $h_t = f_1 \cdot \dots \cdot f_t$, so that h_t is the indicator of $\mathbf{B}_t := \mathbf{A}_1 \cap \mathbf{A}_2 \cap \dots \cap \mathbf{A}_t$. We let $\mathbf{B}_0 := \{0, 1\}^n$ for convenience.

Claim 4.2 For $t = 1, \dots, T$ let $p_t : \{0, 1\}^n \rightarrow [0, 1]$ denote the following distribution:

$$p_t(x) = \Pr[M_t x \in \mathbf{A}_{t, \text{reduced}}].$$

Further, let $p(x) = \prod_{t=1}^T p_t(x)$ for all $x \in \{0, 1\}^n$.

The conditional distribution of X^* given messages $\mathbf{A}_1, \dots, \mathbf{A}_t$ is exactly given by $p(x)/\|p\|_1$ as above.

Proof: Since X^* is uniformly random in $\{0, 1\}^n$ initially, we have

$$\begin{aligned} & \Pr[X^* = x | M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T] \\ &= \frac{\Pr[X^* = x \wedge M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T]}{\Pr[M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T]} \\ &= \frac{\Pr[M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T | X^* = x] \Pr[X^* = x]}{\Pr[M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T]} \\ &= \frac{\Pr[X^* = x]}{\Pr[M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T]} \cdot \prod_{t=1}^T \Pr[M_t x \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T] \\ &= \frac{2^{-n}}{\Pr[M_t X^* \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T]} \cdot \prod_{t=1}^T \Pr[M_t x \in \mathbf{A}_{t, \text{reduced}}, t = 1, \dots, T] \\ &= p(x)/\|p\|_1 \end{aligned}$$

■

4.2 (C, s^*) -bounded sets and their properties

The following definition is crucial for our analysis:

Definition 4.3 ((C, s^*) -bounded set) Let $\mathbf{B} \subset \{0, 1\}^n$ with indicator function h . We say that \mathbf{B} (or h) is (C, s^*) -bounded if

- For all $\ell \leq s^*$ we have

$$\sum_{\substack{v \in \{0, 1\}^n \\ |v|=2\ell}} |\widetilde{\mathbb{1}}_{\mathbf{B}}(v)| \leq \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell;$$

- For all $s^* < \ell < \frac{n}{C^2}$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}}_{\mathbf{B}}(v) \right| \leq \left(\frac{C^2 n}{\ell} \right)^{\ell/2}.$$

Defining the function $\mathbf{U}_{C,s^*}(\ell)$ by

$$\mathbf{U}_{C,s^*}(\ell) = \begin{cases} 1 & \ell = 0; \\ \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell & \ell \in [1 : s^*]; \\ \left(\frac{C^2 n}{\ell} \right)^{\ell/2} & \ell > s^*, \end{cases} \quad (5)$$

we are able to simplify notation, ensuring that an indicator function h is (C, s^*) bounded if and only if for all $\ell < n/C^2$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{h}(v) \right| \leq \mathbf{U}_{C,s^*}(\ell).$$

Remark 4.4 For intuition it is useful to compare our bounds throughout the paper to the bounds on the weight of Fourier coefficients of the simple adaptive component growing algorithm described in Section 5.1 when the latter is given a somewhat larger space budget, specifically $s' := \sqrt{s^*n}$ amount of space.

The following lemma provides the base case of our analysis:

Lemma 4.5 For every s^* , every matching M on $[n]$, every $\mathbf{A} \subseteq \{0, 1\}^M$ such that $|\mathbf{A}|/2^{|\mathbf{A}|} \geq 2^{-s^*}$ the set $\mathbf{B} := \{x \in \{0, 1\}^n : Mx \in \mathbf{A}\}$ is $(3, s^*)$ -bounded.

Proof: By Lemma 3.4 Fourier coefficients of $\mathbb{1}_{\mathbf{B}}$ can be written in terms of those of $\mathbb{1}_{\mathbf{A}}$, namely, we have

$$\widehat{\mathbb{1}}_{\mathbf{B}}(M^T w) = \widehat{\mathbb{1}}_{\mathbf{A}}(w),$$

for every $w \in \{0, 1\}^M$ and $\widehat{\mathbb{1}}_{\mathbf{B}}(v) = 0$ if v is not of the form $M^T w$. Also note that $|M^T w| = 2 \cdot |w|$ and $|\mathbf{B}| = 2^{n-\alpha n} \cdot |\mathbf{A}|$. The latter implies that we also have for every $w \in \{0, 1\}^M$

$$\widetilde{\mathbb{1}}_{\mathbf{B}}(M^T w) = \frac{2^n}{|\mathbf{B}|} \cdot \widehat{\mathbb{1}}_{\mathbf{B}}(M^T w) = \frac{2^{\alpha n}}{|\mathbf{A}|} \cdot \widehat{\mathbb{1}}_{\mathbf{A}}(w) = \widetilde{\mathbb{1}}_{\mathbf{A}}(w).$$

We then apply Lemma 3.3 to infer

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}}_{\mathbf{B}}(v) \right| = \sum_{\substack{w \in \{0,1\}^M \\ |w|=\ell}} \left| \widetilde{\mathbb{1}}_{\mathbf{A}}(w) \right| \leq \sqrt{\binom{\alpha n}{\ell} \left(\frac{4s^*}{\ell} \right)^\ell} \leq \left(\frac{3\sqrt{s^*n}}{\ell} \right)^\ell. \quad (6)$$

For high weights ($\ell > s^*$) we use a trivial bound saying that sum of squares of normalized Fourier coefficients at any level is at most 2^{s^*} . We also use the fact that by Lemma 3.4 the function $\mathbb{1}_{\mathbf{B}}$ has at most $\binom{\alpha n}{\ell}$ non-zero Fourier coefficients at level 2ℓ .

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}}_{\mathbf{B}}(v) \right| \leq \sqrt{\binom{\alpha n}{\ell} 2^{s^*}} \leq \left(\frac{2en}{\ell} \right)^{\ell/2}.$$

■

Remark 4.6 (Intuition for the choice of $\mathbb{U}_{C,s^*}(\ell)$) Lemma 4.5 basically states that the set of possible $x \in \{0, 1\}^n$ consistent with a message of one player is $(3, s^*)$ bounded. This Lemma is never explicitly used in the proof of Theorem 2.1, however, it provides a good intuition for why the definition of (C, s) boundedness should be as it is. Indeed, the fact that in the lemma above Fourier coefficients of $\mathbb{1}_{\mathbf{B}}$ at level 2ℓ correspond to Fourier coefficients of $\mathbb{1}_{\mathbf{A}}$ at level ℓ allows to write

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widetilde{\mathbb{1}_{\mathbf{B}}}^2(v) \leq \left(\frac{4s^*}{\ell}\right)^\ell. \quad (7)$$

In order to convert this into a bound on the L_1 norm we need to know how many non-zero summands we have on the left hand side. For one matching the correspondence with Fourier coefficients of $\mathbb{1}_{\mathbf{A}}$ readily implies that we have at most $\binom{\alpha n}{\ell}$ non-zero summands. When several messages of players are combined this is no longer true and the number of non-zero summands can get as large as $\binom{n}{2\ell}$, however, it turns out that L_1 norm behaves as if there were order $\binom{n}{\ell}$ non-zero summands.

Lemma 4.7 For every $C > 100$, $\delta \in (n^{-1}, 1/2)$ and $\alpha \in (0, 1/100)$ the following condition holds if n is sufficiently large. Let $\mathbf{B} \subset \{0, 1\}^n$ be (C, s^*) -bounded for $s^* \in [10 \ln(n+1), \delta^4 n/C^2]$ as per Definition 4.3, $|\mathbf{B}|/2^n \geq 2^{-s^*}$, and let $h : \{0, 1\}^n \rightarrow \{0, 1\}$ be the indicator of \mathbf{B} . Let M be a uniformly random matching on $[n]$ of size αn . Then with probability at least $1 - \delta$ over the choice of M for any non-negative function q on $\{0, 1\}^M$ one has

$$1 - \delta \leq \frac{\mathbb{E}_{x \sim \text{Uniform}(\mathbf{B})} [q(Mx)]}{\mathbb{E}_{z \sim \text{Uniform}(\{0,1\}^M)} [q(z)]} \leq 1 + \delta.$$

The proof is given in Appendix A.

Remark 1 We note that the proof of Lemma 4.7 does not use the bound on the spectrum of h at levels above s^* that follow from (C, s^*) -bounded property.

Remark 2 Using this lemma we can now work with the uniform measure on $\{0, 1\}^{M_t}$ instead of the one given by $M_t h_{t-1}$. For instance, we may assume that each player receives such bits on M_t that the corresponding function f_t and the part of the cube A_t satisfy $|A_t| > 2^{n-s^*}$ (as the opposite happens with probability at most 2^{-s}).

5 Technical overview of our analysis

In this section we define a simple protocol for $\mathbf{DIHP}(n, \alpha, T)$ and analyze its Fourier spectrum. We show that our protocol does not solve $\mathbf{DIHP}(n, \alpha, T)$ for constant $\alpha \in (0, 1)$ and $T \geq C_1/\alpha$ with a sufficiently large constant $C_1 > 1$, unless the communication budget is at least n/A^T for some constant $A > 1$. We also non-rigorously explain why this protocol solves the problem if the communication budget is at least n/c^T for some constant $c > 1$. Our main theorem (Theorem 2.1) in particular implies that this protocol is close to optimal. More importantly, however, the proof of Theorem 2.1 can be viewed as analyzing an ℓ_1 -relaxation of the simple protocol.

5.1 A component growing protocol for DIHP

In this section we consider very simple communication protocols, where players choose a subset of the edges of the matching that they receive, post the bits on that subset on the board, and ignore the bits on the

other edges (note that the subset of edges should be no larger than the communication bound s). We note that in order to achieve a constant advantage over random guessing for **DIHP** it suffices to ensure that the set of edges whose labels are posted on the board **form a cycle**. Indeed, in the **YES** case the sum of labels over the edges of any cycle is zero, while in the **NO** case it is a uniformly random number in $\{0, 1\}$. Thus, one can distinguish between the two cases with constant probability.

A very simple non-adaptive protocol. The question is of course which subset of edges the players choose. A very basic approach would be to post bits on edges both of whose endpoints have indices between 1 and $\sqrt{s/n} \cdot n$. Note that the expected number of such edges in a given matching is $O(s)$, which is consistent with the communication budget per player. This protocol is non-adaptive in the sense that the edges that the t -th player posts the bits for are independent of the matchings of the other players. It is easy to see that this protocol will not find a cycle as long as $s \ll n/T^2$. Indeed, the graph induced on the first $\sqrt{s/n} \cdot n$ vertices is quite similar to an Erdős-Rényi graph with expected degree $\sqrt{s/n} \cdot T \ll 1$, and hence the graph will contain no cycle with high (constant) probability. Note that this behaviour seems to suggest that Theorem 1.1 could be strengthened significantly, ensuring that the dependence of space on T , the number of matchings, is polynomial as opposed to exponential. However, it turns out that the simple *non-adaptive* protocol above is not a good model for the problem. We now introduce a more interesting, but still quite simple, protocol, which serves as a good model for our general analysis.

The component growing protocol. Let s be the per player communication budget. In order to show that even **adaptive** protocols cannot solve **DIHP** unless $s \gg n/C^T$ for a constant $C > 1$, for every such protocol Π we define a strictly stronger protocol Π' that has a larger communication budget than Π and is more convenient to analyze. This new protocol Π' can be thought of as having budget larger than s , is still unable to solve the problem if $s \ll n/A^T$ for some large $A > 1$. The protocol Π' is defined as follows. The first player simply posts the bits according to the protocol Π . We let F_1 denote the forest created in this way, with edges labeled by numbers in $\{0, 1\}$. For every $t = 2, \dots, T$, the t -th player posts the bits on edges of M_t that have at least one endpoint in a connected component in F_{t-1} (these bits should be thought of as free), as well as the bits on a set of s edges of the matching M_t that do not intersect any component in F_{t-1} so that any edge revealed by Π is also revealed by Π' . Let F_t denote the forest obtained in this way. If at least one of the edges of M_t closes a cycle (i.e. edges of $F_{t-1} \cap M_t$ form a cycle), the t -th player computes the sum of bits on the cycle, and outputs **YES** if that sum is zero, and **NO** otherwise. Note that if the game is in the **YES** case, the players always say **YES**. If the game is in the **NO** case, then if a cycle is found, the players say **NO** with probability $1/2$. Thus, the players obtain advantage $1/4$ over random guessing for **DIHP**. We note that every player posts s bits, and some number of bits (those that intersect existing components in F_{t-1}) are revealed for free, so the communication cost of this protocol is at least s bits per player. We show below that this simple protocol does not find a cycle, and hence does not solve **DIHP**(n, α, T) with high constant probability unless s is close to n .

At the same time, we note that adaptive protocols are quite powerful: they can solve **DIHP**(n, α, T) as long as their communication budget s exceeds n/c^T for some $c > 1$. An example such protocol works as follows. The first player posts $s/2$ bits on arbitrary edges of the matching M_1 . The subsequent players then maintain a collection of connected components C_1, \dots, C_k formed by the edges whose bits are posted (we call such edges *revealed*). At each step a new player reveals edges incident to at least one of the connected components. We then remove the smallest components from the list so that the total size of all components remains at most $s/2$. Note that this ensures that every player posts at most s bits. After a new player reveals edges, most of the components will increase its size by a factor of $(1 + \alpha)$ with high probability. Thus, at each step the average component size will be multiplied by $(1 + \alpha)$ whereas the total size of all components will stay close to s . Note that deleting the smallest components cannot decrease the average component

size. After $T - 1$ steps we will get approximately s/c^T connected components each of size c^T . One can see that if $s \cdot c^T \gg n$ then with high probability one of the edges of M_t will have both endpoints in the same connected component and thus we will solve the problem.

In the rest of this section we first show (Section 5.2) directly that even the powerful protocol Π' (with free edges) defined above still cannot solve **DIHP**(n, α, T) with constant $\alpha \in (0, 1)$ unless $s \gg n/C^T$ for some absolute constant $C > 1$. This analysis is quite simple, but does not generalize to arbitrary protocols. We then illustrate our analysis of general protocols by instantiating the relevant parts of the analysis for the simple protocol above and proving that it does not solve **DIHP**(n, α, T) unless $s \gg n/T^T$ (note the slightly weaker bound) by analyzing its Fourier spectrum in Section 5.3.

5.2 Combinatorial analysis of the component growing protocol

Our analysis uses a potential function defined on the forest F maintained by the protocol. The potential function is simply the sum of squares of sizes of (nontrivial) connected components of F . We refer to this quantity as the weight of F , or $\|F\|$:

Definition 5.1 For a forest F with non-trivial(size more than 1) connected components C_1, \dots, C_k we define its weight by $\|F\| := \sum_{i=1}^k |C_i|^2$.

For $t = 1, \dots, T$ let F_t denote the forest computed by players $1, 2, \dots, t$. Recall that in order to distinguish between the **YES** and **NO** cases using our simple protocol the players must ensure that at least one of the matchings M_t contains an edge with both endpoints inside one of the connected components of F_{t-1} . Let C_1, \dots, C_k denote the collection of connected components in F_{t-1} , and note that for a uniformly random edge $e = (u, v)$ one has

$$\Pr[\exists j = 1, \dots, k \text{ s.t. } u, v \in C_j] \leq \sum_{j=1}^k |C_j|^2/n^2, \quad (8)$$

and hence, taking a union bound over all edges of M_t , we get

$$\Pr[\exists j = 1, \dots, k, e = (u, v) \in M_t \text{ s.t. } u, v \in C_j] \leq |M_t| \cdot \sum_{j=1}^k |C_j|^2/n^2 \leq \sum_{j=1}^k |C_j|^2/n = \|F_{t-1}\|/n.$$

The latter expression does not take into account the fact that as edges are added one after another, the set of components may increase, but if one adds edges of the matching M_t , a similar bound follows. In particular, one can show that the players succeed with probability at most

$$\sum_{t=1}^T O(\|F_t\|/n) = O(T) \cdot (\|F_T\|/n),$$

since $\|F_t\|$ is a non-decreasing function of t . It thus suffices to prove that $\|F_T\| \ll n/T$. We now outline a simple analysis that shows that if F_t is the forest computed by the players in the simple protocol above, then as long as $s \ll n/A^T$ for a sufficiently large $A > 1$, then after t steps one has

$$\|F_t\| \leq s \cdot B^t \quad (9)$$

for some constant $B > 1$. Thus, if the players start with $s \ll n/A^T$ space with $A \geq 100B$, say, then one has $\|F_T\|/n \ll 100^{-T}$, and thus the players do not succeed distinguishing between the **YES** and **NO** cases with any significant advantage over random guessing.

Analyzing the growth of $\|F_t\|$. Fix $t \in [T]$, and let C_1, C_2, \dots, C_k be the nontrivial (size more than 1) connected components of F_{t-1} . Let $e_1, e_2, \dots, e_{\alpha n}$ be edges of M_t . We now analyze the expected increase of the weight $\|F_t\|$ of F_t relative to the weight $\|F_{t-1}\|$ of F_{t-1} .

In order to compare $\|F_t\|$ to $\|F_{t-1}\|$, we define $|M_t|$ intermediate forests, where the j -th such forest is the forest that results from adding edges from the set e_1, \dots, e_j to F_{t-1} . Specifically, let $F_t^0 := F_{t-1}$, and for every $j = 1, \dots, |M_t|$ let

$$F_t^j := \begin{cases} F_t^{j-1} \cup \{e_j\} & \text{if } e_j \text{ intersects with a nontrivial component in } F_t^{j-1} \\ F_t^{j-1} & \text{o.w.} \end{cases}$$

Recall that besides keeping edges of M_t that intersect connected components of F_{t-1} , the t -th player also posts the bits on an arbitrarily chosen subset of edges of M_t on the board. Let E^* denote this set of additional edges (not incident on any nontrivial component in F_{t-1} whose bits the t -th player posts). This means that $F_t \subseteq F_t^{|M_t|} \cup E^*$. Since edges in E^* are not incident to any components in $F_t^{|M_t|}$, their addition simply increases the weight of the forest by $4s$ (since every component is of size 2). We now upper bound $\|F_t^{|M_t|}\|$.

For every $i = 1, \dots, |M_t|$ we upper bound

$$\mathbb{E}_{e_i} [\|F_t^i\| | F_t^{i-1}, e_{[1:i-1]}].$$

Conditioned on edges e_1, e_2, \dots, e_{i-1} , the edge e_i is a uniformly random edge not sharing end-points with e_1, e_2, \dots, e_{i-1} . Thus, for every subset of vertices $C \subseteq [n]$ one has

$$\Pr_{e_i} [|C \cap e_i| = 1 | e_{[1:i-1]}] \leq 4|C|/n, \quad (10)$$

since $|M_t| = \alpha n$ and $\alpha < 1/4$ by assumption. Note that if for a component C in F_t^{i-1} one has $|C \cap e_i| = 1$ and the other endpoint of e_i does not belong to any nontrivial component in F_t^{i-1} , then the increase in the weight of the forest is $(|C| + 1)^2 - |C|^2 = 2|C| + 1$. We call such edges e_i **boundary edges**.

Similarly, for every pair of subset of vertices $C, C' \subseteq [n], C \cap C' = \emptyset$ one has

$$\Pr_{e_i} [|C \cap e_i| = 1 \text{ and } |C' \cap e_i| = 1 | e_{[1:i-1]}] \leq 8|C||C'|/n^2, \quad (11)$$

and we note that if the event above happens, the increase in the weight of the forest due to addition of e_i is $(|C| + |C'|)^2 - |C|^2 - |C'|^2 = 2|C||C'|$. We call such edges e_i **internal edges**. These notions of internal edges and boundary edges will later be crucial in the proof of our main theorem through Fourier analysis (see Section 5.3 below for an outline and then proofs of Lemmas 7.1 and 7.9 for the actual application).

Putting the two observations above (increase of potential due to boundary edges and internal edges) together, we get

$$\begin{aligned} \mathbb{E}_{e_i} [\|F_t^i\| | e_{[1:i-1]}] &\leq \|F_t^{i-1}\| + \sum_{\substack{C \text{ nontrivial component} \\ \text{in } F_t^{i-1}}} (2|C| + 1) \cdot \Pr_{e_i} [|C \cap e_i| = 1 | e_{[1:i-1]}] \\ &+ \sum_{\substack{C, C' \text{ nontrivial components} \\ \text{in } F_t^{i-1}}} (2|C| \cdot |C'|) \cdot \Pr_{e_i} [|C \cap e_i| = 1 \text{ and } |C' \cap e_i| = 1 | e_{[1:i-1]}] \end{aligned} \quad (12)$$

Substituting (10) and (11) into (12), we obtain

$$\begin{aligned}
\mathbb{E}_{e_i} [\|F_t^i\| | e_{[1:i-1]}] &\leq \|F_t^{i-1}\| + \sum_{\substack{C \text{ nontrivial component} \\ \text{in } F_t^{i-1}}} (2|C| + 1) \cdot \frac{4|C|}{n} \\
&+ \sum_{\substack{C, C' \text{ nontrivial components} \\ \text{in } F_t^{i-1}}} (2|C| \cdot |C'|) \cdot \frac{8|C| \cdot |C'|}{n^2} \\
&\leq \|F_t^{i-1}\| + \frac{12\|F_t^{i-1}\|}{n} + \frac{8\|F_t^{i-1}\|^2}{n^2} \\
&= \|F_t^{i-1}\| \cdot \left(1 + \frac{12}{n} + \frac{8\|F_t^{i-1}\|}{n^2}\right).
\end{aligned} \tag{13}$$

Applying this $|M_t| = \alpha n$ times formally requires a careful application of concentration inequalities (the details are provided in Appendix C), but ultimately results in

$$\mathbb{E} [\|F_t^{|M_t|}\|] \approx \|F_{t-1}\| \cdot \left(1 + \frac{12}{n} + \frac{8\|F_{t-1}\|}{n^2}\right)^{|M_t|} \approx \|F_{t-1}\| \cdot (B/2) \tag{14}$$

for some constant B , as long as all intermediate forests satisfy $\|F_t^{i-1}\| \ll n$ (which they do with the appropriate setting of parameters – see Appendix C for details).

Now recall that besides keeping edges of M_t that intersect connected components of F_{t-1} , the t -th player also posts the bits on an arbitrarily chosen subset of edges of M_t that do not share an endpoint with F_{t-1} on the board (we call such edges **external edges**; a similar notion plays a crucial role in our analysis of general protocols see (26) and related discussion). Recall that $F_t \subseteq F_t^{|M_t|} \cup E^*$, where E^* denotes the set of additional edges (not incident on any nontrivial component in F_t) that the t -th player reveals. We thus get $\|F_t\| \leq \|F_t^{|M_t|}\| + 4s$, which, when put together with (14), gives

$$\mathbb{E} [\|F_t\|] \approx \|F_{t-1}\| \cdot B + 4s.$$

Applying the above iteratively for $t = 1, \dots, T$ results in

$$\mathbb{E}[\|F_t\|] \lesssim \|F_0\| \cdot B^t \lesssim s \cdot B^t.$$

This (informally) establishes (9), and shows that the players need $s \gg n/B^t$ in order to solve **DIHP**(n, α, T).

The analysis outlined above is quite simple, but does not generalize to arbitrary communication protocols. In the next section we introduce our Fourier analytic approach, and illustrate some of the main claims by instantiating them on our component growing protocol above.

5.3 Overview of general analysis (proof of Theorem 2.1)

In this section we first make some remarks about the Fourier spectrum of our component growing protocol, and then use it to illustrate our analysis, which is formally presented in Section 6.

5.3.1 Second level Fourier spectrum of the component growing protocol vs combinatorial analysis

Suppose that the players use the simple protocol as described above. In that case the set of possible values of the hidden partition X^* consistent with the players' knowledge at time t can be defined quite easily:

$$\mathbf{B}_t = \{x \in \{0, 1\}^n : \forall e = (a, b) \in F_t, x_a + x_b = w_e\}. \tag{15}$$

Thus, \mathbf{B}_t is simply a linear subspace of $\{0, 1\}^n$ with constraints given by the edges in F_t . We will derive expressions for the Fourier transform of the indicators $h_t := \mathbf{1}_{\mathbf{B}_t}$ of \mathbf{B}_t for this simple protocol. Similarly, set of possible values of the hidden partition X^* consistent with the t -th player's message can be defined quite easily as well:

$$\mathbf{A}_t = \{x \in \{0, 1\}^n : \forall e = (a, b) \in F_t \cap M_t, x_a + x_b = w_e\}. \quad (16)$$

Again, \mathbf{A}_t is simply a linear subspace of $\{0, 1\}^n$ with constraints given by the edges revealed by the t -th player. The normalized Fourier transform of the indicator $f_t := \mathbf{1}_{\mathbf{A}_t}$ of \mathbf{A}_t for this simple protocol is quite simple:

$$\tilde{f}(z) = \begin{cases} (-1)^{\sum w_i} & \text{if } z \text{ is matched by edges } \{e_i\} \text{ of } M_t \\ 0 & \text{o.w.} \end{cases} \quad (17)$$

These expressions will not be directly useful for our proof, but will provide very good intuition.

It can be verified (a calculation is included in appendix E for convenience of the reader) that $\tilde{h}_t(v)$ can only be nonzero if the set v has an even intersection with every cluster in F_t . We refer to such $v \in \{0, 1\}^n$ as *admissible* for brevity. Furthermore, it can be verified that $\tilde{h}_t(v)$ has the following simple form. For each admissible v let $Q(v)$ denote a pairing of vertices of v via edge-disjoint paths in F_t (we associate Q_v with the set of edges on these paths). This is illustrated in Fig. 1(a), where the vertices of $v \in \{0, 1\}^n$ are marked red, and the edges of $Q(v)$ are the green dashed edges. Then we have

$$\tilde{h}_t(v) = \begin{cases} (-1)^{\sum_{e \in Q(v)} w_e} & \text{if } v \text{ is admissible} \\ 0 & \text{o.w.} \end{cases}$$

We refer to appendix E for the proof. For example, the coefficient $\tilde{h}_t(\{a_1, a_2, b_1, b_2, c_1, c_2\})$ is nonzero, equals 1 in absolute value, and its sign is determined by the parities of labels on the green paths connecting a_1 to a_2 , b_1 to b_2 , and c_1 to c_2 (see Fig. 1(a)).

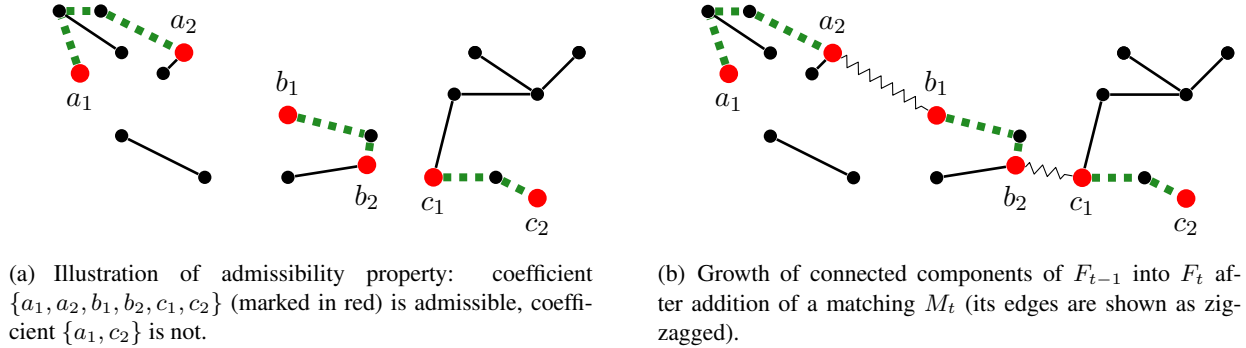


Figure 1: Fourier transform of h_t vs growth of connected components in F_t .

Recall that our direct analysis of the adaptive protocol in Section 5.1 used the *weight* of the forest F_t , defined by

$$\|F_t\| = \sum_{i=1}^k |C_i|^2,$$

where C_1, \dots, C_k are the nontrivial (size strictly larger than 1) connected components of F_t . As we noted in that section, this particular way of analyzing the component growing protocol does not generalize, but

the following reformulation does. Consider weight two Fourier coefficients of h_t . As noted above, one has $\tilde{h}_t(\{a, b\}) \neq 0$ if and only if a and b belong to the same connected component in F_t . We therefore have

$$\begin{aligned} \sum_{v \in \{0,1\}^n: |v|=2} |\tilde{h}_t(v)| &= \frac{1}{2} \sum_{\substack{a,b \in [n], \\ a \neq b}} |\tilde{h}_t(\{a, b\})| \\ &= \frac{1}{2} \sum_{i=1}^k \sum_{\substack{a,b \in C_i, \\ a \neq b}} |\tilde{h}_t(\{a, b\})| \\ &= \sum_{i=1}^k \binom{C_i}{2} \\ &= \Theta(\|F_t\|). \end{aligned}$$

This suggests an analysis that is based on proving that

$$\sum_{v \in \{0,1\}^n: |v|=2} |\tilde{h}_t(v)| \ll n \quad (18)$$

for all $t = 1, \dots, T$, i.e, the sum of absolute values of second level Fourier coefficients stays small throughout the game, hoping that this is a general enough approach for handling arbitrary protocols. Our proof indeed proceeds along similar lines but there are two major difficulties that one needs to overcome in order to make this work. **First**, while for the simple protocol higher order Fourier coefficients are (essentially) determined by weight two Fourier coefficients, i.e., by the collection of connected components in F_t , this is not the case for general protocols. We thus need to generalize (18) to higher weights. **Second**, we need to design techniques for analyzing the equivalent of ‘component growth’ in the combinatorial version of our analysis, in terms of Fourier coefficients.

5.3.2 Evolution of Fourier coefficients

We now present our general analysis, and illustrate it by applying to the component growing protocol from Section 5.1. Recall that after t players have spoken the random bipartition X^* is uniform in the set $\mathbf{B}_t = \mathbf{A}_1 \cap \mathbf{A}_2 \cap \dots \cap \mathbf{A}_t$. The following lemma (see Section 6 for the proof) is crucial for our proof:

Lemma 6.2 *For every constant $\alpha \in (0, 10^{-10})$, integer $n \geq 1$, every $T \in [10, \ln n]$, $s \geq \sqrt{n}$, the following conditions hold if n is sufficiently large. Suppose there exists a protocol Π for **DIHP**(n, α, T) such that $|\Pi| = s < n/(10T)^{10^9 T}$. Let $\delta = 1/(1000T)$. Then there exist events $\mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \dots \supseteq \mathcal{E}_T$ with $\Pr[\mathcal{E}_1] = 1$ and $\Pr[\bar{\mathcal{E}}_{t+1} | \mathcal{E}_t] \leq 1/(100T)$ for $t = 1 \dots T - 1$, such that \mathcal{E}_t depends on $(S_{1:t-1}^Y, M_{1:t})$ only, and conditioned on \mathcal{E}_t one has*

1. \mathbf{B}_{t-1} is $((10^{12}T)^t, 10Ts)$ -bounded (as per Definition 4.3);
2. $|\mathbf{B}_{t-1}|/2^n \geq 2^{-s(t-1) - 10(t-1) \log T}$;
3. for any non-negative function q on $\{0, 1\}^{M_t}$

$$1 - \delta \leq \frac{\mathbb{E}_{x \sim \text{Uniform}(\mathbf{B}_{t-1})} [q(M_t x)]}{\mathbb{E}_{z \sim \text{Uniform}(\{0,1\}^{M_t})} [q(z)]} \leq 1 + \delta.$$

The third part of the lemma helps us conclude that the messages posted on the board do not reveal enough information to distinguish between the **YES** and the **NO** cases: indeed it implies that the posterior distribution of the labels that the t -th player observes on M_t in the **YES** case is pointwise close to uniform. Since we show that this is true for all players $t = 1, \dots, T$, the result follows by simple properties of the total variation distance (see Lemma 6.3 and then proof of Theorem 2.1 in Section 6). The first two parts of Lemma 6.2 are the main invariants on the evolution of the Fourier spectrum of h_t that drive our analysis.

Recall that by Definition 4.3 a set $\mathbf{B} \subseteq \{0, 1\}^n$ is (C, s^*) -bounded if for all $\ell \leq s^*$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}_{\mathbf{B}}}(v) \right| \leq \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell; \quad (19)$$

and for all $s^* < \ell < \frac{n}{C^2}$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}_{\mathbf{B}}}(v) \right| \leq \left(\frac{C^2 n}{\ell} \right)^{\ell/2}. \quad (20)$$

As per (5), defining the function $\mathbf{U}_{C,s^*}(\ell)$ by

$$\mathbf{U}_{C,s^*}(\ell) = \begin{cases} 1 & \ell = 0; \\ \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell & \ell \in [1 : s^*]; \\ \left(\frac{C^2 n}{\ell} \right)^{\ell/2} & \ell > s^*, \end{cases}$$

we are able to simplify notation, ensuring that an indicator function h is (C, s^*) bounded if and only if for all $\ell < n/C^2$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{h}(v) \right| \leq \mathbf{U}_{C,s^*}(\ell).$$

Intuition behind the choice of the bound $\mathbf{U}_{C,s^*}(\ell)$. We note that the bound above is essentially obtained as follows: one first thinks of the bounds on the ℓ_1 norm of the Fourier transform of the indicator f of the set \mathbf{A} in the cube that is consistent with the message of a single player that follow by applying Cauchy-Schwarz to the ℓ_2 norm bounds provided by hypercontractivity (see Lemma F.2 in Appendix F; this is the bound that was used in [GKK⁺08] and follow up works). We then prove by induction that even the Fourier transform of the product of such indicator functions maintains the small ℓ_1 norm property. The fact that $\mathbf{U}_{C,s^*}(\ell)$ provides different bounds for small and large ℓ is a consequence of the fact that hypercontractivity only implies strong bounds on the ℓ_2^2 mass of the Fourier spectrum of an (indicator of a) set \mathbf{A} of density 2^{-s^*} when $\ell \leq s^*$, for larger ℓ one simply uses Parseval's equality. See the proof of Lemma 3.3 for the details of this calculation.

We also note that the bound that we prove on the ℓ_1 norm of the Fourier transform of h is indeed surprisingly strong: it is much stronger than what follows by Cauchy-Schwarz for an arbitrary function of the same ℓ_2^2 norm.

Overall, in our analysis we distinguish between ‘low weight’ Fourier coefficients, namely those with weights between 1 and s^* , the ‘intermediate weight’ Fourier coefficients, namely those between s^* and $\frac{n}{C^2}$, and the ‘high weight’ Fourier coefficients, namely those with weights between $\frac{n}{C^2}$ and n . Here, $s^* = 10Ts$ is an upper bound for the total amount of information revealed by all players.

The low weight bound (19) for the component growing protocol ($\ell \in [1, s^*]$). Note that instantiating the first guarantee of Lemma 6.2 above for $\ell = 1$ leads to a bound of

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2}} \left| \widetilde{\mathbb{1}}_{\mathbf{B}}(v) \right| \leq (10^{12}T)^t \cdot \sqrt{10Ts \cdot n},$$

which is similar to our upper bound of $s \cdot B^t$ on $\|F_t\|$ from Section 5.2. Also note that the bound that we get for general communication protocols with a budget of s bits is similar to what one would get for the simple protocol with $\approx \sqrt{s \cdot n}$ bits (see rhs above). This is a consequence of the fact that our analysis starts with ℓ_2^2 bounds on Fourier coefficients and converts those into ℓ_1 bounds, with an appropriate loss from Cauchy-Schwarz.

The intermediate weight bound (20) for the component growing protocol ($\ell \in (s^*, n/C^2]$). Note that since the middle weight coefficients correspond to weights at least $s^* = 10Ts$, they all vanish for any simple protocol of size at most s . Indeed, since each player reveals at most s edges, the total size of non-trivial connected components in the resulting forest F_T does not exceed $2Ts$.

High weights for the component growing protocol ($\ell > n/C^2$). Similarly, the high weight part of the spectrum is zero for the component growing protocol, since the maximum weight of a nonzero Fourier coefficient is upper bounded by twice the number of edges in the forest F_t , and that number never becomes close to n with the appropriate setting of parameters.

We now outline the proof of Lemma 6.2. Note that the main challenge that we had to overcome in the analysis of the component growing protocol in Section 5.2 is bounding the rate at which connected components of different sizes are merged when edges of the next matching M_t connect two nontrivial components (we refer to these edges as **internal edges**, see Section 5.2) or connect a nontrivial component to an isolated vertex (we refer to these edges as **boundary edges**, see Section 5.2). Our Fourier analytic approach analyzes the component merging process using the convolution theorem: we note that the arrival of internal or boundary edges results in the Fourier transform of h_t being convolved with the Fourier transform of the message f_t that the t -th player sends, and we analyze this process directly.

The proof of Lemma 6.2 is by induction on t , with the inductive step being the main technical lemma of our paper. It is given by

Lemma 6.1 *For every $n, C, s^*, \alpha, \delta$ that satisfy conditions*

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4 \quad \text{(P5)} \delta \in (n^{-1}, 1/2),$$

every $\mathbf{B} \subseteq \{0, 1\}^n$, if \mathbf{B} is (C, s^) -bounded (as per Definition 4.3) and M is a uniformly random matching of size αn , the following conditions hold with probability at least $1 - 5\delta$ over the choice of M .*

For every $\mathbf{A}_{\text{reduced}} \subseteq \{0, 1\}^M$ such that $|\mathbf{A}_{\text{reduced}}|/2^{\alpha n} \geq 2^{-s^}$, if $\mathbf{A} = \{x \in \{0, 1\}^n : Mx \in \mathbf{A}_{\text{reduced}}\}$, then $\mathbf{B}' := \mathbf{B} \cap \mathbf{A}$ is $((10^9/\delta)C, s^*)$ -bounded.*

We now illustrate the proof using the component growing protocol. Let \mathbf{B}_t be as in Definition 4.1. In particular, for the component growing protocol \mathbf{B}_t is explicitly given by (15). For simplicity we write M for M_t , \mathbf{B}' for \mathbf{B}_t , \mathbf{B} for \mathbf{B}_{t-1} , and \mathbf{A} for \mathbf{A}_t . We further let $h := \mathbf{1}_{\mathbf{B}}$, $f := \mathbf{1}_{\mathbf{A}}$, $h' := h \cdot f = \mathbf{1}_{\mathbf{B}'}$. Proving that h' is $((10^9/\delta)C, s^*)$ -bounded involves upper bounding the ℓ_1 norm of Fourier coefficients at various

levels $2\ell, \ell \in [1, n/(2C^2)]$. Convolution theorem (1) together with triangle inequality give

$$\begin{aligned}
\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \tilde{h}'(v) \right| &= \frac{2^n}{|\mathbf{B}'|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \hat{h}'(v) \right| \\
&= \frac{2^n}{|\mathbf{B}|} \frac{2^n}{|\mathbf{A}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \hat{h}'(v) \right| \\
&\leq \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \sum_{z \in \{0,1\}^n} \left| \frac{2^n}{|\mathbf{B}|} \hat{h}(v \oplus z) \frac{2^n}{|\mathbf{A}|} \hat{f}(z) \right| \\
&= \sum_{v \in \{0,1\}^n} \sum_{\substack{z \in \{0,1\}^n \\ |v \oplus z|=2\ell}} \left| \tilde{h}(v) \tilde{f}(z) \right| \\
&= \sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \sum_{\substack{z \in \{0,1\}^n \\ |v \oplus z|=2\ell}} \left| \tilde{f}(z) \right|.
\end{aligned} \tag{21}$$

Here in going from line 2 to line 3 we used the fact that $|\mathbf{B}'|/2^n = (|\mathbf{B}|/2^n) \cdot (|\mathbf{A}|/2^n)$ for our simple protocol, as long as no cycle has been revealed. This in particular proves part 2 of Lemma 6.2 for the component growing protocol without the additive loss in the exponent (we get $|\mathbf{B}_{t-1}|/2^n = 2^{-s(t-1)}$ as opposed to just $|\mathbf{B}_{t-1}|/2^n \geq 2^{-s(t-1)-10(t-1)\log T}$; see proof of Lemma 6.2 in Section 6 for general argument). It is useful to recall at this point (see (16) and (17)) that $\tilde{f}(w) \neq 0$ only if w is perfectly matched by M .

Now note that for the component growing protocol the sum $\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \tilde{h}'(v) \right|$ is the number of sets of 2ℓ vertices that intersect every connected component in F_{t-1} an even number of times. The sum on the rhs of the last line in (21) is over all v that intersect every component in F_{t-1} an even number of times, and subsets z of edges of M_t such that $v \oplus z$ has weight 2ℓ . **For small ℓ** most of the sum is contributed by sets of 2ℓ vertices having two vertices in each connected component of F_{t-1} . In particular, every such v together with $z \in \{0,1\}^n$ that satisfies $|v \oplus z| = 2\ell$ and $\tilde{f}(z) \neq 0$ corresponds to a collection of components in F_{t-1} that are merged into a collection of ℓ components in F_t by the edges in z .

The core of our proof (see Lemma 7.1 and Lemma 7.9) shows that if h is (C, s^*) -bounded, then

$$\sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \left| \tilde{f}(w) \right| \right] \leq \mathbf{U}_{10^9 C, s^*}(\ell),$$

which in turn implies that $h' = h \cdot f$ is $(10^9 C, s^*)$ -bounded on average. Applying Markov's inequality to the above then yields a proof of Lemma 6.1, and applying Lemma 6.1 iteratively leads to a proof of Lemma 6.2.

We now illustrate the main ideas of the proof of the implication above when ℓ is small, namely when $\ell \leq s^*$. For clarity of exposition we sketch the proof of the following bound:

$$\sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{z \in \{0,1\}^n \\ |v \oplus z|=2\ell}} \left| \tilde{f}(z) \right| \right] \leq \binom{20Cs}{\ell}, \tag{22}$$

starting from the assumption that for all $\ell = 1, \dots, n/C^2$

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} |\tilde{h}(v)| \leq \binom{Cs}{\ell}. \quad (23)$$

We note that the condition above implies (C, s) -boundedness.

We first note that $\tilde{f}(z) \neq 0$ only if $z = Mw$ for $w \in \{0,1\}^M$ (see (16) and (17) for the component growing protocol, and Lemma 3.4 for general protocols), and thus for every $v \in \{0,1\}^n$ one has $\mathbb{E}_M \left[\sum_{\substack{z \in \{0,1\}^n \\ |v \oplus z|=2\ell}} |\tilde{f}(z)| \right] = \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw|=2\ell}} |\tilde{f}(Mw)| \right]$. We also note that since our matchings are small ($|M| \leq \alpha n$ for small constant $\alpha \in (0, 1)$, see, e.g. Lemma 6.2), the condition $|v \oplus Mw| = 2\ell$ implies that $|v| < n/100$. Thus, the only terms with a nonzero contribution to the sum that we need to bound are $v \in \{0,1\}^n$ with $|v| = 2k \leq n/100$. Thus, it suffices to bound, for a parameter $k \in [0 : n/200]$ and $v \in \{0,1\}^n$ with $|v| = 2k$, the quantity

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw|=2\ell}} |\tilde{f}(Mw)| \right]. \quad (24)$$

Note that this quantity depends on v only through $k = |v|$. We will later combine our bounds over all $k \in [0 : n/200]$ to obtain the result of the lemma.

Bounding (24) for fixed v : internal and boundary edges. A vector $v \in \{0,1\}^n$ naturally corresponds to a subset of $[n]$ of its non-zero coordinates. Let $\text{Int}_M(v) = \{e_1^{\text{int}}, e_2^{\text{int}}, \dots\}$ be the set of edges $e = (a, b) \in M$ that match points of v , i.e. with $a, b \in v$. Let $\partial_M(v) = \{e_1^{\text{bound}}, e_2^{\text{bound}}, \dots\}$ be the set of boundary edges, i.e. edges $e = (a, b) \in M$ with $a \in v, b \notin v$ or vice versa. Let $\text{Ext}_M(v) = \{e_1^{\text{ext}}, e_2^{\text{ext}}, \dots\}$ be the set of external edges, i.e. edges $e = (a, b) \in M$ with $a, b \in [n] \setminus v$. See Fig. 2 for an illustration.

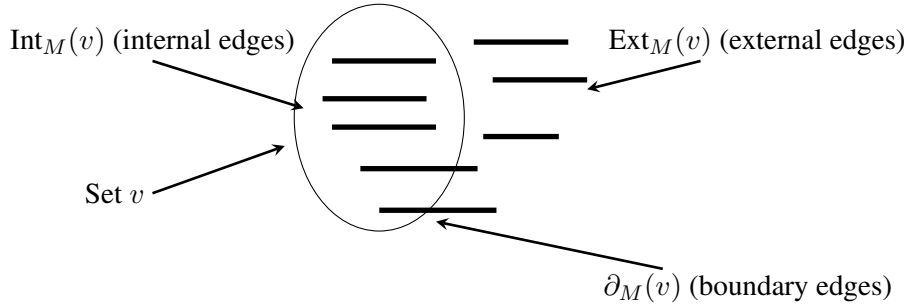


Figure 2: A coefficient v together edges of the matching M classified into internal edges $\text{Int}_M(v)$, external edges $\text{Ext}_M(v)$ and boundary edges $\partial_M(v)$.

We decompose the sum (24) according to the set of boundary edges in w :

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw|=2\ell}} |\tilde{f}(Mw)| \right] = \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{w \in \{0,1\}^M} \mathbb{1}_{\{w \cap \partial_M(v) = S\}} \mathbb{1}_{\{|v \oplus Mw|=2\ell\}} |\tilde{f}(Mw)| \right].$$

It turns out that if $w \cap \partial_M(v) = S$ then the latter indicator function can be rewritten as the indicator of $|w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)$, where $w_S \in \{0, 1\}^M$ is the set of all internal edges $\text{Int}_M(v)$ and all edges in S (see the proof of Lemma 7.1 for details). This gives the following upper bound:

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &= \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} \mathbb{1}_{\{w \cap \partial_M(v) = S\}} \cdot |\tilde{f}(Mw)| \right] \\ &\leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}(Mw)| \right], \end{aligned}$$

where we dropped the indicator function in going from the first line to the second line above. Note that $|\tilde{f}(Mw)| = 1$ if all the edges of w are revealed by the t -th player and zero otherwise. Since all the edges of w_S (i.e. internal edges and some subset of the boundary edges) are revealed we know that $w \oplus w_S$ is also a subset of the revealed edges which means that we have

$$\sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}(Mw)| \leq \binom{\ell - k + |\text{Int}_M(v)|}{s}. \quad (25)$$

Note that here we are crucially using the fact that $w \oplus w_S$ is a set of external edges only. This allows us to bound the sum of Fourier coefficients by a function of s , the communication budget per player. The equivalent statement for general protocols is provided by Lemma 3.3 (see (39)), which bounds the sum of absolute values of Fourier coefficients of a dense subset of the boolean cube by a similar expression to the above.

Since the bound in (25) is independent of S , summing over all possible subsets $S \subseteq \partial_M(v)$, we infer

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \leq \mathbb{E}_M \left[2^{|\partial_M(v)|} \binom{\ell - k + |\text{Int}_M(v)|}{s} \right]. \quad (26)$$

We then bound the sum on the last line above:

$$\begin{aligned} \mathbb{E}_M \left[2^{|\partial_M(v)|} \binom{\ell - k + |\text{Int}_M(v)|}{s} \right] &= \mathbb{E}_M \left[\sum_{i=0}^k \sum_{b=0}^{2k} 2^b \mathbb{1}_{\{|\partial_M(v)|=b \text{ and } |\text{Int}_M(v)|=i\}} \binom{\ell - k + i}{s} \right] \\ &= \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} 2^b q(k, i, b, n) \binom{\ell - k + i}{s}. \end{aligned}$$

In going from line 1 to line 2 above we used the fact that by Definition 3.10 for every $v \in \{0, 1\}^n$ with $|v| = 2k$ one has $\mathbb{E}_M[\mathbb{1}_{\{|\partial_M(v)|=b \text{ and } |\text{Int}_M(v)|=i\}}] = q(k, i, b, n)$, i.e. $q(k, i, b, n)$ is the probability that a uniformly random matching M of size αn is such that i edges of M match points of v (i.e. $|\text{Int}_M(v)| = i$) and b edges of M are boundary edges (i.e. $|\partial_M(v)| = b$). Recall that $|k - \ell|_+ = \max\{0, k - \ell\}$. We note

that $q(k, i, b, n)$ depends on the size αn of the matching M , but we prefer to keep this dependence implicit to simplify notation.

We thus have that for a fixed $v \in \{0, 1\}^n$ with $|v| = 2k$

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \leq \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} q(k, i, b, n) 2^b \binom{s}{\ell - k + i}. \quad (27)$$

We now note that intuitively, $q(k, i, b, n) \approx n^{-i}$, since (at least for small i and k) the probability of having an internal edge in a given set of size $2k$ is approximately $1/n$. In fact, for small k, i we have $q(k, i, b, n) \lesssim 10^k n^{-i}$ (the formal bounds are somewhat different for larger k and i , and are given in Lemma 3.15 and Lemma 3.13).

We now consider two cases.

Case 1: $k \geq \ell$. This case essentially corresponds to analyzing the rate at which collection of k components get merged into collections of ℓ components. Using Lemma 3.15 and Lemma 3.13 we get that the summand in (27) decays exponentially with b and i , and since in this case the sum starts at $i = k - \ell$, we get

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \sum_{i=k-\ell}^k \sum_{b=0}^{2k} q(k, i, b, n) 2^b \binom{s}{\ell - k + i} \\ &\lesssim 10^k \cdot \sum_{i=k-\ell}^k n^{-i} \binom{s}{\ell - k + i} \\ &\approx 10^k \cdot n^{-(k-\ell)}. \end{aligned}$$

This is consistent with the intuition that (at least for constant k and ℓ) the probability that a given collection of k constant size components becomes merged into only $\ell < k$ components is about $n^{-(k-\ell)}$: this is simply because such an event requires at least $k - \ell$ edges of the matching M_t to have both endpoints inside the k components.

Case 2: $k < \ell$. In this case, as in case 1, the sum in (27) above is close to the value of the maximum summand, which gives, since the sum starts with $i = |k - \ell|_+ = 0$,

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \sum_{i=0}^k \sum_{b=0}^{2k} q(k, i, b, n) 2^b \binom{s}{\ell - k + i} \\ &\lesssim 10^k \cdot \binom{s}{\ell - k}. \end{aligned}$$

This is consistent with the intuition that a given collection of k components in F_{t-1} can contribute to about $\binom{s}{\ell-k}$ size ℓ components in F_t : simply consider adding any subset of $\ell - k$ edges of the matching M_t that do

not intersect with nontrivial components in F_{t-1} . Note that this is the part where we analyze the contribution of the bits that the players are actually charged for in the component growing protocol, i.e. external edges.¹

Using the bounds from **Case 1** and **Case 2** above in equation (22), we obtain

$$\begin{aligned} \sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{z \in \{0,1\}^n \\ |v \oplus z| = 2\ell}} \left| \tilde{f}(z) \right| \right] &\lesssim \sum_{k \leq \ell} 10^k \cdot \binom{Cs}{k} \cdot \binom{s}{\ell - k} + \sum_{k > \ell} 10^k \cdot \binom{Cs}{k} \cdot n^{\ell - k} \\ &\lesssim \binom{20Cs}{\ell}. \end{aligned}$$

The sketch above is informal, but Lemma 7.1 (whose proof is given in Section 6) provides the formal version that matches the qualitative conclusion. We state this lemma here for convenience of the reader:

Lemma 7.1 *For every $n, s^*, C > 1, \alpha \in (0, 1)$ that satisfy conditions*

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

if $\mathbf{B} \subseteq \{0, 1\}^n$ is (C, s^) -bounded and M is a uniformly random matching on $[n]$ of size αn , the following conditions hold. For every $\mathbf{A}_{\text{reduced}} \subseteq \{0, 1\}^M$, if $\mathbf{A} = \{x \in \{0, 1\}^n : Mx \in \mathbf{A}_{\text{reduced}}\}$ and f is the indicator of \mathbf{A} , if $|\mathbf{A}|/2^n \geq 2^{-s^*}$, then for every $\ell \in [1, s^*]$*

$$\sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{z \in \{0,1\}^n \\ |v \oplus z| = 2\ell}} \left| \tilde{f}(z) \right| \right] \leq \left(\frac{10^9 C \sqrt{s^* n}}{\ell} \right)^\ell.$$

Similar ideas lead to the proof of Lemma 7.3. In the general case we also need to bound the mass transfer from intermediate and high weights, see Lemmas 7.5 and 7.7.

6 Proof of main theorem (Theorem 2.1)

In this section we give the proof of Theorem 2.1, restated here for convenience of the reader:

Theorem 2.1 *There exists a constant $C_0 > 0$ such that for every $\epsilon \in (0, 1)$ and every sufficiently large n every protocol Π for **DIHP**(n, α, T) with $\alpha = 10^{-11}$, $T = 512/(\alpha\epsilon^2)$ that succeeds with probability at least $2/3$ satisfies $|\Pi| \geq n/(C_0/\epsilon)^{C_0/\epsilon^2}$.*

Section outline. The proof is structured as follows. We first state our main technical lemma (Lemma 6.1, proved in Section 7), which is the core tool behind the proof. Lemma 6.1 analyzes the relation between the properties of the Fourier transform of \mathbf{B}_{t-1} and the Fourier transform of \mathbf{B}_t : we show that if \mathbf{B}_{t-1} is (C, s^*) -bounded (as per Definition 4.3) for some parameter $s^* \approx T \cdot s$ that is essentially the total communication budget of the players and M_t is a uniformly random matching of size αn , then \mathbf{B}_t is $((10^9/\delta)C, s^*)$ -bounded with high probability (note that the lemma is stated without any reference to the index t , but is actually applied iteratively as above in subsequent analysis). We then prove Lemma 6.2 below, which iteratively applies Lemma 6.1 and analyzes the evolution the Fourier coefficients of \mathbf{B}_t . The proof of Theorem 2.1 then follows by combining basic properties of the total variation distance (Lemma 6.3 below) with Lemma 6.2.

The following technical lemma, which is the main result of Section 7, is the main tool in our proof:

¹At the same time, one should note that while the simple intuitive overview of case 1 does not involve external edges, they do have an effect on the actual proof – see Lemmas 7.1 and 7.9 in Section 6. The fact that their effect in this case is second order lets us present the simple intuition for case 1 above.

Lemma 6.1 For every $n, C, s^*, \alpha, \delta$ that satisfy conditions

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4 \quad \text{(P5)} \delta \in (n^{-1}, 1/2),$$

every $\mathbf{B} \subseteq \{0, 1\}^n$, if \mathbf{B} is (C, s^*) -bounded (as per Definition 4.3) and M is a uniformly random matching of size αn , the following conditions hold with probability at least $1 - 5\delta$ over the choice of M .

For every $\mathbf{A}_{\text{reduced}} \subseteq \{0, 1\}^M$ such that $|\mathbf{A}_{\text{reduced}}|/2^{\alpha n} \geq 2^{-s^*}$, if $\mathbf{A} = \{x \in \{0, 1\}^n : Mx \in \mathbf{A}_{\text{reduced}}\}$, then $\mathbf{B}' := \mathbf{B} \cap \mathbf{A}$ is $((10^9/\delta)C, s^*)$ -bounded.

6.1 Evolution of Fourier coefficients

The next lemma bounds the evolution of Fourier coefficients of the sets \mathbf{B}_t for $t = 1, \dots, T$, and forms the main technical part of the proof of Theorem 2.1. The proof of Lemma 6.2 essentially amounts to iteratively applying Lemma 6.1 and keeping track of the density $|\mathbf{B}_t|/2^n$ through a natural inductive hypothesis, using Lemma 4.7 at every step of the induction to argue that $|\mathbf{B}_t|/2^n \approx (|\mathbf{B}_{t-1}|/2^n) \cdot (|\mathbf{A}_t|/2^n)$ (i.e., the densities almost multiply):

Lemma 6.2 For every constant $\alpha \in (0, 10^{-10})$, integer $n \geq 1$, every $T \in [10, \ln n]$, $s \geq \sqrt{n}$, the following conditions hold if n is sufficiently large. Suppose there exists a protocol Π for **DIHP** (n, α, T) such that $|\Pi| = s < n/(10T)^{10^9 T}$. Let $\delta = 1/(1000T)$. Then there exist events $\mathcal{E}_1 \supseteq \mathcal{E}_2 \supseteq \dots \supseteq \mathcal{E}_T$ with $\Pr[\mathcal{E}_1] = 1$ and $\Pr[\bar{\mathcal{E}}_{t+1} | \mathcal{E}_t] \leq 1/(100T)$ for $t = 1 \dots T-1$, such that \mathcal{E}_t depends on $(S_{1:t-1}^Y, M_{1:t})$ only, and conditioned on \mathcal{E}_t one has

1. \mathbf{B}_{t-1} is $((10^{12}T)^t, 10Ts)$ -bounded (as per Definition 4.3);
2. $|\mathbf{B}_{t-1}|/2^n \geq 2^{-s(t-1) - 10(t-1)\log T}$;
3. for any non-negative function q on $\{0, 1\}^{M_t}$

$$1 - \delta \leq \frac{\mathbb{E}_{x \sim \text{Uniform}(\mathbf{B}_{t-1})}[q(M_t x)]}{\mathbb{E}_{z \sim \text{Uniform}(\{0, 1\}^{M_t})}[q(z)]} \leq 1 + \delta. \quad (28)$$

Proof: We let $s^* := 10Ts$ to simplify notation. Our proof is by induction on $t = 1, \dots, T$.

Base ($t = 1$) The event \mathcal{E}_1 happens with probability 1, and the set \mathbf{B}_0 is defined as $\mathbf{B}_0 = \{0, 1\}^n$, so the inductive claim is satisfied since for any matching M_1 a random variable $M_1 x$ with $x \sim \text{Uniform}(\{0, 1\}^n)$ is uniform in $\{0, 1\}^{M_1}$.

Inductive step: $t \rightarrow t+1$ We first define the events \mathcal{E}'_{t+1} and \mathcal{E}''_{t+1} and prove that both of them occur with high probability conditioned on \mathcal{E}_t . We then show that the inductive hypothesis for $t+1$ holds conditioned on \mathcal{E}_{t+1} , which we define to be $\mathcal{E}_t \cap \mathcal{E}'_{t+1} \cap \mathcal{E}''_{t+1} \cap \mathcal{E}'''_{t+1}$, where \mathcal{E}'''_{t+1} is a similar event that ensures that the third condition is satisfied, establishing the inductive step.

Constructing events \mathcal{E}'_{t+1} and \mathcal{E}''_{t+1} . Define $\mathcal{E}'_{t+1} := \{|\mathbf{A}_t| \geq 2^{n-s-9\log T}\}$. Conditioned on \mathcal{E}_t , we have by the inductive hypothesis and Lemma D.1

$$\mathbb{E}_{x \sim \text{Uniform}(\mathbf{B}_{t-1})}[1/|\mathbf{A}_t|] \leq (1 + \delta) \cdot \mathbb{E}_{x \sim \text{Uniform}(\{0, 1\}^{M_t})}[1/|\mathbf{A}_t|] \leq (1 + \delta) \cdot 2^{s-n}.$$

This means that by Markov's inequality for the event we have $\Pr[\mathcal{E}'_{t+1} | \mathcal{E}_t] > 1 - (1 + \delta) \cdot 2^{-9\log T} > 1 - \delta$.

Lower bounding $|\mathbf{B}_t|/2^n$. We start by showing that conditioned on \mathcal{E}_t the set \mathbf{B}_t satisfies

$$\frac{|\mathbf{B}_t|}{2^n} \geq (1 - \delta) \frac{|\mathbf{B}_{t-1}|}{2^n} \frac{|\mathbf{A}_t|}{2^n}. \quad (29)$$

Indeed, it suffices to take for q the indicator function of $\mathbf{A}_{reduced,t}$ and apply (28). Since

$$\mathbb{E}_{x \sim \text{Uniform}(\mathbf{B}_{t-1})}[q(Mx)] = \frac{|\mathbf{B}_{t-1} \cap \mathbf{A}_t|}{|\mathbf{B}_{t-1}|} = \frac{|\mathbf{B}_t|}{|\mathbf{B}_{t-1}|}, \quad \mathbb{E}_{z \sim \text{Uniform}(\{0,1\}^{M_t})}[q(z)] = \frac{|\mathbf{A}_{reduced,t}|}{2^{\alpha n}} = \frac{|\mathbf{A}_t|}{2^n},$$

we have

$$\begin{aligned} \frac{|\mathbf{B}_t|}{2^n} &= \frac{|\mathbf{B}_{t-1}|}{2^n} \cdot \mathbb{E}_{x \sim \text{Uniform}(\mathbf{B}_{t-1})}[q(Mx)] \\ &\geq (1 - \delta) \frac{|\mathbf{B}_{t-1}|}{2^n} \cdot \mathbb{E}_{z \sim \text{Uniform}(\{0,1\}^{M_t})}[q(z)] \quad (\text{by (28)}) \\ &= (1 - \delta) \frac{|\mathbf{B}_{t-1}|}{2^n} \cdot \frac{|\mathbf{A}_t|}{2^n} \end{aligned}$$

establishing (29). Now by conditioning on \mathcal{E}'_{t+1} we have $|\mathbf{A}_t| \geq 2^{n-s-9 \log T}$. Putting this together with (29) yields

$$\begin{aligned} \frac{|\mathbf{B}_t|}{2^n} &\geq (1 - \delta) \frac{|\mathbf{B}_{t-1}|}{2^n} \frac{|\mathbf{A}_t|}{2^n} \\ &\geq (1 - \delta) \frac{|\mathbf{B}_{t-1}|}{2^n} \cdot 2^{-s-9 \log T} \quad (\text{by conditioning on } \mathcal{E}'_{t+1}) \\ &\geq (1 - \delta) 2^{-s(t-1)-10(t-1) \log T} \cdot 2^{-s-9 \log T} \quad (\text{by the inductive hypothesis}) \\ &= (1 - \delta) \cdot 2^{\log T} \cdot 2^{-st-10t \log T} \\ &\geq 2^{-st-10t \log T} \quad (\text{since } \delta < 1/2 \text{ and } T \geq 10) \end{aligned} \quad (30)$$

establishing the lower bound on $|\mathbf{B}_t|$ required for the inductive step.

Proving that \mathbf{B}_t is $((10^{12}T)^{t+1}, 10Ts)$ -bounded. Let \mathcal{E}''_{t+1} be the success event from Lemma 6.1 applied to the sets $\mathbf{B} := \mathbf{B}_{t-1}$ and $\mathbf{A}_{reduced} = \mathbf{A}_{reduced,t}$ with $C = (10^{12}T)^t$, $\delta = 1/(10^3T)$, $s^* = 10Ts$. By Lemma 6.1 (we verify that preconditions are satisfied below) we have $\Pr[\mathcal{E}''_{t+1} | \mathcal{E}_t] \geq 1 - 5\delta$ and conditioned on \mathcal{E}''_{t+1} the set $\mathbf{B}_t = \mathbf{B}_{t-1} \cap \mathbf{A}_t$ is $((10^9/\delta)C, s^*)$ -bounded, as required.

We now verify that the preconditions of Lemma 6.1 are satisfied. We have to check five conditions:

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4 \quad \text{(P5)} \delta \in (n^{-1}, 1/2)$$

First two conditions are clear since $\alpha < 10^{-10}$ and $C = (10^{12}T)^t > 10^6$. Condition **(P3)** is satisfied since

$$s^* = 10Ts < \frac{n}{(10T)^{10^9 T - 1}} \leq \frac{n}{(10^{12}T)^{100T}} \leq \frac{n}{10^9 \cdot (10^{12}T)^{3T}} \leq \frac{n}{10^9 C^3}.$$

Condition **(P4)** is satisfied because $s \geq 1$ implies $n \geq (10T)^{10^9 T}$ and so

$$n \geq (10T)^{10^9 T} > 10^9 \cdot (10^{12}T)^{4T} \geq 10^9 C^4.$$

Condition **(P5)** is satisfied since $1/2 > \frac{1}{10^{3T}} = \delta > n^{-1}$, as $T \leq \ln n$ by assumption.

Proving point-wise closeness of pdfs. Conditioned on $\mathcal{E}_t \cap \mathcal{E}'_{t+1} \cap \mathcal{E}''_{t+1}$ the set \mathbf{B}_t is $((10^{12}T)^{t+1}, s^*)$ -bounded and satisfies $|\mathbf{B}_t|/2^n \geq 2^{-st-10t \log T} \geq 2^{-s^*}$, since $T \leq \ln n$ and $s \geq \sqrt{n}$ by assumption of the theorem. Let \mathcal{E}'''_{t+1} be the event (over the choice of the matching M_{t+1}) that for any non-negative function q on $\{0, 1\}^{M_{t+1}}$

$$1 - \delta \leq \frac{\mathbb{E}_{x \sim \text{Uniform}(\mathbf{B}_{t-1})} [q(M_t x)]}{\mathbb{E}_{z \sim \text{Uniform}(\{0,1\}^{M_t})} [q(z)]} \leq 1 + \delta, \quad (31)$$

that is, the event that the third assumption is satisfied for $t+1$. By Lemma 4.7 invoked with $C = (10^{12}T)^{t+1}$ and $s^* = 10Ts$ and $\delta = 1/(1000T)$ we have $\Pr[\mathcal{E}'''_{t+1} | \mathcal{E}_t \cap \mathcal{E}'_{t+1} \cap \mathcal{E}''_{t+1}] \geq 1 - \delta$. Indeed, preconditions of the lemma are satisfied since we have $C = (10^{13}T)^{t+1} > 100$, $\delta = 1/(10^3T) \in (n^{-1}, 1/2)$, $\alpha < 1/100$ and

$$s^* = 10Ts \leq (10T)n/(10T)^{10^9 T} = n/(10T)^{10^9 T - 1} \leq \left(\frac{1}{1000T}\right)^4 n/(10^{13}T)^{2t} = \delta^4 n/C^2.$$

It remains to note that for $\mathcal{E}_{t+1} := \mathcal{E}_t \cap \mathcal{E}'_{t+1} \cap \mathcal{E}''_{t+1} \cap \mathcal{E}'''_{t+1}$ we have

$$\Pr[\bar{\mathcal{E}}_{t+1} | \mathcal{E}_t] \leq \Pr[\mathcal{E}'_{t+1} | \mathcal{E}_t] + \Pr[\mathcal{E}''_{t+1} | \mathcal{E}_t] + \Pr[\mathcal{E}'''_{t+1} | \mathcal{E}_t \cap \mathcal{E}'_{t+1} \cap \mathcal{E}''_{t+1}] \leq \delta + 5\delta + \delta < 1/(100T).$$

■

6.2 Putting it together

We now combine Lemma 6.2 from the previous section with basic properties of the total variation distance to obtain a proof of Theorem 2.1. Specifically, the proof relies on Lemma 6.3 below, which is essentially the hybrid argument. Informally, the lemma says the following. Consider the **YES** case distribution \mathcal{D}^Y , and for each $t = 1, \dots, T-1$ compare the distribution of S_{t+1}^Y to the distribution of messages of the $(t+1)$ -st player obtained by supplying this player with random labels on their edges as opposed to labels consistent with the hidden bipartition X^* (the latter leads to $r_t(M_{1:t}, S_{1:t-1}^Y, U_t)$). If the resulting distributions are close in total variation distance, then the joint distribution of *all messages and matchings* M_t posted on the board after T rounds in the **YES** case is close to the same distribution in the **NO** case.

Lemma 6.3 *Let $X^* \sim \text{UNIF}(\{0, 1\}^n)$ be a uniformly random binary vector of length n . For each $t = 1, \dots, T$ let $U_t \sim \text{UNIF}(\{0, 1\}^{\alpha n})$ be an independent uniformly random vector of length αn . Let M_1, \dots, M_T be independently chosen random matchings on $[n]$.*

Let $S_0^Y = S_0^N := 0$ and for each $t = 1, \dots, T$ let

$$S_t^Y := r_t(M_{1:t}, S_{1:t-1}^Y, M_t X^*)$$

and

$$S_t^N := r_t(M_{1:t}, S_{1:t-1}^N, U_t)$$

for some functions $r_t, t = 1, \dots, T$. Suppose that there exists a sequence of events $\mathcal{E}_1 \supset \mathcal{E}_2 \supset \dots \supset \mathcal{E}_T$ such that \mathcal{E}_t depends only on $M_{1:t}$ and $S_{1:t-1}^Y$, and \mathcal{E}_1 occurs with probability 1, such that for any fixed $M_{1:t}, S_{1:t-1}^Y$ satisfying \mathcal{E}_t one has for some $\gamma > 0$

$$\|S_t^Y - r_t(M_{1:t}, S_{1:t-1}^Y, U_t)\|_{\text{tvd}} \leq \gamma/T. \quad (*)$$

Suppose further that for each $t = 1, \dots, T$ we have $\Pr[\bar{\mathcal{E}}_t | \mathcal{E}_{t-1}] \leq \gamma/T$. Then

$$\|(M_{1:T}, S_{1:T}^Y) - (M_{1:T}, S_{1:T}^N)\|_{\text{tvd}} \leq 2\gamma.$$

The (simple) proof of the lemma is given in Appendix B.

We now give

Proof of Theorem 2.1: Follows by putting together Lemma 6.3 and Lemma 6.2, as we show below.

Assume by contradiction that for any C_0 there exists a protocol Π for **DIHP**(n, α, T) succeeding with probability at least $2/3$ and satisfying $|\Pi| < n/(C_0/\varepsilon)^{C_0/\varepsilon^2}$. Take $C_0 = 10^{30}$, since $T \leq 10^{14}/\varepsilon^2$, we have

$$|\Pi| < n/(C_0/\varepsilon)^{C_0/\varepsilon^2} \leq n/(10^{30}/\varepsilon)^{10^{16}T} \leq n/(\sqrt{10T})^{10^{16}T} < n/(10T)^{10^9T}.$$

Without loss of generality we may assume that $s = |\Pi|$ satisfies $s \geq \sqrt{n}$, and hence we can apply Lemma 6.2 since $\alpha = 10^{-11} \in (0, 10^{-10})$ and $T \in [10, \ln n]$ for n large enough. We choose \mathcal{E}_t to be the events whose existence is guaranteed by Lemma 6.2. By Lemma 6.2 one has, conditioned on \mathcal{E}_t , for every fixed M_t , that

$$\|M_t X^* - UNIF(\{0, 1\}^{M_t})\|_{tvd} \leq \delta = 1/(1000T) \quad (32)$$

for every $t = 1, \dots, T$. We claim that these events satisfy the preconditions of Lemma 6.3. Indeed, recalling that $S_t^Y = r_t(M_{1:t}, S_{1:t-1}^Y, M_t X^*)$ (see Definition 4.1), we get for any fixed S_{t-1}^Y and $M_{1:t}$ satisfying \mathcal{E}_{t-1}

$$\begin{aligned} \|S_t^Y - r_t(M_{1:t}, S_{1:t-1}^Y, U_t)\|_{tvd} &= \|r_t(M_{1:t}, S_{1:t-1}^Y, M_t X^*) - r_t(M_{1:t}, S_{1:t-1}^Y, U_t)\|_{tvd} \\ &\leq \|M_t X^* - U_t\|_{tvd}, \end{aligned} \quad (33)$$

where we applied Claim 3.7 in the last transition with $f = r_t$, $W = (M_{1:t}, S_{1:t-1}^Y)$ and $X = M_t X^*$ and $Y = U_t$. We stress the fact that here W is deterministic and so the tvds are over the randomness of $X^* \sim \text{Uniform}(\mathbf{B}_{t-1})$, $U_t \sim \text{Uniform}(\{0, 1\}^{M_t})$.

Since by Lemma 6.2 we have $\Pr[\bar{\mathcal{E}}_t | \mathcal{E}_{t-1}] \leq 10\delta$ and $\|M_t X^* - U_t\|_{tvd} \leq \delta$, we can then apply Lemma 6.3 with $\gamma = 10\delta T = 1/100$ to deduce that

$$\|(M_{1:T}, S_{1:T}^Y) - (M_{1:T}, S_{1:T}^N)\|_{tvd} \leq \gamma,$$

which means that it is not possible to distinguish between **YES** and **NO** cases with probability more than $(1 + \gamma)/2$ leading to a contradiction. \blacksquare

7 Proof of main technical lemma (Lemma 6.1)

The main result of this section is a proof of Lemma 6.1, restated here for convenience of the reader:

Lemma 6.1 (Restated) *For every $n, C, s^*, \alpha, \delta$ that satisfy conditions*

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4 \quad \text{(P5)} \delta \in (n^{-1}, 1/2),$$

every $\mathbf{B} \subseteq \{0, 1\}^n$, if \mathbf{B} is (C, s^) -bounded and M is a uniformly random matching of size αn , the following conditions hold with probability at least $1 - 5\delta$ over the choice of M .*

For every $\mathbf{A}_{\text{reduced}} \subseteq \{0, 1\}^M$ such that $|\mathbf{A}_{\text{reduced}}|/2^{\alpha n} \geq 2^{-s^}$, if $\mathbf{A} = \{x \in \{0, 1\}^n : Mx \in \mathbf{A}_{\text{reduced}}\}$, then $\mathbf{B}' := \mathbf{B} \cap \mathbf{A}$ is $((10^9/\delta)C, s^*)$ -bounded.*

We also restate the definition of (C, s^*) -boundedness and the definition of $\mathbf{U}_{C, s^*}(\ell)$ for convenience.

Definition 4.3 *(C, s^*) -bounded set; restated) Let $\mathbf{B} \subset \{0, 1\}$ with indicator function h . We say that \mathbf{B} (or h) is (C, s^*) -bounded if*

- For all $\ell \leq s^*$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}_{\mathbf{B}}}(v) \right| \leq \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell;$$

- For all $s^* < \ell < \frac{n}{C^2}$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{\mathbb{1}_{\mathbf{B}}}(v) \right| \leq \left(\frac{C^2 n}{\ell} \right)^{\ell/2}.$$

As per (5), defining the function $\mathbf{U}_{C,s^*}(\ell)$ by

$$\mathbf{U}_{C,s^*}(\ell) = \begin{cases} 1 & \ell = 0; \\ \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell & \ell \in [1 : s^*]; \\ \left(\frac{C^2 n}{\ell} \right)^{\ell/2} & \ell > s^*, \end{cases}$$

we are able to simplify notation, ensuring that an indicator function h is (C, s^*) bounded if and only if for all $\ell < n/C^2$ we have

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{h}(v) \right| \leq \mathbf{U}_{C,s^*}(\ell).$$

Proof outline. The lemma starts with the assumption that a subset \mathbf{B} of $\{0, 1\}^n$ is (C, s^*) -bounded, and proves that the intersection $\mathbf{B}' = \mathbf{B} \cap \mathbf{A}$ with a subset \mathbf{A} of $\{0, 1\}^n$ (which should be thought of as the typical message corresponding to a player who receives labels MX for X selected uniformly at random from \mathbf{B}) is $(10^9 C, s^*)$ -bounded. It is instructive to contrast the bounds implied by the set \mathbf{B} being (C, s^*) -bounded (see Definition 4.3 and eq. (5), also restated above) with corresponding bounds for our component growing algorithm presented in Section 5 (see Section 5.3). Intuitively, the rhs in the definition of (C, s^*) -boundedness above shows that the amount of Fourier mass that a general protocol can have at some level ℓ is upper bounded by the amount of mass that the component growing protocol with slightly increased budget, namely $\sqrt{s^*n} = s^* \cdot \sqrt{n/s^*} \gg s^*$, can have at levels $\ell \in [1, s^*]$. This increase from s^* to $\sqrt{s^*n}$ is due to our conversion on the ℓ_2^2 norm of the Fourier transform to bounds on the ℓ_1 norm.

The proof of Lemma 6.1 is based on the convolution theorem, and follows quite closely the outline presented in Section 5. The main part of the proof that goes beyond the outline presented in Section 5 is the analysis of contribution to and from weights higher than s^* (specifically, weight levels in $[s^*, n/(2C^2)]$ and $[n/(2C^2), n]$), as such weight levels carry zero Fourier mass for the simple component growing protocol from Section 5. The corresponding analysis is carried out in Lemma 7.1 (mass transfer **to** low weights, possibly from low, intermediate or high weight coefficients) and Lemma 7.9 (mass transfer **to** intermediate weight coefficients). At the same time, we note that most of the final contribution to the Fourier transform of the final function comes from ‘low’ weight levels, namely from $\ell \in [1, s^*]$ (see Lemma 7.3), similarly to the component growing protocol in Section 5, and the dominant bound is thus given by Lemma 7.1.

Proof of Lemma 6.1: By Lemma 4.7 with probability at least $1 - \delta$ over the choice of M we have

$$\frac{2^n}{|\mathbf{B}'|} \leq \frac{2^n}{|\mathbf{B}|} \cdot \frac{2^n}{|\mathbf{A}|} \cdot \frac{1}{1 - \delta}.$$

Indeed we just need to take q the indicator function of $A_{reduced}$ (see the proof of 29 in Lemma 6.2). So we show that conditioned on this the set \mathbf{B}' is $((10^9/\delta)C, s^*)$ -bounded with probability at least $1 - 4\delta$.

We let $h := \mathbf{1}_B$, $f := \mathbf{1}_A$, $h' := h \cdot f = \mathbf{1}_{B'}$. Proving that h' is $((10^9/\delta)C, s^*)$ -bounded involved upper bounding the ℓ_1 norm of Fourier coefficients at various levels 2ℓ , $\ell \in [1, n/(2C^2)]$. Convolution theorem (1) together with triangle inequality give

$$\begin{aligned}
\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widehat{h}'(v) \right| &\leq \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \sum_{w \in \{0,1\}^n} \left| \widehat{h}(v \oplus w) \widehat{f}(w) \right| \\
&= \sum_{\substack{v \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \sum_{w \in \{0,1\}^n} \left| \widehat{h}(v) \widehat{f}(w) \right| \\
&= \sum_{v \in \{0,1\}^n} \left| \widehat{h}(v) \right| \sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \left| \widehat{f}(w) \right|.
\end{aligned} \tag{34}$$

We need to bound $\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{h}'(v) \right| = \frac{2^n}{|\mathbf{B}'|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widehat{h}'(v) \right|$, but it turns out to be more useful to upper bound $\frac{2^n}{|\mathbf{B}'|} \cdot \frac{2^n}{|\mathbf{A}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widehat{h}'(v) \right|$ which is within a factor of $(1 - \delta)$ of what we need. Multiplying both sides of (34) by $\frac{2^n}{|\mathbf{B}'|} \cdot \frac{2^n}{|\mathbf{A}|}$, we get

$$\begin{aligned}
\frac{2^n}{|\mathbf{A}|} \cdot \frac{2^n}{|\mathbf{B}'|} \cdot \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widehat{h}(v) \right| &\leq \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \sum_{w \in \{0,1\}^n} \left| \widetilde{h}(v \oplus w) \widetilde{f}(w) \right| \\
&= \sum_{\substack{v \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \sum_{w \in \{0,1\}^n} \left| \widetilde{h}(v) \widetilde{f}(w) \right| \\
&= \sum_{v \in \{0,1\}^n} \left| \widetilde{h}(v) \right| \sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \left| \widetilde{f}(w) \right|.
\end{aligned}$$

Taking expectation over the choice of M and recalling that $2^n/|\mathbf{B}'| \leq (2^n/|\mathbf{B}|) \cdot (2^n/|\mathbf{A}|)/(1 - \delta)$ we get

$$\frac{2^n}{|\mathbf{B}'|} \mathbb{E}_M \left[\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \left| \widetilde{h}'(v) \right| \right] \leq \frac{1}{1 - \delta} \cdot \sum_{v \in \{0,1\}^n} \left| \widetilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \left| \widetilde{f}(w) \right| \right]. \tag{35}$$

By Lemma 7.1 we have for all $\ell \in [1, s^*]$, using the fact that $|\mathbf{A}|/2^n = |\mathbf{A}_{reduced}|/2^{\alpha n} \geq 2^{-s^*}$ by assumption of the lemma, we get

$$\sum_{v \in \{0,1\}^n} \left| \widetilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \left| \widetilde{f}(w) \right| \right] \leq \left(\frac{10^9 C \sqrt{s^* n}}{\ell} \right)^\ell,$$

and by Lemma 7.9 we have for all $\ell \in [s^*, n/(2C^2)]$, using the fact that $|\mathbf{A}|/2^n = |\mathbf{A}_{reduced}|/2^{\alpha n} \geq 2^{-s^*}$ by assumption of the lemma,

$$\sum_{v \in \{0,1\}^n} \left| \widetilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w|=2\ell}} \left| \widetilde{f}(w) \right| \right] \leq \left(\frac{(10^9 C)^2 n}{\ell} \right)^\ell.$$

Putting these bounds together and recalling the definition of $\mathbf{U}_{10^9 C, s^*}(\ell)$ (see (5), as well as the definition restated above) we have for each $\ell \in [1 : n/(2C^2)]$ and all sets $\mathbf{A}_{reduced}$

$$\sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w| = 2\ell}} \left| \tilde{f}(w) \right| \right] \leq \mathbf{U}_{10^9 C, s^*}(\ell).$$

Thus, by Markov's inequality the probability over the choice of M that

$$\sum_{v \in \{0,1\}^n} \left| \tilde{h}(v) \right| \sum_{\substack{w \in \{0,1\}^n \\ |v \oplus w| = 2\ell}} \left| \tilde{f}(w) \right| \geq \mathbf{U}_{10^9 C/\delta, s^*}(\ell)$$

for some $\ell > 0$ and some $\mathbf{A}_{reduced}$ is at most

$$\sum_{\ell=1}^{n/(2C^2)} \frac{\mathbf{U}_{10^9 C, s^*}(\ell)}{(1-\delta) \cdot \mathbf{U}_{10^9 C/\delta, s^*}(\ell)} = \frac{1}{1-\delta} \cdot \sum_{\ell=1}^{n/(2C^2)} \delta^\ell \leq \frac{\delta}{(1-\delta)^2} \leq 4\delta.$$

Recalling that $2^n/|\mathbf{B}'| \leq (2^n/|\mathbf{B}|) \cdot (2^n/|\mathbf{A}|)/(1-\delta)$ happens with probability at least $1-\delta$ we conclude that \mathbf{B}' is $((10^9/\delta)C, s^*)$ -bounded with probability at least $1-5\delta$. \blacksquare

7.1 Bounding mass transfer to low weight Fourier coefficients

The goal of this subsection is to prove Lemma 7.1 below.

Proof outline. The lemma starts with the assumption that a subset \mathbf{B} of $\{0, 1\}^n$ is (C, s^*) -bounded and upper bounds the ℓ_1 norm of the convolution of the normalized Fourier transform \tilde{h} of \mathbf{B} with the normalized Fourier transform \tilde{f} of a subset \mathbf{A} of $\{0, 1\}^n$ (which should be thought of as the typical message corresponding to a player who receives labels MX for X selected uniformly at random from \mathbf{B}) on coefficients with Hamming weight 2ℓ , for $\ell \in [1, s^*]$.

The proof follows the outline presented in Section 5 (see Section 5.3). Indeed, we first partition the pairs v (a subset of the vertices, or a coefficient of \tilde{h}), and w (a subset of the edges of the matching that \tilde{f} is supported on) classes depending on the number of **boundary**, **internal** and **external** edges of the matching M involved in the mass transfer. See the proof below for the definition of these types of edges M , Fig. 2 for an illustration and Section 5.3 for an illustration of these notions on the simple example of the component growing protocol from Section 5. We then reduce the problem of bounding the Fourier mass to the problem of verifying certain sums that reflect the tradeoffs between the amount of mass at various Fourier levels and basic combinatorics on matchings. A crucial parameter that governs these calculations is the probability, over the choice of a uniformly random matching, that this matching has a given number of internal and boundary edges with respect to a fixed subset of the vertices (i.e. a fixed Fourier coefficient). Convenient upper bounds on such quantities are provided by Lemma 3.13 and Lemma 3.14. Using these lemmas, we reduce the problem to verifying several combinatorial bounds (which at this point are disjoint from any Fourier analytic considerations), which is done in Lemma 7.2, Lemma 7.3, Lemma 7.5 and Lemma 7.7 from Section 7.1.1

Overall, the proof of Lemma 7.1 follows quite closely the outline presented in Section 5, with the main that goes beyond this outline being the analysis of contribution to and from weights higher than s^* (specifically, weight levels in $[s^*, n/(2C^2)]$ and $[n/(2C^2), n]$), as such weight levels carry zero Fourier

mass for the simple component growing protocol from Section 5. This analysis is provided by Lemmas 7.2, Lemma 7.3, Lemma 7.5 and Lemma 7.7 from Section 7.1.1. At the same time, we note that most of the final contribution to the Fourier transform of the final function comes from ‘low’ weight levels, namely from $\ell \in [1, s^*]$ (see Lemma 7.3), similarly to the component growing protocol in Section 5.

Lemma 7.1 (Mass transfer to low weights) *For every $n, s^*, C > 1, \alpha \in (0, 1)$ that satisfy conditions*

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

if $\mathbf{B} \subseteq \{0, 1\}^n$ is (C, s^) -bounded and M is a uniformly random matching on $[n]$ of size αn , the following conditions hold. For every $\mathbf{A}_{reduced} \subseteq \{0, 1\}^M$, if $\mathbf{A} = \{x \in \{0, 1\}^n : Mx \in \mathbf{A}_{reduced}\}$ and f is the indicator of \mathbf{A} , if $|\mathbf{A}|/2^n \geq 2^{-s^*}$, then for every $\ell \in [1, s^*]$*

$$\sum_{v \in \{0, 1\}^n} \left| \tilde{h}(v) \right| \mathbb{E}_M \left[\sum_{\substack{z \in \{0, 1\}^n \\ |v \oplus z| = 2\ell}} \left| \tilde{f}(z) \right| \right] \leq \left(\frac{10^9 C \sqrt{s^* n}}{\ell} \right)^\ell.$$

Proof: We first note that $\tilde{f}(z) \neq 0$ only if $z = Mw$ for $w \in \{0, 1\}^M$, and thus for every $v \in \{0, 1\}^n$ one has $\mathbb{E}_M \left[\sum_{\substack{z \in \{0, 1\}^n \\ |v \oplus z| = 2\ell}} \left| \tilde{f}(z) \right| \right] = \mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} \left| \tilde{f}(Mw) \right| \right]$. We also note that for every $v \in \{0, 1\}^n$ such that $|v \oplus Mw| = 2\ell$ one has $|v| \leq |v \oplus Mw| + |Mw| \leq \ell + |Mw| \leq s^* + 2\alpha n \leq n/200$ since $s^* \leq n/10^9$ and $\alpha < 1/400$ by assumptions **(P1)**, **(P2)** and **(P3)**. Thus, the only terms with a nonzero contribution to the sum that we need to bound are $v \in \{0, 1\}^n$ with $|v| = 2k \leq n/200$. Thus, it suffices to bound, for a parameter $k \in [0 : n/100]$ and $v \in \{0, 1\}^n$ with $|v| = 2k$, the quantity

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} \left| \tilde{f}(Mw) \right| \right]. \quad (36)$$

We will later (see (41) below) combine our bounds over all $k \in [0 : n/100]$ to obtain the result of the lemma.

Let $\text{Int}_M(v) = \{e_1^{\text{int}}, e_2^{\text{int}}, \dots\}$ be the set of edges $e = (a, b) \in M$ that match points of v , i.e. with $a, b \in v$. Let $\partial_M(v) = \{e_1^{\text{bound}}, e_2^{\text{bound}}, \dots\}$ be the set of boundary edges, i.e. edges $e = (a, b) \in M$ with $a, b \in v$. Let $\text{Ext}_M(v) = \{e_1^{\text{ext}}, e_2^{\text{ext}}, \dots\}$ be the set of external edges, i.e. edges $e = (a, b) \in M$ with $a, b \in [n] \setminus v$.

We decompose the sum (36) according to the number of boundary edges in Mw :

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} \left| \tilde{f}(Mw) \right| \right] = \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{w \in \{0, 1\}^M} \mathbb{1}_{\{w \cap \partial_M(v) = S\}} \mathbb{1}_{\{|v \oplus Mw| = 2\ell\}} \left| \tilde{f}(Mw) \right| \right]. \quad (37)$$

We now rewrite the latter indicator function. For a subset $S \subseteq \partial_M(v)$ define $w_S \in \{0, 1\}^M$ as the set of all internal edges $\text{Int}_M(v)$ and all edges in S . We then have $|v \oplus Mw_S| = 2k - 2|\text{Int}_M(v)|$, since adding a boundary edge to v does not change the Hamming weight, and adding an internal edge reduces it by 2. Also note that $|w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)$. Indeed, $|v \oplus Mw_S| = 2k - 2|\text{Int}_M(v)|$, $|v \oplus w| = 2\ell$, w can be

obtained from w_S be removing internal edges and adding external edges and both of these changes increase $|Mw \oplus Mw_S|$ by 2. These observations together with (37) yield the following upper bound on (36):

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} \mathbb{1}_{\{w_S = S\}} \cdot |\tilde{f}(Mw)| \right] \\ &\leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}(Mw)| \right], \end{aligned} \quad (38)$$

where we dropped the indicator function in going from the first line to the second line above. We now apply Lemma 3.3 to the inner summation on the rhs of (38). We now let f' denote the indicator of $\mathbf{A}_{reduced}$ and note that $\tilde{f}'(w) = \tilde{f}(Mw)$ for all $w \in \{0,1\}^M$. Since $\mathbf{A}_{reduced}/2^{\alpha n} = |\mathbf{A}|/2^n \geq 2^{-s^*}$, this, in turn, can be bounded by Lemma 3.3 applied to the set $\mathbf{A}_{reduced} \subset \{0,1\}^M$ (i.e. $m = \alpha n$) with $q = \ell - k + \text{Int}_M(v)$ and $y = w_S$. We obtain

$$\begin{aligned} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}(Mw)| &= \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}'(w)| \\ &\leq \sqrt{\left(\frac{\alpha n}{\ell - k + |\text{Int}_M(v)|} \right) \left(\frac{64s^*}{\ell - k + |\text{Int}_M(v)|} \right)^{\ell - k + |\text{Int}_M(v)|}} \\ &\leq \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + |\text{Int}_M(v)|} \right)^{\ell - k + |\text{Int}_M(v)|}. \end{aligned} \quad (39)$$

Summing over all possible subsets $S \subseteq \partial_M(v)$, we then infer

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + |\text{Int}_M(v)|} \right)^{\ell - k + |\text{Int}_M(v)|} \right] \\ &\leq \mathbb{E}_M \left[2^{|\partial_M(v)|} \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + |\text{Int}_M(v)|} \right)^{\ell - k + |\text{Int}_M(v)|} \right], \end{aligned}$$

where in going from line 1 to line 2 in the equation above we used the fact that the bound is independent of the set S and upper bounded the summation by multiplying by the number of such sets S , i.e. $2^{|\partial_M(v)|}$.

We now bound the sum on the last line above. We have

$$\begin{aligned} &\mathbb{E}_M \left[2^{|\partial_M(v)|} \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + |\text{Int}_M(v)|} \right)^{\ell - k + |\text{Int}_M(v)|} \right] \\ &= \mathbb{E}_M \left[\sum_{i=0}^k \sum_{b=0}^{2k} 2^b \mathbb{1}_{\{|\partial_M(v)|=b \text{ and } |\text{Int}_M(v)|=i\}} \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right] \\ &= \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} 2^b q(k, i, b, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \end{aligned}$$

In going from line 2 to line 3 above we used the fact that by Definition 3.10 for every $v \in \{0, 1\}^n$ with $|v| = 2k$ one has $\mathbb{E}_M[\mathbb{1}_{\{|\partial_M(v)|=b \text{ and } |\text{Int}_M(v)|=i\}}] = q(k, i, b, n)$, i.e. $q(k, i, b, n)$ is the probability that a uniformly random matching M of size αn is such that i edges of M match points of v (i.e. $|\text{Int}_M(v)| = i$) and b edges of M are boundary edges (i.e. $|\partial_M(v)| = b$). Note that we sum over b between 0 and $2k$, as the number of boundary edges of v with respect to M is bounded by its Hamming weight $2k$. Similarly, the number of internal edges i cannot be larger than $|v|/2 = k$, and must be at least $k - \ell$ in order for the binomial coefficient $\binom{\alpha n}{\ell - k + i}$ on the first line above to be nonzero (combinatorially, this means that in order for $v + Mw$ to have weight 2ℓ for some $w \in \{0, 1\}^M$ one must have $|\text{Int}_M(v)| \geq k - \ell$). Recall that $|k - \ell|_+ = \max\{0, k - \ell\}$.

We thus have that for a fixed $v \in \{0, 1\}^n$ with $|v| = 2k$

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \leq \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} q(k, i, b, n) 2^b \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i}$$

Recall that by Lemma 3.14 we have $q(k, i, b, n) \leq q(k, i, n) 20^{-b} 4^{k-i}$ (note that $k \leq n/100$ and $\alpha < 1/100$ by assumption **(P2)**, so the preconditions of the lemma are satisfied). We thus get, substituting into the rhs of the equation above,

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} 4^{k-i} q(k, i, n) 2^{-b} \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \\ &\leq 2 \sum_{i=|k-\ell|_+}^k 4^{k-i} q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \\ &\leq 2 \sum_{i=|k-\ell|_+}^k 4^\ell q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \\ &\leq 8^\ell \sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i}. \end{aligned} \tag{40}$$

Equipped with the upper bound (40), we now sum over possible v of Hamming weight $k \in [0 : n/100]$:

$$\begin{aligned} \sum_{v \in \{0, 1\}^n} |\tilde{h}(v)| \mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \\ \leq 8^\ell \sum_{k=0}^{n/100} \left(\sum_{\substack{v \in \{0, 1\}^n \\ |v| = 2k}} |\tilde{h}(v)| \right) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \end{aligned} \tag{41}$$

Recall that we want to bound the expression above for $\ell \leq s^*$. We now split the sum depending on how

large k is. We consider three intervals. For $k \in [0 : 100s^*]$ and $k \in [100s^* + 1 : n/C^2]$ we use the bound

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)| \leq \mathbf{U}_{C,s^*}(k),$$

but for $k \in [n/C^2 + 1 : n/100]$ we use the bound that follows by Cauchy-Schwarz together with Parseval's inequality: sum of squares of all normalized Fourier coefficients is $2^n/|\mathbf{B}| \leq 2^{s^*}$ due to (C, s^*) -boundedness of \mathbf{B} , and so

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)| \leq \sqrt{\left(\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)|^2 \right)} \cdot \binom{n}{2k} \leq \sqrt{2^{s^*} \binom{n}{2k}}.$$

Thus, it is sufficient to derive strong upper bounds on S_0, S_1, S_2, S_3 that we define below:

$$S_0 := \sum_{k=0} \mathbf{U}_{C,s^*}(k) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^*\alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \quad (42)$$

$$S_1 := \sum_{k=1}^{100s^*} \mathbf{U}_{C,s^*}(k) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^*\alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \quad (43)$$

$$S_2 := \sum_{k=100s^*+1}^{n/C^2} \mathbf{U}_{C,s^*}(k) \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{s^*\alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \quad (44)$$

$$S_3 := \sum_{k=n/C^2+1}^{n/100} \sqrt{2^{s^*} \binom{n}{2k}} \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{s^*\alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \quad (45)$$

Lemma 7.2, Lemma 7.3, Lemma 7.5 and Lemma 7.7 show that $S_0 \leq \mathbf{U}_{15,s^*}(\ell)$, $S_1 \leq \mathbf{U}_{10^8 C, s^*}(\ell)$, $S_2 \leq \mathbf{U}_{10^7 C, s^*}(\ell)$, and $S_3 \leq 1$, which concludes the proof since

$$\begin{aligned} 8^\ell \left(\mathbf{U}_{15,s^*}(\ell) + \mathbf{U}_{10^8 C, s^*}(\ell) + \mathbf{U}_{10^7 C, s^*}(\ell) + 1 \right) &= 8^\ell \left((15/C)^\ell + 10^{8\ell} + 10^{7\ell} + 1 \right) \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell \\ &\leq \left(\frac{10^9 C\sqrt{s^*n}}{\ell} \right)^\ell \end{aligned}$$

We note that the dominant contribution comes from S_1 , i.e. from mass transfer from low weights to low weights, and the amount of mass transfer is consistent with what we would expect for the component growing protocol with the slightly increased communication budget of $\sqrt{s^*n}$ (as opposed to s^*). ■

7.1.1 Bounding S_0, S_1, S_2, S_3 (technical lemmas)

The first lemma bounds transfer of mass from weight zero to low weights:

Lemma 7.2 (Mass transfer from weight zero to low weights) For every n, s^* , every $\ell \in [1 : s^*]$, if parameters C, α, s^*, n satisfy

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

then

$$\sum_{k=0} \mathbf{U}_{C, s^*}(k) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \leq \mathbf{U}_{15, s^*}(\ell).$$

Proof: The sum only contains the summand corresponding to $i = k = 0$; since $q(0, 0, n) = 1$, we get at most $\left(\frac{15\sqrt{s^* \alpha n}}{\ell} \right)^\ell \leq \mathbf{U}_{15, s^*}(\ell)$. ■

The lemma below bounds transfer of mass from low weights to low weights:

Lemma 7.3 (Mass transfer from low weights to low weights) For every n, s^* , every $\ell \in [1 : s^*]$, if parameters C, α, s^*, n satisfy

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

then

$$\sum_{k=1}^{100s^*} \mathbf{U}_{C, s^*}(k) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \leq \mathbf{U}_{10^8 C, s^*}(\ell).$$

Remark 7.4 We note that the contribution of ‘low’ to ‘low’ weights, bounded by Lemma 7.3 is the dominant one in the application in Lemma 7.1. The calculations in this lemma are quite similar to those provided in the overview of the analysis for the component growing protocol in Section 5 (see Section 5.3.2).

Proof of Lemma 7.3: Note that for $k \leq 100s^*$ we have by definition of $\mathbf{U}_{C, s^*}(k)$ (see (5))

$$\begin{aligned} \mathbf{U}_{C, s^*}(k) &\leq \max \left\{ \left(\frac{C\sqrt{s^* n}}{k} \right)^k, \left(\frac{C^2 n}{k} \right)^{k/2} \right\} \\ &= \max \left\{ \left(\frac{C\sqrt{s^* n}}{k} \right)^k, \left(\frac{C\sqrt{s^* n} \cdot \sqrt{k/s^*}}{k} \right)^k \right\} \\ &\leq \left(\frac{10C\sqrt{s^* n}}{k} \right)^k. \end{aligned}$$

We now define, as in (43),

$$S_1 := \sum_{k=1}^{100s^*} \mathbf{U}_{C, s^*}(k) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right)$$

We now get, using the fact that $q(k, i, n) = \binom{\alpha n}{i} \binom{n-2\alpha n}{2(k-i)} \binom{n}{2k}^{-1}$ by Lemma 3.13,

$$\begin{aligned}
S_1 &\leq \sum_{k=1}^{100s^*} \left(\frac{10C\sqrt{s^*n}}{k} \right)^k \sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{15\sqrt{\alpha s^*n}}{\ell+i-k} \right)^{\ell+i-k} \\
&= \sum_{k=1}^{100s^*} \sum_{i=|k-\ell|_+}^k \left(\frac{10C\sqrt{s^*n}}{k} \right)^k \binom{\alpha n}{i} \binom{n-2\alpha n}{2(k-i)} \binom{n}{2k}^{-1} \left(\frac{15\sqrt{\alpha s^*n}}{\ell+i-k} \right)^{\ell+i-k} \\
&\leq \left(\frac{10C\sqrt{s^*n}}{\ell} \right)^\ell \sum_{k=1}^{100s^*} \sum_{i=|k-\ell|_+}^k (10C)^{k-\ell} \left(\sqrt{s^*n} \right)^i n^{i+2(k-i)-2k} \alpha^{i+(\ell+i-k)/2} \cdot \Gamma,
\end{aligned}$$

where

$$\Gamma := \frac{\ell^\ell (2k)!}{i!(2k-2i)!k^k(\ell+i-k)^{\ell+i-k}} 2^{2k} 15^{\ell+i-k} \leq 10^{3(\ell+k+i)}$$

by Lemma D.5. Indeed, we first apply Lemma D.5 with $m = 2$, $a_1 = \ell$, $a_2 = 2k$ to get $\ell^\ell \cdot 2k! \leq (2k + \ell)^{2k+\ell}$, and then apply Lemma D.5 with $m = 4$, $a_1 = i$, $a_2 = 2k - 2i$, $a_3 = k$, $a_4 = \ell + i - k$, obtaining $i!(2k - 2i)!k^k(\ell + i - k)^{\ell+i-k} \geq ((2k + \ell)/12)^{2k+\ell}$. This yields an upper bound of $\Gamma \leq 12^{2k+\ell} \cdot 2^{2k} \cdot 15^{\ell+i-k} \leq 10^{3(\ell+k+i)}$.

The exponent of n is $-i$ which together with the $(\sqrt{s^*n})^i$ factor gives us $(s^*/n)^{i/2}$. So we have

$$\begin{aligned}
S_1 \cdot \left(\frac{10C\sqrt{s^*n}}{\ell} \right)^{-\ell} &\leq \sum_{k=1}^{100s^*} \sum_{i=|k-\ell|_+}^k 10^i C^{k-\ell} (s^*/n)^{i/2} 10^{3(\ell+k+i)} \alpha^{i+(\ell+i-k)/2} \\
&\leq \sum_{k=1}^{100s^*} \sum_{i=|k-\ell|_+}^k 10^i C^{k-\ell-i} \left(C\sqrt{s^*/n} \right)^i 10^{6\ell+6i} \alpha^{i+(\ell+i-k)/2} \\
&\leq 10^{6\ell} \sum_{i=0}^{\infty} \sum_{k=i}^{i+\ell} (\sqrt{\alpha}/C)^{\ell+i-k} \left(10^7 C \alpha \sqrt{s^*/n} \right)^i \\
&\leq 10^{6\ell} \sum_{i=0}^{\infty} \left(10^7 C \alpha \sqrt{s^*/n} \right)^i \sum_{k=i}^{i+\ell} (\sqrt{\alpha}/C)^{\ell+i-k} \\
&\leq 10^{6\ell} \sum_{i=0}^{\infty} 2 \left(10^7 C \alpha \sqrt{s^*/n} \right)^i \\
&\leq 10^{7\ell}
\end{aligned}$$

In going from line 4 to line 5 we used the fact that $\sqrt{\alpha}/C < 1/2$ by assumptions **(P1)** and **(P2)**. Similarly, in going from line 5 to line 6 we used the fact that $10^7 C \alpha \sqrt{s^*/n} < 1/2$ by assumptions **(P1)**, **(P2)**, **(P3)**. ■

The next lemma bounds transfer of mass from intermediate weights to low weights:

Lemma 7.5 (Mass transfer from intermediate to low weights) *For every n, s^* , every $\ell \in [1 : s^*]$, if C, α, s^*, n satisfy*

$$(\mathbf{P1}) \alpha < 10^{-10} \quad (\mathbf{P2}) C > 10^6 \quad (\mathbf{P3}) s^* < \frac{n}{10^9 C^3} \quad (\mathbf{P4}) n > 10^9 C^4,$$

then

$$\sum_{k=100s^*+1}^{n/C^2} \mathbf{U}_{C, s^*}(k) \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right) \leq \mathbf{U}_{10^7 C, s^*}(\ell).$$

Remark 7.6 A careful inspection of the proof of Lemma 7.5 reveals that the upper bound is in fact somewhat stronger than what Lemma 7.1 needs. This is mostly due to the fact that Lemma 7.5 analyzes contribution of weights k starting at $100s^*$ to weights $\ell \leq s^*$. Since the strengthening is not consequential for the analysis, we prefer to keep the bound in the present form for simplicity. However, it is interesting to note that the ‘low weight’ regime (i.e. the regime of Lemma 7.3), for which the component growing protocol from Section 5 is a reasonable illustration, provides the dominant contribution to the Fourier spectrum.

Proof of Lemma 7.3: Since in this sum we have $k > 100s^*$ we must have $i \geq k - \ell > 99s^*$ and also $i \geq k - \ell > k/2$. For $k \in [100s^*, n/C^2]$ we have by (5)

$$\mathbf{U}_{C,s^*}(k) \leq \left(\frac{C^2 n}{k}\right)^{k/2} = \left(\sqrt{\frac{k}{s^*}}\right)^k \left(\frac{C\sqrt{s^* n}}{k}\right)^k. \quad (46)$$

We now define, as in (44),

$$S_2 := \sum_{k=100s^*+1}^{n/C^2} \mathbf{U}_{C,s^*}(k) \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i}\right)^{\ell - k + i} \right).$$

Putting this together with (46), we get

$$\begin{aligned} S_2 &\leq \sum_{k=100s^*+1}^{n/C^2} \left(\sqrt{\frac{k}{s^*}}\right)^k \left(\frac{C\sqrt{s^* n}}{k}\right)^k \sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{\alpha s^* n}}{\ell + i - k}\right)^{\ell + i - k} \\ &= \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k \left(\sqrt{\frac{k}{s^*}}\right)^k \left(\frac{C\sqrt{s^* n}}{k}\right)^k \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k-i)} \binom{n}{2k}^{-1} \left(\frac{15\sqrt{\alpha s^* n}}{\ell + i - k}\right)^{\ell + i - k} \\ &\leq \left(\frac{C\sqrt{s^* n}}{\ell}\right)^\ell \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k \left(\sqrt{\frac{k}{s^*}}\right)^k C^{k-\ell} \left(\sqrt{s^* n}\right)^i n^{i+2(k-i)-2k} \alpha^{i+(\ell+i-k)/2} \cdot \Gamma \end{aligned}$$

where

$$\Gamma := \frac{\ell^\ell (2k)!}{i!(2k-2i)!k^k(\ell+i-k)^{\ell+i-k}} 2^{2k} 15^{\ell+i-k} \leq 10^{3(\ell+k+i)}$$

by Lemma D.5. Indeed, we first apply Lemma D.5 with $m = 2$, $a_1 = \ell$, $a_2 = 2k$ to get $\ell^\ell \cdot 2k! \leq (2k + \ell)^{2k+\ell}$, and then apply Lemma D.5 with $m = 4$, $a_1 = i$, $a_2 = 2k - 2i$, $a_3 = k$, $a_4 = \ell + i - k$, obtaining $i!(2k - 2i)!k^k(\ell + i - k)^{\ell+i-k} \geq ((2k + \ell)/12)^{2k+\ell}$. This yields an upper bound of $\Gamma \leq 12^{2k+\ell} \cdot 2^{2k} \cdot 15^{\ell+i-k} \leq 10^{3(\ell+k+i)}$.

The exponent of n is $-i$ which together with the $(\sqrt{s^* n})^i$ factor gives us $(s^*/n)^{i/2}$. So we have

$$\begin{aligned} S_2 \cdot \left(\frac{C\sqrt{s^* n}}{\ell}\right)^{-\ell} &\leq \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k \left(\sqrt{\frac{k}{s^*}}\right)^k C^{k-\ell} \left(\sqrt{s^*/n}\right)^i 10^{3(\ell+k+i)} \alpha^{i+(\ell+i-k)/2} \\ &\leq \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k \left(\sqrt{\frac{k}{s^*}}\right)^k C^{k-\ell-i} \left(C\sqrt{s^*/n}\right)^i 10^{6\ell+6i} \alpha^{i+(\ell+i-k)/2} \\ &\leq 10^{6\ell} \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k \left(\sqrt{\frac{k}{s^*}}\right)^k (C/\sqrt{\alpha})^{k-\ell-i} \left(10^4 C \alpha \sqrt{s^*/n}\right)^i \\ &\leq 10^{6\ell} \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k \left(10^4 C \alpha \sqrt{k/n}\right)^i \left(\sqrt{\frac{k}{s^*}}\right)^{k-i} \end{aligned} \quad (47)$$

We used the fact that $C/\sqrt{\alpha} > 1$ and $k - \ell - i \leq 0$ for i in the range of summation in going from line 3 to line 4 above. We now note that $10^4 C \alpha \sqrt{k/n} \leq 10^4 \alpha$ since $k \leq n/C^2$ in the range of summation, and $\left(\sqrt{\frac{k}{s^*}}\right)^{k-i} \leq \left(\sqrt{\frac{k}{s^*}}\right)^\ell \leq \left(\sqrt{\frac{k}{s^*}}\right)^{s^*}$ for i in the range of summation, since $\ell \leq s^*$ by assumption of the lemma. We thus get that, substituting these bounds into (47),

$$\begin{aligned}
S_2 \cdot \left(\frac{C\sqrt{s^*n}}{\ell}\right)^{-\ell} &\leq 10^{6\ell} \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k (10^4\alpha)^i \left(\sqrt{\frac{k}{s^*}}\right)^{s^*} \\
&\leq 10^{6\ell} \sum_{k=100s^*+1}^{n/C^2} \sum_{i=k-\ell}^k (10^4\alpha)^i e^{k/2e} \\
&\leq 10^{6\ell} \sum_{i=99s^*+1}^{n/C^2} \sum_{k=i}^{i+\ell} (10^4\alpha)^i e^{k/2e} \\
\text{(P1)} &\leq 10^{6\ell} \sum_{i=99s^*+1}^{n/C^2} \sum_{k=i}^{i+\ell} 10^{-i} \\
&\leq 10^{6\ell} \sum_{i=99s^*+1}^{n/C^2} 2\ell \cdot 10^{-i} \\
&\leq 10^{7\ell}
\end{aligned}$$

We used the fact that $(\sqrt{k/s^*})^s = (k/s^*)^{s^*/2} \leq e^{k/(2e)}$ in going from line 1 to line 2 above (this follows since $(k/s^*)^{s^*}$ is maximized when $s^* = k/e$). The transition from line 3 to line 4 follows since $i \geq k - \ell \geq k/2$ (as $\ell \leq s^*$ and $k \geq 100s^*$), and hence

$$(10^4\alpha)^i e^{k/2e} \leq (10^4\alpha)^i e^{i/e} = \left(10^4 e^{1/e} \alpha\right)^i \leq 10^{-i},$$

since $\alpha < 10^{-6}$ by assumption **(P1)**. The transition from line 5 to line 6 follows since $\ell \leq s^*$ by assumption. ■

Lemma 7.7 (Mass transfer from high weights to low weights) *For every n, s^* , every $\ell \in [1 : s^*]$, if parameters C, α, s^*, n satisfy*

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

$$\sum_{k=n/C^2}^{n/100} \sqrt{2^{s^*} \binom{n}{2k}} \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{s^* \alpha n}}{\ell - k + i}\right)^{\ell - k + i} \right) \leq 1.$$

Remark 7.8 *We note that the bound of Lemma 7.7 is much stronger than what Lemma 7.1 needs. Indeed, a much weaker bound of $\approx (\sqrt{s^*n}/\ell)^\ell$ would have been sufficient for our purposes. The contribution of high weights to low weights is significantly lower than the dominant terms ($\ell \in [1 : s^*]$ and $\ell \in [s^*, n/(2C^2)]$), see Lemma 7.3 and Lemma 7.5) provide, and hence we choose (rather arbitrarily) to prove the upper bound of 1.*

Proof of Lemma 7.7: We first define, as in (45),

$$S_3 := \sum_{k=n/C^2+1}^{n/100} \mathbf{U}_{C,s^*}(k) \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{15\sqrt{s^*\alpha n}}{\ell - k + i} \right)^{\ell - k + i} \right).$$

This allows us to write

$$\begin{aligned} S_3 &\leq \sum_{k=n/C^2+1}^{n/10} \sum_{i=k-\ell}^k \sqrt{2^{s^*} \binom{n}{2k}} \cdot q(k, i, n) \left(\frac{15\sqrt{\alpha s^* n}}{\ell + i - k} \right)^{\ell + i - k} \\ &\leq \sum_{k=n/C^2+1}^{n/10} \sum_{i=k-\ell}^k 4^{s^*} \binom{n}{2k}^{1/2} \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k-i)} \binom{n}{2k}^{-1} \binom{n}{\ell + i - k} \\ &\leq \sum_{k=n/C^2+1}^{n/10} \sum_{i=k-\ell}^k 4^{s^*} \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k-i)} \binom{n}{2k}^{-1/2} \binom{n}{\ell + i - k} \end{aligned}$$

The transition from line 1 to line 2 follows since $15\sqrt{\alpha s^* n} \leq n$ and $k \leq n/10 \leq n/2$ in the range of summation. Since $\ell + i - k \leq \ell \leq s^*$ and $s^* \leq n/4$ by assumptions **(P2)** and **(P3)**, we have $\binom{n}{\ell + i - k} \leq \binom{n}{s^*}$. At the same time, since $k - i \leq \ell \leq s^*$, we also have $\binom{n - 2\alpha n}{2(k-i)} \leq \binom{n}{2s^*}$. Using the bounds above as well as the fact that $(a/b)^b \leq \binom{a}{b} \leq (ea/b)^b$ for all integer $a, b \geq 0$, we get

$$\begin{aligned} S_3 &\leq \sum_{k=n/C^2+1}^{n/10} \sum_{i=k-\ell}^k 4^{s^*} \left(\frac{e\alpha n}{i} \right)^i \binom{n}{2k}^{-k} \binom{n}{s^*} \binom{n}{2s^*} \\ &\leq \sum_{k=n/C^2+1}^{n/10} \sum_{i=k-\ell}^k 4^i \left(\frac{e\alpha n}{i} \right)^i \left(\frac{n}{2i} \right)^{-i} \binom{n}{s^*}^3 \quad (\text{since } i \geq s^*) \\ &\leq \sum_{k=n/C^2+1}^{n/10} \sum_{i=k-\ell}^k (22\alpha)^i e^{3H(s^*/n) \cdot n} \quad (\text{since } \binom{n}{s^*} \leq e^{H(s^*/n)n} \text{ by Lemma D.3}) \\ &\leq \sum_{k=n/C^2+1}^{n/10} 2(22\alpha)^{k-\ell} e^{3H(s^*/n) \cdot n} \quad (\text{summing geometric series}) \\ &\leq 4(22\alpha)^{n/C^2 - s^*} e^{3H(s^*/n) \cdot n} \quad (\text{summing geometric series and using } \ell \leq s^*) \\ (\mathbf{P3}) &\leq 4(22\alpha)^{n/(2C^2)} e^{3H(1/C^3) \cdot n} \quad (\text{since } \ell \leq s^* \leq n/C^3) \\ (\mathbf{P1}, \mathbf{P2}) &\leq n^2 e^{-n/(4C^2)} \\ (\mathbf{P4}) &\leq 1 \end{aligned}$$

■

7.2 Bounding mass transfer to intermediate weight Fourier coefficients

The goal of this subsection is to prove Lemma 7.9 below.

Proof outline. The proof of the lemma is similar to that of Lemma 7.1, with different supporting combinatorial lemmas (Lemma 7.10 and Lemma 7.12 from Section 7.2.1). Unlike the proof of Lemma 7.1, which mostly deals with low weight coefficients, Lemma 7.9 analyzes high weight coefficients, which are zero for the simple component growing protocol from Section 5.

Lemma 7.9 For every $n, s^*, C, \alpha \in (0, 1)$ that satisfy conditions

$$(\mathbf{P1}) \alpha < 10^{-10} \quad (\mathbf{P2}) C > 10^6 \quad (\mathbf{P3}) s^* < \frac{n}{10^9 C^3} \quad (\mathbf{P4}) n > 10^9 C^4,$$

if $\mathbf{B} \subseteq \{0, 1\}^n$ is (C, s^*) -bounded and M is a uniformly random matching on $[n]$ of size αn , the following conditions hold. For every $\mathbf{A}_{reduced} \subseteq \{0, 1\}^M$, if $\mathbf{A} = \{x \in \{0, 1\}^n : Mx \in \mathbf{A}_{reduced}\}$ and f is the indicator of \mathbf{A} , if $|\mathbf{A}|/2^n \geq 2^{-s^*}$, then for all $\ell \in [s^*, n/(2C^2)]$ we have

$$\sum_{v \in \{0, 1\}^n} |\tilde{h}(v)| \mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^n \\ |v \oplus w| = 2\ell}} |\tilde{f}(w)| \right] \leq \left(\frac{(10^9 C)^2 n}{\ell} \right)^{\ell/2}.$$

Proof: We start similarly to the proof of Lemma 7.1.

We first note that $\tilde{f}(z) \neq 0$ only if $z = Mw$ for $w \in \{0, 1\}^M$, and thus for every $v \in \{0, 1\}^n$ one has $\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^n \\ |v \oplus w| = 2\ell}} |\tilde{f}(w)| \right] = \mathbb{E}_M \left[\sum_{\substack{z \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right]$. We also note that for any $w \in \{0, 1\}^M$ one has $|v| \leq |v \oplus Mw| + |Mw| \leq \ell + \alpha n \leq n/(2C^2) + \alpha n \leq n/200$ since $n/(2C^2) \leq n/400$ and $\alpha < 1/400$ by assumptions **(P1)**, **(P2)** and **(P3)**. Thus, the only terms with a nonzero contribution to the sum that we need to bound are $v \in \{0, 1\}^n$ with $|v| = 2k \leq n/200$. Thus, it suffices to bound, for a parameter $k \in [0 : n/100]$ and $v \in \{0, 1\}^n$ with $|v| = 2k$, the quantity

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^n \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right]. \quad (48)$$

We will later (see (53) below) combine our bounds over all $k \in [0 : n/100]$ to obtain the result of the lemma.

Let $\text{Int}_M(v) = \{e_1^{\text{int}}, e_2^{\text{int}}, \dots\}$ be the set of edges $e = (a, b) \in M$ that match points of v , i.e. with $a, b \in v$. Let $\partial_M(v) = \{e_1^{\text{bound}}, e_2^{\text{bound}}, \dots\}$ be the set of boundary edges, i.e. edges $e = (a, b) \in M$ with $a, b \in v$. Let $\text{Ext}_M(v) = \{e_1^{\text{ext}}, e_2^{\text{ext}}, \dots\}$ be the set of external edges, i.e. edges $e = (a, b) \in M$ with $a, b \in [n] \setminus v$.

We decompose the sum (48) according to the number of boundary edges in Mw :

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] = \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{w \in \{0, 1\}^M} \mathbb{1}_{\{w \cap \partial_M(v) = S\}} \mathbb{1}_{\{|v \oplus Mw| = 2\ell\}} |\tilde{f}(Mw)| \right]. \quad (49)$$

We now rewrite the latter indicator function. For a subset $S \subseteq \partial_M(v)$ define $w_S \in \{0, 1\}^M$ as the set of all internal edges $\text{Int}_M(v)$ and all edges in S . We then have $|v \oplus Mw_S| = 2k - 2|\text{Int}_M(v)|$, since adding a boundary edge to v does not change the Hamming weight, and adding an internal edge reduces it by 2. Also note that $|w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)$. Indeed, $|v \oplus Mw_S| = 2k - 2|\text{Int}_M(v)|$, $|v \oplus w| = 2\ell$, w can be

obtained from w_S be removing internal edges and adding external edges and both of these changes increase $|Mw \oplus Mw_s|$ by 2. These observations together with (49) yield the following upper bound on (48):

$$\begin{aligned} \mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} \mathbb{1}_{\{w_S = S\}} \cdot |\tilde{f}(Mw)| \right] \\ &\leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}(Mw)| \right] \end{aligned} \quad (50)$$

where we dropped the indicator function in going from the first line to the second line above. This in turn can be bounded by Cauchy-Schwarz (Lemma D.2) given that the sum of squares of all normalized Fourier coefficient is $2^{\alpha n} / \mathbf{A}_{\text{reduced}} = 2^n / |\mathbf{A}| \leq 2^{s^*}$ by assumption:

$$\sum_{\substack{w \in \{0,1\}^M \\ |w \oplus w_S| = \ell - (k - |\text{Int}_M(v)|)}} |\tilde{f}(Mw)| \leq \sqrt{2^{s^*} \binom{\alpha n}{\ell - k + |\text{Int}_M(v)|}}.$$

Summing over all possible subsets $S \subset \partial_M(v)$ we then infer

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \leq \mathbb{E}_M \left[\sum_{S \subseteq \partial_M(v)} \sqrt{2^{s^*} \binom{\alpha n}{\ell - k + |\text{Int}_M(v)|}} \right].$$

Note that the bound is independent of the set S so we can replace the summation with the multiplication by $2^{|\partial_M(v)|}$ to obtain

$$\mathbb{E}_M \left[\sum_{\substack{w \in \{0,1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \leq \mathbb{E}_M \left[2^{|\partial_M(v)|} \sqrt{2^{s^*} \binom{\alpha n}{\ell - k + |\text{Int}_M(v)|}} \right].$$

We now bound the sum on the last line above. We have

$$\begin{aligned} &\mathbb{E}_M \left[2^{|\partial_M(v)|} \sqrt{\binom{\alpha n}{\ell - k + |\text{Int}_M(v)|} \left(\frac{64s^*}{\ell - k + |\text{Int}_M(v)|} \right)^{\ell - k + |\text{Int}_M(v)|}} \right] \\ &= \mathbb{E}_M \left[\sum_{i=0}^k \sum_{b=0}^{2k} 2^b \mathbb{1}_{\{|\partial_M(v)|=b \text{ and } |\text{Int}_M(v)|=i\}} \sqrt{2^{s^*} \binom{\alpha n}{\ell - k + i}} \right] \\ &= \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} 2^b q(k, i, b, n) \sqrt{2^{s^*} \binom{\alpha n}{\ell - k + i}} \end{aligned} \quad (51)$$

In going from line 2 to line 3 above we used the fact that by Definition 3.10 for every $v \in \{0,1\}^n$ with $|v| = 2k$ one has $\mathbb{E}_M[\mathbb{1}_{\{|\partial_M(v)|=b \text{ and } |\text{Int}_M(v)|=i\}}] = q(k, i, b, n)$, i.e. $q(k, i, b, n)$ is the probability that a

uniformly random matching M of size αn is such that i edges of M match points of v (i.e. $|\text{Int}_M(v)| = i$) and b edges of M are boundary edges (i.e. $|\partial_M(v)| = b$). Note that we sum over b between 0 and $2k$, as the number of boundary edges of v with respect to M is bounded by its Hamming weight $2k$. Similarly, the number of internal edges i cannot be larger than $|v|/2 = k$, and must be at least $k - \ell$ in order for the binomial coefficient $\binom{\alpha n}{\ell - k + i}$ on the first line above to be nonzero (combinatorially, this means that in order for $v + Mw$ to have weight 2ℓ for some $w \in \{0, 1\}^M$ one must have $|\text{Int}_M(v)| \geq k - \ell$). Recall that $|k - \ell|_+ = \max\{0, k - \ell\}$.

We thus get by (51), using the bound $q(k, i, b, n) \leq q(k, i, n)4^{k-i}20^{-b}$ (Lemma 3.13)

$$\begin{aligned}
\mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] &\leq \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} q(k, i, b, n) 2^b \binom{\alpha n}{\ell - k + i}^{1/2} 2^{s^*/2} \\
&\leq 2^{s^*/2} \sum_{i=|k-\ell|_+}^k \sum_{b=0}^{2k} 4^{k-i} q(k, i, n) 2^{-b} \binom{\alpha n}{\ell - k + i}^{1/2} \\
&\leq 2^{s^*} \sum_{i=|k-\ell|_+}^k 4^{k-i} q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell - k + i)/2} \\
&\leq 16^\ell \sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell - k + i)/2}.
\end{aligned} \tag{52}$$

We used the bound $\binom{a}{b} \leq (ea/b)^b \leq (3a/b)^b$ in going from line 2 to line 3, and the fact that $s^* \leq \ell$ by assumption and $k - i \leq k - |k - \ell|_+ \leq \ell$ for i in the range of the summation in going from line 3 to line 4.

Summing (52) over all possible v , we get

$$\begin{aligned}
\sum_{v \in \{0, 1\}^n} |\tilde{h}(v)| \mathbb{E}_M \left[\sum_{\substack{w \in \{0, 1\}^M \\ |v \oplus Mw| = 2\ell}} |\tilde{f}(Mw)| \right] \\
\leq 16^\ell \sum_{k=0}^{n/100} \left(\sum_{\substack{v \in \{0, 1\}^n \\ |v| = 2k}} |\tilde{h}(v)| \right) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell - k + i)/2} \right) \\
= 16^\ell \sum_{k=0}^{n/100} \left(\sum_{\substack{v \in \{0, 1\}^n \\ |v| = 2k}} |\tilde{h}(v)| \right) \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell - k + i)/2} \right),
\end{aligned} \tag{53}$$

where we used the fact that $k \leq n/100$ by assumptions of the lemma that we established earlier.

We now split the sum into two parts and derive upper bounds on $\sum_{\substack{v \in \{0, 1\}^n \\ |v| = 2k}} |\tilde{h}(v)|$ in the two regimes.

Upper bounding $\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)|$ **for** $k \in [1 : n/C^2]$. We show that for every $k \in [1 : n/C^2]$

$$\sum_{\substack{v \in \{0,1\} \\ |v|=2k}} |\tilde{h}(v)| \leq 2^{s^*} \left(\frac{C^2 n}{k} \right)^{k/2}. \quad (54)$$

Indeed, for $k \in [s^* : n/C^2]$ one this follows directly from the assumption that h is (C, s^*) -bounded, whereas for $k \in [0 : s^*]$ this follows by

$$\begin{aligned} \sum_{\substack{v \in \{0,1\} \\ |v|=2k}} |\tilde{h}(v)| &\leq \left(\frac{C\sqrt{s^*n}}{k} \right)^k \quad (\text{since } h \text{ is } (C, s^*)\text{-bounded}) \\ &= \left(\frac{C^2 n}{k} \right)^{k/2} \left(\frac{s^*}{k} \right)^{k/2} \\ &\leq \left(\frac{C^2 n}{k} \right)^{k/2} e^{\frac{s^*}{2e}}, \end{aligned}$$

and the claim follows as $e^{1/2e} \leq 2$.

Upper bounding $\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)|$ **for** $k \in [n/C^2 + 1 : n/100]$. We have, using the assumption that $|\mathbf{B}|/2^n \geq 2^{-s^*}$, Parseval's equality and Cauchy-Schwarz, that

$$\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)| \leq \sqrt{\left(\sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)|^2 \right) \cdot \binom{n}{2k}} \leq \sqrt{2^{s^*} \binom{n}{2k}}. \quad (55)$$

Putting it together. We bound the contribution of $k \in [1 : n/C^2]$ and $k \in [n/C^2 + 1 : n/100]$ separately.

First, substituting (54) into (53) and restricting the summation to $k \in [1 : n/C^2]$, we get

$$\begin{aligned} \sum_{k=1}^{n/C^2} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)| &\leq \sum_{k=1}^{n/C^2} 2^{s^*} \left(\frac{C^2 n}{k} \right)^{k/2} \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell-k+i)/2} \right) \\ &:= T_1 \end{aligned} \quad (56)$$

Similarly, we get, substituting (55) into (53) and restricting the summation to $k \in [n/C^2 + 1 : n/100]$,

$$\begin{aligned} \sum_{k=n/C^2+1}^{n/100} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2k}} |\tilde{h}(v)| &\leq \sum_{k=n/C^2+1}^{n/100} \sqrt{2^{s^*} \binom{n}{2k}} \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell-k+i)/2} \right) \\ &=: T_2. \end{aligned} \quad (57)$$

Lemma 7.10 and Lemma 7.12 below show that $T_1 \leq \mathbf{U}_{10^7 C, s^*}(\ell)$, $T_2 \leq \mathbf{U}_{10^7 C, s^*}(\ell)$ which concludes the proof since in the double sum (53) the only summand corresponding to $k = 0$ is $q(0, 0, n) \left(\frac{3\alpha n}{\ell} \right)^{\ell/2}$

which is at most $\mathbf{U}_{\sqrt{3}, s^*}(\ell)$ since $q(0, 0, n) = 1$; and we clearly have

$$\begin{aligned} 16^\ell \left(\mathbf{U}_{\sqrt{3}, s^*}(\ell) + \mathbf{U}_{10^7 C, s^*}(\ell) + \mathbf{U}_{10^7 C, s^*}(\ell) \right) &= 16^\ell \left((\sqrt{3}/C)^\ell + 10^{8\ell} + 10^{7\ell} \right) \left(\frac{C^2 n}{\ell} \right)^{\ell/2} \\ &\leq \left(\frac{(10^9 C)^2 n}{\ell} \right)^{\ell/2}. \end{aligned}$$

■

7.2.1 Bounding T_1 and T_2 (technical lemmas)

The following two lemmas deal with sum T_1 and T_2 defined in the proof of Lemma 7.9.

Lemma 7.10 *Suppose parameters C, α, s^*, n satisfy*

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

then for every $\ell \in [s^*, n/(2C^2)]$

$$\sum_{k=1}^{n/C^2} 2^{s^*} \left(\frac{C^2 n}{k} \right)^{k/2} \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell - k + i)/2} \right) \leq \mathbf{U}_{10^7 C, s^*}(\ell)$$

Remark 7.11 *We note that Lemma 7.10 provides the dominant contribution in the application in Lemma 7.9 (which is natural, since it bounds the transfer from a range of weights that overlaps with the target range of coefficient weights).*

Proof of Lemma 7.10: Define, similarly to (56),

$$\begin{aligned} T_1 &:= \sum_{k=1}^{n/C^2} 2^{s^*} \left(\frac{C^2 n}{k} \right)^{k/2} \cdot \left(\sum_{i=|k-\ell|_+}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell - k + i)/2} \right) \\ &\leq 2^\ell \sum_{k=1}^{n/C^2} \sum_{i=|k-\ell|_+}^k \left(\frac{C^2 n}{k} \right)^{k/2} \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k - i)} \binom{n}{2k}^{-1} \left(\frac{3\alpha n}{\ell + i - k} \right)^{(\ell + i - k)/2}, \end{aligned}$$

where in going from line 1 to line 2 we used the fact that $\ell \geq s^*$ by assumption of the lemma, as well as the fact that by Lemma 3.13 for every $0 \leq i \leq k \leq n/2$ one has

$$q(k, i, n) = \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k - i)} \binom{n}{2k}^{-1}.$$

We now get, using the fact that $\binom{\alpha n}{i} \leq (\alpha n)^i / i!$, $\binom{n - 2\alpha n}{2(k - i)} \leq n^{2(k - i)} / (2(k - i))!$ and $\binom{n}{2k}^{-1} \geq 2^{2k} n^{2k} / (2k)!$ for $k \leq n/4$ (which is satisfied since $k \leq n/100$ in our setting),

$$T_1 \leq \left(\frac{4C^2 n}{\ell/2} \right)^{\ell/2} \sum_{k=1}^{n/C^2} \sum_{i=|k-\ell|_+}^k C^{k-\ell} n^{k/2 + (\ell + i - k)/2 - \ell/2} n^{i + 2(k - i) - 2k} \alpha^{i + (\ell + i - k)/2} \cdot \Gamma, \quad (58)$$

where

$$\Gamma = \frac{(\ell/2)^{\ell/2} (2k)!}{i! (2k-2i)! (k/2)^{k/2} (\ell+i-k)^{(\ell+i-k)/2}} 2^{2k} 2^{\ell+i-k}.$$

We now bound Γ using Lemma D.5. First, applying the lemma to the numerator with $m = 2$, $a_1 = \ell/2$ and $a_2 = 2k$, getting that the numerator is upper bounded by $(2k + \ell/2)^{2k + \ell/2}$. Next, applying the lemma to the denominator with $m = 4$, $a_1 = i$, $a_2 = 2k - 2i$, $a_3 = k/2$ (we lower bound $k^{k/2}$ by $(k/2)^{k/2}$) and $a_4 = (\ell + i - k)/2$ (we lower bound $(\ell + i - k)^{(\ell+i-k)/2}$ by $(\ell + i - k/2)^{(\ell+i-k)/2}$). By Lemma D.5 we get that the denominator is lower bounded by

$$((2k + \ell/2 + i/2)/12)^{2k + \ell/2 + i/2} \geq 12^{-(2k + \ell/2 + i/2)} ((2k + \ell/2)^{2k + \ell/2} (i/2)^{i/2}),$$

where we used the fact that $(a + b)^{a+b} \geq a^a b^b$ for all integer $a, b \geq 0$. Putting the above bounds together, we get that $\Gamma \leq i^{i/2} 10^{3(\ell+k+i)}$.

Gathering the powers of n and powers of α in the inner summation in (58), we get that the exponent of n is $-i$ and that the exponent of α is at least i . This, together with our upper bound on Γ gives

$$\begin{aligned} T_1 \cdot \left(\frac{4C^2 n}{\ell} \right)^{-\ell/2} &\leq \sum_{k=1}^{n/C^2} \sum_{i=|k-\ell|_+}^k C^{k-\ell} (i/n)^{i/2} 10^{3(\ell+k+i)} \alpha^i \\ &\leq 10^{3\ell} \sum_{k=1}^{n/C^2} \sum_{i=|k-\ell|_+}^k C^{k-\ell-i} \left(\frac{10^8 C^2 \alpha^2 i}{n} \right)^i \end{aligned}$$

Since $k \leq n/(2C^2)$ for all k in the range of the summation, and $i \leq k$, one has $\left(\frac{10^8 C^2 \alpha^2 i}{n} \right)^i \leq (10^8 \alpha)^i \leq 2^{-i}$ by **(P1)**. Substituting into the equation above, we get

$$\begin{aligned} T_1 \cdot \left(\frac{4C^2 n}{\ell} \right)^{-\ell/2} &\leq 10^{3\ell} \sum_{k=1}^{\ell} \sum_{i=0}^k C^{k-\ell-i} 2^{-i} + \sum_{k=\ell+1}^{n/(2C^2)} \sum_{i=k-\ell}^k C^{k-\ell-i} 2^{-i} \\ &\leq 10^{3\ell} \sum_{k=1}^{\ell} C^{k-\ell} \sum_{i=0}^k C^{-i} 2^{-i} + \sum_{k=\ell+1}^{n/(2C^2)} 2^{-(k-\ell)} \sum_{i=0}^{k-\ell} C^{-i} 2^{-i} \\ &\leq 4 \cdot 10^{3\ell} \leq 4 \cdot 10^{6\ell}, \end{aligned}$$

where in going from line 2 to line 3 we used assumption **(P2)**. This implies

$$T_1 \leq \left(\frac{4 \cdot 16 \cdot 10^{12} C^2 n}{\ell} \right)^{\ell/2} \leq \left(\frac{(10^7 C)^2 n}{\ell} \right)^{\ell/2}.$$

Lemma 7.12 Suppose parameters C, α, s^*, n satisfy

$$\text{(P1)} \alpha < 10^{-10} \quad \text{(P2)} C > 10^6 \quad \text{(P3)} s^* < \frac{n}{10^9 C^3} \quad \text{(P4)} n > 10^9 C^4,$$

then for every $\ell \in [s^*, n/2C^2]$

$$\sum_{k=n/C^2+1}^{n/100} \sqrt{2^{s^*} \binom{n}{2k}} \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell-k+i)/2} \right) \leq \left(\frac{(10^8 C)^2 n}{\ell} \right)^{\ell/2}.$$

Remark 7.13 A close inspection of the proof of Lemma 7.12 reveals that a stronger bound than is needed by the application in Lemma 7.9 holds. However, we prefer to keep the bound in present form to simplify presentation.

Proof of Lemma 7.12: Note that we have $\ell \leq n/2C^2$ and $k > n/C^2$, so we must have $i > n/2C^2$. This allows us to write

$$\begin{aligned} T_2 &:= \sum_{k=n/C^2+1}^{n/100} \sqrt{2^{s^*} \binom{n}{2k}} \cdot \left(\sum_{i=k-\ell}^k q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell-k+i)/2} \right) \\ &\leq \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k \sqrt{2^{s^*} \binom{n}{2k}} q(k, i, n) \left(\frac{3\alpha n}{\ell - k + i} \right)^{(\ell-k+i)/2} \end{aligned}$$

Substituting the expression for $q(k, i, n)$ from Lemma 3.13 and using the assumption that $\ell \geq s^*$ to upper bound 2^{s^*} by 2^ℓ , we get

$$\begin{aligned} T_2 &\leq 2^\ell \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k \binom{n}{2k}^{1/2} \binom{\alpha n}{i} \binom{n-2\alpha n}{2(k-i)} \binom{n}{2k}^{-1} \left(\frac{3\alpha n}{\ell + i - k} \right)^{(\ell+i-k)/2} \\ &= 2^\ell \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k \binom{\alpha n}{i} \binom{n-2\alpha n}{2(k-i)} \binom{n}{2k}^{-1/2} \left(\frac{3\alpha n}{\ell + i - k} \right)^{(\ell+i-k)/2} \quad (59) \\ &\leq \left(\frac{16C^2 n}{\ell} \right)^{\ell/2} \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k C^{-\ell} n^{(\ell+i-k)/2 - \ell/2} n^{i+2(k-i)-2k/2} \alpha^{i+(\ell+i-k)/2} \cdot \Gamma, \end{aligned}$$

where

$$\Gamma = \frac{(\ell/2)^{\ell/2} k^k}{i!(2k-2i)!((\ell+i-k)/2)^{(\ell+i-k)/2}} 2^{2k} 2^{\ell+i-k}.$$

In (59) we used the bounds $\binom{\alpha n}{i} \leq (\alpha n)^i / i!$, $\binom{n-2\alpha n}{2(k-i)} \leq n^{2(k-i)} / (2(k-i))!$ and $\binom{n}{2k} \geq 2^{-2k} n^k / (2k)!$ when going from line 2 to line 3. We also upper bounded $\sqrt{(2k)!}$ by $2^k k^k$ to simplify the expression for Γ .

We now show using Lemma D.5 that $\Gamma \leq \left(\frac{k-i}{2}\right)^{-\frac{k-i}{2}} 10^{3(\ell+k+i)}$. We first apply the lemma to the numerator of Γ with $m = 2$, $a_1 = \ell/2$, $a_2 = k$, obtaining an upper bound of $(k + \ell/2)^{k+\ell/2}$. Applying the lemma to the denominator of Γ with $m = 3$, $a_1 = i$, $a_2 = 2k - 2i$, $a_3 = (\ell + i - k)/2$, we obtain a lower bound of $((k + \ell/2)/9)^{k+\ell/2}$. Putting the bounds above together, we get that $\Gamma \leq \left(\frac{k-i}{2}\right)^{-\frac{k-i}{2}} 10^{3(\ell+k+i)}$, as required.

Gathering the powers of n in (59), we obtain $n^{(k-i)/2}$ so this together with factor of $\left(\frac{k-i}{2}\right)^{-\frac{k-i}{2}}$ from the upper bound on Γ gives us $\left(\frac{2n}{k-i}\right)^{\frac{k-i}{2}}$. Finally the power of α is at least i . Putting these bounds together, we get

$$\begin{aligned}
T_2 \cdot \left(\frac{16C^2n}{\ell}\right)^{-\ell/2} &\leq \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k C^{-\ell} \left(\frac{2n}{k-i}\right)^{\frac{k-i}{2}} \alpha^i 10^{3(\ell+k+i)} \\
&\leq 10^{6\ell} C^{-\ell} \left(\frac{2n}{\ell}\right)^{\ell/2} \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k \alpha^i 10^{6i} 10^{k-\ell-i} \\
&\leq 10^{6\ell} \left(\frac{2n/C^2}{\ell}\right)^{\ell/2} \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k (10^6\alpha)^i \\
&\leq 10^{6\ell} e^{n/C^2} \sum_{k=n/C^2+1}^{n/100} \sum_{i=k-\ell}^k (10^6\alpha)^i \\
&\leq 10^{6\ell} e^{n/C^2} \sum_{k=n/C^2+1}^{n/100} 2 \cdot (10^6\alpha)^{k-\ell} \\
&\leq 4 \cdot 10^{6\ell} e^{n/C^2} (10^6\alpha)^{n/(2C^2)} \\
&\leq 10^{6\ell}.
\end{aligned}$$

We used the fact that $k - i \leq \ell \leq n/2$ by **(P3)** and **(P2)** in going from line 1 to line 2, and the fact that $k - \ell - i \leq 0$ for all i in the range of summation in going from line 2 to line 3. In going from line 3 to line 4 we used the fact that $\left(\frac{2n/C^2}{\ell}\right)^{\ell/2} \leq 4^{n/(2C^2)} \leq e^{n/C^2}$ for $\ell \in [s^*, n/(2C^2)]$. In going from line 4 to line 5 we used the fact that $\sum_{i=k-\ell}^k (10^6\alpha)^i \leq 2(10^6\alpha)^{k-\ell}$ by **(P1)**. In going from line 5 to line 6 we used the fact that $\sum_{k=n/C^2+1}^{n/100} (10^6\alpha)^{k-\ell} \leq 2 \cdot (10^6\alpha)^{n/C^2-\ell} \leq 2 \cdot (10^6\alpha)^{n/(2C^2)}$ by **(P1)** as well as the assumption that $\ell \leq n/(2C^2)$ and $k \geq n/C^2$. In going from line 6 to line 7 we used assumption **(P1)**.

Putting the bounds above together, we get

$$T_2 \leq 10^{6\ell} \left(\frac{16C^2n}{\ell}\right)^{\ell/2} \leq \left(\frac{(10^8C)^2 n}{\ell}\right)^{\ell/2},$$

as required. ■

8 Acknowledgements

The authors would like to thank Madhu Sudan for many useful comments on drafts of this manuscript. Michael Kapralov would like to thank Sanjeev Khanna for many useful discussions on precursors to this work.

References

- [AG09] K. Ahn and S. Guha. Graph sparsification in the semi-streaming model. *ICALP*, pages 328–338, 2009.
- [AG11] K. Ahn and S. Guha. Linear programming in the semi-streaming model with application to the maximum matching problem. *ICALP*, pages 526–538, 2011.

- [AG13] K. Ahn and S. Guha. Access to data and number of iterations: Dual primal algorithms for maximum matching under resource constraints. *CoRR*, abs/1307.4359, 2013.
- [AGM12a] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. *SODA*, pages 459–467, 2012.
- [AGM12b] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Graph sketching: Sparsification, spanners, and subgraphs. *PODS*, 2012.
- [AKLY15] Sepehr Assadi, Sanjeev Khanna, Yang Li, and Grigory Yaroslavtsev. Tight bounds for linear sketches of approximate matchings. *CoRR*, 2015.
- [AMS96] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *STOC*, pages 20–29, 1996.
- [CCE⁺15] Rajesh Hemant Chitnis, Graham Cormode, Hossein Esfandiari, MohammadTaghi Hajiaghayi, Andrew McGregor, Morteza Monemizadeh, and Sofya Vorotnikova. Kernelization via sampling with applications to dynamic graph streams. *CoRR*, abs/1505.01731, 2015.
- [EHL⁺15] Hossein Esfandiari, Mohammad Taghi Hajiaghayi, Vahid Liaghat, Morteza Monemizadeh, and Krzysztof Onak. Streaming algorithms for estimating the matching size in planar graphs and beyond. *SODA*, 2015.
- [GKK⁺08] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008.
- [GKK12] A. Goel, M. Kapralov, and S. Khanna. On the communication and streaming complexity of maximum bipartite matching. *SODA*, 2012.
- [GO12] Venkatesan Guruswami and Krzysztof Onak. Superlinear lower bounds for multipass graph processing. *CCC*, 2012.
- [HRVZ15] Zengfeng Huang, Božidar Radunović, Milan Vojnović, and Qin Zhang. Communication complexity of approximate maximum matching in distributed graph data. *STACS*, 2015.
- [Kap13] Michael Kapralov. Better bounds for matchings in the streaming model. *SODA*, 2013.
- [KK14] Dmitry Kogan and Robert Krauthgamer. Sketching cuts in graphs and hypergraphs. *manuscript*, September 2014.
- [KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions (extended abstract). *29th Annual Symposium on Foundations of Computer Science, White Plains, New York, USA, 24-26 October 1988*, pages 68–80, 1988.
- [KKP18] John Kallaugher, Michael Kapralov, and Eric Price. The sketching complexity of graph and hypergraph counting. *FOCS*, 2018.
- [KKS14] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Approximating matching size from random streams. In *25th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.
- [KKS15] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Streaming lower bounds for approximating MAX-CUT. In *26th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2015.

- [KKS^V17] Michael Kapralov, Sanjeev Khanna, Madhu Sudan, and Ameya Velingker. $(1 + \Omega(1))$ -Approximation to MAX-CUT requires linear space. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 1703–1722, 2017.
- [KL11] Jonathan A. Kelner and Alex Levin. Spectral sparsification in the semi-streaming setting. *STACS*, pages 440–451, 2011.
- [KLM⁺14] Michael Kapralov, Yin Tat Lee, Cameron Musco, Christopher Musco, and Aaron Sidford. Single pass spectral sparsification in dynamic streams. *FOCS*, 2014.
- [Kon15] Christian Konrad. Maximum matching in turnstile streams. *CoRR*, abs/1505.01460, 2015.
- [KW14] Michael Kapralov and David Woodruff. Spanners and sparsifiers in dynamic streams. *PODC*, 2014.
- [McG14] Andrew McGregor. Graph stream algorithms: a survey. *SIGMOD Record*, 43(1):9–20, 2014.
- [SJ00] Andrzej Rucinski Svante Janson, Tomasz Luczak. *Random Graph Dynamicss*. Wiley, New York, NY, USA, 2000.
- [VY11] Elad Verbin and Wei Yu. The streaming complexity of cycle counting, sorting by reversals, and other problems. *SODA*, pages 11–25, 2011.

A Proof of Lemma 4.7

Proof of Lemma 4.7: The claim of the lemma is equivalent to saying that with probability at least $1 - \delta$ over the choice of the matching M we have, for any $z_0 \in \{0, 1\}^M$, the following inequality

$$1 - \delta \leq 2^{|M|} \mathbb{P}_{x \sim \text{Uniform}(\mathbf{B})} [Mx = z_0] \leq 1 + \delta,$$

which would in turn follow by Markov inequality from the following fact:

$$\mathbb{E}_M \left[\max_{z_0 \in \{0,1\}^M} \left| 2^{|M|} \mathbb{P}_{x \sim \text{Uniform}(\mathbf{B})} [Mx = z_0] - 1 \right| \right] \leq \delta^2.$$

In order to prove this we express the LHS in terms of the Fourier transform of $h = \mathbb{1}_{\mathbf{B}}$. Define $g_{z_0}(x) := \mathbb{1}_{\{x: Mx = z_0\}}$. We have

$$\begin{aligned} 2^{|M|} \mathbb{P}_{x \sim \text{Uniform}(\mathbf{B})} [Mx = z_0] - 1 &= \frac{2^{|M|} |\mathbf{B} \cap \{x : Mx = z_0\}| - |\mathbf{B}|}{|\mathbf{B}|} \\ &= \frac{2^{|M|} 2^n \widehat{h g_{z_0}}(0) - 2^n \widehat{h}(0)}{|\mathbf{B}|} \\ &= \frac{2^{n+|M|}}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ v \neq 0}} \widehat{h}(v) \widehat{g_{z_0}}(v). \end{aligned}$$

Note that in order for $\widehat{g_{z_0}}(v)$ to be non-zero, $|v|$ must be even. We then use a triangle inequality to obtain

$$\mathbb{E}_M \left[\max_{z_0 \in \{0,1\}^M} \left| 2^{|M|} \mathbb{P}_{x \sim \text{Uniform}(\mathbf{B})} [Mx = z_0] - 1 \right| \right] \leq \sum_{\ell=1}^{n/2} \mathbb{E}_M \left[\max_{z_0 \in \{0,1\}^M} \left| \frac{2^{n+|M|}}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{h}(v) \widehat{g_{z_0}}(v) \right| \right].$$

Recall that $p_\alpha(\ell, n)$ stands for the probability that fixed 2ℓ points are matched by a uniformly random matching of size αn . For $v \in \{0, 1\}^n$ which is perfectly matched by M (i.e. M restricted to v is a perfect matching) let $e(v) \in \{0, 1\}^M$ denote the set of edges induced by v . Using explicit structure of \widehat{g}_{z_0} we find that

$$\begin{aligned} \frac{2^{n+|M|}}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{h}(v) \widehat{g}_{z_0}(v) &= \frac{2^{n+|M|}}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell \\ v \text{ is matched by } M}} \widehat{h}(v) \cdot 2^{-|M|} (-1)^{z_0 \cdot e(v)} \\ &\leq \frac{2^n}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell \\ v \text{ is matched by } M}} |\widehat{h}(v)|, \end{aligned} \quad (60)$$

where we used the fact that $\widehat{g}_{z_0}(v)$ is zero for v 's that are not perfectly matched by M . The bound is independent of z_0 which allows us to write, after taking expectation over M ,

$$\begin{aligned} &\mathbb{E}_M \left[\max_{z_0 \in \{0,1\}^M} \left| 2^{|M|} \mathbb{P}_{x \sim \text{Uniform}(\mathbf{B})} [Mx = z_0] - 1 \right| \right] \\ &\leq \sum_{\ell=1}^{n/2} p(\ell, n) \frac{2^n}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} |\widehat{h}(v)| \\ &\leq \sum_{\ell=1}^{\alpha n} p(\ell, n) \frac{2^n}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} |\widehat{h}(\ell)| \\ &\leq \sum_{\ell=1}^{s^*} p(\ell, n) \mathbf{U}_{C, s^*}(\ell) + \sum_{\ell=s^*+1}^{n/2} p(\ell, n) \sqrt{2^s \binom{n}{2\ell}}. \end{aligned} \quad (61)$$

In going from line 2 to line 3 we used the fact that $\widehat{h}(v) = 0$ for any v that is not perfectly matched by M , as well as the assumption that M is a matching of size αn . In going from line 3 to line 4 we used the fact that $\frac{2^n}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} |\widehat{h}(\ell)| \leq \mathbf{U}_{C, s^*}(\ell)$ for ℓ between 1 and s^* (since \mathbf{B} is (C, s^*) -bounded by assumption, as well as the fact that for every ℓ between 0 and $n/2$

$$\frac{2^n}{|\mathbf{B}|} \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} |\widehat{h}(\ell)| \leq \sqrt{\binom{n}{2\ell}} \frac{2^n}{|\mathbf{B}|} \sqrt{\sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{h}(\ell)^2} \leq \sqrt{2^{s^*} \binom{n}{2\ell}}.$$

The latter bound holds for any subset \mathbf{B} of the cube with $|\mathbf{B}|/2^n \geq 2^{-s^*}$ by Parseval's equality ((2) and Remark 3.1) together with Cauchy-Schwarz.

We now upper bound individual terms in the summation over ℓ in (61) above. We use the expression

$$p(\ell, n) = \binom{\alpha n}{\ell} \binom{n}{2\ell}^{-1} \quad (62)$$

provided by Lemma 3.12.

Upper bounding the contribution from low weights ($1 \leq \ell \leq s^*$). For low weights ($\ell \leq s^*$) we have by the assumption that h is (C, s^*) -bounded

$$\begin{aligned}
p(\ell, n) \mathbf{U}_{C, s^*}(\ell) &= \binom{\alpha n}{\ell} \binom{n}{2\ell}^{-1} \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell \\
&\leq \left(\frac{e\alpha n}{\ell} \right)^\ell \cdot \frac{(2\ell)!}{(n/2)^{2\ell}} \cdot \left(\frac{C\sqrt{s^*n}}{\ell} \right)^\ell \\
&\leq \left(44\alpha C\sqrt{s^*/n} \right)^\ell \\
&\leq (44\alpha)^\ell \delta^{2\ell}.
\end{aligned} \tag{63}$$

In going from line 1 to line 2 above we used the lower bound $\binom{n}{2\ell} \geq (n/2)^{2\ell}/(2\ell)!$, since $\ell \leq s^* \leq n/C^2 \leq n/4$ (as $C \geq 100$ by assumption), as well as the bound $\binom{\alpha n}{\ell} \leq (e\alpha n/\ell)^\ell$. We used the assumption $s^* \leq \delta^4 n/C^2$ to from line 3 to line 4.

Upper bounding the contribution from high weights ($s^* < \ell \leq \alpha n$). We have, using (62),

$$\begin{aligned}
p(\ell, n) \sqrt{2^{s^*} \binom{n}{2\ell}} &= \binom{\alpha n}{\ell} \binom{n}{2\ell}^{-1/2} 2^{s^*/2} \\
&\leq \frac{(\alpha n)^\ell}{\ell!} \left(\frac{(n/2)^{2\ell}}{(2\ell)!} \right)^{-1/2} 2^{\ell/2} \\
&= (2\sqrt{2}\alpha)^\ell \sqrt{\binom{2\ell}{\ell}} \\
&\leq (4\sqrt{2}\alpha)^\ell \\
&\leq e^{-2s^*}.
\end{aligned} \tag{64}$$

In going from line 1 to line 2 we used the fact that $\binom{n}{2\ell} \geq (n/2)^{2\ell}/(2\ell)!$ for any $\ell \leq \alpha n \leq n/4$. In going from line 4 to line 5 we used the assumption that $\alpha \leq 1/100$.

Putting (63) together with (64) and summing over all $\ell \in [1 : \alpha n]$, we get using (61)

$$\begin{aligned}
&\mathbb{E}_M \left[\max_{z_0 \in \{0,1\}^M} \left| 2^{|M|} \mathbb{P}_{x \sim \text{Uniform}(\mathbf{B})} [Mx = z_0] - 1 \right| \right] \\
&\leq \sum_{\ell=1}^{s^*} (44\alpha)^\ell \delta^{2\ell} + \sum_{\ell=s^*+1}^{n/2} e^{-2s^*} \\
&\leq 88\alpha \cdot \delta^2 + n^{-10} \\
&\leq \delta^2.
\end{aligned}$$

In going from line 2 to line 3 we used the assumption that $s^* \geq 10 \ln(n+1)$, and in going from line 3 to line 4 we used the assumption that $\alpha < 1/100$ and $\delta \in (1/n, 1/2)$. An application of Markov's inequality now gives the result. \blacksquare

B Proof of Lemma 6.3

We will use some basic properties of total variation distance, which we now state.

Lemma B.1 Let μ, ν be two probability distributions on the same finite sample space Ω , and consider independent random variables $X, \tilde{X} \sim \mu, Y \sim \nu$ taking values in Ω . Then one has

$$2 \cdot \|\mu - \nu\|_{tvd} = \mathbb{E}_X[|1 - \Pr_Y[Y = X] / \Pr_{\tilde{X}}[\tilde{X} = X]|].$$

Proof: Identify Ω with $\{1, 2, \dots, n\}$ and let $p_k = \Pr[X = k], q_k = \Pr[Y = k]$. Then we have

$$\mathbb{E}_X[|1 - \Pr_Y[Y = X] / \Pr_{\tilde{X}}[\tilde{X} = X]|] = \sum_{k=1}^n p_k \cdot |1 - q_k/p_k| = \sum_{k=1}^n |p_k - q_k| = 2 \cdot \|\mu - \nu\|_{tvd}.$$

■

Lemma B.2 (Substitution lemma) Let X^1, X^2 be random variables taking values on finite sample space Ω_1 . Let Z^1, Z^2 be random variables taking values on sample space Ω_2 , and suppose that Z^2 is independent of X^1, X^2 . Let $f : \Omega_1 \times \Omega_2 \rightarrow \Omega_3$ be a function. Then

$$\|(X^1, f(X^1, Z^1)) - (X^2, f(X^2, Z^2))\|_{tvd} \leq \|(X^1, f(X^1, Z^1)) - (X^1, f(X^1, Z^2))\|_{tvd} + \|X^1 - X^2\|_{tvd}.$$

Proof: By triangle inequality the left hand side is at most

$$\|(X^1, f(X^1, Z^1)) - (X^1, f(X^1, Z^2))\|_{tvd} + \|(X^1, f(X^1, Z^2)) - (X^2, f(X^2, Z^2))\|_{tvd}.$$

It remains to note that the second summand is at most $\|X^1 - X^2\|_{tvd}$ by Claim 3.7. ■

The next lemma also follows easily from the definition of total variation distance (see, e.g. Claim 6.5 in [KKS15] for a proof):

Lemma B.3 For any random variables X, Y taking values on finite sample space Ω_1 , independent random variable Z taking values on finite sample space Ω_2 and any function $f : \Omega_1 \times \Omega_2 \rightarrow \Omega_3$ one has $\|f(X, Z) - f(Y, Z)\|_{tvd} \leq \|X - Y\|_{tvd}$.

Proof of Lemma 6.3: The proof is by induction on $t = 1, \dots, T$. We prove that for all t one has

$$\|(M_{1:t}, S_{1:t}^Y) - (M_{1:t}, S_{1:t}^N)\|_{tvd} \leq \gamma t/T + \sum_{j=1}^t \Pr[\bar{\mathcal{E}}_j | \mathcal{E}_{j-1}].$$

Base: $t = 1$ We have, conditional on the event \mathcal{E}_1 ,

$$\begin{aligned} \|(M_1, S_1^Y) - (M_1, S_1^N)\|_{tvd} &= \|(M_1, S_1^Y) - (M_1, r_1(M_1, U_1, S_0^N))\|_{tvd, \mathcal{E}_1} \\ &= \|(M_1, S_1^Y) - (M_1, r_1(M_1, U_1, S_0^Y))\|_{tvd, \mathcal{E}_1} \end{aligned}$$

where we used the fact that $S_0^Y = S_0^N$. By assumption Eq. (*) of the lemma and the fact that total variation distance is bounded by 1 we have

$$\|(M_1, S_1^Y) - (M_1, r_1(M_1, U_1, S_0^N))\|_{tvd, \mathcal{E}_1} \leq \gamma/T + \Pr[\bar{\mathcal{E}}_1] = \gamma/T$$

as required. This proves the base of the induction.

Inductive step: $t - 1 \rightarrow t$ We condition on \mathcal{E}_t in what follows, and write $\|\cdot\|_{tvd, \mathcal{E}_t}$ to denote the total variation distance between conditional distributions. We have

$$\begin{aligned} & \|(M_{1:t}, S_{1:t}^Y) - (M_{1:t}, S_{1:t}^N)\|_{tvd, \mathcal{E}_t} \\ = & \|(M_{1:t-1}, S_{1:t-1}^Y, M_t, r_t(M_{1:t-1}, S_{1:t-1}^Y, M_t, M_t X^*)) - (M_{1:t-1}, S_{1:t-1}^N, r_t(M_{1:t-1}, S_{1:t-1}^N, M_t, U_t))\|_{tvd, \mathcal{E}_t} \end{aligned}$$

We would like to apply Lemma B.2 to the expression above. To that effect we define

$$Q_{t-1}^Y = (M_{1:t-1}, S_{1:t-1}^Y) \text{ and } Q_{t-1}^N = (M_{1:t-1}, S_{1:t-1}^N).$$

With this notation in place we have

$$\begin{aligned} & \|(M_{1:t-1}, S_{1:t-1}^Y, M_t, r_t(M_{1:t-1}, S_{1:t-1}^Y, M_t, M_t X^*)) - (M_{1:t-1}, S_{1:t-1}^N, r_t(M_{1:t-1}, S_{1:t-1}^N, M_t, U_t))\|_{tvd, \mathcal{E}_t} \\ = & \|(Q_{t-1}^Y, M_t, r_t(Q_{t-1}^Y, M_t, M_t X^*)) - (Q_{t-1}^N, M_t, r_t(Q_{t-1}^N, M_t, U_t))\|_{tvd, \mathcal{E}_t}. \end{aligned} \tag{65}$$

We now apply the substitution lemma (Lemma B.2) to Eq. (65). The parameters are as follows. We let $X^1 = Q_{t-1}^Y$ and $X^2 = Q_{t-1}^N$. The variables Z^1 and Z^2 are set as $Z^1 = (M_t, M_t X^*)$ (recall that X^* is the hidden bipartition) and $Z^2 = (M_t, U_t)$. Note that this setting of Z^2 satisfies the preconditions of Lemma B.2: $Z^2 = (M_t, U_t)$ is independent of X^1, X^2 , as required.

In order to apply Lemma B.2, it remains to define the function f that maps tuples (X, Z) to some universe so that

$$(X^1, f(X^1, Z^1)) = (Q_{t-1}^Y, f(Q_{t-1}^Y, (M_t, M_t X^*)))$$

equals $(Q_{t-1}^Y, M_t, r_t(Q_{t-1}^Y, M_t, M_t X^*))$ and

$$(X^2, f(X^2, Z^2)) = (Q_{t-1}^N, f(Q_{t-1}^N, (M_t, U_t)))$$

equals $(Q_{t-1}^N, M_t, r_t(Q_{t-1}^N, M_t, U_t))$. For that, it is sufficient to let f be the function that maps input tuple $(X, (B, C))$ to $(B, r_t(X, B, C))$.

Applying Lemma B.2 with these settings, we get

$$\begin{aligned} & \|(Q_{t-1}^Y, M_t, r_t(Q_{t-1}^Y, M_t, M_t X^*)) - (Q_{t-1}^N, r_t(Q_{t-1}^N, M_t, U_t))\|_{tvd, \mathcal{E}_t} \\ \leq & \|Q_{t-1}^Y - Q_{t-1}^N\|_{tvd} + \|(M_t, r_t(Q_{t-1}^Y, M_t, M_t X^*)) - (M_t, r_t(Q_{t-1}^N, M_t, U_t))\|_{tvd, \mathcal{E}_t} \end{aligned}$$

The first term is bounded by $\gamma(t-1)/T + \sum_{j=1}^{t-1} \Pr[\bar{\mathcal{E}}_j | \mathcal{E}_{j-1}]$ by the inductive hypothesis. Using the assumption Eq. (*) of the lemma, we get

$$\begin{aligned} & \|(M_t, r_t(Q_{t-1}^Y, M_t, M_t X^*)) - (M_t, r_t(Q_{t-1}^N, M_t, U_t))\|_{tvd, \mathcal{E}_t} \\ = & \mathbf{E}_{(M_t, Q_{t-1}^Y) \in \mathcal{E}_t} [\|(M_t, r_t(Q_{t-1}^Y, M_t, M_t X^*)) - (M_t, r_t(Q_{t-1}^N, M_t, U_t))\|_{tvd}] \\ \leq & \gamma/T, \end{aligned}$$

where the total variation distance in the second line is over $X^* \sim UNIF(S_{t-1}^Y)$. Putting the two bounds together yields

$$\|(M_{1:t}, S_{1:t}^Y) - (M_{1:t}, S_{1:t}^N)\|_{tvd} \leq \gamma t/T + \sum_{j=1}^t \Pr[\bar{\mathcal{E}}_j | \mathcal{E}_{j-1}]$$

as required. Substituting $t = T$ we get

$$\|(M_{1:T}, S_{1:T}^Y) - (M_{1:T}, S_{1:T}^N)\|_{\text{tvd}} \leq \gamma + \sum_{j=1}^T \Pr[\bar{\mathcal{E}}_j | \mathcal{E}_{j-1}] \leq \gamma + T \cdot \gamma/T = 2\gamma,$$

thus proving the lemma. ■

C Details omitted from Section 5

Here we formally prove that if $s \ll n/B^T$ for large enough B then with high probability we will have $\|F_T\| \ll n$. We start by the following Lemma.

Lemma C.1 *Let m, T be positive integers and let $\{X_k\}_{k \in [mT]}$ be a sequence of positive random variables satisfying $X_0 < m/100^T$, and for $k \in [mT]$ we have*

$$\mathbb{E}[X_k | X_{k-1}] \leq X_{k-1} \cdot \left(1 + \frac{1}{m} + \frac{X_{k-1}}{m^2}\right).$$

Then $\Pr[\max_{i \in [mT]} X_i > n/2^T] < 2^{-T}$.

Proof: Consider stopping time $\tau := \inf\{t : X_t > m/2^T\}$, if $X_{mT} \leq m/2^T$ we define $\tau := mT$. Note that for $k \leq \tau$ we have

$$1 + \frac{1}{m} + \frac{X_{k-1}}{m^2} < 1 + \frac{2}{m},$$

which means that the process $Y_k := X_k / (1+2/m)^k$ stopped at time τ is a supermartingale (i.e. $\mathbb{E}[Y_k | Y_{k-1}] \leq Y_{k-1}$). By Markov's inequality applied to Y_τ we have

$$\Pr\left[\max_{i \in [mT]} X_i > n/2^T\right] = \Pr[X_\tau > n/2^T] \leq \Pr\left[Y_\tau > \frac{m}{2^T \cdot (1+2/m)^{mT}}\right] \leq \frac{\mathbb{E}[Y_\tau]}{\left(\frac{m}{2^T \cdot (1+2/m)^{mT}}\right)}.$$

Optional stopping theorem implies that $\mathbb{E}[Y_\tau] \leq \mathbb{E}[Y_0] = \mathbb{E}[X_0] < m/100^T$. So we have

$$\Pr\left[\max_{i \in [mT]} X_i > n/2^T\right] \leq \frac{m/100^T}{\left(\frac{m}{2^T \cdot (1+2/m)^{mT}}\right)} \leq \frac{m/100^T}{m/(2e^2)^T} < 6^{-T}.$$

We then deduce the following lemma ■

Lemma C.2 *Suppose the communication budget is $s < n/800^T$, then with probability at least $1 - 2^{-T}$ the forest F_T formed by all edges revealed by players satisfies $\|F_T\| \leq n/2^T$.*

Proof: Recall that we consider forests $F_k, k \in [0 \dots T]$ formed by edges revealed by first k players. We also define F_t^i for $i \in [0 \dots \alpha n]$ inductively as

$$F_t^j := \begin{cases} F_t^{j-1} \cup \{e_j\} & \text{if } e_j \text{ intersects with a nontrivial component in } F_t^{j-1} \\ F_t^{j-1} & \text{o.w.} \end{cases}$$

We then deduced that (see (13))

$$\mathbb{E}_{e_i} [\|F_t^i\| | e_{[1:i-1]}] \leq \|F_t^{i-1}\| \cdot \left(1 + \frac{12}{n} + \frac{8\|F_t^{i-1}\|}{n^2}\right). \quad (66)$$

We refer to Section 5.2 for more details. Let $m = \alpha n + 1$ and define a sequence of random variables $X_k, k \in [0 \dots mT]$ by

$$X_k := \|F_{t+1}^i\| \cdot 2^{T-t},$$

where $i = 0 \dots m - 1$ and $k = m \cdot t + i$. We now check that X_k satisfies the assumption of Lemma C.1. For k not divisible by m this trivially follows from (66) since $m < n/12$ and $X_k \geq \|F_{t+1}^i\|$. For $k = m \cdot t$ we have $X_k = \|F_t^0\| \cdot 2^{T-t} = \|F_{t-1}\| \cdot 2^{T-t} = (\|F_{t-1}^{\alpha n}\| + 4s) \cdot 2^{T-t}$ and $X_{k-1} = F_{t-1}^{\alpha n} \cdot 2^{T-t+1}$. Since we definitely have $\|F_{t-1}^{\alpha n}\| \geq \|F_0\| = 4s$, this implies that $X_k \leq X_{k-1}$ deterministically, and so the condition of Lemma C.1 is satisfied in this case as well. Note also that $X_0 = 4s \cdot 2^T < n/100^T$. Lemma C.1 then implies that $\Pr[X_{mT} > n/2^T] < 2^{-T}$. So we infer

$$\Pr[\|F_T\| > n/2^T] = \Pr[X_{mT} > n/2^T] < 2^{-T}.$$

■

D Useful facts

D.1 Concentration inequalities

We prove Lemma 2.5, restated here for convenience of the reader:

Lemma 2.5 *Let $X = \sum_{i=1}^n X_i$, where X_i are Bernoulli 0/1 random variables satisfying, for every $k \in [n]$, $\mathbb{E}[X_k | X_1, \dots, X_{k-1}] \leq p$ for some $p \in (0, 1)$. Let $\mu = np$. Then for all $\Delta > 0$*

$$\Pr[X \geq \mu + \Delta] \leq \exp\left(-\frac{\Delta^2}{2\mu + 2\Delta}\right).$$

Proof: Note that for any $u > 0$ we have

$$\mathbb{E}[e^{uX}] = \mathbb{E}\left[\prod_{k=1}^n e^{uX_k}\right] \leq (1 - p + p \cdot e^u) \cdot \mathbb{E}\left[\prod_{k=1}^{n-1} e^{uX_k}\right] \leq \dots \leq (1 - p + p \cdot e^u)^n.$$

By Markov inequality we then have

$$\Pr[X \geq \mu + \Delta] \leq \min_{u>0} \frac{(1 - p + p \cdot e^u)^n}{e^{u(\mu+\Delta)}} = \min_{v>0} \frac{(1 + pv)^{\mu/p}}{(1 + v)^{\mu+\Delta}},$$

where in the last equality we made a substitution $v = e^u - 1$. Using the fact that $e^{-x^2/2} \leq 1 + x \leq e^x$ for all positive x we then infer

$$\Pr[X \geq \mu + \Delta] \leq \min_{v>0} \frac{e^{pv \cdot \mu/p}}{e^{(v-v^2/2) \cdot (\mu+\Delta)}} = \min_{v>0} \exp(-v \cdot \Delta + v^2/2 \cdot (\mu + \Delta)).$$

The minimum is achieved at $v = \frac{\Delta}{\mu+\Delta}$ and equals $-\frac{\Delta^2}{2(\mu+\Delta)}$, as desired. ■

We also prove Lemma 2.6, restated here for convenience of the reader:

Lemma 2.6 Let G be a multigraph with n vertices and m edges (counted with multiplicities) in which each edge has multiplicity at most k . Let $S \subset [n]$ be a uniformly random subset of vertices and X be the number of edges crossing (S, \bar{S}) . Then for any $\delta > 0$ we have

$$\Pr[X < m/2 \cdot (1 - \delta)] \leq \frac{k}{\delta^2 m}.$$

Proof: Let m_1, m_2, \dots, m_s be multiplicities of edges of G and let $\{p_i\}_{i \in [s]}$ be 0/1 random variables indicating if the corresponding edge crosses the cut or not. Note that $\mathbb{E}[p_i] = 1/2$ and $\mathbb{E}[p_i p_j] = 1/4$ for $i \neq j$. We then infer that

$$\mathbb{E}[X] = \sum_{i=1}^s m_i \cdot \mathbb{E}[p_i] = m/2, \quad \mathbb{E}[X^2] = \sum_{i=1}^s \sum_{j=1}^s m_i m_j \cdot \mathbb{E}[p_i p_j] = m^2/4 + \frac{1}{4} \sum_{i=1}^s m_i^2.$$

By Chebyshev inequality we then have

$$\Pr[X < m/2 \cdot (1 - \delta)] \leq \frac{(\sum m_i^2)/4}{(\delta m/2)^2} \leq \frac{k \cdot m}{\delta^2 m^2} = \frac{k}{\delta^2 m}.$$

■

We now prove Lemma 2.4 restated here for convenience of the reader:

Lemma 2.4 For every $\epsilon, \alpha \in (n^{-1/10}, 1)$, $\alpha < 1/4$ and $T = 512/(\alpha \epsilon^2)$ if $G'_T = (V, E'_T)$, $|V| = n$, $|E'| = m$ be generated according to the process above, then for sufficiently large n there exists $m_0 = m_0(n, \alpha, T)$ such that in the **YES** case the MAX-CUT value is at least m_0 , and in the **NO** case the MAX-CUT value is at most $m_0/(2 - \epsilon)$, both with probability at least $1 - 1/\sqrt{n}$.

Proof: We let $m_0 = \frac{\alpha n T}{2} \cdot (1 - \delta)$ with $\delta = \epsilon/100$. In the **YES** case the graph is bipartite so the value of MAX-CUT is equal to the number of edges in the graph. Since in the **YES** case we only keep those edges of the matchings which cross a fixed random bipartition, and in the union of matchings every edge has multiplicity at most T , Lemma 2.6 ensures that the probability that the number of edges is smaller than m_0 is at most

$$\frac{T}{\delta^2 \alpha n T} = \frac{1}{\delta^2 \alpha n} \leq 1/\sqrt{n},$$

since $\epsilon, \alpha > n^{-1/10}$ by assumption of the lemma.

We now consider the **NO** case. Since every edge of the matchings is kept with probability $1/2$ independently of the others, by Lemma 2.5 with probability at least $1 - \exp\left(-\frac{\delta^2 \alpha n T}{4(1+\delta)}\right)$ (we invoke Lemma 2.5 with $\mu = \alpha n T/2$ and $\Delta = \delta \mu$) the number m of edges in the graph will be at most $\frac{\alpha n T}{2} \cdot (1 + \delta)$. In the following we assume that $m \leq \frac{\alpha n T}{2} \cdot (1 + \delta)$.

Fix a cut (S, \bar{S}) where $S \subseteq V$. Let $k := |S|$. Consider the edges in E_t , that is the matching generated at step t . Consider an arbitrary order of the edges in E_t . Conditioned on all previous edges, an edge e in E_t is a uniformly random edge with endpoints in a set V of vertices not covered by the previous edges. Clearly, $|V| \geq n - 2\alpha n > n/2$. The probability that the edge e crosses the cut (S, \bar{S}) is

$$\frac{|S \cap V| \cdot |\bar{S} \cap V|}{\binom{|V|}{2}} \leq \frac{|V|^2/4}{|V| \cdot (|V| - 1)/2} = \frac{1}{2} \cdot \frac{|V|}{|V| - 1} \leq \frac{1 + 3/n}{2}.$$

We then use Lemma 2.5 with $p = \frac{1+3/n}{2}$ and $\mu = p \cdot m$ applied to the random variables indexed by edges of the graph and equal 1 if the edge crosses the cut (S, \bar{S}) and 0 otherwise. Then for the random variable Y_S which is the number of edges crossing the cut (S, \bar{S}) , Lemma 2.5 gives

$$\Pr\left[Y_S > \frac{m_0}{2 - \epsilon}\right] \leq \exp\left(-\frac{(\frac{m_0}{2 - \epsilon} - p \cdot m)^2}{2m_0/(2 - \epsilon)}\right).$$

We have

$$\frac{m_0}{2-\epsilon} - p \cdot m \geq \frac{\alpha n T \cdot (1-\delta)}{2 \cdot (2-\epsilon)} - \frac{1+3/n}{2} \cdot \frac{\alpha n T}{2} \cdot (1+\delta) \geq \frac{\alpha n T}{4} \cdot \left(\frac{1-\delta}{1-\epsilon/2} - (1+\delta)^2 \right) \geq \alpha n T \epsilon / 16,$$

and

$$2m_0/(2-\epsilon) = \frac{\alpha n T \cdot (1-\delta)}{2-\epsilon} \leq \alpha n T.$$

Putting this together we get

$$\Pr \left[Y_S > \frac{m_0}{2-\epsilon} \right] \leq \exp \left(-\frac{(\alpha n T \epsilon / 16)^2}{\alpha n T} \right) = \exp(-\alpha n T \epsilon^2 / 256) \leq \exp(-2n).$$

where we used the assumption that $T = 512/(\alpha \epsilon^2)$. Taking a union bound over all 2^n possible cuts completes the proof. \blacksquare

D.2 Combinatorics

Lemma D.1 *Let $\mathbf{P}_1 \cup \mathbf{P}_2 \cdots \cup \mathbf{P}_m$ be a partition of the cube $\{0, 1\}^n$. For $x \in \{0, 1\}^n$ let \mathbf{P}_x be the unique part \mathbf{P}_i which contains x . Then*

$$\mathbb{E}_{x \sim \text{Uniform}(\{0,1\}^n)}[1/|\mathbf{P}_x|] = m/2^n.$$

Proof: Indeed, by the definition of the expectation we have

$$\mathbb{E}_{x \sim \text{Uniform}(\{0,1\}^n)}[1/|\mathbf{P}_x|] = \sum_{i=1}^m \mathbb{P}_{x \sim \text{Uniform}(\{0,1\}^n)}[x \in \mathbf{P}_i] / |\mathbf{P}_i| = \sum_{i=1}^m |\mathbf{P}_i| \cdot 2^{-n} / |\mathbf{P}_i| = m/2^n. \quad \blacksquare$$

We now give proof of Lemmas from Section 3.3 which we restate for the convenience of the reader.

Lemma 3.12 *For every integer n and every $0 \leq \ell \leq n/2$*

$$p(\ell, n) = \binom{\alpha n}{\ell} \binom{n}{2\ell}^{-1}.$$

Proof: Let M be a random matching of size αn . The expected number of sets of size 2ℓ which are matched by M is $\binom{\alpha n}{\ell}$ (in fact, it is always equal to $\binom{\alpha n}{\ell}$). The statement of the lemma easily follows. \blacksquare

Lemma 3.13 *For every integer n and every $0 \leq i \leq k \leq n/2$*

$$q(k, i, n) = \binom{\alpha n}{i} \binom{n-2\alpha n}{2(k-i)} \binom{n}{2k}^{-1}.$$

Proof: For any fixed matching M of size αn the number of sets A of cardinality $2k$ for which exactly $2i$ points are matched and there are no boundary edges is $\binom{\alpha n}{i} \binom{n-2\alpha n}{2(k-i)}$. Indeed, to construct such a set A we need to choose i edges of M which match points of A and also choose $2(k-i)$ points of A which are not matched by M . The statement of the lemma easily follows. \blacksquare

Lemma 3.14 *For all non-negative integers n, i, k, b satisfying $k \leq n/2$ and $2i + b \leq 2k$*

$$q(k, i, b, n) = \binom{\alpha n}{i} \binom{\alpha n - i}{b} 2^b \binom{n-2\alpha n}{2(k-i)-b} \binom{n}{2k}^{-1}.$$

Proof: For any fixed matching M of size αn the number of sets A of size with b boundary edges and i inner edges is

$$\binom{\alpha n}{i} \binom{\alpha n - i}{b} 2^b \binom{n - 2\alpha n}{2(k-i) - b}.$$

Indeed, we need to choose i inner edges, then chose b boundary edges, then for each boundary edge decide which of its end-points belongs to A , and choose remaining $2k - 2i - b$ points of A which are not matched by M . The statement of the lemma easily follows. \blacksquare

Lemma 3.15 *Suppose we have $k < n/10$, $\alpha < 1/100$, and $2i + b \leq 2k$ then*

$$q(k, i, b, n) \leq q(k, i, n) 20^{-b} 4^{k-i}.$$

Proof: By Lemma 3.14 and Lemma 3.13 we have

$$q(k, i, b, n) = \binom{\alpha n}{i} \binom{\alpha n - i}{b} 2^b \binom{n - 2\alpha n}{2(k-i) - b} \binom{n}{2k}^{-1}, \quad q(k, i, n) = \binom{\alpha n}{i} \binom{n - 2\alpha n}{2(k-i)} \binom{n}{2k}^{-1}.$$

Consequently, we have

$$\frac{q(k, i, b, n)}{q(k, i, n)} = 2^b \binom{\alpha n - i}{b} \frac{\binom{n - 2\alpha n}{2(k-i) - b}}{\binom{n - 2\alpha n}{2(k-i)}}. \quad (67)$$

Since we have

$$\frac{\binom{n - 2\alpha n}{2(k-i) - b}}{\binom{n - 2\alpha n}{2(k-i)}} = \frac{(2k - 2i)! (n - 2\alpha n - 2(k-i))!}{(2(k-i) - b)! (n - 2\alpha n - 2(k-i) + b)!} = \frac{\binom{2k - 2i}{b}}{\binom{n - 2\alpha n - 2(k-i) + b}{b}},$$

we can rewrite (67) as

$$\frac{q(k, i, b, n)}{q(k, i, n)} = 2^b \frac{\binom{\alpha n - i}{b} \binom{2k - 2i}{b}}{\binom{n - 2\alpha n - 2(k-i) + b}{b}}.$$

We then use the following bounds to handle factors in the numerator:

$$\binom{\alpha n - i}{b} \leq \frac{(\alpha n - i)^b}{b!} \leq \frac{(\alpha n)^b}{b!}, \quad \binom{2k - 2i}{b} \leq 2^{2k - 2i}.$$

To lower bound the denominator we note that $n - 2\alpha n - 2(k-i) \geq n - n/50 - n/5 > n/2$ and so

$$\binom{n - 2\alpha n - 2(k-i) + b}{b} \geq \frac{(n - 2\alpha n - 2(k-i))^b}{b!} \geq \frac{(n/2)^b}{b!}.$$

Putting this together we obtain

$$\frac{q(k, i, b, n)}{q(k, i, n)} \leq \frac{(\alpha n)^b}{b!} 2^b \frac{2^{2(k-i)}}{(n/2)^b / b!} = (4\alpha)^b 4^{k-i} \leq 20^{-b} 4^{k-i},$$

which completes the proof. \blacksquare

D.3 Inequalities

Since we deal with sums when we know much more about sum of squares, we repeatedly use the following fact.

Lemma D.2 *Let a_1, a_2, \dots, a_m be real numbers. Then*

$$\sum_{i=1}^m |a_i| \leq \sqrt{m \sum_{i=1}^m a_i^2}.$$

Proof: Indeed, we have

$$\left(\sum_{i=1}^m |a_i| \right)^2 = m \cdot \sum_{i=1}^m a_i^2 - \sum_{1 \leq i < j \leq m} (a_i - a_j)^2 \leq m \cdot \sum_{i=1}^m a_i^2.$$

We also need the following inequality for binomial coefficients in terms of the entropy function. ■

Lemma D.3 *Let $H(x) = -x \log x - (1-x) \log(1-x)$ be the entropy function. Then for all $0 \leq k \leq n$ we have*

$$\frac{e^{nH(k/n)}}{n+1} \leq \binom{n}{k} \leq e^{nH(k/n)}.$$

Proof: For the lower bound note that for every $t > 1$ we have

$$I(t) := \int_0^1 (xt + (1-x))^n dx = \int_0^1 (1+x(t-1))^n dx = \frac{t^{n+1} - 1}{(n+1)(t-1)} = \frac{1+t+\dots+t^n}{n+1}.$$

On the other hand,

$$I(t) = \sum_{i=0}^n t^i \binom{n}{i} \int_0^1 x^i (1-x)^{n-i} dx.$$

Comparing coefficients at t^k we get

$$\int_0^1 x^k (1-x)^{n-k} dx = \frac{1}{(n+1) \binom{n}{k}}.$$

Apply AM-GM inequality to $x_1 = x_2 = \dots = x_k = x/k, x_{k+1} = \dots = x_n = (1-x)/(n-k)$ to get

$$\frac{x^k (1-x)^{n-k}}{k^k \cdot (n-k)^{n-k}} = \prod_{i=1}^n x_i \leq \left(\frac{x_1 + x_2 + \dots + x_n}{n} \right)^n = 1/n^n.$$

This implies

$$\frac{1}{(n+1) \binom{n}{k}} = \int_0^1 x^k (1-x)^{n-k} dx \leq \max_{x \in [0,1]} x^k (1-x)^{n-k} \leq \frac{k^k \cdot (n-k)^{n-k}}{n^n} = e^{-nH(k/n)},$$

thus showing the lower bound.

For the upper bound let X be a uniformly random subset of $\{1, 2, \dots, n\}$ of size k . We view X as a bit string (x_1, x_2, \dots, x_n) where $x_i = \mathbb{1}_{i \in X}$. Note that $\Pr[x_i = 1] = k/n$. We then have

$$\log \binom{n}{k} = H(X) \leq H(x_1) + \dots + H(x_n) = n \cdot H(x_1) = n \cdot H(k/n),$$

which implies the upper bound. ■

We frequently deal with the expressions of the form $(m/x)^x$ or $(m/x)^{x/2}$. The following lemma gives an upper bound.

Lemma D.4 For any $m > 0$ and any $x > 0$ the following inequalities hold true

$$\left(\frac{m}{x}\right)^x \leq e^{m/e}, \quad \left(\frac{m}{x}\right)^{x/2} \leq e^{\frac{m}{2e}}.$$

Proof: The second inequality follows from the first one by taking a square root. To prove the first inequality note that for $f(t) = \frac{\log t}{t}$ we have

$$f'(t) = \frac{1 - \log t}{t^2},$$

which means that $f(t)$ is increasing on $(0, e)$ and decreasing on (e, ∞) . Thus, $f(t) \leq f(e)$ for every $t > 0$ and so

$$\left(\frac{m}{x}\right)^x = \left((m/x)^{x/m}\right)^m = \left(e^{f(m/x)}\right)^m \leq \left(e^{f(e)}\right)^m = e^{m/e}.$$

We also deal with factorials and functions of the form ℓ^ℓ . The following lemma says that up to exponentially large factors they are the same and enjoy the property $f(x+y) \approx f(x)f(y)$. ■

Lemma D.5 Let a_1, \dots, a_m be positive integers and $S = \sum a_i$. Let f_1, \dots, f_m be functions, each of them is either $x \mapsto x^x$ or $x \mapsto x!$. Then

$$\left(\frac{S}{3m}\right)^S \leq \prod_{i=1}^m f_i(a_i) \leq S^S.$$

Proof: For the upper bound note that

$$\prod_{i=1}^m f_i(a_i) \leq \prod_{i=1}^m a_i^{a_i} \leq \prod_{i=1}^m S^{a_i} = S^S.$$

For the lower bound we have

$$\prod_{i=1}^m f_i(a_i) \geq \prod_{i=1}^m a_i! \geq \prod_{i=1}^m \left(\frac{a_i}{e}\right)^{a_i} = e^{-S} \prod_{i=1}^m a_i^{a_i}.$$

Consider a multiset of S integers where each a_k appears a_k times. Sum of reciprocals of numbers in this set is then m , which means that the harmonic mean is S/m . The geometric mean is at least as large as the harmonic mean so

$$\sqrt[S]{\prod_{i=1}^m a_i^{a_i}} \geq \frac{S}{m}.$$

This implies that

$$\prod_{i=1}^m f_i(a_i) \geq e^{-S} \left(\frac{S}{m}\right)^S > \left(\frac{S}{3m}\right)^S.$$

Remark 1 In all the applications we have $m \leq 4$. We also frequently apply the lemma for a quotient of two products, in which case what remains (up to exponential factor) is $f(\sum a_i - \sum b_j)$ where a_i 's are in the numerator and b_j 's are on the denominator. ■

Remark 2 If applying the lemma we have expression of the form $\ell^{\ell/2}$ it should be replaced with $(\ell/2)^{\ell/2} \cdot 2^{\ell/2} = f(\ell/2) \cdot 2^{\ell/2}$.

E Fourier spectrum of the component growing protocol from section 5.1

Suppose that the players use the component growing protocol described in section 5.1. For each $t = 1, \dots, T$ let E_t^* denote the set of edges of the forest F_t . In that case the set of possible values of X^* consistent with the players' knowledge at time t can be defined quite easily:

$$\mathbf{B}_t = \{x \in \{0, 1\}^n : \forall e = (a, b) \in E_t^*, x_a + x_b = w_e\}.$$

Thus, \mathbf{B}_t is simply a linear subspace of $\{0, 1\}^n$, where the constraints are given by the edges in E_t^* . Recall that we denote the indicator of \mathbf{B}_t by h_t . We now derive a characterization of \widehat{h}_t . We call a coefficient $v \in \{0, 1\}^n$ *admissible* if it has an even intersection with every connected component in E_t^* (see Fig. 1(a), where the elements of an admissible v are marked red). For each admissible v let $Q(v)$ denote the unique pairing of vertices of v via edge-disjoint paths in E_t^* (we associate Q_v with the set of edges on these paths). This is illustrated in Fig. 1, where the vertices of $v \in \{0, 1\}^n$ are marked red, and the edges of $Q(v)$ are the green dashed edges. Note that since edges of E_t^* form a forest, this pairing is indeed unique for every admissible v .

We show that $\widehat{h}_t(v)$ has the following simple form:

$$\widehat{h}_t(v) = \begin{cases} \frac{|\mathbf{B}_t|}{2^n} \cdot (-1)^{\sum_{e \in Q(v)} w_e} & \text{if } v \text{ is admissible} \\ 0 & \text{o.w..} \end{cases} \quad (68)$$

Recall that w_e are the labels on the edges of the graph G_t formed by first t matchings.

We now prove Eq. (68). We first prove that $\widehat{h}_t(v) = 0$ for any inadmissible v . By definition of the Fourier transform

$$\begin{aligned} \widehat{h}_t(v) &= \frac{1}{2^n} \sum_{w \in \{0, 1\}^n} h_t(w) (-1)^{v \cdot w} = \frac{1}{2^n} \sum_{w \in \mathbf{B}_t} (-1)^{v \cdot w} \\ &= \frac{|\{w \in \mathbf{B}_t : v \cdot w \text{ is even}\}| - |\{w \in \mathbf{B}_t : v \cdot w \text{ is odd}\}|}{2^n}. \end{aligned}$$

Now suppose that $v \in \{0, 1\}^n$ has an odd intersection with at least one of the connected components in E_t^* . Denote the set of vertices in this component by $\mathcal{C}_* \subseteq [n]$, and let $\mathbf{1}_{\mathcal{C}_*} \in \{0, 1\}^n$ denote the indicator vector of vertices in \mathcal{C}_* . We now note that for any $w \in \mathbf{B}_t$ one necessarily has that $w + \mathbf{1}_{\mathcal{C}_*} \in \mathbf{B}_t$. Indeed, adding 1 to every vertex in \mathcal{C}_* could only violate those constraints (edges) that have exactly one endpoint in \mathcal{C}_* . But there are no such edges since \mathcal{C}_* is a connected component by definition, so $w + \mathbf{1}_{\mathcal{C}_*} \in \mathbf{B}_t$ as required. On the other hand, v has an odd intersection with \mathcal{C}_* , we have $v \cdot \mathbf{1}_{\mathcal{C}_*} = 1$, so for any $w \in \{0, 1\}^n$

$$v \cdot (w + \mathbf{1}_{\mathcal{C}_*}) = v \cdot w + 1.$$

This means that $|\{w \in \mathbf{B}_t : v \cdot w \text{ is even}\}| = |\{w \in \mathbf{B}_t : v \cdot w \text{ is odd}\}|$, since the map $w \rightarrow w + \mathbf{1}_{\mathcal{C}_*}$ is an involution on \mathbf{B}_t , and hence $\widehat{h}_t(v) = 0$ as required.

Now suppose that v is admissible. To derive the equation for \widehat{h}_t given in Eq. (68), we note that \mathbf{B}_t can be alternatively characterized as follows. Pick any element $w^* \in \mathbf{B}_t$, and let $\mathcal{C}_1, \dots, \mathcal{C}_k$ denote the connected components in E_t^* (so that each singleton node is a connected component of its own). Then

$$\mathbf{B}_t = \{w^* + \sum_{i=1}^k \lambda_i \mathbf{1}_{\mathcal{C}_i} : \lambda \in \{0, 1\}^k\}.$$

Noting that $v \cdot \mathbf{1}_{\mathcal{C}_i} = 0$ for all $i = 1, \dots, k$, we note that for any $w \in \{0, 1\}^n$

$$v \cdot w = v \cdot (w^* + \sum_{i=1}^k \lambda_i \mathbf{1}_{\mathcal{C}_i}) = v \cdot w^*.$$

We thus have

$$\widehat{h}_t(v) = \frac{1}{2^n} \sum_{w \in \mathbf{B}_t} (-1)^{v \cdot w} = \frac{|\mathbf{B}_t|}{2^n} (-1)^{v \cdot w^*}.$$

Let Λ be the the unique pairing of vertices of v by edge disjoint paths in F_t . For any pair $\{i, j\} \in \Lambda$ let $\mathcal{P}_{i,j}$ denote the path from i to j in F_t . Note that

$$\sum_{e \in \mathcal{P}_{i,j}} w_e = \sum_{e=(a,b) \in \mathcal{P}_{i,j}} (X_a^* + X_b^*) = X_i^* + X_j^*.$$

To complete the proof, it suffices to note that

$$v \cdot w^* = \sum_{\{i,j\} \in \Lambda} X_i^* + X_j^* = \sum_{\{i,j\} \in \Lambda} \sum_{e \in \mathcal{P}_{i,j}} w_e = \sum_{e \in Q(v)} w_e$$

as required.

We now turn to the Fourier transform of f_t for the component growing protocol. Let $M_t^* := M_t \cap E_t^*$ denote the set of edges of M_t whose bits are revealed by the t -th player. We have $E_t^* = E_{t-1}^* \cup M_t^*$ for all $t = 1, \dots, T$. Note that the simple rule that defines M_t^* immediately specifies the set \mathbf{A}_t for our component growing protocol (recall that \mathbf{A}_t was defined in Definition 4.1). For completeness, we instantiate both the definitions of $\mathbf{A}_{t, \text{reduced}}$ and \mathbf{A}_t for our component growing protocol. The set $\mathbf{A}_{t, \text{reduced}}$ is simply

$$\mathbf{A}_{\text{reduced}, t} = \{z \in \{0, 1\}^{M_t} : z_e = (w_t)_e \text{ for all } e \in M_t^*\}.$$

Thus, $\mathbf{A}_{\text{reduced}, t}$ is a subcube of the boolean hypercube $\{0, 1\}^{M_t}$ obtained by fixing coordinates in M_t^* to their specified values. The set \mathbf{A}_t is defined as $\mathbf{A}_t = \{x \in \{0, 1\}^n : M_t x \in \mathbf{A}_{t, \text{reduced}}\}$ (see Definition 4.1), and in our case is a linear subspace of $\{0, 1\}^n$:

$$\mathbf{A}_t = \{x \in \{0, 1\}^n : \text{for all } e = (a, b) \in M_t^* \ x_a + x_b = (w_t)_e\}.$$

Recall that the function f_t is defined as the indicator of \mathbf{A}_t (see Definition 4.1). Since \mathbf{A}_t is a linear subspace just like \mathbf{B}_t , the Fourier transform of f_t is also quite easy to understand. We note that the same derivation as for h_t shows that the Fourier transform of \widehat{f}_t is supported on edges of M_t . We now say that $v \in \{0, 1\}^n$ is admissible if it has even intersection with every connected component in M_t^* . But this can only happen when v is a union of edges of M_t^* , as required.

F Proof of Lemmas 3.3 and 3.4

F.1 General bound for the L_1 mass of the Fourier transform

Here we prove Lemma 3.3. The starting point is the following Lemma which can be found in [KKL88].

Lemma F.1 *Let f be a function $f : \{0, 1\}^m \rightarrow \{-1, 0, 1\}$ and let $\mathbf{A} = f^{-1}(\{-1, 1\})$. Let $|s|$ denote the Hamming weight of $s \in \{0, 1\}^m$. Then for every $\delta \in [0, 1]$*

$$\sum_{s \in \{0, 1\}^m} \delta^{|s|} \widehat{f}^2(s) \leq \left(\frac{|\mathbf{A}|}{2^m} \right)^{\frac{2}{1+\delta}}.$$

We then deduce

Lemma F.2 Let f be a function on $\{0, 1\}^m$ taking values in $\{-1, 0, 1\}$. Define a set $\mathbf{A} = f^{-1}(\{-1, 1\})$. Then if $|\mathbf{A}| \geq 2^{m-d}$ and $q \leq d$ then

$$\left(\frac{2^m}{|\mathbf{A}|}\right)^2 \sum_{\substack{x \in \{0,1\}^m \\ |x|=q}} \widehat{f}^2(x) \leq \left(\frac{4d}{q}\right)^q.$$

Proof: By the Lemma above, for every $\delta \in [0, 1]$ we have

$$\left(\frac{2^m}{|\mathbf{A}|}\right)^2 \sum_{\substack{x \in \{0,1\}^m \\ |x|=q}} \widehat{f}^2(x) \leq \frac{2^{2m}}{|\mathbf{A}|^2} \delta^{-q} \left(\frac{|\mathbf{A}|}{2^m}\right)^{\frac{2}{1+\delta}} = \delta^{-q} \left(\frac{2^m}{|\mathbf{A}|}\right)^{\frac{2\delta}{1+\delta}} \leq \delta^{-q} \left(\frac{2^m}{|\mathbf{A}|}\right)^{2\delta} \leq \frac{2^{2\delta d}}{\delta^q}.$$

Plugging in $\delta = \lambda q/d$ with $\lambda \in [0, 1]$ (which ensures that $\delta \in [0, 1]$) we obtain

$$\left(\frac{2^m}{|\mathbf{A}|}\right)^2 \sum_{\substack{x \in \{0,1\}^m \\ |x|=q}} \widehat{f}^2(x) \leq \frac{2^{2\delta d}}{\delta^q} = \left(\frac{2^{2\lambda d}}{\lambda q}\right)^q.$$

It remains to note that for $\lambda = \frac{1}{2 \log 2}$ we have $2^{2\lambda}/\lambda = 2e \log 2 < 4$. ■

Lemma F.3 Let $\mathbf{A} \subset \{0, 1\}^m$ be a set of cardinality at least 2^{m-d} with indicator function f . Then for every $y \in \{0, 1\}^m$ and every $q \leq d$ one has

$$\sum_{\substack{x \in \{0,1\}^m \\ |x \oplus y|=q}} \widetilde{f}^2(x) \leq \left(\frac{4d}{q}\right)^q.$$

Proof: Recall that $\widetilde{f}(x) := \frac{2^m}{|\mathbf{A}|} \cdot \widehat{f}(x)$, see definition 3.2. So the statement is equivalent to the following inequality:

$$\left(\frac{2^m}{|\mathbf{A}|}\right)^2 \sum_{\substack{x \in \{0,1\}^m \\ |x \oplus y|=q}} \widehat{f}^2(x) \leq \left(\frac{4d}{q}\right)^q.$$

Consider a function $g(z) := f(z)(-1)^{z \cdot y}$. For this function we have

$$\widehat{g}(x) = 2^{-m} \sum_{z \in \{0,1\}^m} g(z)(-1)^{z \cdot x} = 2^{-m} \sum_{z \in \{0,1\}^m} f(z)(-1)^{z \cdot (x \oplus y)} = \widehat{f}(x \oplus y).$$

It remains to apply the previous lemma to the function $g(z)$. ■

Since there are exactly $\binom{m}{q}$ different $x \in \{0, 1\}^m$ for which $|x \oplus y| = q$, the above Lemma together with Lemma D.2 imply Lemma 3.3.

F.2 Structure of the Fourier transform of a single player's message

Here we prove Lemma 3.4.

Proof: We compute the Fourier transform of $f(x)$. For $z \in \{0, 1\}^n$ let $x(z) \in \{0, 1\}^n$ be defined by setting, for each edge $(u, v) \in M$,

$$x(z)_u = z_e \text{ and } x(z)_v = 0$$

and $x(z)_w = 0$ if w is not matched by M . Note that $x(z)$ is a particular solution of $Mx = z$ and the set of solutions is given by

$$\{x(z) + Ns : s \in \{0, 1\}^{n-\alpha n}\}, \quad (69)$$

where N is a basis for the kernel of M . Without loss of generality suppose that M contains the edges $(2i - 1, 2i), i = 1, \dots, \alpha n$. Then the matrix $N \in \{0, 1\}^{n \times (n-\alpha n)}$ may be taken as

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix},$$

where first αn columns form M^T , the bottom right submatrix is the $(n - 2\alpha n) \times (n - 2\alpha n)$ identity, all the other entries are zero.

The Fourier transform of f at $v \in \{0, 1\}^n$ is given by

$$\begin{aligned} \hat{f}(v) &= \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} f(x) \cdot (-1)^{x \cdot v} \\ &= \frac{1}{2^n} \sum_{z \in \mathbf{A}_{reduced}} \sum_{s \in \{0, 1\}^{n-\alpha n}} (-1)^{(x(z) + Ns) \cdot v} \\ &= \frac{1}{2^n} \sum_{z \in \mathbf{A}_{reduced}} (-1)^{x(z) \cdot v} \sum_{s \in \{0, 1\}^{n-\alpha n}} (-1)^{(v^T N) \cdot s} \end{aligned}$$

First note that

$$\sum_{s \in \{0, 1\}^{n-\alpha n}} (-1)^{(v^T N) \cdot s} = \mathbf{1}_{v^T N = 0} \cdot 2^{n-\alpha n},$$

so $\hat{f}(v) = 0$ unless $v^T N = 0$. Note that all such v are of the form $v = M^T r$ for some $r \in \{0, 1\}^{\alpha n}$.

Thus,

$$\hat{f}(M^T r) = \frac{2^{n-\alpha n}}{2^n} \sum_{z \in \mathbf{A}_{reduced}} (-1)^{x(z) \cdot M^T r} = \frac{2^{n-\alpha n}}{2^n} \sum_{z \in \mathbf{A}_{reduced}} (-1)^{z \cdot r} = \hat{q}(r)$$

and $\hat{f}(v) = 0$ for all v not of the form $M^T r$. Here we use the fact that $x(z) \cdot M^T r = z \cdot r$ for all z and r .

Note that Fourier coefficients of f only have even weight, and weight k Fourier coefficients of q are in direct correspondence with weight $2k$ coefficients of f (since $|M^T r| = 2|r|$ for all $r \in \{0, 1\}^{\alpha n}$). ■