$(1+\Omega(1))$-Approximation to MAX-CUT Requires Linear Space

Michael Kapralov* Sanjeev Khanna† Madhu Sudan‡ Ameya Velingker§

Abstract

We consider the problem of estimating the value of MAX-CUT in a graph in the streaming model of computation. We show that there exists a constant $\epsilon_* > 0$ such that any randomized streaming algorithm that computes a $(1+\epsilon_*)$-approximation to MAX-CUT requires $\Omega(n)$ space on an n vertex graph. By contrast, there are algorithms that produce a $(1+\epsilon)$-approximation in space $O(n/\epsilon^2)$ for every $\epsilon > 0$. Our result is the first linear space lower bound for the task of approximating the max cut value and partially answers an open question from the literature [2]. The prior state of the art ruled out $(2-\epsilon)$-approximation in $\tilde{O}(\sqrt{n})$ space or $(1+\epsilon)$-approximation in $n^{1-O(\epsilon)}$ space, for any $\epsilon > 0$.

Previous lower bounds for the MAX-CUT problem relied, in essence, on a lower bound on the communication complexity of the following task: Several players are each given some edges of a graph and they wish to determine if the union of these edges is ϵ-close to forming a bipartite graph, using one-way communication. The previous works proved a lower bound of $\Omega(\sqrt{n})$ for this task when $\epsilon = 1/2$, and $n^{1-O(\epsilon)}$ for every $\epsilon > 0$, even when one of the players is given a candidate bipartition of the graph and the graph is promised to be bipartite with respect to this partition or ϵ-far from bipartite. This added information was essential in enabling the previous analyses but also yields a weak bound since, with this extra information, there is an $n^{1-O(\epsilon)}$ communication protocol for this problem. In this work, we give an $\Omega(n)$ lower bound on the communication complexity of the original problem (without the extra information) for $\epsilon = \Omega(1)$ in the three-player setting. Obtaining this $\Omega(n)$ lower bound on the communication complexity is the main technical result in this paper. We achieve it by a delicate choice of distributions on instances as well as a novel use of the convolution theorem from Fourier analysis combined with graph-theoretic considerations to analyze the communication complexity.

*School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland. Email: michael.kapralov@epfl.ch

†Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104. Email: sanjeev@cis.upenn.edu. Supported in part by National Science Foundation grants CCF-1116961, CCF-1552909, CCF-1617851, and IIS-1447470.

‡Harvard John A. Paulson School of Engineering and Applied Sciences, 33 Oxford Street, Cambridge, MA 02138, USA. Email: madhu@cs.harvard.edu. Supported in part by NSF Award CCF 1565641.

§School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland. Email: avelingk@cs.cmu.edu. This work was partly done while the author was at Carnegie Mellon University. Supported in part by National Science Foundation grant CCF-0963975.

1 Introduction

In this paper, we consider the space complexity of approximating MAX-CUT in the streaming model of computation. We elaborate on these terms and describe our main result below.

The input to the MAX-CUT problem is an undirected graph, and the goal is to find a bipartition of the vertices of this graph (or a *cut*) that maximizes the number of edges that cross the bipartition. The size of a MAX-CUT on graph G, denoted MAX-CUT(G), is the number of edges that cross the optimal bipartition. An algorithm A is said to produce an α-approximation to the size of the MAX-CUT if for every graph G, the algorithm's output $A(G)$ satisfies $\text{MAX-CUT}(G)/\alpha \leq A(G) \leq \text{MAX-CUT}(G)$.

In this paper, we study the space complexity of approximating MAX-CUT in the streaming model of computation. The streaming model of computation, formally introduced in the seminal work of [8] and motivated by applications in processing massive datasets, is an extremely well-studied model for designing sublinear space algorithms. For the MAX-CUT problem in this model, the edges of the input graph G are presented as a stream to a (randomized) algorithm, which must output an α-approximation to MAX-CUT(G). The complexity measure is the space complexity, namely, the number of bits of memory used by the streaming algorithm, measured as a function of n, the number of vertices in G.

Our main result is a strong lower bound (optimal to within polylogarithmic factors) on the space required for a strong approximation to the MAX-CUT size. Specifically, we show that there is an $\alpha > 1$ such that every α-approximation algorithm in the streaming model must use $\Omega(n)$ space (see Theorem 1.1).

Context and Significance. There are two basic algorithmic results for MAX-CUT in the streaming model: On the one hand, the trivial algorithm that counts the number, say m, of edges in G and outputs $m/2$ is a 2-approximation that uses $O(\log n)$ space. On the other

hand, if one has $\tilde{O}(n)$ space[1], one can get an *approximation scheme*, i.e., a $(1+\epsilon)$-approximation algorithm for every $\epsilon > 0$, by building a "cut-sparsifier" [10, 28].

Given just the two algorithms above, it is possible to envision three possible scenarios for improving the approximability of MAX-CUT: (1) Perhaps MAX-CUT has an approximation scheme in polylogarithmic space? (2) Perhaps MAX-CUT admits a space-approximation tradeoff, i.e., for every $\alpha > 1$, there is a $\beta < 1$ such that an α-approximation can be computed in n^β space? (3) Perhaps there is an $\alpha < 2$ and an algorithm using n^β space for some $\beta < 1$ that can compute an α-approximation to MAX-CUT. (Note that the scenarios are nested with (1) $\Rightarrow$ (2) $\Rightarrow$ (3).)

Previous works [25, 21] have ruled out scenario (1) above, making progress on an open question from [1]. In particular, these works have showed that for every $\beta < 1$, there exists $\alpha > 1$ such that a streaming algorithm with space n^β cannot compute an α-approximation to MAX-CUT. The work of [21] also shows that $\beta < 1/2$ and $\alpha < 2$ are not simutaneously achievable. These results still allow for either scenario (2) or (3). Our result achieves the next level of understanding by ruling out scenario (2) as well:

THEOREM 1.1. (MAIN RESULT) *There exists $\epsilon^* > 0$ such that every randomized single-pass streaming algorithm that yields a $(1+\epsilon^*)$-approximation to the MAX-CUT size with probability at least $9/10$ must use $\Omega(n)$ space, where n denotes the number of vertices in the input graph.*

This step has also been suggested as an open problem in the Bertinoro workshop [2], though we settle their question only partially since the question suggests a particular approach to proving the lower bound, which we do not follow. Eventually we suspect that even scenario (3) is not achievable, but ruling this out involves more technical challenges. Indeed, one of the hopes of this work is to introduce some techniques that may be useful in the eventual resolution of this problem.

Techniques. As with most lower bounds in streaming, ours is obtained by a reduction from a communication complexity problem. However, the communication problem and even communication model in this paper are somewhat new, so we describe our model and then explain why the novelty is necessary and useful.

Roughly, our paper considers a T-player sequential communication game, that we call the **Implicit Hidden Partition Problem**, where player P_i, for $1 \le i \le T$, is given a set of edges E_i on vertex set $[n]$, and the players wish to determine whether $\cup_i E_i$ forms a bipartite graph or is ϵ-far from being bipartite. (To be more precise, in our actual game the players also get some "non-edges" F_i and they also need to verify that (most of) the edges of F_i do not cross the bipartition, but we ignore this distinction here since it is not conceptually significant.) The communication is one-way and player P_i is only allowed to broadcast a message based on its own input and broadcast messages from players P_j for $1 \le j < i$. We show that for $T = 3$ and some $\epsilon > 0$, there is a distribution on inputs for which this task requires $\Omega(n)$ communication.

The communication problems from previous works included an additional player P_0 whose input was a bipartition of the vertices of the graph, and later players needed to verify that the graph was bipartite with respect to this bipartition. The presence of this additional player was essential to previous analyses. These analyses roughly suggested that when the input graph is far from being bipartite, conditioned on not discovering a violating edge, the information of the first i players is effectively dominated by the information of P_0 — i.e., knowledge of the partition subsumes all other knowledge. This suggests a reduction from the T (or $T+1$) player communication problem to several two-player games involving player P_0 and P_i for $1 \le i \le T$, and this two player game can be analyzed as in [14, 29]. Implementing this reduction does take technical work, but the intuition works!

For our purposes, the presence of the 0-th player poses an insurmountable obstacle—with this player, there is a $O(\sqrt{n} \cdot \text{poly}(1/\epsilon))$ communication protocol (based on the "birthday paradox") to distinguish bipartite graphs that are ϵ-far from being bipartite! Indeed, one can just send information about the classification of about $\sqrt{n}$ vertices with respect to the bipartition and check how many edges violate the bipartition. Harder communication complexity problems (e.g. the Boolean Hidden Hypermatching Problem of [29]—see [25, 21]) have been considered, leading to stronger $n^{1-O(\epsilon)}$ lower bounds on testing ϵ-closeness to bipartite, but they still use an explicit candidate bipartition and admit $n^{1-O(\epsilon)}$ protocols for any constant ϵ. This forces us to remove the 0-th player, thereby leading to the (in retrospect, more natural) "Implicit Hidden Partition" problem that we introduce explicitly in this paper.

The removal of the 0-th player, however, forces us to introduce new mechanisms to cope with the leakage of information as the protocol evolves. We do so by changing the communication model to allow for some "public inputs" and some "private inputs". All inputs to player i are selected after the transmission of the message of player $i-1$, and the public input becomes

[1] Throughout this paper we use the notation $\tilde{O}(f(n))$ to denote the set $\cup_{c>0} O(f(n)(\log(f(n))^c))$.

known to all players while the private input is known only to player i. (In our case, the public input is a superset of the edges E_i and the private input is the set E_i.) This separation brings back a little flexibility into our analysis, but the task of bounding the flow of relevant information as the protocol evolves remains challenging and, indeed, we are only able to carry out such an analysis for $T = 3$, by a careful choice of input distributions and paramters.

One major challenge is the task of finding the right set of hard instances for the problem. Natural candidates (for example the one suggested in [2]) would involve random bipartite graphs and random graphs; however, the presence of vertices of degree larger than 2 in these graphs poses obstacles to our analysis. So we pick a delicate distribution in which the graph formed by $E_1 \cup E_2$ has no cycles and no vertices of degree > 2 (so $E_1 \cup E_2$ is a union of paths). Of course, this implies that the resulting graph is bipartitite, thereby allowing the final edge set E_3 to come into play. Our final edge set E_3 is chosen to be either a random graph consistent with this bipartition (the **YES** case), or a random sufficiently dense graph (the **NO** case) so that the resulting graph $(E_1 \cup E_2 \cup E_3)$ is $\Omega(1)$-far from being bipartite. The choice of parameters is delicate—we need to ensure that the distributions of E_3 in the **YES** and **NO** cases are statistically close while still ensuring that $E_1 \cup E_2 \cup E_3$ is far from bipartite in the case of NO instances.

Finally, we are left with the task of actually analyzing the communication protocols aiming to solve the communication problem on the aforementioned distribution. As with previous works [14], we make use of Fourier analysis. We specifically analyze the set of bipartitions that are consistent with the set of public inputs and messages broadcast thus far and then look at the Fourier coefficients of the indicator function of this set. We employ relatively elementary methods (at least given previous works) to analyze this set after the player P_1 speaks. To analyze the set after player P_2 speaks, we perform some combinatorial analysis involving the special distributions on E_1 and E_2 and then incorporate this combinatorics into the Fourier language, while finally combining the effects of the two steps using the convolution theorem in Fourier analysis. While the use of this theorem is natural in our setting (involving a composition of many messages, that corresponds to a product of various indicator functions), the fact that the convolved coefficients can be subjected to spectral analysis appears somewhat novel, and we hope it will spur further progress on this and other questions.

Related work. 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 [27]. 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 [3, 24, 7, 22], spanning trees [6], matchings [4, 5, 15, 19, 16, 17, 26, 9], spanners [7, 23]. 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. [20, 13, 11]). 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. We formally define our communication problem and describe its connection to streaming algorithms for approximating MAX-CUT value in Section 2. We then state the main technical lemmas and prove the main theorem in Section 3. The proof of the main technical lemma of our communication lower bound is given in Section 4. The gap analysis is omitted in this paper but can be found in the full version of this paper.

2 Communication problem and hard distribution

In this section, we introduce a multi-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 sequential communication model where T players sequentially receive *public inputs* M_t and *private inputs* w_t, for $t \in [T]$. A problem in this model is specified by an $F(M_1, \ldots, M_T; w_1, \ldots, w_T)$ and the goal of the players is to compute this function. A protocol for this problem Π is specified by a sequence of functions $\Pi = (r_1, \ldots, r_t)$. At stage $t \in [T]$, the t-th player announces its message $a_t = r_t(M_1, \ldots, M_t; a_1, \ldots, a_{t-1}; w_t)$, and the message a_T is defined to be the output of the protocol Π. The complexity of Π, denoted $|\Pi|$, is the maximum length of the messages $\{a_t\}_{t\in[T]}$. We consider the distributional setting, i.e., where the inputs are drawn from some distribution μ and the error of the protocol is the probability that its output does not equal $F(M_1, \ldots, M_T; w_1, \ldots, w_T)$. By Yao's minmax principle, we assume, without loss of generality, that the communication protocol is deterministic. Also, for the remainder of the paper, addition over $\{0,1\}^n$ and matrix multiplication occurs modulo 2.

We now describe the specific communication problem that we consider in this work.

Implicit Hidden Partition (IHP) Problem. The

T-player Implicit Hidden Partition problem **IHP**(n) for positive integer n is defined as follows: The public inputs are sets of edges, $M_1, \ldots, M_T$, on vertex set $[n]$, while the private inputs $w_1, \ldots, w_T$ are $\{0,1\}$-colorings of the corresponding sets of edges. The goal is to distinguish the case in which the colorings are *valid* (i.e., there exists a cut such that every edge of $\cup_t M_t$ is colored 1 if and only if it crosses the cut) from the case in which no such cut exists. A convenient representation of the inputs will be to represent the edges M_t as incidence matrices $M_t \in \{0,1\}^{m_t \times n}$ and the coloring by $w_t \in \{0,1\}^{m_t}$, for $t \in [T]$, where m_t denotes the number of edges of M_t. In this representation a coloring $x \in \{0,1\}^n$ is valid if and only if $M_t x = w_t$ for every $t \in [T]$.

In the instances we use, we will set $T = 3$, while M_1 and M_2 will be (incidence matrices of) matchings so that their rows sum to 2 and columns sum to at most 1. Also, $M_3 \in \{0,1\}^{m_3 \times n}$ will be the edge incidence matrix of a suitable cycle-free subgraph of an Erdős-Rényi graph below the threshold for emergence of a giant component.

Distributional Implicit Hidden Partition (DIHP) Problem. In this work, we will actually deal with a *distributional* version of **IHP** with $T = 3$ that we denote **DIHP**. **DIHP** has three parameters: a positive even integer Δ, a positive integer n divisible by Δ, and a real number α with $0 < \alpha < 1$. **DIHP**(n, Δ, α) is defined to be **IHP**(n) on inputs chosen from a distribution $\mathcal{D} = \frac{1}{2}(\mathcal{D}^Y + \mathcal{D}^N)$, where $\mathcal{D}^Y$ and $\mathcal{D}^N$ are defined as follows: In both the distributions $\mathcal{D}^Y$ and $\mathcal{D}^N$, the triples (M_1, M_2, M_3) are chosen identically from a process $\mathcal{P}_{n,\Delta,\alpha}$ that we describe below shortly. In $\mathcal{D}^Y$, the private inputs w_1, w_2, w_3 are chosen by sampling $X^* \in \{0,1\}^n$ uniformly and settting $w_t = M_t X^*$ for $t \in \{1,2,3\}$. Note that the distribution $\mathcal{D}^Y$ is supported on **YES** instances. In the distribution $\mathcal{D}^N$, the w_t's are uniformly random strings chosen independently of each other. As we show later, the distribution $\mathcal{D}^N$ is mostly supported on **NO** instances that are, in fact, far from **YES** instances, where distance is measured in terms of the number of edges that have to be removed in order to produce a valid coloring.

Although the notation M_t denotes an $m_t \times n$ edge incidence matrix, we will often use M_t to denote the corresponding graph as well. However, the sense in which M_t is used will be clear from context. Furthermore, we will use E_t to denote the set of edges specified by M_t.

Edge Sampling Process $\mathcal{P}_{n,\Delta,\alpha}$. We now specify the process $\mathcal{P}_{n,\Delta,\alpha}$, which is used to sample the graphs (edge incidence matrices) M_1, M_2, M_3 in both $\mathcal{D}^Y$ and $\mathcal{D}^N$. The set M_1 is a deterministic perfect matching that matches vertex i to $i + n/2$ for every $i \in [n/2]$. The set M_2 is a also a matching sampled as follows: We sample a permutation $\pi : [n/2] \to [n/2]$ uniformly and then match the vertex $\pi(i)$ to the vertex $\pi(i+1) + n/2$ for every i that is *not* divisible by $\Delta/2$. (Note that by this process, the union of the graphs $M_1 \cup M_2$ is a collection of disjoint paths, each of length $\Delta - 1$.) Finally, we sample M_3 in three steps:

Step 1. We first sample a random graph M_3' from the Erdős-Rényi model with parameter α/n, i.e., every possible edge is included independently with probably α/n.

Step 2. We remove all edges in M_3' that have already been included in $M_1 \cup M_2$ to get a subgraph M_3''.

Step 3. We now consider the connected components of M_3'' and, for every component that contains a cycle, we remove all edges of that component. The resulting subgraph is M_3.

Note that since α is close to 1, the graph M_3 (or M_3' for that matter) is subcritical and most of its components are of constant size. At most a constant number of edges of M_3' appear in $M_1 \cup M_2$ and another small constant appear in cycles. Thus, for all practical purposes, M_3 behaves like M_3'. In particular, as we show later, the fraction of invalidly colored edges in a random coloring of the edges remains nearly the same in $M_1 \cup M_2 \cup M_3$ as in $M_1 \cup M_2 \cup M_3'$.
The following theorem is the main technical contribution of the paper:

THEOREM 2.1. *There exist constants $\Delta^* > 0$ and $0 < \alpha^* < 1$ such that for every even integer $\Delta \geq \Delta^*$ and every $\alpha \in (\alpha^*, 1)$, there exists $c > 0$ such that the following holds: For every sufficiently large integer n that is divisible by Δ, every protocol Π for* **DIHP**(n, Δ, α) *that succeeds with probability at least $2/3$ satisfies $|\Pi| \geq cn$.*

We accompany the above theorem with a reduction from **DIHP** to MAX-CUT:

THEOREM 2.2. (REDUCTION FROM **DIHP** TO MAX-CUT) *There exist constants $\Delta^* > 0$ and $0 < \alpha^* < 1$ such that for every even integer $\Delta \geq \Delta^*$ and every $\alpha \in (\alpha^*, 1)$, there exists $\epsilon^* > 0$ such that the following holds: If there exists a single-pass streaming $(1 + \epsilon^*)$-approximation algorithm for MAX-CUT with space complexity $s(n)$ that succeeds with probability at least 9/10, then there exists a protocol Π for* **DIHP**(n, Δ, α) *with $|\Pi| \leq s(n) + O(\log n)$ that succeeds with probability at least $2/3$.*

Central to both of the above theorems is a combinatorial analysis that establishes that $\mathcal{D}^N$ is supported mostly on **NO** instances and that, furthermore, these instances generate MAX-CUT instances (under the reduction used in Theorem 2.2) whose optimum is bounded away from the total number of edges by a constant fraction. The following definition gives the (simple) reduction which simply outputs the edges of the **DIHP** instance that are labelled 1, and then the lemma establishes the above formally.

DEFINITION 2.1. *Given $\mathcal{I} = (M_1, M_2, M_3; w_1, w_2, w_3)$, the reduction $R(\mathcal{I})$ outputs the stream containing edges of M_1 that are labelled 1 in w_1, followed by the edges of M_2 labelled 1 in w_2, followed by the edges of M_3 labelled 1 in w_3. (Within each M_t, the order of the edges in the stream is arbitrary.)*

LEMMA 2.1. *There exist constants $\Delta^* > 0$ and $0 < \alpha^* < 1$ such that for every $\alpha \in (\alpha^*, 1)$ and even integer $\Delta \geq \Delta^*$, there is a constant $\epsilon^* > 0$ for which the following conditions hold for the reduction R from Definition 2.1:*

(1) *If $\mathcal{I} = (M_1, M_2, M_3; w_1, w_2, w_3)$ is sampled from $\mathcal{D}^Y$ of $\mathbf{DIHP}(n, \Delta, \alpha)$, then $R(\mathcal{I})$ is a bipartite graph.*

(2) *If $\mathcal{I}$ is sampled from $\mathcal{D}^N$, then with probability at least 95/100, $R(\mathcal{I})$ is a graph on m edges with MAX-CUT value at most $(1 - \epsilon^*)m$.*

Our main theorem (Theorem 1.1) follows immediately from Theorem 2.1 and Theorem 2.2. The proof of Lemma 2.1 is omitted in this paper but appears in the full version of this paper. Theorem 2.2 is simple to prove using Lemma 2.1. We devote the rest of this section to providing this proof, as well as a proof of Theorem 1.1. The rest of the paper focuses on proving Theorem 2.1.

Reduction from DIHP to MAX-CUT. We now provide a proof of Theorem 2.2.

Proof. [of Theorem 2.2] Let R be the reduction from Definition 2.1. Let α^* and Δ^* be the constants guaranteed by Lemma 2.1. We fix an $\alpha \in (\alpha^*, 1)$ as well as an even integer $\Delta \geq \Delta^*$. Let $\epsilon^* > 0$ be the constant from Lemma 2.1 for this choice of α and Δ.

It is easy to see that for instances $\mathcal{I}$ sampled from $\mathcal{D}^Y$, the MAX-CUT value of $G = R(\mathcal{I})$ is m, the number of edges of G since G is bipartite. Moreover, by Lemma 2.1, the MAX-CUT value of $R(\mathcal{I})$ for instances $\mathcal{I}$ sampled from $\mathcal{D}^N$ is at most $(1-\epsilon^*)m$ with probability at least 95/100.

Now suppose **ALG** is a one-pass streaming algorithm with space complexity $s(n)$ that produces a $(1 - \epsilon^*)$-approximation to the MAX-CUT value with success probability at least 9/10. Consider the following protocol Π for $\mathbf{DIHP}(n, \Delta, \alpha)$, which makes use of **ALG** as a subroutine: Augment **ALG** with a counter m for the total number of edges presented to it. This takes $O(\log n)$ additional bits of space for simple input graphs on n vertices. Now, for each $t \in \{1, 2, 3\}$, let player t (1.) run (the augmented) **ALG** on the state posted by player $t-1$ with the stream of edges formed by enumerating all edges in M_t for which the corresponding value in w_t is 1 and, (2.) if $t \in \{1, 2\}$, pass on the resulting state of **ALG** to the next player. In other words, the players simulate **ALG** on the stream $R(\mathcal{I})$. The last player then takes the ending state of **ALG** and checks whether the output MAX-CUT value of **ALG** is at least $m/(1 + \epsilon^*)$. If so, the player outputs **YES**; otherwise, the player outputs **NO** .

It is clear that the aforementioned simulation succeeds on $\mathbf{DIHP}(n, \Delta, \alpha)$ with probability at least 2/3. Moreover, the amount of communication $|\Pi|$ in Π is at most the amount of space used for our augmented **ALG**. Thus, $|\Pi| \leq s(n) + O(\log n)$, as desired.

Given Theorem 2.2 and Theorem 2.1, our main theorem follows easily and the proof is included below for completeness.

Proof. [of Theorem 1.1] Let α_1^* and Δ_1^* be the constants guaranteed by Theorem 2.1, and let α_2^* and Δ_2^* be the constants of Theorem 2.2. Let Δ be the smallest even integer larger than $\max\{\Delta_1^*, \Delta_2^*\}$ and choose $\alpha \in (\max\{\alpha_1^*, \alpha_2^*\}, 1)$. Let ϵ^* be the constant given by Theorem 2.2 for this choice of α and Δ.

Now, suppose there exists a randomized single-pass streaming algorithm **ALG** that yields a $(1 + \epsilon^*)$-approximation to MAX-CUT with probability at least 9/10. Let $s(n)$ be the amount of space used by **ALG** on input graphs with n nodes. By Theorem 2.2, there is a protocol Π for $\mathbf{DIHP}(n, \Delta, \alpha)$ with $|\Pi| \leq s(n) + O(\log n)$ such that Π succeeds with probability at least 2/3.

Now, Theorem 2.1 implies that $|\Pi| \geq c'n$ for some constant c'. Hence, $s(n) \geq c'n - O(\log n) \geq cn$ for some constant $c > 0$ and sufficiently large n, which completes the proof.

3 Analysis of communication problem via Fourier techniques

In this section, we first review Fourier analysis on the boolean hypercube, then review relevant communication complexity techniques that were developed in prior work [14], explain why they do not suffice for our result, and give an outline of our approach.

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

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

We will use the relation between multiplication of functions in the time domain and convolution in the frequency domain to analyze the Fourier spectrum of $f_1 \cdot f_2$. With our normalization of the Fourier transform the convolution identity is

$$\widehat{(p \cdot q)}(v) = (\widehat{p} * \widehat{q})(v) = \sum_{x \in \{0,1\}^n} \widehat{p}(x)\widehat{q}(x+v). \tag{3.1}$$

The main object of our analysis will be the Fourier transform of $h_2 = f_1 \cdot f_2$ (these functions are defined later in Definition 3.1). By (3.1), we have $\widehat{h}_2 = \widehat{f}_1 * \widehat{f}_2$. This identity will form the basis of our proof. We will also need Parseval's equality, which, with our normalization, takes the form

$$\begin{aligned} ||\widehat{p}||^2 &= \sum_{v \in \{0,1\}^n} \widehat{p}(v)^2 \\ &= \sum_{v \in \{0,1\}^n} \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} p(x) \cdot (-1)^{x \cdot v} \right)^2 \qquad (3.2) \\ &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} p(x)^2 = \frac{1}{2^n} ||p||^2. \end{aligned}$$

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

3.2 The basic setup We use the notation $X_{i:j}$ to denote $(X_i, X_{i+1}, \ldots, X_j)$. Recall that the messages posted by the players are denoted by $a_t = r_t(M_{1:t}, a_{1:t-1}, w_t)$, where M_t are public $m_t \times n$ edge incidence matrices and w_t are private inputs to players. We use s to denote the maximum of the bit lengths of messages posted by the players. Our goal is to show that if $s \ll n$, then the total variation distance between the distribution of the publicly shared information (messages a_1, a_2, a_3 and graphs M_1, M_2, M_3) in the **YES** and **NO** cases is small. As we show, this task can be simplified as follows. It suffices to consider the **YES** case only and show that if $s \ll n$, then the distribution of $w_t = M_t X^*$ conditional on the publicly posted content up to time t (namely, $a_1, \ldots, a_{t-1}$ and $M_1, \ldots, M_t$) is close to the uniform distribution in total variation distance for $t = 1, 2, 3$ (recall that w_t is actually uniformly distributed in the **NO** case). Our proof of this fact relies on Fourier analytic techniques for reasoning about the distribution of $M_t X^*$ conditioned on typical communication history.

More specifically, our goal is to show that the total variation distance between the distribution of $(M_{1:3}, a_{1:3})$ for the **YES** and **NO** instances is vanishingly small. It suffices to consider the **YES** case only. Fix $t \in \{1,2,3\}$ and let $X^* \in \{0,1\}^n$ denote a uniform random vector conditioned on the graphs $M_{1:t}$ and messages $a^Y_{1:t-1}$. In Lemma 3.2, we show that it suffices to show that with high probability, for each $t = 1, 2, 3$, the distribution of $M_t X^*$ is close to uniform in $\{0,1\}^{m_t}$ and is, hence, indistinguishable from the **NO** case.

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 now define notation that lets us reason about the distribution of X^* at each step t. Since we assume that the protocol is deterministic and the prior distribution of X^* is uniform over $\{0,1\}^n$, the distribution of X^* conditioned on the publicly posted content thus far is uniform over some set $\mathbf{B}_t \subseteq \{0,1\}^n$. We prove the desired claim by analyzing the Fourier spectrum of the indicator function of $\mathbf{B}_t$. It turns out to be convenient to represent $\mathbf{B}_t$ as the intersection of simpler subsets $\mathbf{A}_t$ of the hypercube, where each $\mathbf{A}_t$ essentially conveys the information that the t-th player's message gives about X^*. We give formal definitions below.

DEFINITION 3.1. (SETS $\mathbf{A}_t, \mathbf{B}_t$ AND THEIR INDICATOR FUNCTIONS f_t, h_t) *Fix $\alpha \in (0,1)$ and integers $n \geq 1$ and $t \in \{1,2,3\}$. Consider a* **YES** *instance $(M_{1:3}, w_{1:3})$ of* **DIHP**(n, Δ, α) *with X^* being the (random) hidden partition (so that $w_t = M_t X^*$). Recall that $a_t = r_t(M_{1:t}, a_{1:t-1}, w_t)$.*

We define $\mathbf{A}_{reduced,t} \subseteq \{0,1\}^{m_t}$ as the set of possible values of $w_t = M_t X^*$ that lead to the message a_t, and we define $\mathbf{A}_t$ to be the set of values of $X^* \in \{0,1\}^n$ that correspond to $\mathbf{A}_{reduced,t}$. Formally, letting $g_t(\cdot) := r_t(M_{1:t}, a_{1:t-1}, \cdot) : \{0,1\}^{m_t} \to \{0,1\}^s$, we define

$$\begin{aligned} \mathbf{A}_{reduced,t} &= g_t^{-1}(a_t) \subseteq \{0,1\}^{m_t} \\ \mathbf{A}_t &= \{x \in \{0,1\}^n : M_t x \in \mathbf{A}_{reduced,t}\} \end{aligned}. \tag{3.3}$$

Moreover, for each $t = 1, 2, 3$, let $f_t : \{0,1\}^n \to \{0,1\}$ denote the indicator function of $\mathbf{A}_t$, and let $h_t = f_1 f_2 \cdots f_t$, so that h_t is the indicator of $\mathbf{B}_t := \mathbf{A}_1 \cap \mathbf{A}_2 \cap \ldots \cap \mathbf{A}_t$. We let $\mathbf{B}_0 := \{0,1\}^n$ for convenience.

Our proof of near-uniformity of $M_t X^*$ conditioned on a typical history of communication in $\mathbf{DIHP}(n, \Delta, \alpha)$ is inspired by the work of [14], which used Fourier analysis to give a communication lower bound on the (explicit) hidden partition problem (where Alice is given X^*, Bob gets (M, w), and Bob needs to check whether $w = MX^*$). In our setting, their results translate to showing that if X^* is uniform in $\mathbf{B} \subseteq \{0,1\}^n$, where $|\mathbf{B}|/2^n \geq 2^{-s}$ with $s = O(\sqrt{n})$, and the indicator function h of $\mathbf{B}$ satisifies

$$\left(\frac{2^n}{|\mathbf{B}|}\right)^2 \sum_{v \in \{0,1\}^n, |v|=2\ell} \widehat{h}_t(v)^2 \leq (4\sqrt{2}s/\ell)^{2\ell} \tag{3.4}$$

for all $0 \leq \ell \leq s$, where $|v|$ denotes the Hamming weight of v, then the distribution of MX^* is close to uniform for a random sparse graph M (a random matching in [14]). This translates to a lower bound of $\Omega(\sqrt{n})$ on the communication complexity of the explicit hidden partition problem, but this is too weak for our purposes.

To improve this bound we need to replace the right hand side of the inequality above to a form $(O(s)/\ell)^\ell$ from $(O(s)/\ell)^{2\ell}$. Unfortunately, such an improvement is not possible for the explicit hidden partition problem, which stems from the fact that X^* is known to Alice. In our case, X^* is not known to any player, but we need an analysis that can take advantage of this key fact. We now outline our approach for doing so.

Our first observation is that if the bound in (3.4) could be strengthened by replacing the exponent on the righthand side with ℓ (i.e., reducing the exponent by a factor of 2), an $\Omega(n)$ lower bound would follow. This observation is formalized in Lemma 3.1, which is stated below. Although we do not prove Lemma 3.1 in this paper, the proof appears in the full version of this paper.

LEMMA 3.1. *Let $\Delta > 0$ be an even integer. Then, for every $0 < \alpha < 1$, there exists a constant $0 < c < 1$ such that for every $\delta \in (n^{-1/10}, c)$, the following conditions hold if n is any sufficiently large multiple of Δ:*

(1) *Let $\mathbf{B} = \mathbf{A}_1$, as defined in Definition 3.1. Then, for every choice of matchings M_1, M_2 sampled according to $\mathcal{P}_{n,\Delta,\alpha}$, the distribution of $M_2 x$ is uniform over $\{0,1\}^{m_2}$ when x is uniformly random in $\mathbf{B}$.*

(2) *Let $\mathbf{B} \subseteq \{0,1\}^n, |\mathbf{B}| = 2^{n-z}$ for $z \leq \delta^4 n$, and let $h : \{0,1\}^n \to \{0,1\}$ be the indicator of $\mathbf{B}$. If $\left(\frac{2^n}{|\mathbf{B}|}\right)^2 \sum_{v:|v|=2\ell} \hat{h}(v)^2 \leq \left(\frac{64\delta^4 n}{\ell}\right)^\ell$ holds for all $\ell \leq \delta^4 n$, then the following conditions hold: Let M_1, M_2, M_3 be sampled according to $\mathcal{P}_{n,\Delta,\alpha}$. Then, with probability at least $1 - O(\delta)$ over the choice of M_3, the total variation distance between the distribution of $M_3 x$, where x is uniformly random in $\mathbf{B}$, and the uniform distribution over $\{0,1\}^{m_3}$ is $O(\delta/\sqrt{1-\alpha})$. In particular, one can take $c = \min\left\{\left(\frac{1-\alpha}{512}\right)^{1/4}, \left(\frac{e^{-\alpha}\log_2(32/(31+\alpha))}{32}\right)^{1/4}\right\}$.*

We note that such a strengthening of (3.4) is **impossible** for an indicator function $f : \{0,1\}^n \to \{0,1\}$ of an **arbitrary** subset $\mathbf{B} \subseteq \{0,1\}^n$ with $|\mathbf{B}| = 2^{n-z}$, $z \leq \delta^2 n$—a subcube of appropriate size shows that (3.4) is essentially the best possible bound. Our improvement crucially uses the fact that unlike in the boolean hidden matching problem, in **DIHP**, the players only have **indirect access** to X^* via linear functions $M_t X^*$. In particular, the sets whose indicator functions we analyze are of a special form (see Definition 3.1).

If we could prove that the preconditions of Lemma 3.1 hold w.h.p. for h_2, we would be done by Lemma 3.1. It turns out that one can prove that these preconditions are satisfied for $h_1 = f_1$ rather directly (see Theorem 4.1) using the fact that the compression function g_1 (see Definition 3.1) is applied to the parities of $x_a + x_b, (a, b) \in M_1$. Proving a similar result for the function $h_2 = f_1 \cdot f_2$ is challenging, and this proof is the main technical contribution of our paper. In order to do that, we need to analyze the Fourier transform $\widehat{h_2} = \widehat{f_1} \cdot \widehat{f_2}$, which we do using the convolution identity $\widehat{h_2} = \widehat{f_1} * \widehat{f_2}$. Our main bound on the Fourier transform of $f_1 \cdot f_2$ is stated below.

LEMMA 3.2. *There exists $C > 1$ such that for every even integer $\Delta > 2$, $\gamma > n^{-1/5}$ smaller than an absolute constant, and $\alpha \in (0,1)$, the following conditions hold for sufficiently large n divisible by Δ: Let Π be a protocol for $\mathbf{DIHP}(n, \Delta, \alpha)$ such that $|\Pi| =: s$, where $s = s(n) = \omega(\sqrt{n})$ and $s(n) \leq \frac{1}{2048C\Delta^2}\gamma^5 n$. Then, there exists an event $\mathcal{E}$ that only depends on X^*, M_1, M_2 and occurs with probability at least $1 - O(\gamma)$ over $\mathcal{P}_{n,\Delta,\alpha}$ and the choice of $X^* \in \{0,1\}^n$ such that, conditioned on $\mathcal{E}$, one has*

(1) $|\mathbf{B}_2|/2^n \geq 2^{-\gamma^4 n}$.

(2) $\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{v \in \{0,1\}^n, |v|=2\ell} \widehat{h}_2(v)^2 \leq (C\Delta^2\gamma^4 n/\ell)^\ell$ *for all* $\ell \leq \gamma^4 n$.

Before we present the proof of Theorem 2.1, we require one simple lemma about total variation distance of two probability distributions, which appears with proof in [21].

LEMMA 3.3. (LEMMA 5.6 IN [21]) *Let* (X, Y^1), (X, Y^2) *be random variables taking values on a finite sample space* $\Omega = \Omega_1 \times \Omega_2$. *For any* $x \in \Omega_1$, *let* Y_x^i, $i = 1, 2$ *denote the conditional distribution of* Y^i *given* $X = x$. *Then,*

$$\|(X, Y^1) - (X, Y^2)\|_{tvd} = \mathbf{E}_X \left[\|Y_X^1 - Y_X^2\|_{tvd} \right].$$

Proof. [of Theorem 2.1] Suppose $\Delta > 0$ is an even integer and $0 < \alpha < 1$. Then, we choose $\delta \in (0,1)$ as well as $\gamma \in (0,1)$ such that $\gamma < (64/C\Delta^2)^{1/4}\delta$. Moreover, we pick δ and γ to be sufficiently small such they obey the upper bounds in the hypotheses of Lemmas 3.1 and 3.2. Also, assume n is a sufficiently large multiple of Δ (in particular, $n^{-1/10} < \delta$ and $n^{-1/5} < \gamma$) so that δ and γ obey the lower bounds in the hypotheses of Lemmas 3.1 and 3.2. Moreover, assume γ is sufficiently small so that the event $\mathcal{E}$ in Lemma 3.2 occurs with probability greater than $1/2$.

We now assume that Π is a protocol for $\mathbf{DIHP}(n, \Delta, \alpha)$ that uses less than $\frac{1}{2048C\Delta^2}\gamma^5 n$ bits of communication, where $C > 0$ is the constant in Lemma 3.2.

Recall that the first player posts the message $a_1 = r_1(M_1, w_1)$. We now consider the distribution of (M_1, M_2, a_1, w_2). Let D_1^Y and D_1^N be the distributions of (M_1, M_2, a_1, w_2) on **YES** and **NO** instances, respectively. Thus, $D_1^Y = (M_1, M_2, a, p_{M_1,M_2,a})$, where $p_{M_1,M_2,a}$ is the distribution of $M_2 x$ conditional on $r_1(M_1, x) = a$. For any M_1, M_2, a, we let $D^Y_{(M_1,M_2,a)} = p_{M_1,M_2,a}$ and $D^N_{(M_1,M_2,a)} = U_{M_2}$ denote the distribution of w_2 given the message a and edge incidence matrices M_1, M_2 for the **YES** and **NO** instances, respectively. (Here, U_r denotes the uniform distribution on $\{0,1\}^r$.) Moreover, note that the distribution of (M_1, M_2, a_1) is identical in both the **YES** and **NO** cases. Thus, by Lemma 3.3 and part (1) of Lemma 3.1, we have

$$\begin{aligned}\|D_1^Y - D_1^N\|_{tvd} &= \mathbf{E}_{M_1,M_2,a}\left[\|D^Y_{(M_1,M_2,a)} - D^N_{(M_1,M_2,a)}\|\right] \\ &= 0.\end{aligned}$$

Moreoever, since $a_2 = r_2(M_1, M_2, a_1, w_2)$, another simple application of Lemma 3.3 implies that

$$\left\|D_2^Y - D_2^N\right\|_{tvd} = 0,$$

where D_2^Y and D_2^N denote the distributions of (M_1, M_2, a_1, a_2) for the **YES** and **NO** instances, respectively.

Now, let $\mathcal{E}$ be the event for the **YES** case that is guaranteed by Lemma 3.2. Recall that $\mathcal{E}$ occurs with probability $1 - O(\gamma)$ over $\mathcal{P}_{n,\Delta,\alpha}$ and the random choice of $X^* \in \{0,1\}^n$. Moreover, Lemma 3.2 implies that for any **YES** instance conditioned on $\mathcal{E}$, we have that for all $\ell \le \gamma^4 n$,

$$\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v \in \{0,1\}^n \\ |v| = 2\ell}} \widehat{h}_2(v)^2 \le \left(\frac{C\Delta^2\gamma^4 n}{\ell}\right)^\ell \le \left(\frac{64\delta^4 n}{\ell}\right)^\ell,$$

where h_2, $\mathbf{B}_2$ are defined as in Definition 3.1. Thus, letting $q := q_{M_1,M_2,M_3,a_1,a_2}$ denote the distribution on $M_3 x$ conditioned on $r_1(M_1, x) = a_1$ and $r_2(M_1, M_2, a_1) = a_2$, we see that part (2) of Lemma 3.1 implies that, given the occurrence of $\mathcal{E}$,

$$\|q - U_{M_3}\|_{tvd} = O(\delta/\sqrt{1-\alpha}). \tag{3.5}$$

with probability $p \ge 1 - \frac{O(\delta)}{\mathbf{Pr}[\mathcal{E}]} \ge 1 - O(\delta)$ over the choice of M_3 (since γ was chosen small enough to guarantee that $\mathbf{Pr}[\mathcal{E}] \ge 1/2$). Therefore, since $\mathcal{E}$ only depends on X^*, M_1, M_2, Lemma 3.3 and (3.5) imply that

$$\begin{aligned}&\|D_3^Y - D_3^N\|_{tvd} \\ &\quad= \mathbf{Pr}[\mathcal{E}] \cdot \mathbf{E}_{Z|\mathcal{E}}\left[\mathbf{E}_{M_3}\left[\|q - U_{M_3}\|_{tvd}\right]\right] + \mathbf{Pr}[\overline{\mathcal{E}}] \cdot 1 \\ &\quad\le \mathbf{Pr}[\mathcal{E}] \cdot \mathbf{E}_{Z|\mathcal{E}}\left[p \cdot O(\delta/\sqrt{1-\alpha}) + (1-p)\cdot 1\right] \\ &\qquad + \mathbf{Pr}[\overline{\mathcal{E}}] \\ &\quad\le \mathbf{Pr}[\mathcal{E}](1 - p(1 - O(\delta/\sqrt{1-\alpha}))) + (1 - \mathbf{Pr}[\mathcal{E}]) \\ &\quad\le 1 - \mathbf{Pr}[\mathcal{E}] \cdot p(1 - O(\delta/\sqrt{1-\alpha})) \\ &\quad\le 1 - (1 - O(\gamma))(1 - O(\delta))(1 - O(\delta/\sqrt{1-\alpha})) \\ &\quad= O(\gamma) + O(\delta) + O(\delta/\sqrt{1-\alpha}),\end{aligned}$$

where Z denotes (M_1, M_2, a_1, a_2), while D_3^Y and D_3^Y denote the distribution of $(M_1, M_2, M_3, a_1, a_2, w_3)$ in the **YES** and **NO** instances, respectively. We choose δ, γ to be small enough so that the above total variation distance is less than $1/3$.

Finally, observe that since $a_3 = r_3(M_1, M_2, M_3, a_1, a_2, w_3)$, the total variation distance of the distributions of $(M_1, M_2, M_3, a_1, a_2, a_3)$ in the **YES** and **NO** cases is also less than $1/3$, which means that Π cannot distinguish the **YES** and **NO** cases with advantage more than $1/6$ over random guessing, i.e., the success probability of Π is less than $2/3$.

Hence, it follows that any algorithm Π for **DIHP** that succeeds with probability at least $2/3$ must use at least cn bits of communication, for $c = \frac{1}{2048C\Delta^2}\gamma^5$. This completes the proof of the claim.

4 Proof of main lemma (Lemma 3.2)

The main result of this section is a proof of Lemma 3.2. The main idea behind the proof is to use the convolution

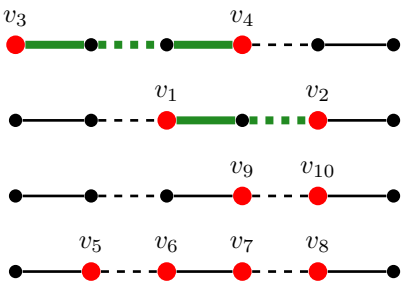

Figure 1: Illustration of $P^*(v)$, where $v = \{v_1, \ldots, v_{10}\}$ (marked red). Edges of M_1 are shown as solid lines, edges of M_2 as dashed lines. The set of paths $P(v)$ is the set of edges between the marked nodes. The paths $v_5 - v_6$, $v_7 - v_8$ and $v_9 - v_{10}$ consist only of an edge of M_2, and hence are not grounded. Grounded paths $P^*(v)$ are marked green (paths $v_1 - v_2$ and $v_3 - v_4$).

identity to express the Fourier transform of h_2 in terms of the Fourier transform of f_1 and f_2. Specifically, for every $v \in \{0,1\}^n$, we have, by the convolution identity,

$$\widehat{h_2}(v) = \widehat{f_1 \cdot f_2}(v) = \sum_{w \in \{0,1\}^n} \widehat{f_1}(w) \cdot \widehat{f_2}(w+v). \tag{4.6}$$

Besides the convolution identity, we use the structure of the Fourier transform of f_1 and f_2. Specifically, we use the fact that $\widehat{f_1}$ and $\widehat{f_2}$ are supported on edges of M_1 and M_2, respectively (equivalently, they are zero except on the column span of $(M_1; M_2)$). This allows us to classify the terms $\widehat{f_1}(w) \cdot \widehat{f_2}(w+v)$ on the rhs of (4.6) according to the weight of w and $w+v$. We would like to show that only very few large weight coefficients $\widehat{f_1}(w)$ can contribute to $\widehat{h_2}(v)$ for a low weight v. Note that this is intuitively necessary for the proof, as according to our bounds the ℓ_2^2 mass of coefficients of $\widehat{f_1}$ or $\widehat{f_2}$ grows with the weight level. We prove that a high weight coefficient is unlikely to appear on the rhs of (4.6) if the coefficient on the lhs is low weight in section 4.1 (see Lemma 4.1). Then in section 4.2, we show how these bounds imply that not too much ℓ_2^2 mass of $\widehat{f_1}$ can be trasferred from high weight levels to low weight levels (see Lemma 4.4). Finally, in section 4.3, we put the developed results together into a proof of Lemma 3.2.

4.1 Useful definitions and basic claims The following definitions form the basis of our analysis.

Definition 4.1. *Given matchings M_1, M_2 such that $M_1 \cup M_2$ is a union of paths, a vector $v \in \{0,1\}^n$ is called* admissible *with respect to M_1, M_2 if v has an even number of nonzeros on every path in $M_1 \cup M_2$.*

Definition 4.2. (Path decomposition of admissible coefficients) *Given M_1, M_2 such that $M_1 \cup M_2$ is a union of paths, for any $v \in \{0,1\}^n$ admissible wrt M_1, M_2, let $P(v)$ denote the unique set of vertex disjoint paths in $M_1 \cup M_2$ whose endpoints are exactly the nonzeros of v.*

Claim 4.1. *The path decomposition is well defined for any admissible $v \in \{0,1\}^n$.*

Proof. It suffices to show that for any admissible v the set of paths $P(v)$ exists and is unique. Existence follows immediately from definition of admissibility. Uniqueness follows since $M_1 \cup M_2$ is a collection of simple vertex disjoint paths.

Definition 4.3. *Given M_1, M_2 such that $M_1 \cup M_2$ is a union of paths, for any $v \in \{0,1\}^n$ admissible wrt M_1, M_2, let $P^*(v) \subseteq P(v)$ denote the set of paths in $P(v)$ that contain at least one edge of M_1. We refer to $P^*(v)$ as the* core *of the path decomposition of v.*

Note that paths in $P(v) \setminus P^*(v)$ are all of length one, i.e. edges of M_2. See Fig. 1 for an illustration.

We will often associate matchings M with the sets of vertices that they match. For example, for $w \in \{0,1\}^n$, we will write $w \subseteq M_1$ to denote the fact that w is a subset of the vertices matched by M_1. We will say that w is supported on edges of M_1 if for every $e = \{u, v\} \in M_1$, one has either $w \cap \{u, v\} = \emptyset$ or $w \cap \{u, v\} = \{u, v\}$. The following claim is crucial to our subsequent analysis:

Lemma 4.1. *For every even integer $\Delta > 2$ and $\alpha \in (0,1)$, if matchings M_1, M_2 are sampled from $\mathcal{P}_{n,\Delta,\alpha}$, then the following conditions hold for every $\ell, k \geq 0$. Conditioned on M_1, for every subset $w \subseteq M_1$ such that $|w| = 2k$, we have the following:*

(1) $\mathbf{Pr}_{M_2}[\exists M' \subseteq M_2 \text{ s.t. } |P^*(w + M')| = \ell \mid M_1] \leq (O(\Delta))^\ell \binom{n/2}{\ell} \binom{n/2}{k}^{-1}$.

(2) *For every $M' \subseteq M_2$, one has $|P^*(w + M')| \geq |w|/\Delta$.*

Proof. The second claim follows by recalling that our input distribution on matchings is such that $M_1 \cup M_2$ does not contain cycles, and the largest path length in the graph induced by $M_1 \cup M_2$ is not larger than Δ.

We now prove the first claim. We first upper bound the number of $w \subseteq M_1$ such that $|P^*(w + M')| = \ell$ for some $M' \subseteq M_2$, i.e. the core of $w + M'$ contains ℓ paths. We then show that since the distribution of M_2 is invariant under permutation of edges of M_1, this gives the result.

We now upper bound the number of sets of ℓ paths that each contain at least one edge of M_1, given M_1 and M_2 (we refer to such paths as *grounded*). Given M_1, M_2, in order to select a grounded set of paths, it suffices to first select ℓ edges from M_1, one per path (at most $\binom{n/2}{\ell}$ choices). Then order these edges arbitrarily, and for each $t = 1, \ldots, \ell$,

- choose whether the path starts with an edge of M_1 or an adjacent edge of M_2 (three choices);
- choose a direction to go on the corresponding path in $M_1 \cup M_2$ (at most 2 choices);
- choose a number of steps to go for (at most 2Δ choices).

Putting the bounds above together, we get that for any M_1, M_2, the number of grounded sets of k paths is bounded by $(12\Delta)^\ell \binom{n/2}{\ell}$.

Next, we recall that the matchings M_1, M_2 are generated as follows (our description here is somewhat more detailed than in Section 2, and results in exactly the same distribution; this formulation is more convenient for our analysis):

- Let M_1 be a perfect matching that matches, for each $i = 1, \ldots, n/2$, vertex i to vertex $i + n/2$. Note that edges of M_1 are naturally indexed by $[n/2]$: the i-th edge matches i to $i + n/2$, for $i \in [n/2] = \{1, 2, \ldots, n/2\}$.
- Choose a permutation π of $[n/2] = \{1, 2, \ldots, n/2\}$ uniformly at random. Partition edges of M_1 into $r = n/\Delta$ sets $S_1, \ldots, S_r$ with $\Delta/2$ edges each, where Δ is an even integer that divides n, by letting

$$S_j = \left\{ \pi\left(\frac{\Delta}{2} \cdot (j-1) + 1\right), \pi\left(\frac{\Delta}{2} \cdot (j-1) + 2\right), \ldots, \pi\left(\frac{\Delta}{2} \cdot (j-1) + \frac{\Delta}{2}\right) \right\}$$

  for each $j = 1, \ldots, n/\Delta$.
- For each $j = 1, \ldots, n/\Delta$, let $M_{2,j}$ match, for each $i = 1, \ldots, \Delta/2 - 1$, the node $\pi\left(\frac{\Delta}{2} \cdot (j-1) + i\right)$ to the node $\pi\left(\frac{\Delta}{2} \cdot (j-1) + i + 1\right) + n/2$. Note that $|M_{2,j}| = \frac{\Delta}{2} - 1$ for each j.

Let $M_1 := \bigcup_{j=1}^r M_{1,j}$ and $M_2 := \bigcup_{j=1}^r M_{2,j}$.

By the derivation above, we have that for any permutation π, the number of grounded sets of k paths in the union $M_1 \cup M_2$ generated by our process is bounded by $(12\Delta)^\ell \binom{n/2}{\ell}$. Denote this set by $\mathcal{P}^\ell(\pi)$ and note that for every P there exists a unique $w \subseteq M_1$ such that $|P^*(w + M')| = \ell$ for some $M' \subseteq M_2$. Specifically, $w = P \cap M_1$ satisfies these constraints. Let $\mathcal{S}(\pi) := \{P \cap M_1 : P \in \mathcal{P}(\pi)\}$. Thus, we have $|\mathcal{S}(\pi)| = (12\Delta)^\ell \binom{n/2}{\ell}$. We now note that $S(\mathrm{id})$ is hence a fixed set of at most $(12\Delta)^\ell \binom{n/2}{\ell}$ subsets of edges. At the same time for every permutation π of $[n/2]$ one has

$$\mathcal{S}(\pi) = \pi^{-1}(\mathcal{S}(\mathrm{id})). \tag{4.7}$$

Since π is uniformly random, we thus get for every $w \in \{0,1\}^n$ with $|w| = 2k$,

$$\begin{aligned} \mathbf{Pr}_\pi[w \in \mathcal{S}(\pi)] &= \mathbf{Pr}_\pi[w \in \pi^{-1}(\mathcal{S}(\mathrm{id}))] \\ &= \mathbf{Pr}_\pi[\pi(w) \in \mathcal{S}(\mathrm{id})] \\ &= |\mathcal{S}(\mathrm{id})| / \binom{n/2}{k} \\ &= (12\Delta)^\ell \binom{n/2}{\ell} \binom{n/2}{k}^{-1}, \end{aligned} \tag{4.8}$$

where we used the fact that $\pi(w)$ is uniformly random in the set of unordered k-tuples of edges of M_1 when π is uniformly random. This completes the proof.

4.2 Bounds on expected transfer of Fourier mass In this section, we use the convolution identity (4.6) to bound the contribution of Fourier transforms $\widehat{f_1}$ and $\widehat{f_2}$ to the Fourier transform $\widehat{h_2}$ of $h_2 = f_1 \cdot f_2$ (see Definition 3.1). The main result of this section is Lemma 4.4. The more basic bounds on the Fourier transform of f_1 and f_2 are provided by Theorem 4.1, whose proof appears in the full version of this paper. Part **(1)** of the theorem shows that $\widehat{f_1}$ and $\widehat{f_2}$ are supported on edges of matchings M_1 and M_2 respectively, while parts **(2)** and **(3)** use this fact to derive upper bounds of the form $(O(s)/\ell)^\ell$ (i.e., with the improved exponent of ℓ as opposed to 2ℓ that we are looking for) for the amount of mass on weight level ℓ in $\widehat{f_1}$ and $\widehat{f_2}$, respectively.

THEOREM 4.1. *Let $M \in \{0,1\}^{m \times n}$ 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 $g : \{0,1\}^m \to \{0,1\}^s$ for some $s > 0$. Let $a \in \{0,1\}^s$ and let $\mathbf{A}_{reduced} := \{z \in \{0,1\}^m : g(z) = a\}$. Furthermore, let $f : \{0,1\}^n \to \{0,1\}$ denote the indicator of the set*

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

Suppose that $|\mathbf{A}| = 2^{n-d}$ for some $d \in [0, n]$. Then,

(1) *The only nonzero Fourier coefficients of $\widehat{f}$ are of the form $\widehat{f}(M^T w)$ for some $w \in \{0,1\}^m$.*

(2) *For all $\ell \in [0,d]$ and every $Q \subseteq M$,*

$$2^{2d} \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell+|Q| \\ v\supseteq Q}} \widehat{f}(v)^2 \le 2^{|Q|}(64d/\ell)^\ell,$$

where $|Q|$ denotes the number of vertices in Q.

(3) $2^{2d}\sum_{v\in\{0,1\}^n} \widehat{f}(v)^2 = 2^d$ *(Parseval's equality).*

The proof of Theorem 4.1 is omitted but appears in the full version of this paper.

LEMMA 4.2. *For any $v \in \{0,1\}^n$, one has $(\widehat{f_1 \cdot f_2})(v) = 0$ if v is not admissible with respect to M_1, M_2, and $(\widehat{f_1 \cdot f_2})(v) = \widehat{f}_1(P(v) \cap M_1) \cdot \widehat{f}_2(P(v) \cap M_2)$ otherwise.*

Proof. By the convolution identity (4.6) we have $(\widehat{f_1 \cdot f_2})(v) = \sum_{x\in\{0,1\}^n} \widehat{f}_1(x)\widehat{f}_2(v+x)$. By Theorem 4.1, **(1)** applied to the sets $\mathbf{A}_i$, messages a_i, functions g_i, $i \in \{1,2\}$ (as per Definition 3.1) we also have that $\widehat{f}_1(x) \neq 0$ only if x is a union of edges of M_1, and $\widehat{f}_2(v+x) \neq 0$ if $v+x$ is a union of edges of M_2. We can thus write $(\widehat{f_1 \cdot f_2})(v) = \sum_{\substack{M_1'\subseteq M_1, M_2'\subseteq M_2 \\ M_1'+M_2'=v}} \widehat{f}_1(M_1')\widehat{f}_2(M_2')$. Since M_1 and M_2 are edge disjoint and $M_1 \cup M_2$ is a union of paths, we have that for every admissible $v \in \{0,1\}^n$, there exists a unique pair $M_1' \subseteq M_1, M_2' \subseteq M_2$ such that $v = M_1' + M_2'$.

LEMMA 4.3. *For any $w \subseteq M_1$ with $|w| = 2k$, the number of $v \in \{0,1\}^n$ with $|v| = 2\ell$ and $|P^*(v)| = \ell$ such that $v = w + M_2'$ for some $M_2' \subseteq M_2$ is upper bounded by 2^{2k}.*

Proof. For each path $M_1 \cup M_2$, designate one endpoint to be the left endpoint and the other to be the right endpoint arbitrarily. Note that for each path, this fixes an ordering of vertices (left to right). We associate two binary variables with each of the two endpoints of each edge $e \in w$. Denote these binary variables by $L(e)$ and $R(e)$. Then for each $v \in \{0,1\}^n$ and every $e \in w$, we let $L(e) = 1$ if $P(v)$ extends beyond the left endpoint of e, and 0 otherwise. Similarly, $R(e) = 1$ if $P(v)$ extends beyond the right endpoint of e, and 0 otherwise. Note that the collection of variables $\{(L(e), R(e))\}_{e\in w}$ uniquely determines $P(v)$. On the other hand, the number of possible assignments of $L(e), R(e)$ for $e \in w$ is upper bounded by 2^{2k}, proving the lemma.

We now state and prove Lemma 4.4. For an event $\mathcal{E}$, we let $\mathbf{I}[\mathcal{E}]$ denote the indicator function of $\mathcal{E}$.

LEMMA 4.4. *For every even integer $\Delta > 2$, every $\alpha \in (0,1)$, every $s \le n/256$, and any protocol Π for $\mathbf{DIHP}(n,\Delta,\alpha)$, the following conditions hold for sufficiently large n. If $f_1, f_2 : \{0,1\}^n \to \{0,1\}$ are indicator functions of $\mathbf{A}_1$ and $\mathbf{A}_2$, respectively, then for every $0 \le \ell \le s$, $0 \le k \le n/2$, and $w \in \{0,1\}^n$ with $|w| = 2k$, the following conditions hold for every $M_1, \mathbf{A}_1$.*

(1) *If $k \le \ell$, then*

$$\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}' \cdot \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \le 4^\ell (O(\Delta))^k \left(\frac{64s}{\ell-k}\right)^{\ell-k},$$

where $\mathbf{I}' = \mathbf{I}\left[\frac{|\mathbf{A}_2|}{2^n} \ge 2^{-s}\right]$.

(2) *If $k \ge \ell$, then*

$$\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}' \cdot \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \le (O(\Delta))^\ell 8^k \left(\frac{k-\ell}{n/2}\right)^{k-\ell},$$

where $\mathbf{I}' = \mathbf{I}\left[\frac{|\mathbf{A}_2|}{2^n} \ge 2^{-s}\right]$.

Proof. We classify elements v according to the size of the core $P^*(v)$. Let $w = M_1' \subseteq M_1$. For any $v \in \{0,1\}^n, |v| = 2\ell$ admissible wrt M_1, M_2, note that $|P(v)| = \ell$, as every path in $P(v)$ contributes 2 to the weight of v via its two endpoints. Note that $P^*(v) \subseteq P(v)$, so $|P^*(v)|$ is between 0 and ℓ:

$$\sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f}_2(w+v)^2 = \sum_{r=0}^{\ell} \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell \\ |P^*(v)|=r \\ P^*(v)\cap M_1 = w}} \widehat{f}_2(w+v)^2.$$

Since paths in $P(v) \setminus P^*(v)$ are all of length 1 and correspond to edges of M_2, any admissible v can be represented uniquely as $v = v' + x$, where $P^*(v) = P^*(v') = P(v')$ and $x \subseteq M_2$ is supported on edges of M_2 and is disjoint from $P^*(v)$ (see Fig. 1 for an illustration of $P^*(v)$). Substituting this into the rhs of the equation

above, we get

$$\begin{aligned}
&\sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2 \\
&\quad= \sum_{r=0}^{\ell} \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell\\|P^*(v)|=r}} \widehat{f}_2(w+v)^2 \\
&\quad= \sum_{r=0}^{\ell} \sum_{\substack{v'\in\{0,1\}^n\\|v'|=2r\\|P^*(v')|=r\\P^*(v')\cap M_1=w}} \sum_{\substack{x\subseteq M_2\\x\cap P^*(v')=\emptyset\\|x|=\ell-r}} \widehat{f}_2(w+v'+x)^2 \\
&\quad=: Y_1.
\end{aligned}$$

Let $\mathbf{I}' = \mathbf{I}\left[\frac{|\mathbf{A}_2|}{2^n} \geq 2^{-s}\right]$. By Theorem 4.1, **(2)** invoked with $\mathbf{A} = \mathbf{A}_2, f = f_2$, $g = g_2$, $M = M_2$, $Q = P^*(v') \cap M_2$, $k = \ell - r$, and $d = \log_2\left(\frac{2^n}{|\mathbf{A}_2|}\right)$, we get

$$\begin{aligned}
&\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}' \sum_{\substack{x\subseteq M_2\\x\cap P^*(v')=\emptyset\\|x|=\ell-r}} \widehat{f}_2(w+v'+x)^2 \\
&\qquad\leq 2^{|Q|}(64s/(\ell-r))^{\ell-r} \\
&\qquad\leq 2^{2k}(64s/(\ell-r))^{\ell-r},
\end{aligned} \tag{4.9}$$

where we have used the fact that $|Q| \leq 2\ell$ (the set $P^*(v')$ is a disjoint union of edges of M_1 and M_2 that form paths; the number of edges of M_2 on each such path is no more than a factor of 2 larger than the number of edges of M_1). Putting the bounds above together, and taking an expectation over M_2 conditional on M_1 and $\mathbf{A}_1$, we get that $\mathbf{E}_{M_2}[Y_1]$ is bounded from above by

(4.10)

$$\begin{aligned}
&\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}' \cdot \sum_{r=0}^{\min\{k,\ell\}} \sum_{\substack{v'\in\{0,1\}^n\\|v'|=2r\\|P^*(v')|=r\\P^*(v')\cap M_1=w}} \cdot \sum_{\substack{x\subseteq M_2\\x\cap P^*(v')=\emptyset\\|x|=\ell-r}} \widehat{f}_2(w+v'+x)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \\
&= \mathbf{E}_{M_2}\left[\sum_{r=0}^{\min\{k,\ell\}} \sum_{\substack{v'\in\{0,1\}^n\\|v'|=2r\\|P^*(v')|=r\\P^*(v')\cap M_1=w}} \left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}' \right. \\
&\qquad\left. \cdot \sum_{\substack{x\subseteq M_2\\x\cap P^*(v')=\emptyset\\|x|=\ell-r}} \widehat{f}_2(w+v'+x)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \\
&\leq \mathbf{E}_{M_2}\left[2^{2k} \cdot \sum_{r=0}^{\min\{k,\ell\}} \left(\frac{64s}{\ell-r}\right)^{\ell-r} \right. \\
&\qquad\left. \sum_{\substack{v'\in\{0,1\}^n\\|v'|=2r\\|P^*(v')|=r}} \mathbf{I}[P^*(v')\cap M_1 = w] \,\middle|\, M_1, \mathbf{A}_1\right] \\
&\leq \mathbf{E}_{M_2}\left[2^{2k}\cdot 2^{2k} \cdot \sum_{r=0}^{\min\{k,\ell\}} \left(\frac{64s}{\ell-r}\right)^{\ell-r} \right. \\
&\qquad\left. \cdot \mathbf{I}[\exists M_2' \subseteq M_2 : |P^*(w+M_2')| = r] \,\middle|\, M_1, \mathbf{A}_1\right] \\
&= 2^{4k} \sum_{r=0}^{\min\{k,\ell\}} \left(\frac{64s}{\ell-r}\right)^{\ell-r} \\
&\qquad \cdot \mathbf{Pr}_{M_2}\left[\exists M_2' \subseteq M_2 : |P^*(w+M_2')| = r \,|\, M_1, \mathbf{A}_1\right] \\
&=: Y_2,
\end{aligned}$$

where $\mathbf{I}[\mathcal{E}]$ stands for the indicator of event $\mathcal{E}$. We have used (4.9) to go from the second line to the third, as well as Lemma 4.3 to conclude that $\sum_{\substack{v' \in \{0,1\}^n \\ |v'|=2r \\ |P^*(v')|=r}} \mathbf{I}[P^*(v') \cap M_1 = w] \leq 2^{2k}$ and obtain the fourth line. Note that the summation above is over r between 0 and $\min\{k, \ell\}$. To see that the size of the core $P^*(w + M_2') = P^*(v)$ cannot be larger than 2ℓ, note that each path in the core contributes two distinct endpoints to the weight of v. To see that the size of the core $P^*(w + M_2') = P^*(v)$ cannot be larger than k, note that every path in the core must contain at least one edge in M_2 that belongs to w, and these edges are disjoint.

By Lemma 4.1, we have that

$$\mathbf{Pr}\left[\exists M_2' \subseteq M_2 : |P^*(w + M_2')| = r \mid M_1\right] \leq (O(\Delta))^r \binom{n/2}{r} \binom{n/2}{|w|/2}^{-1}.$$

Substituting this bound into the equation above, we get

$$\begin{aligned} Y_2 &\leq 16^k \cdot \sum_{r=0}^{\min\{k,\ell\}} (64s/(\ell - r))^{\ell - r} \\ &\quad \cdot \mathbf{Pr}\left[\exists M_2' \subseteq M_2 : |P^*(w + M_2')| = r \mid M_1\right] \\ &\leq 16^k \cdot \sum_{r=0}^{\min\{k,\ell\}} (64s/(\ell - r))^{\ell - r} \\ &\quad \cdot (O(\Delta))^r \binom{n/2}{r} \binom{n/2}{k}^{-1} \\ &=: Y_3. \end{aligned} \tag{4.11}$$

We now consider two cases, depending on whether $k \leq \ell$ or $k \geq \ell$ (the cases overlap, giving us two rather similar bounds for $k = \ell$).

Case 1: $k \leq \ell$**.** Using the bound $(n/k)^k \leq \binom{n}{k} \leq (en/k)^k$ in (4.11), we obtain

$$\begin{aligned} Y_3 &= 16^k \sum_{r=0}^{\min\{k,\ell\}} \left(\frac{64s}{\ell - r}\right)^{\ell - r} (O(\Delta))^r \binom{n/2}{r} \binom{n/2}{k}^{-1} \\ &= 16^k \sum_{r=0}^{k} \left(\frac{64s}{\ell - r}\right)^{\ell - r} (O(\Delta))^r \binom{n/2}{r} \binom{n/2}{k}^{-1} \\ &\leq 16^k \left(\frac{64s}{\ell - k}\right)^{\ell - k} \sum_{r=0}^{k} (64s)^{k-r} \left[\frac{(\ell - k)^{\ell - k}}{(\ell - r)^{\ell - r}}\right] \\ &\quad \cdot (O(\Delta))^r \left(\frac{en}{2r}\right)^r \left(\frac{(n/2)}{k}\right)^{-k} \\ &\leq 4^\ell \cdot (O(\Delta))^k (64s/(\ell - k))^{\ell - k} \sum_{r=0}^{k} (128s/n)^{k-r} \\ &\quad \cdot \frac{(\ell - k)^{\ell - k}(k - r)^{k-r}}{(\ell - r)^{\ell - r}} \cdot \frac{k^k}{r^r (k - r)^{k-r}} \\ &=: Y_4. \end{aligned} \tag{4.12}$$

We now note that $\frac{a^a b^b}{(a+b)^{a+b}} = \exp(a \ln a + b \ln b - (a + b) \ln(a + b)) \leq 1$ for all $a \geq 0, b \geq 0$, by convexity of the function $x \ln x$. Furthermore, for fixed $a + b$, the maximum of $\frac{(a+b)^{a+b}}{a^a b^b}$ is achieved when $a = b$ and equals 2^{a+b}. Applying the first bound with $a = \ell - k, b = k - r$ gives

$$\frac{(\ell - k)^{\ell - k}(k - r)^{k-r}}{(\ell - r)^{\ell - r}} \leq 1, \tag{4.13}$$

and applying the second bound with $a = r, b = k - r$ gives

$$\frac{k^k}{r^r (k - r)^{k-r}} \leq 2^k. \tag{4.14}$$

Substituting these bounds into (4.12) yields

$$
\begin{aligned}
Y_4 &= 4^\ell \cdot (O(\Delta))^k \left(\frac{64s}{\ell-k}\right)^{\ell-k} \\
&\quad \cdot \sum_{r=0}^{k} \left(\frac{128s}{n}\right)^{k-r} \cdot \frac{(\ell-k)^{\ell-k}(k-r)^{k-r}}{(\ell-r)^{\ell-r}} \\
&\quad \cdot \frac{k^k}{r^r(k-r)^{k-r}} \\
&\le 4^\ell \cdot (O(\Delta))^k (64s/(\ell-k))^{\ell-k} \\
&\quad \cdot \sum_{r=0}^{k} (128s/n)^{k-r} \cdot \frac{k^k}{r^r(k-r)^{k-r}} \\
&\le 4^\ell \cdot (O(\Delta))^k \left(\frac{64s}{\ell-k}\right)^{\ell-k} \sum_{r=0}^{k} \left(\frac{128s}{n}\right)^{k-r} \\
&\le 4^\ell (O(\Delta))^k \left(\frac{64s}{\ell-k}\right)^{\ell-k},
\end{aligned}
$$

where we have used (4.13) and (4.14). Substituting this into (4.11) and then in (4.10), we get the result for the case $k \le \ell$ (Case 1).

Case 2: $k \ge \ell$**.** Using the bound $(n/k)^k \le \binom{n}{k} \le (en/k)^k$ in (4.11), we obtain

$$
\begin{aligned}
Y_3 &= 16^k \sum_{r=0}^{\min\{k,\ell\}} \left(\frac{64s}{\ell-r}\right)^{\ell-r} (O(\Delta))^r \binom{n/2}{r}\binom{n/2}{k}^{-1} \\
&= 16^k (O(\Delta))^\ell \sum_{r=0}^{\ell} \left(\frac{64s}{\ell-r}\right)^{\ell-r} \cdot \binom{n/2}{r}\binom{n/2}{k}^{-1} \\
&\le 16^k (O(\Delta))^\ell \sum_{r=0}^{\ell} \left(\frac{64s}{\ell-r}\right)^{\ell-r} (n/2)^{r-k} k^k / r^r \\
&\le 16^k (O(\Delta))^\ell \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\quad \cdot \sum_{r=0}^{\ell} \left(\frac{64s}{\ell-r}\right)^{\ell-r} (n/2)^{r-\ell} \frac{k^k}{r^r(k-\ell)^{k-\ell}} \\
&\le 16^k (O(\Delta))^\ell \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\quad \cdot \sum_{r=0}^{\ell} \left(\frac{128s}{n}\right)^{\ell-r} \frac{k^k}{r^r(k-\ell)^{k-\ell}(\ell-r)^{\ell-r}} \\
&\le 16^k (O(\Delta))^\ell \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\quad \cdot \sum_{r=0}^{\ell} \left(\frac{128s}{n}\right)^{\ell-r} \frac{k^k \ell^\ell}{r^r(\ell-r)^{\ell-r}(k-\ell)^{k-\ell}\ell^\ell} \\
&\le 16^k (O(\Delta))^\ell \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\quad \cdot \sum_{r=0}^{\ell} \left(\frac{128s}{n}\right)^{\ell-r} \frac{\ell^\ell}{r^r(\ell-r)^{\ell-r}} \frac{k^k}{(k-\ell)^{k-\ell}\ell^\ell} \\
&=: Y_5.
\end{aligned}
$$

Again, by convexity arguments as in Case 1, we have $\frac{\ell^\ell}{r^r(\ell-r)^{\ell-r}} \frac{k^k}{(k-\ell)^{k-\ell}\ell^\ell} \le 2^{\ell+k}$. Substituting this in the derivation above, we get

$$
\begin{aligned}
Y_5 &= 16^k (O(\Delta))^\ell \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\quad \cdot \sum_{r=0}^{\ell} (128s/n)^{\ell-r} \frac{\ell^\ell}{r^r(\ell-r)^{\ell-r}} \frac{k^k}{(k-\ell)^{k-\ell}\ell^\ell} \\
&\le 16^k (O(\Delta))^\ell 2^{\ell+k} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \sum_{r=0}^{\ell} (128s/n)^{\ell-r} \\
&\le 32^k (O(\Delta))^\ell \left(\frac{k-\ell}{n/2}\right)^{k-\ell},
\end{aligned}
$$

since $s < n/256$, by assumption of the lemma.

4.3 Putting it together

We now present a proof of Lemma 3.2, which we restate here for convenience of the reader:

Lemma 3.2 *There exists $C > 1$ such that for every even integer $\Delta > 2$, $\gamma > n^{-1/5}$ smaller than an absolute constant, and $\alpha \in (0,1)$, the following conditions hold for sufficiently large n divisible by Δ: Let Π be a protocol for $\mathbf{DIHP}(n,\Delta,\alpha)$ such that $|\Pi| =: s$, where $s = s(n) = \omega(\sqrt{n})$ and $s(n) \le \frac{1}{2048C\Delta^2}\gamma^5 n$. Then, there exists an event $\mathcal{E}$ that only depends on X^*, M_1, M_2 and occurs with probability at least $1 - O(\gamma)$ over $\mathcal{P}_{n,\Delta,\alpha}$ and the choice of $X^* \in \{0,1\}^n$ such that, conditioned on $\mathcal{E}$, one has*

(1) $|\mathbf{B}_2|/2^n \ge 2^{-\gamma^4 n}$.

(2) $\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{v\in\{0,1\}^n, |v|=2\ell} \widehat{h}_2(v)^2 \le (C\Delta^2\gamma^4 n/\ell)^\ell$ *for all* $\ell \le \gamma^4 n$.

Proof. We denote

(4.15)

$$
\begin{aligned}
\mathcal{E}_1 &:= \left\{ |\mathbf{A}_1|/2^n \ge 2^{-s-\log_2(2/\gamma)} \right\} \\
\mathcal{E}_2 &:= \left\{ |\mathbf{A}_t|/2^n \ge 2^{-s-\log_2(2/\gamma)} \ \text{ for } \ t \in \{1,2\} \right\}.
\end{aligned}
$$

Note that $\mathcal{E}_2 \subseteq \mathcal{E}_1$. We will later show that for every $t \in \{1,2\}$,

$$\mathbf{Pr}[\mathcal{E}_t] \geq 1 - O(\gamma). \tag{4.16}$$

Note that neither $\mathcal{E}_1$ nor $\mathcal{E}_2$ coincides with the event $\mathcal{E}$—we define $\mathcal{E}$ at the end of the proof as the intersection of $\mathcal{E}_2$ and the success event for an application of Markov's inequality (see Eq. (4.25) and Eq. (4.26)).

We prove that if matchings M_1, M_2 are selected according to the random process $\mathcal{P}_{n,\Delta,\alpha}$, then the following conditions hold:

(1) $\frac{|\mathbf{B}_2|}{2^n} = \frac{|\mathbf{A}_1|}{2^n} \cdot \frac{|\mathbf{A}_2|}{2^n}$ for all choices of M_1, M_2, f_1, f_2;

(2) Conditioned on $\mathcal{E}_1$ for all $\ell \in [1, 2s]$,

$$\left(\frac{2^n}{|\mathbf{B}_1|}\right)^2 \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{f_1}(v)^2 \leq (128s/\ell)^\ell.$$

(3) Conditioned on $\mathcal{E}_2$ for all $\ell \in [1, 2s]$,

$$\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{f_1 \cdot f_2}(v)^2 \leq (O(\Delta^2/\gamma) \cdot s/\ell)^\ell.$$

We prove claims above, then put them together to get the proof of the lemma. Claims **(1)** and **(2)** are simple, and the proof of the lemma from the claims is simple as well. The bulk of the proof is in **(3)**. We now give the proof of the lemma assuming the claims above.

We now combine **(1)-(3)** to obtain the result of the lemma. Recall that $h_2 = f_1 \cdot f_2$.

First, for $\ell \in [1, 2s]$, we have that **(3)** implies

$$\begin{aligned}\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{f_1 \cdot f_2}(v)^2 &\leq (O(\Delta^2/\gamma)s/\ell)^\ell \\ &\leq (C\Delta^2\gamma^4 n/\ell)^\ell\end{aligned}$$

for sufficiently large C, since $s \leq \frac{1}{2048C\Delta^2}\gamma^5 n$ by assumption of the lemma.

It remains to show that this bound holds for all $\ell \leq \gamma^4 n$, i.e., we need to consider ℓ in the range $[2s, \gamma^4 n]$. We note that, conditioned on $\mathcal{E}$, one has

$$\frac{2^n}{|\mathbf{B}_2|} = \frac{2^n}{|\mathbf{A}_1|} \cdot \frac{2^n}{|\mathbf{A}_2|} \leq (2^{2s})^2 \leq 2^{4s},$$

where we have combined **(1)** with the fact that conditioned on $\mathcal{E} \subseteq \mathcal{E}_2$, one has $|\mathbf{A}_t|/2^n \geq 2^{-s-\log_2(2/\gamma)} \geq 2^{-2s}$ for every $t \in \{1,2\}$ and sufficiently large n, since $\gamma > n^{-1/5}$ and $s = s(n) = \omega(\sqrt{n})$. Thus, by Theorem 4.1, **(3)** (Parseval's equality), we have

$$\begin{aligned}\left(\frac{2^n}{|\mathbf{B}_t|}\right)^2 \sum_{\substack{v \in \{0,1\}^n \\ |v|=2\ell}} \widehat{h}_t(v)^2 \leq \left(\frac{2^n}{|\mathbf{B}_t|}\right)^2 \sum_{v \in \{0,1\}^n} \widehat{h}_t(v)^2 \\ \leq \frac{2^n}{|\mathbf{B}_t|} \\ \leq 2^{4s}.\end{aligned} \tag{4.17}$$

for $t \in \{1,2\}$ and all ℓ.

We now show that the rhs above is dominated by $(C\Delta^2\gamma^4 n/\ell)^\ell$ for $\ell \in [2s, \gamma^4 n]$, provided that $C > 0$ is a sufficiently large absolute constant. Indeed, recalling that Δ is a positive integer, we note that as long as $C \geq e$, we have that $(C\Delta^2\gamma^4 n/\ell)^\ell$ is monotonically increasing[2] for $\ell \in [2s, \gamma^4 n]$. Thus, the smallest value is achieved when $\ell = 2s$ and equals

$$\begin{aligned}(C\Delta^2\gamma^4 n/(2s))^{2s} &\geq (4C\Delta^2\gamma^4 n/(\gamma^5 n))^{2s} \\ &\geq (4C\Delta^2/\gamma)^{2s} \\ &\geq 2^{4s},\end{aligned}$$

where we have used the assumption that $s \leq \frac{1}{2048C\Delta^2}\gamma^5 n \leq \frac{1}{8}\gamma^5 n$. This establishes part **(2)** of the lemma statement. Also, note that **(1)** of the lemma statement holds, since

$$\frac{|\mathbf{B}_2|}{2^n} \geq 2^{-4s} \geq 2^{-\gamma^4 n},$$

since $4s \leq 4 \cdot (\gamma^5 n/2048C\Delta^2) \leq \gamma^4 n$. This completes the proof of the lemma assuming claims **(1)-(3)** above.

We now prove the claims.

First, we establish Claim **(1)**, which follows from the fact that $M_1 \cup M_2$ does not contain cycles. Indeed, by (4.6), we have

$$\begin{aligned}\widehat{h_2}(0) &= \sum_{w \in \{0,1\}^n} \widehat{f_1}(w) \cdot \widehat{f_2}(w) \\ &= \widehat{f_1}(0^n) \cdot \widehat{f_2}(0^n) + \sum_{w \in \{0,1\}^n \setminus 0^n} \widehat{f_1}(w) \cdot \widehat{f_2}(w),\end{aligned}$$

and by Theorem 4.1, **(1)**, all $w \in \{0,1\}^n \setminus 0^n$ such that $\widehat{f_1}(w) \neq 0$ and $\widehat{f_2}(w) \neq 0$ can be perfectly matched by both M_1 and M_2. Let $M_1' \subseteq M_1$ denote the set of edges of M_1 that perfectly match elements of w to each other, and let $M_2' \subseteq M_2$ denote the set of edges of M_2 that perfectly match elements of w to each other.

[2] Since the function $(ea/b)^b$ is monotone increasing for any $b \in (0, b]$.

However, this implies that $M_1' \cup M_2'$ must be a union of cycles (note that M_1' and M_2' do not share edges by our construction), which is impossible as $M_1 \cup M_2$ does not contain cycles. Thus, the second term on the rhs of the equation above is zero, and we get

$$\frac{|\mathbf{B}_2|}{2^n} = |\widehat{h_2}(0^n)| = |\widehat{f_1}(0^n)| \cdot |\widehat{f_2}(0^n)| = \frac{|\mathbf{A}_1|}{2^n} \cdot \frac{|\mathbf{A}_2|}{2^n},$$

as desired. This establishes Claim **(1)**.

Let us now concentrate on Claim **(2)**. Note that the claim only applies to $\ell \geq 1$, which will be useful for simplifying calculations somewhat below.

Typical messages. First, note that for each $t \in \{1,2\}$, the function g_t induces a partition $K_1^t, K_2^t, \ldots, K_{2^s}^t$ of $\{0,1\}^{m_t}$, where s is the bit length of the message a_t (recall that we assume wlog that messages are the same length for all t). The number of points in $\{0,1\}^{m_t}$ that belong to sets K_i^t of size less than $\gamma 2^{m_t - s}$ is bounded by $2^s \cdot \gamma 2^{m_t - s} < \gamma 2^{m_t}$, i.e., at least a $1-\gamma$ fraction of $\{0,1\}^{m_t}$ is contained in large sets K_i^t, whose size is at least $\gamma 2^{m_t - s}$. We call a message m *typical* if $|K_m^t| \geq \gamma 2^{m_t - s}$. Moreover, we say that $a_t = g_t(M_t x)$ is typical if $M_t x$ is typical. We have that $a_t = g_t(z)$ is not typical with probability at most γ if z is uniformly random in $\{0,1\}^{m_t}$. Letting $d := \log_2(2^n/|\mathbf{A}_1|)$, we now conclude that with probability at least $1 - \gamma/2$ over the choice of $X^* \in \{0,1\}^n$, one has $d \leq s + \log_2(2/\gamma)$. Since $\gamma > n^{-1/5}$ and $s = \omega(\sqrt{n})$ by assumption of the lemma, we have $d \leq s + \log_2(2/\gamma) \leq 2s$ for sufficiently large n. We now invoke Theorem 4.1, **(2)** on the function f_1 with $d \leq s + \log_2(2/\gamma) \leq 2s$, which establishes Claim **(2)**.

Finally, we establish Claim **(3)**. First note that by Lemma 3.1, **(1)** applied to $\mathbf{A}_1$ and M_2, we get that $M_2 X^*$ is uniformly distributed over $\{0,1\}^{m_2}$ when X^* is uniformly distributed over $\mathbf{A}_1$. We thus have that the argument on 'typical' sets from the above paragraph applies even when we condition on M_1 and the first player's message a_1 (equivalently, on the set $\mathbf{A}_1$). Thus, with probability $1 - O(\gamma)$, we have that $\log_2(2^n/|\mathbf{A}_2|) \leq s + \log_2(2/\gamma)$, which establishes (4.16).

Thus, assume $\mathcal{E}_2$ holds. Recall that $d = \log_2(2^n/|\mathbf{A}_1|) \leq s + \log_2(1/\gamma) \leq 2s$. We now claim that for every $k \leq 2s$,

$$\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w \in \{0,1\}^n \\ |w| = 2k}} \widehat{f_1}(v)^2 \leq (128s/k)^k \tag{4.18}$$

and

$$\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w \in \{0,1\}^n \\ |w| \geq 4s}} \widehat{f_1}(w)^2 \leq 2^{4s}. \tag{4.19}$$

Indeed, (4.19) holds by Theorem 4.1, **(3)**:

$$\begin{aligned}\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w \in \{0,1\}^n \\ |w| \geq 4s}} \widehat{f_1}(w)^2 &\leq 2^{2d} \sum_{w \in \{0,1\}^n} \widehat{f_1}(w)^2 \\ &= 2^d \\ &\leq 2^{4s}.\end{aligned}$$

For (4.18), note that if $k \leq d$, then Theorem 4.1, **(2)** implies that

$$\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w \in \{0,1\}^n \\ |w| = 2k}} \widehat{f_1}(w)^2 \leq (64d/k)^k \leq (128s/k)^k,$$

as desired, while if $d < k \leq 2s$, then Theorem 4.1, **(3)** implies that

$$\begin{aligned}\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w \in \{0,1\}^n \\ |w| = 2k}} \widehat{f_1}(w)^2 &\leq 2^d \\ &\leq (128s/d)^d \\ &\leq (128s/k)^k,\end{aligned}$$

since $(128s/k)^k$ is a monotonically increasing function in k for $k \leq 2s$. This establishes (4.18).

Next, by Lemma 4.2, we have that for any $v \in \{0,1\}^n$, $\widehat{(f_1 \cdot f_2)}(v) = 0$ if v is not admissible with respect to M_1, M_2, while $\widehat{(f_1 \cdot f_2)}(v) = \widehat{f_1}(P(v) \cap M_1) \cdot \widehat{f_2}(P(v) \cap M_2)$ otherwise. Thus, for any $\ell \geq 0$,

$$\begin{aligned}\sum_{\substack{v \in \{0,1\}^n \\ |v| = 2\ell}} \widehat{f_1 \cdot f_2}(v)^2 &= \sum_{\substack{v \in \{0,1\}^n \\ |v| = 2\ell \\ v \text{ admissible} \\ \text{wrt } M_1, M_2}} \widehat{f_1}(P(v) \cap M_1)^2 \\ &\qquad \cdot \widehat{f_2}(P(v) \cap M_2)^2 \\ &= \sum_{\substack{v \in \{0,1\}^n \\ |v| = 2\ell}} \sum_{w \in \{0,1\}^n} \widehat{f_1}(w)^2 \widehat{f_2}(w+v)^2.\end{aligned}$$

Note that the second line follows from the first by letting $w := P(v) \cap M_1$ (so that $w + v = (P(v) \cap M_1) + v = P(v) \cap M_2$ and, thus, $\widehat{f_1}(w)^2 \cdot \widehat{f_2}(w+v)^2 = \widehat{f_1}(P(v) \cap M_1)^2 \cdot \widehat{f_2}(P(v) \cap M_2)^2$) as well as noting that there exists at most one $w \in \{0,1\}^n$ such that $\widehat{f_1}(w)^2 \cdot \widehat{f_2}(w+v)^2 \neq 0$ (see the proof of Lemma 4.2).

We now further partition the set of $w \in \{0,1\}^n$ in the inner summation on the rhs above according to weight and obtain

(4.20)

$$\begin{aligned}
&\sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f_1\cdot f_2}(v)^2 \\
&\quad= \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \sum_{w\in\{0,1\}^n} \widehat{f}_1(w)^2\cdot\widehat{f}_2(w+v)^2 \\
&\quad= \sum_{k=0}^{\Delta\cdot\ell} \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2\cdot\widehat{f}_2(w+v)^2 \\
&\quad= \sum_{k=0}^{\Delta\cdot\ell} \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2\cdot\left(\sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2\right).
\end{aligned}$$

Note that we have restricted the summation over k to the range $[0,\Delta\cdot\ell]$ in line 3, as this is justified by Lemma 4.1, **(2)**, which implies that $|P^*(w+M')|\geq |w|/\Delta$ for all $M'\subseteq M_2$, and so, $|v|=|w+(v+w)|=|w+M'|\geq |w|/\Delta$, or $k=|w|/2\leq\Delta\cdot\ell$ for all v,w such that $\widehat{f}_1(w)\widehat{f}_2(v+w)\neq 0$.

Taking the expectation of (4.20) with respect to M_2 (conditional on M_1, $\mathbf{A}_1$, and $\mathcal{E}_2$), we obtain

$$\begin{aligned}
&\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f_1\cdot f_2}(v)^2 \,\middle|\, M_1,\mathbf{A}_1,\mathcal{E}_2\right] \\
&\quad= \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \\
&\qquad\cdot\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f_1\cdot f_2}(v)^2 \,\middle|\, M_1,\mathbf{A}_1,\mathcal{E}_2\right] \\
&\quad\leq \sum_{k=0}^{\Delta\cdot\ell}\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2 \\
&\qquad\cdot\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1,\mathbf{A}_1,\mathcal{E}_2\right].
\end{aligned}$$

In what follows, we apply Lemma 4.4 to the inner summation on last line above. In order to reason about 'typical' messages as defined above, we let

$$\mathbf{I}_1^\star := \mathbf{I}\left[\frac{|\mathbf{A}_1|}{2^n}\geq 2^{-2s}\right] \quad\text{and}\quad \mathbf{I}_2^\star := \mathbf{I}\left[\frac{|\mathbf{A}_2|}{2^n}\geq 2^{-2s}\right].$$

Note that

(4.21)

$$\begin{aligned}
\mathbf{I}_1^\star &\geq \mathbf{I}\left[\frac{|\mathbf{A}_1|}{2^n}\geq 2^{-s-\log_2(2/\gamma)}\right] \\
\mathbf{I}_2^\star &\geq \mathbf{I}\left[\frac{|\mathbf{A}_2|}{2^n}\geq 2^{-s-\log_2(2/\gamma)}\right].
\end{aligned}$$

Specifically, we have for that for any $\ell\leq 2s$,

$$\begin{aligned}
&\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{k=0}^{\Delta\cdot\ell} \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2\cdot\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \right. \\
&\qquad\left.\cdot\sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1,\mathbf{A}_1,\mathcal{E}_2\right] \\
&\quad= \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{k=0}^{\Delta\cdot\ell} \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2\cdot\mathbf{E}_{M_2}\left[\mathbf{I}_1^\star\cdot\mathbf{I}_2^\star \right. \\
&\qquad\left.\cdot\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1,\mathbf{A}_1,\mathcal{E}_2\right] \\
&\quad\leq \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \frac{1}{\mathbf{Pr}[\mathcal{E}_2]} \sum_{k=0}^{\Delta\cdot\ell} \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2\cdot\mathbf{E}_{M_2}\left[\right. \\
&\qquad\left.\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2\cdot\mathbf{I}_1^\star\cdot\mathbf{I}_2^\star\cdot\sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1,\mathbf{A}_1\right] \\
&\quad= 2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2\cdot\mathbf{I}_1^\star\cdot\sum_{k=0}^{\Delta\ell} \sum_{\substack{w\in\{0,1\}^n\\|w|=2k}} \widehat{f}_1(w)^2\cdot\mathbf{E}_{M_2}\left[\right. \\
&\qquad\left.\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2\cdot\mathbf{I}_2^\star\cdot\sum_{\substack{v\in\{0,1\}^n\\|v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1,\mathbf{A}_1\right],
\end{aligned}$$

where we have used (4.21) to conclude that both $\mathbf{I}_1^\star$ and $\mathbf{I}_2^\star$ equal 1 when $\mathcal{E}_2$ occurs, as well as the fact that $\mathbf{Pr}[\mathcal{E}_2]=1-O(\gamma)\geq 1/2$ (when γ is smaller than an absolute constant) by (4.16) and the fact that $\mathbf{I}_1^\star$ is independent of M_2.

We now apply Lemma 4.4 to the expectation over M_2 in the last line above. Since Lemma 4.4 provides two bounds (one for $\ell\leq k$ and another for $\ell\geq k$), we split the summation into two and apply the respective

part of the lemma to each summation. Specifically, we have

$$
\begin{aligned}
&2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \mathbf{I}_1^\star \cdot \sum_{k=0}^{\Delta\ell} \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \cdot \mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \right. \\
&\qquad \left. \cdot \mathbf{I}_2^\star \cdot \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \\
&\le 2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \mathbf{I}_1^\star \cdot \sum_{k=0}^{\ell} \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \cdot \mathbf{E}_{M_2}\left[\right. \\
&\qquad \left.\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}_2^\star \cdot \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \\
&\quad + 2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \mathbf{I}_1^\star \cdot \sum_{k=\ell+1}^{\Delta\ell} \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \cdot \mathbf{E}_{M_2}\left[\right. \\
&\qquad \left.\left(\frac{2^n}{|\mathbf{A}_2|}\right)^2 \cdot \mathbf{I}_2^\star \cdot \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f}_2(w+v)^2 \,\middle|\, M_1, \mathbf{A}_1\right] \\
&= S_1 + S_2,
\end{aligned}
$$

where we let

$$
\begin{aligned}
S_1 &= 2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \mathbf{I}_1^\star \cdot \sum_{k=0}^{\ell} 4^\ell (O(\Delta))^k \left(\frac{64(2s)}{\ell-k}\right)^{\ell-k} \\
&\quad \cdot \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \\
S_2 &= 2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \mathbf{I}_1^\star \cdot \sum_{k=\ell+1}^{\Delta\ell} (O(\Delta))^\ell 8^k \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\quad \cdot \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2.
\end{aligned}
$$

We now proceed to bound the terms S_1 and S_2 separately.

Bounding S_1. We have

(4.22)

$$
\begin{aligned}
S_1 &= 2\left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \mathbf{I}_1^\star \cdot \sum_{k=0}^{\ell} 4^\ell (O(\Delta))^k \left(\frac{64(2s)}{(\ell-k)}\right)^{\ell-k} \\
&\quad \cdot \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \\
&= 2\sum_{k=0}^{\ell} 4^\ell (O(\Delta))^k \left(\frac{64(2s)}{(\ell-k)}\right)^{\ell-k} \cdot \mathbf{I}_1^\star \cdot \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \\
&\quad \cdot \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \\
&\le 4^{\ell+1} \sum_{k=0}^{\ell} (O(\Delta))^k \left(\frac{64(2s)}{(\ell-k)}\right)^{\ell-k} \cdot \left(\frac{64(2s)}{k}\right)^k \\
&= (O(\Delta))^\ell \sum_{k=0}^{\ell} \left(\frac{128s}{\ell-k}\right)^{\ell-k} \cdot \left(\frac{128s}{k}\right)^k \\
&= (O(\Delta))^\ell \left(\frac{128s}{\ell}\right)^\ell \sum_{k=0}^{\ell} \frac{\ell^\ell}{(\ell-k)^{\ell-k}k^k} \\
&= (O(\Delta))^\ell \left(\frac{128s}{\ell}\right)^\ell \sum_{k=0}^{\ell} 2^\ell \\
&\le \left(\frac{128s}{\ell}\right)^\ell (O(\Delta))^\ell \\
&\le \left(\frac{O(\Delta)s}{\ell}\right)^\ell,
\end{aligned}
$$

where we have used (4.18) as well as the fact that $\frac{(a+b)^{a+b}}{a^a b^b} \le 2^{a+b}$ for all $a, b > 0$. Note that we have absorbed the factor of $4^{\ell+1}$ into $(O(\Delta))^\ell$ crucially using the assumption that $\ell > 0$.

Bounding S_2. Observe that

$$
S_2 = \mathbf{I}_1^\star \cdot \sum_{k=\ell+1}^{\Delta\cdot\ell} (O(\Delta))^\ell 8^k \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \cdot \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2.
$$

We split this summation further into two summations, one over $k \in [\ell+1, 2s]$ and the other over $k \in [2s, \Delta\cdot\ell]$ (assuming that the second range is nonempty).

Case 1: $k \in [\ell+1, 2s]$**.** We have

(4.23)
$$\begin{aligned}
&\sum_{k=\ell+1}^{2s} (O(\Delta))^\ell 8^{\ell+(k-\ell)} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \cdot \mathbf{I}_1^\star \\
&\qquad \cdot \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \\
&\leq \sum_{k=\ell+1}^{s} (O(\Delta))^\ell 8^{\ell+(k-\ell)} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \left(\frac{64(2s)}{k}\right)^k \\
&\leq \left(\frac{O(\Delta)s}{\ell}\right)^\ell \sum_{k=\ell+1}^{s} \left(\frac{2048s}{n}\right)^{k-\ell} \frac{(k-\ell)^{k-\ell}\ell^\ell}{k^k} \\
&\leq \left(\frac{O(\Delta)s}{\ell}\right)^\ell \sum_{k=\ell+1}^{s} \left(\frac{2048s}{n}\right)^{k-\ell} \\
&\leq \left(\frac{O(\Delta)s}{\ell}\right)^\ell \sum_{k=\ell+1}^{s} \left(\frac{2048s}{n}\right)^{k-\ell} \\
&\leq \left(\frac{O(\Delta)s}{\ell}\right)^\ell,
\end{aligned}$$

using (4.18), as well as the fact that $a^a b^b/(a+b)^{a+b} \leq 1$ for all $a, b > 0$ and $s < n/4096$.

Case 2: $k \in [2s, \Delta\cdot\ell]$**.** Note that increasing the upper limit in the summation to $\Delta \cdot 2s \geq \Delta \cdot \ell$ may only increase the sum since the summands are non-negative. We upper bound the sum of k in this range as follows:

$$\begin{aligned}
&\sum_{k=2s}^{2\Delta\cdot s} (O(\Delta))^\ell 8^{\ell+(k-\ell)} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \\
&\qquad \cdot \mathbf{I}_1^\star \cdot \left(\frac{2^n}{|\mathbf{A}_1|}\right)^2 \sum_{\substack{w\in\{0,1\}^n \\ |w|=2k}} \widehat{f}_1(w)^2 \\
&\leq \sum_{k=2s}^{2\Delta\cdot s} (O(\Delta))^\ell 8^{\ell+(k-\ell)} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \cdot 2^{4s} \\
&\leq \sum_{k=2s}^{2\Delta\cdot s} (O(\Delta))^\ell 8^{\ell+(k-\ell)} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \left(\frac{8\Delta s}{k}\right)^k,
\end{aligned}$$

where we have used (4.19) and the fact that

$$\left(\frac{8\Delta s}{k}\right)^k \geq \left(\frac{8\Delta s}{2s}\right)^{2s} \geq (4\Delta)^{2s} \geq 2^{4s}.$$

We now upper bound the expression on the last line above as follows:

(4.24)
$$\begin{aligned}
&\sum_{k=2s}^{2\Delta\cdot s} (O(\Delta))^\ell 8^{\ell+(k-\ell)} \left(\frac{k-\ell}{n/2}\right)^{k-\ell} \cdot \left(\frac{8\Delta s}{k}\right)^k \\
&\leq \left(\frac{O(\Delta^2)s}{\ell}\right)^\ell \sum_{k=2s}^{2\Delta\cdot s} \left(\frac{O(\Delta)s}{n}\right)^{k-\ell} \frac{(k-\ell)^{k-\ell}\ell^\ell}{k^k} \\
&\leq \left(\frac{O(\Delta^2)s}{\ell}\right)^\ell \sum_{k=2s}^{\infty} \left(\frac{O(\Delta)s}{n}\right)^{k-\ell} \\
&\leq \left(\frac{O(\Delta^2)s}{\ell}\right)^\ell \sum_{k=\ell+1}^{\infty} \left(\frac{O(\Delta)s}{n}\right)^{k-\ell} \\
&\leq \left(\frac{O(\Delta^2)s}{\ell}\right)^\ell,
\end{aligned}$$

where we have used the fact that $a^a b^b/(a+b)^{a+b} \leq 1$ for all $a, b > 0$.

Putting Eq. (4.22), Eq. (4.23) and Eq. (4.24) together, we get that for every $0 < \ell \leq 2s$,

$$\begin{aligned}
S_1 + S_2 &\leq (O(\Delta)s/\ell)^\ell + (O(\Delta)s/\ell)^\ell + (O(\Delta^2)s/\ell)^\ell \\
&= (O(\Delta^2)s/\ell)^\ell,
\end{aligned}$$

where we again have used the assumption that $\ell > 0$ to absorb a constant factor into the $O(\Delta)$ term. Substituting this bound in the derivations above, we note that for every $0 < \ell \leq 2s$,

$$\begin{aligned}
&\mathbf{E}_{M_2}\left[\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f_1\cdot f_2}(v)^2 \,\middle|\, M_1, \mathbf{A}_1, \mathcal{E}_2\right] \\
&\qquad = \left(\frac{O(\Delta^2)s}{\ell}\right)^\ell.
\end{aligned}$$

Thus, Markov's inequality implies that with probability at least $1 - O(\gamma)$, one has that for every $0 < \ell \leq 2s$, there exists an absolute constant $K > 0$ such that

$$\begin{aligned}
&\mathbf{Pr}_{M_2}\left[\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f_1\cdot f_2}(v)^2 \right. \\
&\qquad \left. > (K(\Delta^2/\gamma)s/\ell)^\ell \,\middle|\, M_1, \mathbf{A}_1, \mathcal{E}_2\right] \leq \gamma^\ell. \qquad (4.25)
\end{aligned}$$

Therefore, by a union bound over $0 < \ell \leq 2s$,

(4.26)

$$\mathbf{Pr}_{M_2}\left[\left(\frac{2^n}{|\mathbf{B}_2|}\right)^2 \sum_{\substack{v\in\{0,1\}^n \\ |v|=2\ell}} \widehat{f_1 \cdot f_2}(v)^2 > (K(\Delta^2/\gamma)s/\ell)^\ell \right.$$
$$\left. \text{for some } \ell \in [1, 2s] \,\middle|\, M_1, \mathbf{A}_1, \mathcal{E}_2 \right]$$
$$\leq \sum_{\ell \geq 1} \gamma^\ell = O(\gamma),$$

since γ is bounded from above by an absolute constant. We now define the event $\mathcal{E}$ (promised by the lemma) as the intersection of $\mathcal{E}_2$ and the success event for the application of Markov's inequality above. This completes the proof of Claim **(3)**, as desired.

References

[1] *Bertinoro workshop 2011, problem 45,* `http://sublinear.info/45`.

[2] *Bertinoro workshop 2014, problem 67,* `http://sublinear.info/67`.

[3] K. Ahn and S. Guha, *Graph sparsification in the semi-streaming model*, ICALP, (2009), pp. 328–338.

[4] ———, *Linear programming in the semi-streaming model with application to the maximum matching problem*, ICALP, (2011), pp. 526–538.

[5] 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).

[6] K. J. Ahn, S. Guha, and A. McGregor, *Analyzing graph structure via linear measurements*, SODA, (2012), pp. 459–467.

[7] ———, *Graph sketching: Sparsification, spanners, and subgraphs*, PODS, (2012).

[8] N. Alon, Y. Matias, and M. Szegedy, *The space complexity of approximating the frequency moments*, in STOC, 1996, pp. 20–29.

[9] S. Assadi, S. Khanna, Y. Li, and G. Yaroslavtsev, *Tight bounds for linear sketches of approximate matchings*, CoRR, (2015).

[10] A. A. Benczúr and D. R. Karger, *Approximating s-t minimum cuts in $\tilde{O}(n^2)$ time*, Proceedings of the 28th annual ACM symposium on Theory of computing, (1996), pp. 47–55.

[11] R. H. Chitnis, G. Cormode, H. Esfandiari, M. Hajiaghayi, A. McGregor, M. Monemizadeh, and S. Vorotnikova, *Kernelization via sampling with applications to dynamic graph streams*, CoRR, abs/1505.01731 (2015).

[12] R. Durrett, *Random Graph Dynamics (Cambridge Series in Statistical and Probabilistic Mathematics)*, Cambridge University Press, New York, NY, USA, 2006.

[13] H. Esfandiari, M. T. Hajiaghayi, V. Liaghat, M. Monemizadeh, and K. Onak, *Streaming algorithms for estimating the matching size in planar graphs and beyond*, SODA, (2015).

[14] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf, *Exponential separation for one-way quantum communication complexity, with applications to cryptography*, SIAM J. Comput., 38 (2008), pp. 1695–1708.

[15] A. Goel, M. Kapralov, and S. Khanna, *On the communication and streaming complexity of maximum bipartite matching*, SODA, (2012).

[16] V. Guruswami and K. Onak, *Superlinear lower bounds for multipass graph processing*, CCC, (2012).

[17] Z. Huang, B. Radunović, M. Vojnović, and Q. Zhang, *Communication complexity of approximate maximum matching in distributed graph data*, STACS, (2015).

[18] J. Kahn, G. Kalai, and N. 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, (1988), pp. 68–80.

[19] M. Kapralov, *Better bounds for matchings in the streaming model*, SODA, (2013).

[20] M. Kapralov, S. Khanna, and M. Sudan, *Approximating matching size from random streams*, in 25th ACM-SIAM Symposium on Discrete Algorithms (SODA), 2014.

[21] ———, *Streaming lower bounds for approximating MAX-CUT*, in 26th ACM-SIAM Symposium on Discrete Algorithms (SODA), 2015.

[22] M. Kapralov, Y. T. Lee, C. Musco, C. Musco, and A. Sidford, *Single pass spectral sparsification in dynamic streams*, FOCS, (2014).

[23] M. Kapralov and D. Woodruff, *Spanners and sparsifiers in dynamic streams*, PODC, (2014).

[24] J. A. Kelner and A. Levin, *Spectral sparsification in the semi-streaming setting*, STACS, (2011), pp. 440–451.

[25] D. Kogan and R. Krauthgamer, *Sketching cuts in graphs and hypergraphs*, ITCS, (2015).

[26] C. Konrad, *Maximum matching in turnstile streams*, CoRR, abs/1505.01460 (2015).

[27] A. McGregor, *Graph stream algorithms: a survey*, SIGMOD Record, 43 (2014), pp. 9–20.

[28] D. Spielman and N. Srivastava, *Graph sparsification by effective resistances*, STOC, (2008), pp. 563–568.

[29] E. Verbin and W. Yu, *The streaming complexity of cycle counting, sorting by reversals, and other problems*, SODA, (2011), pp. 11–25.