Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 86 (2021)

Linear Space Streaming Lower Bounds for Approximating CSPs

Chi-Ning Chou* Alexander Golovnev† Madhu Sudan‡ Ameya Velingker§
Santhoshini Velusamy¶

Abstract

We consider the approximability of constraint satisfaction problems in the streaming setting. For every constraint satisfaction problem (CSP) on n variables taking values in $\{0, \ldots, q-1\}$, we prove that improving over the trivial approximability by a factor of q requires $\Omega(n)$ space even on instances with $O(n)$ constraints. We also identify a broad subclass of problems for which any improvement over the trivial approximability requires $\Omega(n)$ space. The key technical core is an optimal, $q^{-(k-1)}$-inapproximability for the Max k-LIN-mod q problem, which is the Max CSP problem where every constraint is given by a system of $k-1$ linear equations mod q over k variables.

Our work builds on and extends the breakthrough work of Kapralov and Krachun (Proc. STOC 2019) who showed a linear lower bound on any non-trivial approximation of the Max-Cut problem in graphs. MaxCut corresponds roughly to the case of Max k-LIN-mod q with $k = q = 2$. For general CSPs in the streaming setting, prior results only yielded $\Omega(\sqrt{n})$ space bounds. In particular no linear space lower bound was known for an approximation factor less than $1/2$ for *any* CSP. Extending the work of Kapralov and Krachun to Max k-LIN-mod q to $k > 2$ and $q > 2$ (while getting optimal hardness results) is the main technical contribution of this work. Each one of these extensions provides non-trivial technical challenges that we overcome in this work.

*School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported by NSF grants DMS-2134157 and CCF-1565264, DARPA grant W911NF2010021, DOE grant DE-SC0022199, and the Simons foundation. Email: chiningchou@g.harvard.edu.

†Department of Computer Science, Georgetown University. Email: alexgolovnev@gmail.com.

‡School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported in part by a Simons Investigator Award and NSF Awards CCF 1715187 and CCF 2152413. Email: madhu@cs.harvard.edu.

§Google Research, USA. Email: ameyav@google.com.

¶School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported in part by a Google Ph.D. Fellowship, a Simons Investigator Award to Madhu Sudan, and NSF Awards CCF 1715187 and CCF 2152413. Email: svelusamy@g.harvard.edu.

ISSN 1433-8092

Contents

1 Introduction

In this work we consider the *approximability of constraint satisfaction problems (CSPs)* by *streaming algorithms* with sublinear space. We give tight inapproximability results for a broad class of CSPs, while giving somewhat weaker bounds on the approximability of every CSP. We introduce these terms below.

1.1 Background

We consider the general class of constraint satisfaction problems with finite constraints over finite-valued variables. A *problem* in this class, denoted $\mathsf{Max\text{-}CSP}(\mathcal{F})$, is given by positive integers q and k and a family of functions $\mathcal{F} \subseteq \{f : \mathbb{Z}_q^k \to \{0,1\}\}$. An *instance* of the problem consists of m constraints placed on n variables that take values in the set $\mathbb{Z}_q = \{0, \ldots, q-1\}$, where each constraint is given by a function $f \in \mathcal{F}$ and k distinct indices of variables $j_1, \ldots, j_k \in [n]$. Given an instance Ψ of $\mathsf{Max\text{-}CSP}(\mathcal{F})$, the goal is to compute the *value* val_Ψ defined to be the maximum, over all assignments to n variables, of the fraction of constraints satisfied by the assignment. For $\alpha \in [0,1]$, the goal of the α-approximate version of the problem is to compute an estimate η such that $\alpha \cdot \mathsf{val}_\Psi \leq \eta \leq \mathsf{val}_\Psi$.

In this work we consider the space complexity of approximating $\mathsf{Max\text{-}CSP}(\mathcal{F})$ by a single pass (potentially randomized) streaming algorithm that is presented the instance Ψ one constraint at a time. We consider "non-trivial" approximation algorithms for $\mathsf{Max\text{-}CSP}(\mathcal{F})$, where we first dismiss two notions of "triviality". First note that since we only consider space restrictions but not time restrictions, one can sample $O(n)$ constraints of Ψ and solve the $\mathsf{Max\text{-}CSP}(\mathcal{F})$ problem on the sampled constraint optimally to get a $(1-\varepsilon)$-approximation algorithm for every constant $\varepsilon > 0$ in $\widetilde{O}(n)$ space. Thus for this paper we view non-trivial algorithms to be those that run in $o(n)$ space.[1] The other form of "triviality" we dismiss is in the approximation factor. Given a family $\mathcal{F}$, let $\rho_{\min}(\mathcal{F})$ denote the infimum, over all instances Ψ of $\mathsf{Max\text{-}CSP}(\mathcal{F})$, of the value val_Ψ. Note that the algorithm that outputs the constant $\rho_{\min}(\mathcal{F})$ is a ($O(1)$-space!) $\rho_{\min}(\mathcal{F})$ approximation algorithm for $\mathsf{Max\text{-}CSP}(\mathcal{F})$. Thus we consider $\rho_{\min}(\mathcal{F})$ to be the "trivial" approximation factor for a family $\mathcal{F}$. With these two notions of "triviality" in mind, we define $\mathsf{Max\text{-}CSP}(\mathcal{F})$ to be *α-approximable* (in the streaming setting) if α is the largest constant such that there exists an α-approximation algorithm for $\mathsf{Max\text{-}CSP}(\mathcal{F})$ using $o(n)$ space. We say that $\mathsf{Max\text{-}CSP}(\mathcal{F})$ is simply *approximable* (in the streaming setting) if it is α-approximable for some $\alpha > \rho_{\min}$. We define a problem to be *approximation-resistant* (in the streaming setting) otherwise.

1.2 Results

Our first main result in this paper gives a sufficient condition for a problem to be approximation resistant in the streaming setting. We say that $f : \mathbb{Z}_q^k \to \{0,1\}$ is a *wide* constraint if there exists $\mathbf{a} \in \mathbb{Z}_q^k$ such that for every $i \in \mathbb{Z}_q$ we have $f(\mathbf{a} + i^k) = 1$ where $i^k = (i, i, \ldots, i)$ and addition is performed in the group $\mathbb{Z}_q^k$. We say that a family $\mathcal{F}$ is *wide* if every function $f \in \mathcal{F}$ is wide.

Theorem 1.1. *For every q, k and every wide family $\mathcal{F}$, $\mathsf{Max\text{-}CSP}(\mathcal{F})$ is approximation-resistant.*

Many natural CSPs are wide, including Boolean problems such as Max k-SAT and Max q-colorability. Others, such as Max k-LIN(q) and the "Unique Games" problem, contain wide subfamilies with the same "trivial" approximation factor, and thus Theorem 1.1 implies these are

[1] We note that there is a gap between the $o(n)$ space we allow and the $O(n \log n)$ space that is trivial, but we are not able to get sharp enough lower bounds to address this gap.

also approximation resistant. We elaborate on some of these examples in Section 4.1. However, clearly wideness does not capture all CSPs. For general CSPs, while we do not pin down the approximability exactly, we do manage to pin it down up to a multiplicative factor of q.

Theorem 1.2. *For every q, k and every family $\mathcal{F}$, if $\mathcal{F}$ is α-approximable then $\alpha \in [\rho_{\min}(\mathcal{F}), q \cdot \rho_{\min}(\mathcal{F})]$.*

Both Theorems 1.1 and 1.2 follow from our more detailed Theorem 4.3. In Section 4.1 we give a few examples illustrating how our theorems give tight lower bounds for some commonly studied CSPs including Max q-coloring, Unique Games, and Max Linear Systems.

1.3 Prior work

There have been a number of works in the broad area of approximations for streaming constraint satisfaction problems and lower bound techniques for those [GKK+09, VY11, KKS15, AKL16, KKSV17, GVV17, GT19, KK19, CGV20, AKSY20, AN21, CGSV21a, SSV21]. Among these our work is the *first work to aim to get tight inapproximability results for a broad class of CSPs for almost linear space single-pass streaming algorithms.* Previous works either did not get tight approximation factors or were aimed at specific problems or only got $\Omega(\sqrt{n})$-space lower bounds, though some do target multi-pass streaming algorithms [AKSY20, AN21] — which we do not do here. We describe the state of the art prior to our work below. (More detailed descriptions of prior works can be found in [CGSV21a].)

On the front of general lower bounds, Chou, Golovnev, Sudan and Velusamy [CGSV21a] explored the same set of CSP problems as we do, i.e, $\mathsf{Max\text{-}CSP}(\mathcal{F})$ for arbitrary q, k and $\mathcal{F}$. Their focus is on looser space lower bounds: specifically, they focus on problems that require $n^{\Omega(1)}$ space vs. those where $n^{o(1)}$ space suffices. They give a complete dichotomy for sketching algorithms, a special class of streaming algorithms. They also give sufficient conditions for approximation resistance with respect to sub-polynomial space general streaming algorithms. Theorem 2.9 in their paper shows that families $\mathcal{F}$ where the satisfying assignments of every function in the class support a one-wise independent distribution are approximation resistant. This theorem is incomparable with our Theorem 1.1 in that they give approximation resistance for a broader collection of problems (all wide families support one-wise independence) but the space lower bound is weaker — they give an $\Omega(\sqrt{n})$ lower bound and we get $\Omega(n)$ lower bounds for wide families. [CGSV21a] does not give an analogue of our Theorem 1.2, though such a result (with the weaker $\Omega(\sqrt{n})$ space lower bound) can be derived from their theorems equally easily. Indeed, our Section 4 is based on their work.

Turning to linear space lower bounds the breakthrough work here is due to Kapralov and Krachun [KK19], who show that approximating Max Cut (which translates in our setting to $\mathsf{Max\text{-}CSP}(\mathcal{F})$ for $\mathcal{F} = \{\oplus_2\}$ where $\oplus_2 : \{0,1\}^2 \to \{0,1\}$ is the binary XOR function) to within a factor $\frac{1}{2} + \varepsilon$ requires $\Omega(n)$ space for every $\varepsilon > 0$. Indeed, our work builds on their work and we compare our techniques later. Prior to the work of Kapralov and Krachun, there was a weaker result due to Kapralov, Khanna, Sudan and Velingker [KKSV17] showing that there exists $\varepsilon > 0$ such that $(1-\varepsilon)$-approximation for Max Cut requires linear space. Finally, Chou, Golovnev and Velusamy [CGV20] get a tight inapproximability for Max Exact 2-SAT (corresponding to $\mathsf{Max\text{-}CSP}(\mathcal{F})$ for $\mathcal{F} = \{\vee_2\}$, where $\vee_2 : \{0,1\}^2 \to \{0,1\}$ is the binary OR function) for linear space algorithms, by a reduction from Max Cut.

Thus, prior to our work it was conceivable (though of course extremely unlikely) that every $\mathsf{Max\text{-}CSP}(\mathcal{F})$ allowed a 1/2-approximating streaming algorithm using $o(n)$ space. Our work is the first to prove inapproximability $\alpha \leq 1/2$ for any $\mathsf{Max\text{-}CSP}(\mathcal{F})$. Indeed, we get inapproximabilities

going to 0 either as $q \to \infty$ (e.g., for the Unique Games problem) or as $k \to \infty$ (e.g., for the Max k-equality problem with $q = 2$ as defined later in Section 1.4).

The main contribution of our work is to extend the techniques of [KK19] to problems beyond Max Cut. Indeed the bulk of our proof takes the tour-de-force proof in [KK19] and finds the correct replacements in our setting. In the process, we arguably even present cleaner abstractions of their work. We elaborate on this further in the next section but first comment on why we feel the extensions are not straightforward given [KK19]. First we note that the exact class of problems we are able to deal with in Theorem 1.1 is not the fullest extension one may hope for. At the very least we have expected to cover the same set of problems as [CGSV21a, Theorem 2.9], i.e., families supporting one-wise independent distributions, but this remains open. Indeed to get our extensions we have to formulate a new communication problem which generalizes the one in [KK19] and is different from the many variations considered in [CGSV21b] and [CGSV21a]. In particular we are forced to work with a less expressive set of communication problems that already forces a "linear-algebraic" restriction on the core problems we work with. (We do believe a slight extension of our results to "families containing one-wise independent cosets of $\mathbb{Z}_q^k$" should be more feasible.) Having identified the right set of problems, carrying out the proof of Kapralov and Krachun is still non-trivial. In particular one has to be careful to ensure that the improvement in the exponent of the space bound (from $n^{1/2}$ to n) is by a full factor of 2 and not a factor of $k/(k-1)$, which is what one natural extension would lead to! We comment on these improvements in greater detail in the following.

Finally we point out that an extension of the lower bounds in [CGSV21a] to $\Omega(n)$ space lower bounds may actually be false. In particular, there is a candidate algorithm for one of the problems (Max 2-AND) that might improve on the approximation factors with $\omega(\sqrt{n})$ space. It certainly works better on the hard instances from previous reductions, but we do not have an improved analysis on all graphs.

1.4 Techniques and New Contributions

There are two lines of previous work that seem relevant to this work and we discuss our technical contributions relative to those here. We start with quick comparison with the previous work [CGSV21a] that gives $\Omega(\sqrt{n})$ lower bounds for a broader subset of problems than those addressed in this paper. We then move on to the work [KK19] which is much closer to our work and needs more detailed comparison.

Comparison with [CGSV21a]. While there is some obvious overlap in the set of problems considered in [CGSV21a] and this paper (and also in the set of authors) we claim that, beyond this aspect, the overlap in techniques is minimal. Both papers do use lower bounds on communication problems to establish lower bounds on streaming CSPs (which is standard in the context of streaming lower bounds). But the exact set of communication problems is different, and the tools used to establish the lower bounds are also different. In particular, [CGSV21a] create roughly a new communication problem for every γ, β and $\mathcal{F}$ and the main technical contributions there are lower bounds for these problems achieved mainly through a rich set of reductions among these communication problems. In our work we essentially work with one communication problem (once we fix k and q) and the core of our work is proving a lower bound for this problem. (This lower bound is based on extending [KK19] and we will elaborate on this later.) We use this one problem to get hardness for many different γ, β and $\mathcal{F}$ — this part is arguably related to the work of [CGSV21a] but we feel this is the obvious part of their work as well as our work. Finally, turning to the communication problems, the natural communication problems used to analyze streaming

complexity involves one way communication among a large constant number of players. The exact problem of this type that we focus on is different from the ones considered in [CGSV21a] due to a concept we call "folding". Folding makes our problems too restrictive to work for [CGSV21a] (i.e., would prevent them for addressing every $(\gamma, \beta) - \mathsf{Max\text{-}CSP}(\mathcal{F})$), whereas we do not know how to get our lower bounds without folding. We also note that [CGSV21a] derive their multiplayer lower bounds from lower bounds for a corresponding 2-player game and all their reductions work only for these 2-player games, which are inherently limited to yielding $\Theta(\sqrt{n})$ space lower bounds.

We now turn to the more significant comparison, with [KK19]. We start with a quick review of the main steps of [KK19] and then describe our analysis and conclude with a summary of the differences/new contributions relative to [KK19].

Summary of [KK19]. [KK19] work with a distributional T-player one-way communication game for some constant T. The game also has a parameter $\alpha > 0$. In instances of length n of this game, T players $P_1, \ldots, P_T$ get partial matchings $M_1, \ldots, M_T$ on the vertex set $[n]$ along with respective binary labels $\mathbf{z}_1, \ldots, \mathbf{z}_T$ on the edges of the matchings, i.e., player t receives input $(M_t, \mathbf{z}_t)$. Each matching contains αn edges, while each corresponding label $\mathbf{z}_t$ is an element of $\{0,1\}^{\alpha n}$. In the communication game, the players sequentially broadcast messages as follows. Player $t \in [T-1]$ computes a small message c_t which is a function of $M_t, \mathbf{z}_t$ and all "previous messages" $c_1, \ldots, c_{t-1}$,[2] after which the Tth player outputs a single 0/1 bit that is said to be the output of the communication protocol. The complexity of the protocol is the maximum over $t \in [T]$ of the message length c_t, and the goal of the players is to distinguish input instances drawn according to a **YES** distribution from those drawn according to a **NO** distribution, defined as follows.

In instances chosen from the **NO** distribution, the matchings $M_1, \ldots, M_T$ are chosen uniformly and independently from the set of matchings containing αn edges on the vertex set $[n]$. Furthermore, the vectors $\mathbf{z}_1, \ldots, \mathbf{z}_T$ are chosen uniformly and independently from $\{0,1\}^{\alpha n}$. In the **YES** distribution, the matchings are chosen as in the **NO** distribution, but in order to generate $\mathbf{z}_1, \ldots, \mathbf{z}_T$, we choose a common hidden vector $\mathbf{x}^* \in \{0,1\}^n$ uniformly at random and set each $\mathbf{z}_t$ as $\mathbf{z}_t(e) = x^*_a \oplus_2 x^*_b$ for every edge $e = (a, b)$. Thus, the label $\mathbf{z}_t$ can be viewed as specifying which edges of the i-th matching cross the cut determined by $\mathbf{x}^*$. If $T \gg \frac{1}{\alpha}$ then it can be seen that the **YES** and **NO** distributions are very far. The key theorem shows that for every $\alpha > 0$ and T, any protocol distinguishing **YES** instances from **NO** instances with constant advantage requires $\Omega(n)$ space. With this lower bound a space lower bound on Max Cut is straightforward.

Turning to the communication lower bound, the focus of the analysis are the sets $B_1, \ldots, B_T \subseteq \{0,1\}^n$ corresponding to the purported hidden vector $\mathbf{x}^*$ that are consistent with the messages $c_1, \ldots, c_T$. Specifically for $t \in [T]$, B_t is the set of all vectors $\mathbf{x}^*$ that are consistent with the first t matchings $M_{1:t}$ and the first t messages $c_{1:t}$. [KK19] argue that the sets B_t are not shrinking too fast (in either the **YES** case or the **NO** case) using a property that they term "C-boundedness," defined by the Fourier spectrum of the indicator function of B_t (the function in $\{0,1\}^n$ to $\{0,1\}$ that is 1 on B_t). We do not give the exact definition of boundedness here but roughly describe it as follows: Given an arbitrary set B of size S and a Fourier weight w, the total Fourier mass (strictly the ℓ_1-mass) of the wth level Fourier coefficients of B is well-known (by classical Fourier analysis) to be bounded by some amount $U(w) = U_{S,n}(w)$. For C-bounded sets, the corresponding Fourier mass is required to be at most $C^w U(w/2)$. The factor of two gained here in the argument of U is the crux to improvement in the space lower bound from $\sqrt{n}$ to n. (If the right hand side had been of the form $C^w U(\alpha w)$ then the space lower bound would be $\Omega(n^{1/(2\alpha)})$.) This factor of

[2] For technical reasons the lower bounds are proved in the stronger model where player t get $M_1, \ldots, M_{t-1}$ as well, but this difference is not crucial for the current discussion.

two, in turn, is attributable to the fact that the $\mathbf{z}_t$ only contain information about pairs of bits of $\mathbf{x}^*$. Their analysis shows that, for every t, B_t is C_t-bounded for some constant C_t. (The proof is inductive on t but the inductive hypothesis is complex and we won't reproduce it here.) They further show that if B_T is C-bounded for some constant C, then the distinguishing probability is at most $o(1)$.

Our Analysis. The core of our paper focuses on one problem for every given q and k, which we call Max k-EQ(q). This is the problem given by Max-CSP($\mathcal{F}$) for $\mathcal{F} = \{f_{b_2,\ldots,b_k} : \mathbb{Z}_q^k \to \{0,1\}\}$, where $f_{b_2,\ldots,b_k}(a_1,\ldots,a_k) = 1$ if and only if $a_t = a_1 + b_t \bmod q$ for every $t \in \{2,\ldots,k\}$. All our lower bounds effectively come from a tight $q^{-(k-1)}$-inapproximability of this problem for every q and k.

To study this problem we introduce a T-player communication problem that we call the "Implicit Randomized Mask Detection Problem" (IRMD) described as follows: There are T players each of whom receives an αn k-hypermatching M_t (i.e., a set of αn k-uniform hyperedges on $[n]$ that are pairwise disjoint). Additionally, the players receive a label in $\mathbb{Z}_q^k$ for every hyperedge they see. Thus the ith player's input is $(M_t, \mathbf{z}_t)$ where $\mathbf{z}_t \in (\mathbb{Z}_q^k)^{\alpha n}$. In the **NO** distribution the $\mathbf{z}_t$'s are drawn uniformly. In the **YES** distribution a vector $\mathbf{x}^* \in [q]^n$ is drawn uniformly and the label associated with an edge $\mathbf{j} = (j_1,\ldots,j_k)$ is $(x^*_{j_1} + a_{\mathbf{j}},\ldots,x^*_{j_k} + a_{\mathbf{j}})$ where $a_{\mathbf{j}} \in [q]$ is chosen uniformly and independently for each edge in each matching. The goal of the players is to distinguish between the **YES** and **NO** distributions with minimal communication (with one-way communication from the $t-1$th player to the tth player, as before).

To lower bound the communication complexity of IRMD we consider a folded version of the problem we call IFRMD where the labels associated with an edge are from $\mathbb{Z}_q^{k-1}$ and obtained by mapping an IRMD label $\mathbf{z} = (z^{(1)},\ldots,z^{(k)}) \in \mathbb{Z}_q^k$ to the label $\tilde{\mathbf{z}} = (z^{(2)} - z^{(1)},\ldots,z^{(k)} - z^{(1)})$. With this folding we recover the same communication problem as [KK19] for the case of $k = q = 2$ and the main focus of our work is proving lower bounds for higher k and q.

Our analysis of the communication complexity of IFRMD follows the same sequence of steps (with imitation even within the steps) as [KK19]. In particular we also use the same sets $B_1,\ldots,B_T$ and use the same notion of boundedness.

Turning to the induction and the analysis of boundedness of B_t for general t, we are able to extract a clean lemma (Lemma 5.18) that makes the induction completely routine. To explain this contribution note that B_t is the intersection of B_{t-1} with a set say A_t where A_t is of the same type as B_t (both are obtained by looking at the vector $\mathbf{x}^*$ projected to a matching followed by some folding). Thus both B_{t-1} and A_t are bounded sets. To complete the induction it would suffice to prove that the intersection of bounded sets is bounded, but alas this is not true! To get that B_t is bounded, we need to use the fact that the matching M_t is random and chosen independently of B_{t-1} but it turns out that that is all that is needed. This is exactly what we show in Lemma 5.18 — and of course this only happens with high probability over the choice of M_t.

Incremental contribution over [KK19]. Given that our result closely follows [KK19] we now focus on some key differences, and why these contributions are conceptually significant.

1. The analysis of [KK19] is intricate and it is not a priori clear what problems it may extend to. Our choice of Max k-EQ(q) is not the obvious choice, and was not our first choice. More natural choices would be to go for more general linear systems, or even functions supporting "one-wise independence", but we are unable to push the analysis to more general cases. Our choice reflects an adequate one to get coarse bounds on the approximability of every problem while getting tight ones for many natural ones.

2. The choice of the communication problems to work with is also not obvious: Indeed working with both IRMD and IFRMD seems necessary for our approach — the former is more useful for our final inapproximability results whereas the latter is the one we are able to analyze.

3. The exact notion of boundedness that is necessary and sufficient for our results is also not completely obvious. It is only in hindsight, after carrying out the entire analysis, does it become clear that the notion that works is exactly the same as the one in [KK19]. Part of the challenge is that in the inductive proof of boundedness even the base case (which is quite simple in [KK19]) is not obvious in our case, and nor is the inductive step.

    - With respect to the base case we note that if we had adopted a weaker notion of boundedness allowing wth level Fourier mass to grow roughly as $U((k-1)w/k)$ boundedness would have been easier to prove but the result would not be optimal. Getting a bound of $U(w/2)$ is not technically hard, but involves a non-trivial randomization in the choice of folding purely for analysis purposes. (So there is an implicit passing back and forth between the IRMD and IFRMD problems in this technical step.)
    - We also feel that it is important that we are able to extract an induction lemma (Lemma 5.18) that clearly separates the (Fourier and combinatorial) analytic ingredients from the probabilistic setup. We believe the lemma is clarifying even when applied to the proof of [KK19].

4. Finally we note that the underlying combinatorics are made significantly more intricate due to the need to work with $k > 2$. A conceptual difference from [KK19] here is that whereas they explore the distribution of the number of edges in a random matching that intersect with a fixed set of vertices, we have to explore the distribution of edges that have an odd intersection (or non-zero mod q intersection) with a random hypermatching. Indeed this part is clarifying the role of some of the quantities explored in the previous work. Additionally, we note that the number of parameters we have to track is much larger (and indeed it is fortunate that the number of parameters remains a constant independent of k), and managing these in our inequalities is a non-trivial technical challenge (even given the heavy lifting in [KK19]).

Organization of the rest of the paper. We start with some background material in Section 2. We introduce our communication problems (IRMD and IFRMD) in Section 3 and state our lower bounds for these. We use these lower bounds to prove our streaming lower bounds in Section 4. Section 5 introduces the notion of bounded sets and proves our lower bound on the communication problems modulo some lemmas on the boundedness of sets encountered by the protocol. Section 6 proves these lemmas on boundedness concluding the proofs.

2 Preliminaries

We use the following notations throughout the paper. Let $\mathbb{N} = \{1, \dots\}$ denote the set of natural numbers and let $[n] = \{1, 2, \dots, n\}$. For a discrete set X and a function $f : X \to \mathbb{R}$, we denote $\|f\|_p = (\sum_{x \in X} |f(x)|^p)^{1/p}$ for every $p > 0$ and $\|f\|_0 = \sum_{x \in X} \mathbf{1}_{f(x) \neq 0}$.

2.1 Total variation distance

In our analysis we will use the total variation distance between probability distributions, and several bounds on it presented in this section.

Definition 2.1 (Total variation distance of discrete random variables)**.** *Let Ω be a finite probability space and X, Y be random variables with support Ω. The total variation distance between X and Y is defined as follows.*

$$\|X - Y\|_{tvd} := \frac{1}{2} \sum_{\omega \in \Omega} |\Pr[X = \omega] - \Pr[Y = \omega]| \,.$$

We will use the triangle and data processing inequalities for the total variation distance.

Proposition 2.2 (E.g.,[KKS15, Claim 6.5])**.** *For random variables X, Y and W:*

- *(Triangle inequality) $\|X - Y\|_{tvd} \geq \|X - W\|_{tvd} - \|Y - W\|_{tvd}$.*
- *(Data processing inequality) If W is independent of both X and Y, and f is a function, then $\|f(X, W) - f(Y, W)\|_{tvd} \leq \|X - Y\|_{tvd}$.*

Lemma 2.3. *Let X, Y, W be random variables and let f be a function. If there exists $\delta > 0$ such that for every fixed x in the support of X, we have*

$$\|f(x, Y) - f(x, W)\|_{tvd} \leq \delta \,,$$

then the following holds:

$$\|(X, f(X, Y)) - (X, f(X, W))\|_{tvd} \leq \delta \,.$$

Proof. Consider any statistical test T distinguishing the joint distributions $(X, f(X, Y))$ and $(X, f(X, W))$. It suffices to prove that

$$\mathbb{E}_{X,Y}[T(X, f(X, Y))] - \mathbb{E}_{X,W}[T((X, f(X, W)))] \leq \delta \,.$$

We have

$$\begin{aligned}
&\mathbb{E}_{X,Y}[T(X, f(X, Y))] - \mathbb{E}_{X,W}[T((X, f(X, W)))] \\
&= \mathbb{E}_{x \sim X}\left[\mathbb{E}_{y \sim Y|X=x}[T(x, f(x, y))]\right] - \mathbb{E}_{x \sim X}\left[\mathbb{E}_{w \sim W|X=x}[T(x, f(x, w))]\right] \\
&= \mathbb{E}_{x \sim X}\left[\mathbb{E}_{y \sim Y|X=x}[T(x, f(x, y))] - \mathbb{E}_{w \sim W|X=x}[T(x, f(x, w))]\right] \\
&\leq \mathbb{E}_{x \sim X}[\delta] = \delta \,,
\end{aligned}$$

where the last step follows from the hypothesis that for every fixed x, we have

$$\|f(x, Y) - f(x, W)\|_{tvd} \leq \delta \,.$$

□

We will also need the following lemma from [KK19].

Lemma 2.4 ([KK19] Lemma B.2)**.** *Let X^1, X^2 be random variables taking values on finite sample space Ω_1. Let Z^1, Z^2 be random variables taking values on sample space Ω_2, and suppose that Z^2 is independent of X^1, X^2. Let $f : \Omega_1 \times \Omega_2 \to \Omega_3$ be a function. Then*

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

2.2 Concentration inequality

We will use the following concentration inequality from [KK19] which is essentially an Azuma-Hoeffding style inequality for submartingales.

Lemma 2.5 ([KK19, Lemma 2.5]). *Let $X = \sum_{i\in[N]} X_i$ where X_i are Bernoulli random variables such that for every $k \in [N]$, $\mathbb{E}[X_k \mid X_1, \ldots, X_{k-1}] \leq p$ for some $p \in (0,1)$. Let $\mu = Np$. For every $\Delta > 0$, we have:*

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

2.3 Fourier analysis

In this paper, we will use Fourier analysis over $\mathbb{Z}_q$ (see, for instance, [O'D14, GT19]). For a function $f : \mathbb{Z}_q^n \to \mathbb{C}$, its Fourier coefficients are defined by $\widehat{f}(\mathbf{u}) = \frac{1}{q^n}\sum_{\mathbf{a}\in\mathbb{Z}_q^n} f(\mathbf{a}) \cdot \overline{\omega^{\mathbf{u}^\top \mathbf{a}}}$, where $\mathbf{u} \in \mathbb{Z}_q^n$ and $\omega = e^{2\pi i/q}$ is the primitive q-th root of unity. In particular, for every $\mathbf{a}$, $f(\mathbf{a}) = \sum_{\mathbf{u}\in\mathbb{Z}_q^n} \widehat{f}(\mathbf{u}) \cdot \omega^{\mathbf{u}'^\top \mathbf{a}}$. Later we will use the three following important tools. Note that here we define the p-norm of f as $\|f\|_p^p = \sum_{\mathbf{x}\in\mathbb{Z}_q^n} |f(\mathbf{x})|^p$ rather than the standard definition which uses expectation. This is for future notational convenience.

Lemma 2.6 (Parseval's identity). *For every function $f : \mathbb{Z}_q^n \to \mathbb{C}$,*

$$\|f\|_2^2 = \sum_{\mathbf{a}\in\mathbb{Z}_q^n} f(\mathbf{a})^2 = q^n \sum_{\mathbf{u}\in\mathbb{Z}_q^k} \widehat{f}(\mathbf{u})^2.$$

Note that for every distribution f on $\mathbb{Z}_q^n$, $\widehat{f}(0^n) = q^{-n}$. For the uniform distribution U on $\mathbb{Z}_q^n$, $\widehat{U}(\mathbf{u}) = 0$ for every $\mathbf{u} \neq 0^n$. Thus, by Lemma 2.6, for any distribution f on $\mathbb{Z}_q^n$:

$$\|f - U\|_2^2 = q^n \sum_{\mathbf{u}\in\mathbb{Z}_q^n} \left(\widehat{f}(\mathbf{u}) - \widehat{U}(\mathbf{u})\right)^2 = q^n \sum_{\mathbf{u}\in\mathbb{Z}_q^n\setminus\{0^n\}} \widehat{f}(\mathbf{u})^2. \tag{2.7}$$

We now introduce some standard facts about how convolutions interact with the Fourier transform operation. For functions $f, g\colon \mathbb{Z}_q^n \to \mathbb{C}$, their convolution $f \star g\colon \mathbb{Z}_q^n \to \mathbb{C}$ is defined as $(f \star g)(\mathbf{a}) = \sum_{\mathbf{v}\in\mathbb{Z}_q^n} f(\mathbf{v})g(\mathbf{a} - \mathbf{v})$. The first lemma is the so-called "convolution theorem," which essentially states that, up to normalization factors, the Fourier transform of the convolution of two functions is equal to the product of the individual Fourier transforms.

Lemma 2.8 (Convolution Theorem). *For $f, g : \mathbb{Z}_q^n \to \mathbb{C}$, we have*

$$\widehat{f \star g}(\mathbf{u}) = q^n \cdot \widehat{f}(\mathbf{u}) \cdot \widehat{g}(\mathbf{u}).$$

for all $\mathbf{u} \in \mathbb{Z}_q^n$.

Proof. Note that

$$\begin{aligned}
\widehat{f \star g}(\mathbf{u}) &= \frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} (f \star g)(\mathbf{a}) \cdot \overline{\omega^{\mathbf{u}^\top \mathbf{a}}} \\
&= \frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} \left(\sum_{\mathbf{v} \in \mathbb{Z}_q^n} f(\mathbf{v}) g(\mathbf{a} - \mathbf{v}) \right) \overline{\omega^{\mathbf{u}^\top \mathbf{a}}} \\
&= \frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} \sum_{\mathbf{v} \in \mathbb{Z}_q^n} f(\mathbf{v}) \overline{\omega^{\mathbf{u}^\top \mathbf{v}}} \cdot g(\mathbf{a} - \mathbf{v}) \overline{\omega^{\mathbf{u}^\top (\mathbf{a} - \mathbf{v})}} \\
&= \frac{1}{q^n} \sum_{\mathbf{v} \in \mathbb{Z}_q^n} f(\mathbf{v}) \overline{\omega^{\mathbf{u}^\top \mathbf{v}}} \cdot \sum_{\mathbf{a} \in \mathbb{Z}_q^n} g(\mathbf{a} - \mathbf{v}) \overline{\omega^{\mathbf{u}^\top (\mathbf{a} - \mathbf{v})}} \\
&= q^n \cdot \frac{1}{q^n} \sum_{\mathbf{v} \in \mathbb{Z}_q^n} f(\mathbf{v}) \overline{\omega^{\mathbf{u}^\top \mathbf{v}}} \cdot \frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} g(\mathbf{a}) \overline{\omega^{\mathbf{u}^\top \mathbf{a}}} \\
&= q^n \cdot \widehat{f}(\mathbf{u}) \cdot \widehat{g}(\mathbf{u}),
\end{aligned}$$

as desired. □

We will also need the following lemma, which states that the Fourier transform of the *product* of two functions is given by the convolution of the individual Fourier transforms.

Lemma 2.9 (Fourier transform of product of functions)**.** *For every $f, g : \mathbb{Z}_q^n \to \mathbb{C}$, and $\mathbf{u} \in \mathbb{Z}_q^n$, we have*

$$\widehat{f \cdot g}(\mathbf{u}) = \sum_{\mathbf{u}' \in \mathbb{Z}_q^n} \widehat{f}(\mathbf{u}') \cdot \widehat{g}(\mathbf{u} - \mathbf{u}') \,.$$

Furthermore, for every $h \in [n]$,

$$\sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} \widehat{f \cdot g}(\mathbf{u}) = \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u} + \mathbf{u}'\|_0 = h}} \widehat{f}(\mathbf{u}) \cdot \widehat{g}(\mathbf{u}') \,.$$

Proof. For every $\mathbf{u} \in \mathbb{Z}_q^n$, we have

$$\begin{aligned}
\widehat{f \cdot g}(\mathbf{u}) &= \frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} f(\mathbf{a}) \cdot g(\mathbf{a}) \cdot \overline{\omega^{\mathbf{u}^\top \mathbf{a}}} \\
&= \frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} \left(\sum_{\mathbf{u}' \in \mathbb{Z}_q^n} \widehat{f}(\mathbf{u}') \cdot \omega^{\mathbf{u}'^\top \mathbf{a}} \right) \cdot g(\mathbf{a}) \cdot \overline{\omega^{\mathbf{u}^\top \mathbf{a}}} \\
&= \sum_{\mathbf{u}' \in \mathbb{Z}_q^n} \widehat{f}(\mathbf{u}') \cdot \left(\frac{1}{q^n} \sum_{\mathbf{a} \in \mathbb{Z}_q^n} g(\mathbf{a}) \cdot \overline{\omega^{(\mathbf{u} - \mathbf{u}')^\top \mathbf{a}}} \right) \\
&= \sum_{\mathbf{u}' \in \mathbb{Z}_q^n} \widehat{f}(\mathbf{u}') \cdot \widehat{g}(\mathbf{u} - \mathbf{u}') \,.
\end{aligned}$$

Next, for every $h \in [n]$,

$$\sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} \widehat{f \cdot g}(\mathbf{u}) = \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} \sum_{\mathbf{u}' \in \mathbb{Z}_q^n} \widehat{f}(\mathbf{u}') \cdot \widehat{g}(\mathbf{u} - \mathbf{u}')$$

Letting $\mathbf{w} = \mathbf{u} - \mathbf{u}'$ and switching the order of the summations, the equation becomes

$$= \sum_{\mathbf{u}' \in \mathbb{Z}_q^n} \sum_{\substack{\mathbf{w} \in \mathbb{Z}_q^n \\ \|\mathbf{w}+\mathbf{u}'\|_0 = h}} \widehat{f}(\mathbf{u}') \cdot \widehat{g}(\mathbf{w}) \,,$$

which, after renaming variables, proves the furthermore part of the lemma. □

The hypercontractivity theorem states that the 2-norm of a function after the application of a noise operator can be nicely upper bounded.

Lemma 2.10 (Hypercontractivity Theorem [O'D14, Page 278])**.** *Let $f : \mathbb{Z}_q^n \to \mathbb{C}$ be a square-integrable function and let $1 < p < 2$, $0 \le \rho \le \frac{1}{\sqrt{p-1}}(1/q)^{1/2-1/p}$, we have*

$$\|T_\rho f\|_2 \le \|f\|_p \,,$$

where T_ρ is the noise operator defined by $T_\rho f(\mathbf{x}) = \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \widehat{f}(\mathbf{u}) \rho^{\|\mathbf{u}\|_0} \omega^{\mathbf{u}^\top \mathbf{x}}$.

Next, we prove the following consequence of the hypercontractivity theorem.

Lemma 2.11. *There exists $\zeta > 0$ such that for every q, every $f : \mathbb{Z}_q^n \to \{a \in \mathbb{C} \,|\, |a| \le 1\}$ and $B = \{\mathbf{a} \in \mathbb{Z}_q^n \,|\, f(\mathbf{a}) \ne 0\}$ the following holds: If $|B| \ge q^{n-b}$ for some $b \in \mathbb{N}$, then for every $\mathbf{v} \in \mathbb{Z}_q^n$ and every $h \in \{1, \ldots, 4b\}$, we have*

$$\frac{q^{2n}}{|B|^2} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{v}\|_0 = h}} |\widehat{f}(\mathbf{u})|^2 \le \left(\frac{\zeta \cdot b}{h}\right)^h .$$

Proof. First, let us consider $\mathbf{v} = 0^n$ and $f : \mathbb{Z}_q^n \to \{a \in \mathbb{C} \,|\, |a| \le 1\}$. Let us pick $p = 1 + \frac{h}{\zeta b}$ and $\rho = \frac{1}{\sqrt{p-1}}(1/q)^{1/2-1/p}$, where $\zeta > 4$ is a constant to be specified later. Assume $|B| \ge q^{n-b}$.

The choices of p and ρ satisfy the preconditions of Lemma 2.10, and so applying Lemma 2.10 we have

$$\sum_{\mathbf{u} \in \mathbb{Z}_q^n} \rho^{2\|\mathbf{u}\|_0} |\widehat{f}(\mathbf{u})|^2 = \|T_\rho f\|_2^2 \le \|f\|_p^2 = \left(\frac{1}{q^n} \sum_{\mathbf{x} \in \mathbb{Z}_q^n} |f(\mathbf{x})|^p\right)^{2/p} \le \left(\frac{|B|}{q^n}\right)^{2/p} .$$

Now, suppose $h \in \{1, \ldots, 4b\}$. Noting that $\rho^{2h} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} |\widehat{f}(\mathbf{u})|^2 \le \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \rho^{2\|\mathbf{u}\|_0} |\widehat{f}(\mathbf{u})|^2$, we have

$$\begin{aligned}
\frac{q^{2n}}{|B|^2} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} |\widehat{f}(\mathbf{u})|^2 &\le \frac{1}{\rho^{2h}} \left(\frac{q^n}{|B|}\right)^{2-2/p} \\
&\le \frac{1}{\rho^{2h}} q^{(2-2/p)b} \\
&= \frac{q^{\left(1 + \frac{2b}{h} - \frac{2}{p} - \frac{2b}{hp}\right)h}}{(p-1)^h} \\
&= \left(\frac{\zeta b}{h} \cdot q^{1 + \frac{2b}{h} - \frac{2}{p} - \frac{2b}{hp}}\right)^h , \qquad (2.12)
\end{aligned}$$

where the first equality above is by our choice of ρ and the second by our choice of p. Observe that the exponent of q in the final expression above can be bounded as follows:

$$\begin{aligned}
1+\frac{2b}{h}-\frac{2}{p}-\frac{2b}{ph} &= 1+\frac{2b}{h}-\frac{2\left(1+\frac{b}{h}\right)}{1+\frac{h}{\zeta b}} \\
&= \left(1+\frac{h}{\zeta b}\right)^{-1}\left(\frac{2}{\zeta}+\frac{h}{\zeta b}-1\right) \\
&\leq \left(1+\frac{h}{\zeta b}\right)^{-1}\left(\frac{6}{\zeta}-1\right). \qquad (2.13)
\end{aligned}$$

Note that if we set $\zeta = 6$, then it follows that (2.13) is ≤ 0. For this choice of ζ, (2.12) can be bounded from above by $\left(\frac{\zeta b}{h}\right)^h$, implying that

$$\frac{q^{2n}}{|B|^2}\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0=h}} |\widehat{f}(\mathbf{u})|^2 \leq \left(\frac{\zeta b}{h}\right)^h.$$

In order to extend the above to sums over translational shifts, i.e., $\mathbf{u}$ such that $\|\mathbf{u}+\mathbf{v}\|_0 = h$ for an arbitrary $\mathbf{v}\in\mathbb{Z}_q^n$, consider the function $g(\mathbf{x}) = f(\mathbf{x})\cdot\omega^{\mathbf{x}^\top\mathbf{v}}$. We have for every $\mathbf{x}\in\mathbb{Z}_q^n$,

$$\widehat{g}(\mathbf{u}) = q^{-n}\sum_{\mathbf{a}\in\mathbb{Z}_q^n} g(\mathbf{a})\overline{\omega^{\mathbf{a}^\top\mathbf{u}}} = q^{-n}\sum_{\mathbf{a}\in\mathbb{Z}_q^n} f(\mathbf{a})\overline{\omega^{\mathbf{a}^\top(\mathbf{u}-\mathbf{v})}} = \widehat{f}(\mathbf{u}-\mathbf{v}).$$

By applying the above analysis on g, we have

$$\frac{q^{2n}}{|B|^2}\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{v}\|_0=h}} |\widehat{f}(\mathbf{u})|^2 = \frac{q^{2n}}{|B|^2}\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0=h}} |\widehat{g}(\mathbf{u})|^2 \leq \left(\frac{\zeta b}{h}\right)^h,$$

as desired. □

3 Communication Problems

Throughout this paper, we will be dealing with k-hypermatchings on vertices from the set $[n]$, i.e., a set of edges $e_1,\ldots,e_m$ where $e_i\subseteq[n]$, $|e_i| = k$ and $e_i\cap e_j = \emptyset$ for every $i\neq j\in[m]$. We let $e_i = \{(e_i)_1,\ldots,(e_i)_k\}$. The direct encoding of a matching $M = \{e_1,\ldots,e_m\}$ will be given by a *hypermatching matrix* $A\in\{0,1\}^{km\times n}$ where $A_{k(i-1)+\ell,j} = 1$ if and only if $j = (e_i)_\ell$. (Thus, A is a matrix with row sums being 1 and column sums being at most 1. Note that A also depends on the ordering of $e_1,e_2,\ldots,e_m$ as well as the ordering of the nodes within each e_i.)

We will also find it convenient to refer to edges by their indicator vectors in $\mathbb{Z}_q^n$. For an edge e_i, we will use the boldface notation $\mathbf{e}_i\in\mathbb{Z}_q^n$ to refer to this vector, i.e., $(\mathbf{e}_i)_j = 1$ if $j = (e_i)_\ell$ for some $\ell\in[k]$, while $(\mathbf{e}_i)_j = 0$ otherwise.

We are now ready to define the communication game, which we term the Implicit Randomized Mask Detection (IRMD) problem:

Definition 3.1 (Implicit Randomized Mask Detection (IRMD) Problem)**.** *Let $q,k,n,T\in\mathbb{N}$ and $\alpha\in(0,1/k)$ be parameters. Let $\mathcal{D}_Y$ and $\mathcal{D}_N$ be distributions over $\mathbb{Z}_q^k$. In the $(\mathcal{D}_Y,\mathcal{D}_N)$-IRMD$_{\alpha,T}$ game, there are T players and a hidden q-coloring encoded by a random $\mathbf{x}^*\in\mathbb{Z}_q^n$. The t-th player*

has two inputs: (a.) $A_t \in \{0,1\}^{\alpha k n \times n}$, the hypermatching matrix (see above) corresponding to a random hypermatching M_t of size αn and (b.) a vector $\mathbf{z}_t \in \mathbb{Z}_q^{\alpha k n}$ that can be generated from one of two different distributions:

- *(Yes) $\mathbf{z}_t = A_t \mathbf{x}^* + \mathbf{b}_t$ where $\mathbf{b}_t \in \mathbb{Z}_q^{\alpha k n}$ is of the form $\mathbf{b}_t = (\mathbf{b}_{t,1}, \ldots, \mathbf{b}_{t,\alpha n})$ and each $\mathbf{b}_{t,i} \in \mathbb{Z}_q^k$ is sampled from $\mathcal{D}_Y$.*
- *(No) $\mathbf{z}_t = A_t \mathbf{x}^* + \mathbf{b}_t$ where $\mathbf{b}_t \in \mathbb{Z}_q^{\alpha k n}$ is of the form $\mathbf{b}_t = (\mathbf{b}_{t,1}, \ldots, \mathbf{b}_{t,\alpha n})$ and each $\mathbf{b}_{t,i} \in \mathbb{Z}_q^k$ is sampled from $\mathcal{D}_N$.*

This is a one-way game where the t-th player can send a private message to the $(t+1)$-th player after receiving a message from the previous player. The goal is for the T-th player to be able to decide whether the $\{\mathbf{z}_t\}$ have been chosen from the "Yes" distribution or "No" distribution. The advantage of a protocol (in which the T-th player outputs either "Yes" or "No") is defined as $|\Pr_{\mathcal{D}_Y}[\text{the } T\text{-th player outputs Yes}] - \Pr_{\mathcal{D}_N}[\text{the } T\text{-th player outputs Yes}]|$.

Remark. *We remark that the inputs to the T players in the IRMD problem can be viewed as a stream $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{(1)} \circ \cdots \circ \boldsymbol{\sigma}^{(T)}$, where the t-th player's input $(A_t, \mathbf{z}_t)$ is converted to a stream $\boldsymbol{\sigma}^{(t)} = (\sigma^{(t)}(i) | i \in [\alpha n])$ where the elements of the stream are of the form $\sigma^{(t)}(i) = (\mathbf{j}^{(t)}(i), \mathbf{z}^{(t)}(i))$ with $\mathbf{j}^{(t)}(i) \in [n]^k$ is a sequence of k distinct elements of $[n]$ and $\mathbf{z}^{(t)}(i) \in \mathbb{Z}_q^k$. This "streaming" representation will be used when we relate the complexity of IRMD to the approximability of various Max-CSP($\mathcal{F}$) problems in Theorem 4.3.*

We suppress the subscripts α and T when they are clear from context. Furthermore, we simply use IRMD to refer to $(\mathcal{D}_Y, \mathcal{D}_N)$-IRMD with $\mathcal{D}_Y$ being the uniform distribution over $\{0^k, 1^k, \ldots, (q-1)^k\}$ and $\mathcal{D}_N$ being the uniform distribution over $\mathbb{Z}_q^k$. The following theorem shows that in this special case, the IRMD problem requires linear communication. We remark that the theorem could hold for other pairs of distributions and leave the question of when such a lower bound holds as an interesting open problem.

Theorem 3.2 (Linear lower bound for IRMD)**.** *For every $q, k \in \mathbb{N}$ and $\delta \in (0, 1/2)$, $\alpha \in (0, 1/k)$, $T \in \mathbb{N}$ there exists $n_0 \in \mathbb{N}$ and $\tau \in (0,1)$ such that the following holds. If $\mathcal{D}_Y, \mathcal{D}_N$ are the uniform distributions over $\{0^k, 1^k, \ldots, (q-1)^k\}$ and $\mathbb{Z}_q^k$ respectively and $n \geq n_0$ then every protocol for $(\mathcal{D}_Y, \mathcal{D}_N)$-IRMD$_{\alpha,T}$ with advantage δ requires τn bits of communication.*

We prove the hardness of IRMD by first proving the hardness of a *folded* version of IRMD. In the folded version of the communication problem, we augment each hyperedge with an associated *center* $c \in e$. Given a k-hypermatching $M = (e_1, \ldots, e_m)$ and a sequence of centers $\mathbf{c} = (c_1, \ldots, c_m)$ with $e_i = ((e_i)_1, \ldots, (e_i)_k = c_i)$, the $\mathbf{c}$-centered folded encoding of M is the matrix $A_{\mathbf{c}} \in \mathbb{Z}_q^{(k-1)m \times n}$ given by

$$(A_{\mathbf{c}})_{(k-1)(i-1)+\ell, j} = \begin{cases} 1 & \text{, if } j \in \{(e_i)_\ell\} \text{ and } \ell \in [k-1] \\ -1 & \text{, if } j = c_i \text{ and } \ell \in [k-1] \\ 0 & \text{, otherwise} \end{cases}$$

See Figure 2 for an example. We define the folded version of the IRMD problem below (note that all the arithmetic is over $\mathbb{Z}_q$):

Definition 3.3 (Implicit Folded Randomized Mask Detection (IFRMD) Problem)**.** *Let $q, k, n, T \in \mathbb{N}$ and $\alpha \in (0, 1/k)$ be parameters. In the IFRMD game, there are T players and a hidden q-coloring encoded by a random $\mathbf{x}^* \in \mathbb{Z}_q^n$. The t-th player has a pair of inputs $(A_{t,\mathbf{c}_t}, \mathbf{w}_t)$ given as follows. $A_{t,\mathbf{c}_t} \in \mathbb{Z}_q^{\alpha(k-1)n \times n}$ gives a $\mathbf{c}_t$-centered folded encoding of a random hypermatching M_t of size αn, and $\mathbf{w}_t \in \mathbb{Z}_q^{\alpha(k-1)n}$ is a vector that can be generated from two different distributions:*

- *(**YES**)* $\mathbf{w}_t = A_{t,\mathbf{c}_t}\mathbf{x}^*$.
- *(**NO**)* $\mathbf{w}_t$ *is uniform over* $\mathbb{Z}_q^{\alpha(k-1)n}$.

*This is a one-way game where the t-th player can send a private message to the $(t+1)$-th player after receiving message from the previous player. The goal is to decide (by the T-th player) whether the $\{\mathbf{w}_t\}$ are coming from the **YES** distribution or the **NO** distribution. The advantage of a protocol is defined as*

$$\left| \Pr_{(A_{t,c_t},\mathbf{w}_t)_{t\in T}\sim \boldsymbol{YES}}[\textit{the } T\textit{-th player outputs } \textit{Yes}] - \Pr_{(A_{t,c_t},\mathbf{w}_t)_{t\in T}\sim \boldsymbol{NO}}[\textit{the } T\textit{-th player outputs } \textit{Yes}] \right|.$$

The main technical theorem of this paper is the following $\Omega(n)$ communication lower bound for IFRMD.

Theorem 3.4 (Linear lower bound for IFRMD)**.** *For every $q, k \in \mathbb{N}$ and $\delta \in (0, 1/2)$, there exists $\alpha_0 \in (0, 1/k)$ such that for every $\alpha \in (0, \alpha_0]$ and every $T \in \mathbb{N}$ and every $\delta \in (0, 1)$, there exists $\tau \in (0, 1)$ such that the following holds. When $n \in \mathbb{N}$ is large enough, any protocol for IFRMD with advantage δ requires τn bits of communication.*

The proof of Theorem 3.4 is given in the beginning of Section 5.

We now prove a lemma establishing a reduction from IFRMD to IRMD that preserves the communication complexity. Note that by this lemma, Theorem 3.2 will be an immediate corollary of Theorem 3.4.

Lemma 3.5. *Let n, k, α be the parameters. Suppose there exists a protocol for IRMD using at most s bits communication with advantage δ, then there exists a protocol for IFRMD using at most s bits communication with advantage δ.*

Proof. Suppose we have an instance of IFRMD with input $(A_{t,\mathbf{c}_t}, \mathbf{w}_t)$ to the t-th player. We show how to transform this into an instance of IRMD. For each t, the t-th player performs the following computations on his/her input:

1. Use $A_{t,\mathbf{c}_t}$ to compute the underlying hypermatching M_t (by identifying the set of nonzero columns for each block of $k-1$ rows of $A_{t,\mathbf{c}_t}$) and compute the corresponding matrix Π_t.
2. For each $i \in [\alpha n]$, sample $a_{t,i} \in \mathbb{Z}_q$ uniformly at random. Let $z_t \in \mathbb{Z}_q^{\alpha k n}$ be defined by $(z_t)_{(i-1)k+j} = (w_t)_{(i-1)k+j} + a_{t,i}$ for each $j = 1, 2, \ldots, k-1$ and $z_{t,ik} = a_{t,i}$.

We claim that the inputs $(A_t, \mathbf{z}_t)$ correspond to an instance of IRMD. It suffices to show that if $(\{(A_{t,\mathbf{c}_t}, \mathbf{w}_t)\}_{t\in[T]}, \mathbf{x}^*)$ follows the **YES** (resp. **NO**) distribution of IFRMD, then $(\{(A_t, \mathbf{z}_t)\}_{t\in[T]}, \mathbf{x}^*)$ follows the **YES** (resp. **NO**) distribution of IRMD.

Let $m = \alpha n$. For each t, let $e_1^{(t)}, e_2^{(t)}, \ldots, e_m^{(t)}$ be the hyperedges corresponding to $A_{t,\mathbf{c}_t}$ (in order), with $(e_i^{(t)})_k = c_{t,i}$.

We first focus on the **YES** case. Then, note that for $j = 1, 2, \ldots, k-1$, we have

$$(z_t)_{(i-1)k+j} = (w_t)_{(i-1)k+j} + a_{t,i} = (x^*_{(e_i^{(t)})_j} + x^*_{c_{t,i}}) + a_{t,i} = x^*_{(e_i^{(t)})_j} + (x^*_{c_{t,i}} + a_{t,i}).$$

Moreover,

$$(z_t)_{ik} = a_{t,i} = x^*_{c_{t,i}} + (x^*_{c_{t,i}} + a_{t,i}) = x^*_{(e_i^{(t)})_k} + (x^*_{c_{t,i}} + a_{t,i}).$$

Thus, it follows that $\mathbf{z}_t = \Pi_t \mathbf{x}^* + \mathbf{b}_t$, where $\mathbf{b}_t = (\mathbf{b}_{t,1}, \ldots, \mathbf{b}_{t,\alpha n})$ is given by $\mathbf{b}_{t,i} = (x^*_{c_{t,i}} + a_{t,i}) \cdot \mathbf{1}_k$ where $\mathbf{1}_k$ is the all 1 vector of length k. Since $a_{t,i}$ is uniform over $\mathbb{Z}_q$, this takes care of the **YES** case.

The **NO** case is easier to see: Π_t encodes a random k-hypermatching of size αn and $\mathbf{z}_t$ is uniform over $\mathbb{Z}_q^{\alpha k n}$. □

Proof of Theorem 3.2 using Theorem 3.4. For the sake of contradiction, suppose there exists a protocol for IRMD with advantage δ using fewer than τn bits of communication. Then by Lemma 3.5 there exists a protocol for IFRMD with advantage δ using fewer than τn bits of communication, which contradicts Theorem 3.4. This completes the proof of Theorem 3.2. □

In the following section we show how Theorem 3.2 yields the claimed hardness of streaming problems. In the rest of this paper, we focus on the proof of Theorem 3.4, i.e., the linear communication lower bound for IFRMD.

4 Streaming Problems and Hardness

In this section we state and prove our main technical theorem establishing linear space lower bounds for the approximability of many CSPs.

Below we define the two crucial constants associated with a family $\mathcal{F}$ which lay out the "trivial" approximability, and the inapproximability that we prove. In particular we define the notion of a width $\omega(\mathcal{F}) \in [1/q, 1]$ for every family $\mathcal{F}$. The notion of a wide family from Theorem 1.1 corresponds to a family with maximum width, i.e., $\omega(\mathcal{F}) = 1$.

Definition 4.1 (Minimum value, Width of $\mathcal{F}$)**.** *For a family $\mathcal{F}$, we define its minimum value $\rho_{\min}(\mathcal{F})$ to be the infimum over all instances Ψ of Max-CSP($\mathcal{F}$) of val_Ψ. For $\mathbf{b} \in \mathbb{Z}_q^k$ and $f : \mathbb{Z}_q^k \to \{0,1\}$ we define $\mathbf{b}$-width of f, denoted $\omega_{\mathbf{b}}(f)$ to be the quantity $\frac{|\{a \in \mathbb{Z}_q \mid f(\mathbf{b}+a^k)=1\}|}{q}$. The* width *of f, denoted $\omega(f)$, is given by $\omega(f) = \max_{\mathbf{b} \in \mathbb{Z}_q^k}\{\omega_{\mathbf{b}}(f)\}$. Finally for a family $\mathcal{F}$, we define its* width *to be $\omega(\mathcal{F}) = \min_{f \in \mathcal{F}}\{\omega(f)\}$. We say that a family $\mathcal{F}$ is* wide *if $\omega(\mathcal{F}) = 1$.*

As described above $\rho_{\min}(\mathcal{F})$ may not even be computable given $\mathcal{F}$, but as pointed out in [**?**] it is a computable function. Key to this assertion is the following equivalent definition of $\rho_{\min}(\mathcal{F})$ which follows from Definition 2.4 and Proposition 2.5 of [**?**].

Proposition 4.2 ([**?**, Proposition 2.4])**.** *For every $k, q, \mathcal{F} \subseteq \{f : \mathbb{Z}_q^k \to \{0,1\}\}$ we have*

$$\rho_{\min}(\mathcal{F}) = \rho(\mathcal{F}) := \min_{\mathcal{D}_{\mathcal{F}} \in \Delta(\mathcal{F})} \left\{ \max_{\mathcal{D} \in \Delta([q])} \left\{ \mathop{\mathbb{E}}_{f \sim \mathcal{D}_{\mathcal{F}}, \mathbf{a} \sim \mathcal{D}^k} [f(\mathbf{a})] \right\} \right\}.$$

We are now ready to prove the main theorem of the paper on the approximability of CSPs by applying Theorem 3.2.

Theorem 4.3 (Linear Space Inapproximability of CSPs)**.** *For every $k, q, \mathcal{F} \subseteq \{f : \mathbb{Z}_q^k \to \{0,1\}\}$ and every $\varepsilon \in (0, 1/10)$ we have the following: Every randomized single-pass streaming $(1+\varepsilon)\cdot\frac{\rho(\mathcal{F})}{\omega(\mathcal{F})}$-approximation algorithm for Max-CSP($\mathcal{F}$) requires $\Omega(n)$ space.*

Proof. Given $\mathcal{F}$ and $\varepsilon \in (0, 1/10)$, we let $\alpha = \varepsilon/(100k^2q)$ and T be some large enough constant that only depends on $q, k, \mathcal{F}, \varepsilon, \alpha$. Let **ALG** be a space s algorithm distinguishing instances from the set $\{\Psi \mid \mathsf{val}_\Psi \geq (1-\varepsilon/3)\omega(\mathcal{F})\}$ from instances from the set $\{\Psi \mid \mathsf{val}_\Psi \leq (1+\varepsilon/3)\rho(\mathcal{F})\}$ with success

probability at least 2/3. We show how to use **ALG** to device an s-bit communication protocol for $\mathsf{IRMD} = \mathsf{IRMD}_{\alpha,T}$ with advantage at least 1/6.

For $f \in \mathcal{F}$, let $\mathbf{b}_f \in \mathbb{Z}_q^k$ be a sequence maximizing $\omega_{\mathbf{b}_f}(f)$ and let $S_f = \{\mathbf{b}_f + a^k \mid a \in \mathbb{Z}_q\}$. Further let $\mathcal{D}_{\mathcal{F}} \in \Delta(\mathcal{F})$ be a distribution achieving the minimum in the equivalent definition of $\rho(\mathcal{F})$ from Proposition 4.2. Let $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_m)$ be an instance of IRMD with T players, so that $m = T\alpha n$ and $\sigma_i = (\mathbf{j}(i), \mathbf{z}(i))$ where $\mathbf{j}(i) \in [n]^k$ is a sequence of k distinct elements of $[n]$ and $\mathbf{z}(i) \in \mathbb{Z}_q^k$. For each σ_i we either generate 0 or 1 constraint of $\mathsf{Max\text{-}CSP}(\mathcal{F})$ as follows: We sample $f(i) \sim \mathcal{D}_{\mathcal{F}}$ and output the constraint $(f(i), \mathbf{j}(i))$ if $\mathbf{z}(i) \in S_{f(i)}$ and output no constraint otherwise. Applying this step independently to each σ_i generates an instance Ψ of $\mathsf{Max\text{-}CSP}(\mathcal{F})$ with $\tilde{m} \leq m$ constraints on n variables. We make the following claims about Ψ.

(1) $\Pr_{\mathbf{YES}}[\tilde{m} > (1+\varepsilon/10) \cdot q^{-(k-1)} \cdot m] = o(1)$ and $\Pr_{\mathbf{NO}}[\tilde{m} < (1-\varepsilon/10) \cdot q^{-(k-1)} \cdot m] = o(1)$, i.e., the number of constraints $\tilde{m}$ does not deviate (in the wrong direction) from its expectation $q^{-(k-1)} \cdot m$ with too high a probability.

(2) If $\boldsymbol{\sigma}$ is generated from the **YES** distribution with hidden vector $\mathbf{x}^*$ then with high probability the number of constraints of Ψ satisfied by $\mathbf{x}^*$ is at least $(\omega(\mathcal{F}) - \varepsilon/10) \cdot q^{-(k-1)} \cdot m$. In particular, $\Pr_{\mathbf{YES}}[\mathsf{val}_\Psi \leq (1-\varepsilon/3) \cdot \omega(\mathcal{F})] = o(1)$.

(3) If $\boldsymbol{\sigma}$ is generated from the **NO** distribution with hidden vector $\mathbf{x}^*$ then with high probability for every $\boldsymbol{\nu}$ the number of constraints of Ψ satisfied by $\boldsymbol{\nu}$ is at most $(\rho(\mathcal{F}) + \varepsilon/10) \cdot q^{-(k-1)} \cdot m$. In particular, $\Pr_{\mathbf{NO}}[\mathsf{val}_\Psi \geq (1+\varepsilon/3) \cdot \rho(\mathcal{F})] = o(1)$.

With the above claims in hand, it is straightforward to convert **ALG** into an $O(s)$-bit communication protocol for IRMD with advantage at least 1/6 — the t-th player gets the state of **ALG** after processing constraints corresponding to the first $t-1$ blocks from the $(t-1)$-th player; generates the constraints corresponding to the t-th block of the stream $\boldsymbol{\sigma}$, and simulates **ALG** on this part of the stream corresponding to Ψ, and passes the resulting state on to the $(t+1)$-th player. The T-th player outputs 1 if **ALG** outputs 1 and 0 otherwise. It is straightforward to see that if **ALG** is correct on every input with probability 2/3 and Claims (1)-(3) above hold, then the resulting communication protocol achieves advantage at least $1/3 - o(1) \geq 1/6$ on IRMD. Finally, we invoke Theorem 3.2 and conclude that $s = \Omega(n)$.

We thus turn to proving claims (1)-(3). Given $\sigma_1, \ldots, \sigma_m$ and $\boldsymbol{\nu} \in \mathbb{Z}_q^n$, we create a collection of related variables as follows: For $i \in [m]$, let $X_i = 1$ if σ_i results in a constraint and 0 otherwise. Further, let $Y_i(\boldsymbol{\nu}) = 1$ if $X_i = 1$ and the resulting constraint is satisfied by the assignment $\boldsymbol{\nu}$. (Note all these are random variables depending on $\boldsymbol{\sigma}$). Below, we bound the expectations of the sums of these random variables in the **YES** and **NO** cases, and also argue that these variables are close to their expectations (or at least give bounds on deviating from the expectation in one direction). This will suffice to prove claims (1)-(3) and thus the theorem.

Proof of Claim (1). We start with $\tilde{m} = \sum_{i=1}^m X_i$ in the **NO** case: In this case $\mathbb{E}[X_i] = |S_f|/q^k = q^{-(k-1)}$ (note that $|S_f| = q$ for every f). Furthermore the X_i's are independent since $\mathbf{z}(i)$'s are uniform and independent of each other. Thus X is sharply concentrated around $q^{-(k-1)} \cdot m$ and we get that $\Pr_{\mathbf{NO}}[\tilde{m} \notin (1 \pm \varepsilon/10) \cdot q^{-(k-1)} \cdot m] = o(1)$.

Turning to the **YES** case, since $\mathbf{z}(i)$'s are no longer independent, the X_i's are correlated. To enable the analysis, we define a vector $\mathbf{x}^*$ to be *γ-good* for $\gamma > 0$ if for every $\tau \in \mathbb{Z}_q$ we have $\Pr_{i \in [n]}[\mathbf{x}^*_i = \tau] \in (1 \pm \gamma)(1/q)$. Note that for every constant $\gamma > 0$, the probability that $\mathbf{x}^*$ is not γ-good is $o(1)$. Fix $\mathbf{x}^*$ that is γ-good. We claim that in this case, $\mathbb{E}[X_i \mid X_{1:i-1}] \leq q^{-(k-1)} \cdot (1 + \gamma + \alpha q k)^k$. To see this note that the effect of conditioning on $X_{1:i-1}$ only affects X_i

due to the fact that now $\mathbf{j}(i)$ is chosen from a smaller set of variables and not all of $[n]$. Let $t \in [T]$ denote the block containing i (i.e., $i \in ((t-1)\alpha n, t\alpha n]$). Let S denote the set of variables that do not participate in the edges $\mathbf{j}((t-1)\alpha n+1), \ldots, \mathbf{j}(i-1)$. Note $|S| \geq (1-k\alpha)n$ and so for every $\tau \in \mathbb{Z}_q$ we have $\Pr_{\ell \in S}[\mathbf{x}^*_\ell = \tau] \leq (1+\gamma+\alpha kq)/q$. We conclude that the probability $\Pr[\mathbf{x}^*|_{\mathbf{j}(i)} \in S_f \,|\, X_{1:i-1}] \leq |S_f| \cdot ((1+\gamma+\alpha kq)/q)^k = q^{-(k-1)} \cdot (1+\gamma+\alpha qk)^k$. Setting $\gamma = \varepsilon/(100k)$ and using $\alpha \leq \varepsilon/(100k^2q)$, we conclude $\mathbb{E}[X_i \,|\, X_{1:i-1}] \leq q^{-(k-1)} \cdot (1+\varepsilon/(50k))^k \leq q^{-(k-1)} \cdot (1+\varepsilon/20)$. Applying Lemma 2.5 we conclude that here again we get that $\Pr_{\mathbf{YES}}[\tilde{m} = \sum_i X_i > (1+\varepsilon/10)q^{-(k-1)}m] = o(1)$.

Proof of Claim (2). Now we analyze the number of satisfiable constraints of the resulting instance Ψ in the **YES** case, where we argue that $\mathbf{x}^*$ satisfies a large fraction of constraints with high probability. Again with probability $1-o(1)$ we have that $\mathbf{x}^*$ is γ-good. Now an argument similar to the one in the analysis of X in the **YES** case shows that for every $\mathbf{b} \in \mathbb{Z}_q^k$, $\Pr[\mathbf{x}^*|_{\mathbf{j}(i)} = \mathbf{b} \,|\, Y_{1:i-1}] \geq (1-\varepsilon/50) \cdot q^{-k}$. Fix $f(i)$ and let $T = S_{f(i)} \cap f(i)^{-1}(1)$. Note by definition of $\omega(\mathcal{F})$ that $|T| \geq \omega(\mathcal{F}) \cdot q$. The event that the i-th constraint is satisfied by $\mathbf{x}^*$ is equivalent to the event that $\mathbf{x}^*_{\mathbf{j}(i)} \in T$ and the probability of this event, conditioned on $Y_{1:i-1}$ is at least $|T| \cdot (1-\varepsilon/50) \cdot q^{-k} \geq (1-\varepsilon/50) \cdot \omega(\mathcal{F}) \cdot q^{-(k-1)}$. Using Lemma 2.5 we conclude again that $\Pr[Y(\mathbf{x}^*) = \sum_{i=1}^m Y_i(\mathbf{x}^*) \leq (1-\varepsilon/10) \cdot \omega(\mathcal{F}) \cdot q^{-(k-1)} \cdot m] = o(1)$. Combining this with the lower bound on $\tilde{m}$ from Claim (1) we conclude that $\Pr[\mathsf{val}_\Psi \leq (1-\varepsilon/3) \cdot \omega(\mathcal{F})] = o(1)$.

Proof of Claim (3). Finally we analyze the number of satisfiable constraints in the **NO** case. Fix $\boldsymbol{\nu} \in \mathbb{Z}_q^k$ and let $\mathcal{D} \in \Delta(\mathbb{Z}_q)$ be the distribution obtained by sampling a uniformly random $\ell \in [n]$ and outputting $\boldsymbol{\nu}_\ell$. By Proposition 4.2 we have that $\mathbb{E}_{f \sim \mathcal{D}_\mathcal{F}, \mathbf{b} \sim \mathcal{D}^k}[f(\mathbf{b})] \leq \rho(\mathcal{F})$. We use this to prove that for every $i \in [m]$, $\mathbb{E}[Y_i(\boldsymbol{\nu}) | Y_{1:i-1}(\boldsymbol{\nu})] \leq (1+\varepsilon/50) \cdot \rho(\mathcal{F}) \cdot q^{-(k-1)}$.

First, as in the proof for Claim (2) we have that the total variation distance between $\mathbf{b} \sim \mathcal{D}^k$ and $\{\boldsymbol{\nu}_{\mathbf{j}(i)} | Y_{1:i-1}(\boldsymbol{\nu})\}$ is at most $k^2\alpha$. (In particular, this is upper bounded by the probability that k uniformly and independently chosen elements of $[n]$ either collide or fall in a set of size at most $k(\alpha n - 1)$.) We conclude that the probability that the i-th "potential constraint" (given by $(f(i), \mathbf{j}(i))$) is satisfied is at most $\rho(\mathcal{F}) + k^2\alpha$. Next, note that the event $X_i = 1$ (i.e., the i-th constraint is chosen in Ψ) is independent of $Y_i(\boldsymbol{\nu})$ since in the **NO** case $\mathbf{z}(i) \in \mathbb{Z}_q^k$ is uniform and independent of all other random variables. We conclude that $\mathbb{E}[Y_i(\boldsymbol{\nu}) | Y_{1:i-1}(\boldsymbol{\nu})] \leq (1+\varepsilon/50) \cdot \rho(\mathcal{F}) \cdot q^{-(k-1)}$. Finally, we apply Lemma 2.5 again to conclude that $\Pr[Y(\boldsymbol{\nu}) = \sum_{i=1}^m Y_i(\boldsymbol{\nu}) > (1+\varepsilon/10) \cdot \rho(\mathcal{F}) \cdot q^{-(k-1)} \cdot m] \leq c^{-m}$ where $c > 1$ depends on $q, k, \mathcal{F}, \alpha, \varepsilon$ but not on T or n. Thus by setting T large enough, we can bound $c^{-m} \leq q^{-2n}$. This allows us to use the union bound to conclude that the probability that there exists $\boldsymbol{\nu} \in \mathbb{Z}_q^n$ such that $Y(\boldsymbol{\nu}) > (1+\varepsilon/10) \cdot \rho(\mathcal{F}) \cdot q^{-(k-1)} \cdot m$ is at most $q^{-n} = o(1)$. Combining with the lower bound on $\tilde{m}$ from Claim (1) we get that with probability $1-o(1)$ we have $\mathsf{val}_\Psi \leq (1+\varepsilon/3) \cdot \rho(\mathcal{F})$ in this case.

This concludes the proofs of the claims and thus the proof of Theorem 4.3.

□

Theorems 1.1 and 1.2 follow immediately from Theorem 4.3 as we show below.

Proof of Theorem 1.1. The theorem follows from the fact that for a wide family $\omega(\mathcal{F}) = 1$ and in this case Theorem 4.3 asserts that a $\rho(\mathcal{F}) + \varepsilon$ approximation requires linear space. □

Proof of Theorem 1.1. The theorem follows from the fact that for every non-zero function f we have $\omega(f) \geq 1/q$ and so for every family $\mathcal{F}$ also we have $\omega(\mathcal{F}) \geq 1/q$. Thus Theorem 4.3 asserts that a $\rho(\mathcal{F}) \cdot q + \varepsilon$ approximation requires linear space, where $\rho(\mathcal{F})$ approximation is trivial. □

4.1 Some examples

We now give some examples illustrating the power of Theorem 4.3. Our first example is the familiar q-coloring problem.

Example 1 (Max-qCol).

Let $k = 2$ and $q \geq 2$. Let $\mathcal{F} = \{f : \mathbb{Z}_q^2 \to \{0,1\}\}$ where $f(u,v) = 1$ if and only if $u \neq v$. The "Max q-Coloring" problem is defined to be Max-qCol = Max-CSP($\mathcal{F}$). It is easy to verify $\rho(\mathcal{F}) = 1 - 1/q$ and $\omega(\mathcal{F}) = 1$. We thus conclude by Theorem 1.1 that Max-qCol is approximation resistant.

Next we turn to the Unique Games Problem.

Example 2 (Max-qUG).

Let $k = 2$ and $q \geq 2$. Let $\mathcal{F} = \{f : \mathbb{Z}_q^2 \to \{0,1\} \mid f^{-1}(1) \text{ is a bijection}\}$. The "$q$-ary Unique Games" problem is defined to be Max-qUG = Max-CSP($\mathcal{F}$). We show below that $\rho(\mathcal{F}) = 1/q$. We also show that there exists $\mathcal{F}' \subseteq \mathcal{F}$ such that $\rho(\mathcal{F}') = 1/q$ and $\omega(\mathcal{F}') = 1$. Applying Theorem 1.1 to $\mathcal{F}'$ we get that $1/q + \varepsilon$ approximating Max-CSP($\mathcal{F}'$) requires linear space and the same holds for Max-qUG = Max-CSP($\mathcal{F}$) by monotonicity.
We define the family $\mathcal{F}'$ to be $\mathcal{F}' = \{f_a | a \in \mathbb{Z}_q\}$ where $f_a(u,v) = 1$ if and only if $u = v + a$. Let $\mathcal{D} = \mathsf{Unif}(\mathbb{Z}_q)$. For every $f \in \mathcal{F}$ we have that $\mathbb{E}_{(u,v)\sim\mathcal{D}^2}[f(u,v)] = 1/q$. So for every $\mathcal{D}_\mathcal{F} \in \Delta(\mathcal{F})$ we have $\mathbb{E}_{f\sim\mathcal{D}_\mathcal{F}} \mathbb{E}_{(u,v)\sim\mathcal{D}^2}[f(u,v)] = 1/q$. This proves $\rho(\mathcal{F}), \rho(\mathcal{F}') \geq 1/q$. To get the upper bound we let $\mathcal{D}_\mathcal{F}$ be uniform over $\mathcal{F}'$. For every $(u,v) \in \mathbb{Z}_q^2$ we have $\mathbb{E}_{f\sim\mathcal{D}_\mathcal{F}}[f(u,v)] = 1/q$ and so for every distribution $\mathcal{D} \in \Delta(\mathbb{Z}_q^k)$ (which is more than we need) we have $\mathbb{E}_{f\sim\mathcal{D}_\mathcal{F}} \mathbb{E}_{(u,v)\sim\mathcal{D}}[f(u,v)] \leq 1/q$. This proves $\rho(\mathcal{F}'), \rho(\mathcal{F}) = 1/q$ (since $\mathcal{D}_\mathcal{F}$ is supported on $\mathcal{F}'$).
Now turning to $\omega(\mathcal{F}')$, note that for every $f_a \in \mathcal{F}'$ we have $\{(b+a, b) | b \in \mathbb{Z}_q\} \subseteq f_a^{-1}(1)$. Thus $\omega(f_a) \geq \omega_{(a,0)}(f_a) = 1$. It follows that $\omega(\mathcal{F}') = 1$.

Our third example talks about constraints that are general linear systems.

Example 3 (Max-Lin$_{k,r,q}$).

For $k \geq 2$ and prime q and $0 \leq r < k$, we define Max-Lin$_{k,r,q}$ = Max-CSP($\mathcal{F}$) for $\mathcal{F} = \mathcal{F}_{k,r,q} = \{f_{A,\mathbf{b}} : \mathbb{Z}_q^k \to \{0,1\} | A \in \mathbb{Z}_q^{r\times k}, \mathbf{b} \in \mathbb{Z}_q^k\}$ where $f_{A,\mathbf{b}}(x) = 1$ if and only if $Ax = Ab$. (Thus constraints are systems of satisfiable linear equations with solutions of dimension at least $k - r$.) Let $\mathcal{F}'_{k,r,q} = \{f_{A.\mathbf{b}} \in \mathcal{F}_{k,r,q} | A \cdot 1 = 0\}$. It is easy to verify that for every k, r, q, $\rho(\mathcal{F}'_{k,r,q}) \geq \rho(\mathcal{F}_{k,r,q}) \geq q^{-r}$. By choosing $\mathcal{D}'_\mathcal{F}$ to be uniform over $f_{A,\mathbf{b}}$ with full rank matrices A satisfying $A \cdot 1 = 0$, we get $\rho(\mathcal{F}_{k,r,q}), \rho(\mathcal{F}'_{k,r,q}) = q^{-r}$. For $r < k$, we also get $\omega(\mathcal{F}') = 1$ and thus, applying Theorem 1.1 to $\mathcal{F}'$ we get that Max-CSP($\mathcal{F}'$) is approximation-resistant. The same holds for Max-qUG = Max-CSP($\mathcal{F}$) by monotonicity.[a]

[a] We believe this system is not approximation resistant for $r = k$. This is proved for $q = 2$ in [CGSV21b, Lemma 2.14]. The case of general q may not have been explicitly resolved in previous work.

Finally we mention one more problem. This problem arises in the work of Singer, Sudan and Velusamy [SSV21] who use it to show the approximation resistance of the "maximum acyclic

subgraph" problem to $o(\sqrt{n})$ space algorithms. We suspect the improved space lower bound should improve their work to rule out $o(n)$ space algorithms.

Example 4 (Max-Less-Than$_q$).

For $k = 2$ and $q \geq 2$ we define $\mathcal{F} = \{<_q\}$ where $<_q \colon \mathbb{Z}_q^2 \to \{0,1\}$ is given by $<_q (u,v) = 1$ if and only if $u < v$. It is possible to show $\rho(\mathcal{F}) = \frac{1}{2}(1 - 1/q)$. Also $\omega_{(0,1)}(<_q) = 1 - 1/q$ and this can be used to show that $\omega(\mathcal{F}) = 1 - 1/q$. By Theorem 4.3 it follows that $1/2 + \varepsilon$-approximating Max-CSP($\mathcal{F}$) requires linear space.

5 Lower bound on the Communication Complexity

In this section we prove a linear lower bound on the communication complexity of IFRMD (Theorem 3.4). Our proof is via a hybrid argument which starts with all players receiving inputs from the **NO** distribution, and switching the players' input distribution one at a time starting with Player 1 to the **YES** distribution. We state a key "hybrid lemma" (Lemma 5.1) which asserts that any one step of switching does not alter the distribution of the message output by the switched player.

To state our lemma we recall some notations and set up a few new ones. Let $\alpha, n, k, q, T, m = \alpha n \in \mathbb{N}$ denote the usual parameters of IFRMD. Recall that the player t gets as input a matrix $A_{t,\mathbf{c}_t} \in \mathbb{Z}_q^{(k-1)m \times n}$ corresponding to a k-uniform hypermatching M_t consisting of m hyperedges folded over the center vector $\mathbf{c}_t$ and a vector $\mathbf{w}_t \in \mathbb{Z}_q^{(k-1)m}$. For notational convenience, we will separate the input A_{t,c_t} into a matrix $A_t \in \mathbb{Z}_q^{(k-1)m \times n}$ and the center $\mathbf{c}_t$. For a sequence of objects $O_1, O_2, \ldots, O_T$, we denote $O_{1:t} = \{O_1, O_2, \ldots, O_t\}$ for every $t \in [T]$. With this notation we have that the message S_t sent by the t-th player is a function of $A_{1:t}, \mathbf{c}_{1:t}, \mathbf{w}_t$ and $S_{1:t-1}$.[3] Next, note that by Yao's principle [Yao77], we may assume that the messages sent by the players in IFRMD are all deterministic. Namely, a protocol for IFRMD can be specified by deterministic message functions $r_1, r_2, \ldots, r_T$ so that $S_t = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}, \mathbf{w}_t)$ denotes the message sent by the t-th player. The communication complexity of a protocol is defined as the largest output length of r_t. When $(A_{1:T}, \mathbf{c}_{1:T}, \mathbf{w}_{1:T})$ is drawn from the **YES** distribution (resp. the **NO** distribution), we denote $S^Y_{1:T}$ (resp. $S^N_{1:T}$) to be the resulting messages. Without loss of generality S_T is just a bit "'Yes/No" indicating the output of the protocol. Thus, to prove Theorem 3.4 we need to show that S^Y_T and S^N_T are close in total variation distance. For the induction we prove the much stronger statement that $(A_{1:T}, \mathbf{c}_{1:T}, S^Y_{1:T})$ and $(A_{1:T}, \mathbf{c}_{1:T}, S^N_{1:T})$ are close in total variation distance, i.e.,

$$\|(A_{1:T}, \mathbf{c}_{1:T}, S^Y_{1:T}) - (A_{1:T}, \mathbf{c}_{1:T}, S^N_{1:T})\|_{tvd} \leq \delta\,.$$

The following lemma provides the key step in this analysis. Roughly it says that if the first $t-1$ players's inputs are according to the **YES** distribution then the t-th player's output on the **YES** input is typically distributed very similarly to the output on the **NO** distribution (even conditioned on all previously announced hypermatchings, centers and messages). Formally, the lemma identifies a sequence of events $\mathcal{E}_1 \supset \mathcal{E}_2 \supset \cdots \supset \mathcal{E}_T$ such that (i) $\mathcal{E}_t$ enforces a "typicality" restriction on the messages and inputs that the t-th player receives and (ii) if the messages and input received by the t-th player are typical then the player cannot distinguish whether its input is sampled from the **YES** distribution or the **NO** distribution (assuming all previous players' inputs were from the **YES** distribution).

[3] Note that even though the t-th player does not have access to $A_{1:t-1}, \mathbf{c}_{1:t-1}$, and $S_{1:t-2}$, allowing them to see these only makes our lower bound stronger.

Lemma 5.1 (Hybrid lemma)**.** *For every $q, k \in \mathbb{N}$, there exists $\alpha_0 \in (0, 1/k)$ such that for every $\alpha \in (0, \alpha_0]$, $T \in \mathbb{N}$, and $\delta \in (0,1)$, there exist $n_0 \in \mathbb{N}$, and $\tau \in (0,1)$ such that for every $n \geq n_0$ the following holds:*

Let $\Pi = (r_1, \ldots, r_T)$ be a deterministic protocol for IFRMD *where each message function r_t outputs a message of at most τn bits. Let $x \sim \mathsf{Unif}(\mathbb{Z}_q^n)$ and let $M_1, \ldots, M_T$ be independent random hypermatching of size αn over $[n]$. Let $(A_t, \mathbf{c}_t)$ be an independent random folded encoding of M_t for all $t \in [T]$. Let S_t^Y and S_t^N be the Yes and No message of the t-th player defined previously for message function r_t for all $t \in [T]$. Then there exists a sequence of events $\mathcal{E}_1 \supset \mathcal{E}_2 \supset \cdots \supset \mathcal{E}_T$ such that (i) $\mathcal{E}_t$ only depends on $(A_{1:t}, \mathbf{c}_{1:t})$ and $S_{1:t-1}^Y$, (ii) $\Pr[\mathcal{E}_1] \geq 1 - \delta/T$, (iii) $\Pr[\overline{\mathcal{E}}_t \,|\, \mathcal{E}_{t-1}] \leq \delta/T$ for all $t = 2, 3, \ldots, T$, and (iv) for every fixed $(A_{1:t}, \mathbf{c}_{1:t})$ and $S_{1:t-1}^Y$ satisfying $\mathcal{E}_t$, one has*

$$\|S_t^Y - r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, U_t)\|_{tvd} \leq \delta/T, \tag{5.2}$$

where $U_t \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$.

The proof of Lemma 5.1 will be provided in Section 5.4. Theorem 3.4 follows almost immediately from Lemma 5.1 as shown in Section 5.1. In the rest of this section and the following sections we prove Lemma 5.1. Here we give an overview of this part of the proof.

The general idea behind the proof of Lemma 5.1 is to argue that information about $\mathbf{x}^*$ "leaked" by the messages of the first $t-1$ players (i.e., $S_{1:t-1}$) is not sufficient for the t-th player to distinguish between the case where $\mathbf{w}_t = A_{t,c_t}\mathbf{x}^*$ (the **YES** case) and the case where $\mathbf{w}_t$ is uniform. The earlier proofs of this type (in particular as in [KKS15]) simply counted the total information gleaned about $\mathbf{x}^*$ which is bounded by the total communication. Such proofs are inherently limited to achieving only a $\sqrt{n}$ lower bound. To go further, as in [KK19], one needs to argue about the structure of the information learned about $\mathbf{x}^*$, and in particular note that no player sees $\mathbf{x}^*$ directly, and the t'-th player only sees $A_{t',c_{t'}} \cdot \mathbf{x}^*$. (In particular no coordinate of $\mathbf{x}^*$ is revealed directly, though the sum of many pairs of coordinates are directly revealed.) Thus the information about $\mathbf{x}^*$ comes from a "reduced space" and we would like to capture and exploit the structural restriction imposed by this restriction. Information-theoretic tools seem to fail to capture this restriction and the key to the work of [KK19] is to give a Fourier analytic condition, that they call "boundedness", that captures this restriction.

The boundedness condition applies to what we call the "posterior distribution" of $\mathbf{x}^*$, i.e., the distribution of $\mathbf{x}^*$ conditioned on the first t messages. This distribution turns out to be the uniform distribution over a set $B_t \subseteq \mathbb{Z}_q^n$ (see Lemma 5.8). The boundedness condition places restrictions on the Fourier spectrum of the indicator function of this set. (See Definition 5.13.) To use this condition we need three ingredients elaborated below, which we abstract as lemma statements in this section and prove in later section. Given these three lemmas the proof of Lemma 5.1 follows and is given in Section 5.4.

The first ingredient we need is that boundedness of B_{t-1} does imply that the t-th player is unable to distinguish between its input being from the **YES** distribution or the **NO** distribution. This is stated as Lemma 5.17. Next we need to show that given information about $A_{t,c_t}\mathbf{x}^*$, the posterior distribution of $\mathbf{x}^*$ is indeed bounded, and we assert this in Lemma 5.16. Note that this also serves as the base case of our induction. Finally we argue that if B_{t-1} is bounded, then for most matchings A_t (and every center $\mathbf{c}_t$ of A_t) the resulting set B_t is bounded. This is asserted in Lemma 5.18. See also Figure 1 for a pictorial overview for the proof structure of Lemma 5.1.

In the rest of this section, after showing that Lemma 5.1 implies Theorem 3.4 in Section 5.1, we introduce the posterior sets and discuss their basic properties in Section 5.2, we introduce boundedness and state the three lemmas above in Section 5.3, and finally conclude with the proof of Lemma 5.1 in Section 5.4.

5.1 Proof of Theorem 3.4

We now show how the lemma suffices to prove Theorem 3.4. The proof is analogous to the proof of Lemma 6.3 in [KK19]. We remark that the lemma is not immediate and effectively depends on the fact that players can jointly sample from the **NO** distribution on their own. (Note the players can't jointly sample from the **YES** distribution since these samples are correlated by the hidden vector $\mathbf{x}^*$. So the proof is inherently asymmetric visavis the treatment of the **YES** and **NO** distributions.)

Proof of Theorem 3.4. For the sake of contradiction, assume that there exists a protocol $\Pi = (r_1, \ldots, r_T)$ that solves IFRMD with advantage more than δ and less than τn bits of communication for some $n \geq n_0$. In what follows, we will show that $\|(A_{1:T}, \mathbf{c}_{1:T}, S^Y_{1:T}) - (A_{1:T}, \mathbf{c}_{1:T}, S^N_{1:T})\|_{tvd} \leq \delta$, which implies that the advantage of the protocol cannot be greater than δ, hence producing a contradiction.

For every $q, k \in \mathbb{N}$, we set $\alpha_0 \in (0, 1/k)$ and $\tau_0 \in (0, 1)$ as in Lemma 5.1. For every $\alpha \in (0, \alpha_0]$, $T \in \mathbb{N}$, and $\delta' = \delta/2$, we set $n_0 \in \mathbb{N}$ and $\tau \in (0, 1)$ as in Lemma 5.1.

Let $\mathcal{E}_1 \supset \mathcal{E}_2 \supset \cdots \supset \mathcal{E}_T$ be the sequence of events guaranteed by Lemma 5.1 such that $\Pr\left[\overline{\mathcal{E}_t} \,|\, \mathcal{E}_{t-1}\right] \leq \delta'/T$ for $t = 2, 3 \ldots, T$. Note that by the properties of these events, with probability at least $1 - \delta'$, we have $\|S^Y_t - r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, U_t)\|_{tvd} \leq \delta'/T$ for all $t \in [T]$. We use $\|\cdot\|_{tvd,\mathcal{E}_t}$ to denote the total variation distance of distributions conditioned on $\mathcal{E}_t$. We inductively show that for every $t \in [T]$,

$$\|(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t}) - (A_{1:t}, \mathbf{c}_{1:t}, S^N_{1:t})\|_{tvd,\mathcal{E}_t} \leq \frac{t\delta'}{T}. \qquad \text{(Induction hypothesis)}$$

First, we prove the base case $t = 1$. Recalling that $S^Y_0 = S^N_0$, we have

$$\begin{aligned}\|(A_1, \mathbf{c}_1, S^Y_1) - (A_1, \mathbf{c}_1, S^N_1)\|_{tvd,\mathcal{E}_1} &= \|(A_1, \mathbf{c}_1, S^Y_1) - (A_1, \mathbf{c}_1, r_1(M_1, \mathbf{c}_1, S^N_0, U_1))\|_{tvd,\mathcal{E}_1} \\ &= \|(A_1, \mathbf{c}_1, S^Y_1) - (A_1, \mathbf{c}_1, r_1(M_1, \mathbf{c}_1, S^Y_0, U_1))\|_{tvd,\mathcal{E}_1}.\end{aligned}$$

Observe that for every fixed $A_1, \mathbf{c}_1$ and S^Y_0 satisfying $\mathcal{E}_1$, we have $\|S^Y_1 - r_1(M_1, \mathbf{c}_1, S^Y_0, U_1)\|_{tvd} \leq \frac{\delta'}{T}$, where the randomness is over S^Y_1 and U_1. It follows from Lemma 2.3 that

$$\|(A_1, \mathbf{c}_1, S^Y_1) - (A_1, \mathbf{c}_1, r_1(M_1, \mathbf{c}_1, S^Y_0, U_1))\|_{tvd,\mathcal{E}_1} \leq \frac{\delta'}{T},$$

which completes the base case.

Next, we tackle the induction step. For every $t = 2, \ldots, T$, we have

$$\begin{aligned}&\|(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t}) - (A_{1:t}, \mathbf{c}_{1:t}, S^N_{1:t})\|_{tvd,\mathcal{E}_t} \\ &= \|(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{t-1}, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (A_{1:t}, \mathbf{c}_{1:t}, S^N_{1:t-1}, r_t(A_{1:t}, \mathbf{c}_{1:t}, S^N_{t-1}, U_t))\|_{tvd,\mathcal{E}_t}.\end{aligned}$$

Let us define $Q^Y_{t-1} = (A_{1:t-1}, \mathbf{c}_{1:t-1}, S^Y_{1:t-1})$ and $Q^N_{t-1} = (A_{1:t-1}, \mathbf{c}_{1:t-1}, S^N_{1:t-1})$. Then, we can rewrite the above expression for total variation distance in terms of the new notation as follows:

$$\begin{aligned}&\|(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{t-1}, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (A_{1:t}, \mathbf{c}_{1:t}, S^N_{1:t-1}, r_t(A_{1:t}, \mathbf{c}_{1:t}, S^N_{t-1}, U_t))\|_{tvd,\mathcal{E}_t} \\ &\quad = \|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t((Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (Q^N_{t-1}, A_t, \mathbf{c}_t, r_t(Q^N_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd,\mathcal{E}_t}. \qquad (5.3)\end{aligned}$$

We now apply Lemma 2.4 to Equation 5.3. Applying this lemma with $X^1 = Q^Y_{t-1}$, $X^2 = Q^N_{t-1}$, $Z^1 = (A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)$, $Z^2 = ((A_t, \mathbf{c}_t, U_t))$, and f as the function that maps the tuple $(X, (B, C))$ to

$(B, r_t(X, B, C))$, we get

$$\begin{aligned}
&\|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (Q^N_{t-1}, A_t, \mathbf{c}_t, r_t(Q^N_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd,\mathcal{E}_t} \\
&\leq \|Q^Y_{t-1} - Q^N_{t-1}\|_{tvd,\mathcal{E}_t} + \|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd,\mathcal{E}_t} \\
&= \|Q^Y_{t-1} - Q^N_{t-1}\|_{tvd,\mathcal{E}_{t-1}} + \|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd,\mathcal{E}_t}\,,
\end{aligned} \tag{5.4}$$

where the last equality follows from the fact that $\mathcal{E}_t \subset \mathcal{E}_{t-1}$ and condition (i) of Lemma 5.1 which states that $\mathcal{E}_{t-1}$ only depends on $(A_{1:t-1}, \mathbf{c}_{1:t-1})$ and $S^Y_{1:t-2}$.

Now, by applying the induction hypothesis, we have that

$$\|Q^Y_{t-1} - Q^N_{t-1}\|_{tvd,\mathcal{E}_{t-1}} \leq \frac{(t-1)\delta'}{T}. \tag{5.5}$$

Next, we bound the second term on the right hand side of (5.4), i.e.,

$$\|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd,\mathcal{E}_t},$$

by applying condition (iv) from Lemma 5.1. According to this condition, for every *fixed* $(A_{1:t}, \mathbf{c}_{1:t})$ and $S^Y_{1:t-1}$ satisfying $\mathcal{E}_t$, we have

$$\|r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, A_{t,\mathbf{c}_t}\mathbf{x}^*) - r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, U_t)\|_{tvd} \leq \frac{\delta'}{T},$$

where $U_t \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$. Thus, by Lemma 2.3, it follows that

$$\|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) - (Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd,\mathcal{E}_t} \leq \frac{\delta'}{T}\,. \tag{5.6}$$

Combining Eqs. (5.3) to (5.6), we have

$$\|(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t}) - (A_{1:t}, \mathbf{c}_{1:t}, S^N_{1:t})\|_{tvd,\mathcal{E}_t} \leq \frac{\delta' t}{T},$$

which completes the induction.

Substituting $t = T$, we conclude that

$$\|(A_{1:T}, \mathbf{c}_{1:T}, S^Y_{1:T}) - (A_{1:T}, \mathbf{c}_{1:T}, S^N_{1:T})\|_{tvd,\mathcal{E}_T} \leq \delta'.$$

Finally, by removing the conditioning on $\mathcal{E}_T$, we have

$$\begin{aligned}
|(A_{1:T}, \mathbf{c}_{1:T}, S^Y_{1:T}) - (A_{1:T}, \mathbf{c}_{1:T}, S^N_{1:T})\|_{tvd} &\leq |(A_{1:T}, \mathbf{c}_{1:T}, S^Y_{1:T}) - (A_{1:T}, \mathbf{c}_{1:T}, S^N_{1:T})\|_{tvd,\mathcal{E}_T} + \Pr[\overline{\mathcal{E}_T}] \\
&\leq \delta' + \delta' \leq \delta.
\end{aligned}$$

This implies that Π cannot have advantage more than δ, which contradicts the assumptions of the theorem statement. Therefore, we conclude that any protocol for IFRMD with advantage δ requires τn bits of communication, as desired. □

5.2 Posterior sets and functions

The main challenge in proving Lemma 5.1 lies in the condition (iv), i.e., requiring the closeness of the Yes message (i.e., $S_t^Y = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*)$) and the hybrid No message (i.e., $r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, U_t)$). Intuitively, if $\mathbf{x}^* \sim \mathsf{Unif}(\mathbb{Z}_q^n)$ and is independent of the other arguments, then $A_{t,\mathbf{c}_t}\mathbf{x}^*$ is uniformly distributed over $\mathbb{Z}_q^{(k-1)\alpha n}$ and hence S_t^Y follows the same distribution as $r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, U_t)$. However, $\mathbf{x}^*$ is correlated[4] with the previous messages $S_{1:t-1}^Y$ so the above ideal situation would not happen in general. Nevertheless, we are able to analyze the conditional distribution of $A_{t,\mathbf{c}_t}\mathbf{x}^*$ on the previous messages by explicitly characterizing the *posterior distribution* of $\mathbf{x}^*$ after receiving the messages from the first $t-1$ players. That is, the conditional distribution of $A_{t,\mathbf{c}_t}\mathbf{x}^*$ can be described by first sampling $\mathbf{x}^*$ from the posterior distribution and then applying $A_{t,\mathbf{c}_t}$.

For every fixed $A_{1:t}, \mathbf{c}_{1:t}$ and $S_{1:t}$, we would like to identify a distribution $\mathcal{D}_t$ over $\mathbb{Z}_q^n$ such that $\mathcal{D}_t$ is the conditional distribution of $\mathbf{x}^*$ given messages $S_{1:t}$. Note that by the choice of the No case, the conditional distribution of $\mathbf{x}^*$ given messages $S_{1:t}$ is simply the uniform distribution over $\mathbb{Z}_q^n$. Thus, we only need to worry about the Yes case.

Definition 5.7 (Posterior sets and functions)**.** *Under the setting described above, for each t and fixed $A_{1:t}$, $\mathbf{c}_{1:t}$, and $S_{1:t}$, define*

- *(Reduced posterior set) $B_{r,t} \subseteq \mathbb{Z}_q^{(k-1)m}$ be the set of possible values of $z_t = A_{t,\mathbf{c}_t}\mathbf{x}$ that leads to message S_t; Note that $B_{r,t}$ should be thought of as a function on A_t, $\mathbf{c}_t$, and S_t in the sense that $B_{r,t} = g_t^{-1}(S_t)$ where $g_t(\cdot) = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}, \cdot)$. Let q be the indicator function of $B_{r,t}$.*
- *(Posterior set and function) Let*

$$B_t := \{\mathbf{x} \in \mathbb{Z}_q^n \mid A_{t,\mathbf{c}_t}\mathbf{x} \in B_{r,t}\}.$$

  Also, let $\mathbf{1}_{B_t} : \mathbb{Z}_q^n \to \{0,1\}$ be the indicator function of B_t.

- *(Aggregated posterior set and function) Let*

$$B_{1:t} := \{\mathbf{x} \in \mathbb{Z}_q^n \mid A_{t',\mathbf{c}_{t'}}\mathbf{x} \in B_{r,t'},\ \forall t' = 1, \ldots, t\} = \bigcap_{t'=1}^{t} B_{t'}.$$

  Also, let $\mathbf{1}_{B_{1:t}} : \mathbb{Z}_q^n \to \{0,1\}$ be the indicator function of $B_{1:t}$. Namely, $\mathbf{1}_{B_{1:t}} = \prod_{t'=1}^{t} \mathbf{1}_{B_{t'}}$.

Now, we show that $\mathbf{1}_{B_{1:t}}$ captures the posterior distribution (i.e., the conditional distribution) of $\mathbf{x}$ given messages $S_1, S_2, \ldots, S_t$:

Lemma 5.8 (Posterior function $\mathbf{1}_{B_{1:t}}$ captures the posterior distribution.)**.** *For every $t \in [T]$, the conditional distribution of $\mathbf{x}$ given messages $S_1, S_2, \ldots, S_t$ is exactly given by $\mathbf{1}_{B_{1:t}}(\mathbf{x})/\|\mathbf{1}_{B_{1:t}}\|_1$. In particular, for fixed $A_{1:t}$, $\mathbf{c}_{1:t}$, and $S_{1:t-1}^Y$, we have $S_t^Y = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*)$, where $\mathbf{x}^* \sim \mathsf{Unif}(B_{1:t})$.*

Proof. Wlog, let us focus on the case of $t = T$. The proof is done by some direct manipulations of

[4] In particular, $\mathbf{x}^*$ has to be consistent with the previous messages $S_{1:t-1}^Y$.

Bayes rule. By definition, the conditional probability of $\mathbf{x}$ given messages $S_1, S_2, \ldots, S_T$ would be

$$\begin{aligned}
&\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[\mathbf{x}^* = \mathbf{x} \mid A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T]] \\
&= \frac{\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[\mathbf{x}^* = \mathbf{x} \wedge A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T]]}{\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T]]} \\
&= \frac{\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T] \mid \mathbf{x}^* = \mathbf{x}] \cdot \Pr_{\mathbf{x}^*}[\mathbf{x}^* = \mathbf{x}]}{\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T]]} \\
&= \frac{\Pr_{\mathbf{x}^*}[\mathbf{x}^* = \mathbf{x}]}{\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T]]} \cdot \prod_{t=1}^{T} \Pr_{\mathbf{x}^*, A_t, \mathbf{c}_t}[A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t} \mid \mathbf{x}^* = \mathbf{x}] \\
&= \frac{q^{-n}}{\Pr_{\mathbf{x}^*, A_{1:T}, \mathbf{c}_{1:T}}[A_{t,\mathbf{c}_t}\mathbf{x}^* \in B_{r,t}\ \forall t \in [T]]} \cdot \prod_{t=1}^{T} \Pr_{A_t, \mathbf{c}_t}[A_{t,\mathbf{c}_t}\mathbf{x} \in B_{r,t}]\,.
\end{aligned}$$

As the above quantity is proportional to $\mathbf{1}_{B_{1:T}}(\mathbf{x})$, we conclude that the conditional probability of $\mathbf{x}$ given messages $S_1, S_2, \ldots, S_T$ is given by $\mathbf{1}_{B_{1:T}}(\mathbf{x})/\|\mathbf{1}_{B_{1:T}}\|_1$.

Next, by definition we have $S_T^Y = r_T(A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y, A_{t,\mathbf{c}_T}\mathbf{x}^*)$ where $\mathbf{x}^* \sim \mathsf{Unif}(\mathbb{Z}_q^n)$. Thus, the conditional distribution of S_T^Y on $A_{1:T}$, $\mathbf{c}_{1:T}$, and $S_{1:T-1}^Y$ is

$$\begin{aligned}
&\Pr[S_T^Y = S \,|\, A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y] = \frac{\Pr[S_T^Y = S, A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y]}{\Pr[A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y]} \\
&= \frac{\Pr_{\mathbf{x}^* \sim \mathsf{Unif}(\mathbb{Z}_q^n)}[S = r_T(A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y, A_{t,\mathbf{c}_T}\mathbf{x}^*), A_{1:T}, \mathbf{c}_{1:T}, S_t^Y = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*)\ \forall t \in [T-1]]}{\Pr[A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y]} \\
&= \frac{\Pr_{\mathbf{x}^* \sim \mathsf{Unif}(\mathbb{Z}_q^n)}[S = r_T(A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y, A_{t,\mathbf{c}_T}\mathbf{x}^*), A_{1:T}, \mathbf{c}_{1:T}, \mathbf{x}^* \in B_{1:T}]}{\Pr[A_{1:T}, \mathbf{c}_{1:T}, S_{1:T-1}^Y]}\,.
\end{aligned}$$

Namely, $S_t^Y = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*)$, where $\mathbf{x}^* \sim \mathsf{Unif}(B_{1:t})$. □

Note that we have a characterization of the posterior distribution of $\mathbf{x}^*$, the following corollary shows that Equation 5.2 (i.e., the condition (iv) of Lemma 5.1) can be simplified to bounding the total variation distance between the posterior distribution and the uniform distribution.

Corollary 5.9 (Reducing Equation 5.2)**.** *Let $r_t, S_{1:t-1}^Y, A_{1:t}, \mathbf{c}_{1:t}, B_{1:t}, U_t$ be defined as before, we have*

$$\|r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*) - r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, U_t)\|_{tvd} \le \|(A_{t,\mathbf{c}_t}\mathbf{x}^*) - U_t\|_{tvd}$$

where $\mathbf{x}^* \sim \mathsf{Unif}(B_{1:t})$.

Proof. By Lemma 5.8, we have

$$S_t^Y = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*)$$

where $\mathbf{x}^* \sim \mathsf{Unif}(B_{1:t})$. Note that when we fix $A_{1:t}$, $\mathbf{c}_{1:t}$, and $S_{1:t-1}^Y$ (hence $B_{1:t}$ is also fixed), by data processing inequality (see item 2 of Proposition 2.2) we have

$$\|r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, A_{t,\mathbf{c}_t}\mathbf{x}^*) - r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, U_t)\|_{tvd} \le \|(A_{t,\mathbf{c}_t}\mathbf{x}^*) - U_t\|_{tvd}\,.$$

□

Namely, Equation 5.2 (i.e., the condition (iv) of Lemma 5.1) can be replaced with $\|(A_{t,\mathbf{c}_t}\mathbf{x}^*) - U_t\|_{tvd} \le \gamma/T$, i.e., after applying a random folded hypermatching matrix $A_{t,\mathbf{c}_t}$ to the posterior distribution $\mathsf{Unif}(B_{1:t})$, the distribution of the resulting string is close to the uniform distribution $\mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$.

Finally, the following lemma shows that when the amount of communication is small, the posterior set is large with high probability.

Lemma 5.10 (Posterior set is large)**.** *Let $\Pi = (r_1, \dots, r_T)$ be a deterministic protocol for IFRMD where each message function r_t outputs a message of length at most s bits for some $1 \le s \le n$. Let B_t be the posterior set defined in Definition 5.7 for every $t \in [T]$. For every $\delta \in (0,1)$ and $t \in [T]$, we have $|B_t| \ge \delta \cdot q^{n-s}$ with probability at least $1-\delta$ over the randomness of $\mathbf{x} \in \mathbb{Z}_q^n$.*

Proof. Fix a hypermatching M and centers $\mathbf{c}$, the t-th message function induces a partition $P_1 \cup P_2 \cup \cdots \cup P_{2^s}$ of $\mathbb{Z}_q^n$. For each $\mathbf{x} \in \mathbb{Z}_q^n$, we define $P(\mathbf{x})$ to be the part that contains $\mathbf{x}$, i.e, if $\mathbf{x} \in P_i$, then $P(\mathbf{x}) = P_i$. Note that

$$\mathop{\mathbb{E}}_{\mathbf{x}\in\mathbb{Z}_q^n}\left[\frac{1}{|P(\mathbf{x})|}\right] = \sum_{i=1}^{2^s} \frac{\Pr_{\mathbf{x}\in\mathbb{Z}_q^n}[\mathbf{x}\in P_i]}{|P_i|} = \sum_{i=1}^{2^s} \frac{|P_i|\cdot q^{-n}}{|P_i|} = \frac{2^s}{q^n} \le q^{s-n}.$$

By Markov's inequality, we have $|P(\mathbf{x})| < \delta \cdot q^{n-s}$ with probability at most δ as desired. □

5.3 Fourier analytic conditions

In this subsection, we define and analyze Fourier-analytic properties of the posterior set B and show that these properties are sufficient for the condition (iv) (i.e., Corollary 5.9) of Lemma 5.1.

Recall that given a matching $M = (e_1, \dots, e_m)$ and centers $\mathbf{c} = (c_1, \dots, c_m)$, $A_{\mathbf{c}}$ is the $\mathbf{c}$-centered folded encoding of M. We are going to define three properties for sets B in $\mathbb{Z}_q^n$. First, we say a set $B \subseteq \mathbb{Z}_q^n$ is $(M, \mathbf{c})$-restricted if B is restricted to a union of shifted null spaces of $A_{\mathbf{c}}$.

Definition 5.11 (Restricted set)**.** *Let M be a k-hypermatching of size m and $\mathbf{c}$ be centers. We say a set $B \subseteq \mathbb{Z}_q^n$ is $(M,\mathbf{c})$-restricted if there exists a ("reduced") set $B_r \subseteq \mathbb{Z}_q^{(k-1)m}$ such that $B = \{\mathbf{x} \in \mathbb{Z}_q^n \mid A_{\mathbf{c}}\mathbf{x} \in B_r\}$.*

Next, we say a set B is *bounded* if the Fourier spectrum of the indicator function $\mathbf{1}_B$ can be properly bounded in an appropriate range of the spectrum. This is analogous to Definition 4.3 in [KK19]. First, we introduce some notation:

$$U_{C,s}(h) := \begin{cases} 1, & h = 0 \\ \left(\frac{C\sqrt{sn}}{h}\right)^{h/2}, & 1 \le h \le s \\ \left(\frac{2q^2e^2n}{h}\right)^{h/2}, & h > s. \end{cases} \tag{5.12}$$

Definition 5.13 (Bounded set)**.** *Let $n, q \in \mathbb{N}$, $0 \le s \le n$, $C > 0$, and $B \subset \mathbb{Z}_q^n$. We say B (as well as its indicator function $\mathbf{1}_B$) is (C,s)-bounded if, for every $h \in [s]$,*

$$\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} \frac{q^n}{|B|}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \le U_{C,s}(h). \tag{5.14}$$

Remark. *As we keep track of posterior sets that are inductively refined, we will need the* entire *Fourier spectrum of the corresponding indicator functions to be bounded from above by the function $U_{C,s}$ (for appropriate $C, s > 0$), which is defined piecewise on the low, medium, and high regimes. This allows us to show that $A_{\mathbf{c}}\mathbf{x}$ is close to the uniform distribution on $\mathbb{Z}_q^{(k-1)\alpha n}$ when $\mathbf{x}$ is drawn from such a posterior set $B \subset \mathbb{Z}_q^n$ (see Lemma 5.17). However, the upper bound given by $U_{C,s}(h)$ in the* high regime $h > s$ *is guaranteed automatically as long as B is large enough (see Lemma 6.5). Thus, we only need to keep track of the Fourier spectrum for weights in the* middle *regime; hence, the (C, s)-boundedness property that we maintain inductively only concerns Fourier weights in this regime.*

More specifically, if a set $B \subset \mathbb{Z}_q^n$ is (C, s)-bounded and satisfies $|B| \geq q^{n-s}$, then by Lemma 6.5, we have that

$$\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0=h}} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \leq U_{C,s}(h)$$

for all $0 \leq h \leq n$.

Finally, in what follows we will show that the intersection of a bounded set with a "restricted set" is also bounded and this will be the core of our induction. To do this we need to understand the Fourier behavior of restricted sets. It turns out that restricted sets satisfy a property stronger than being bounded, which we term "reduced"-ness below.

Definition 5.15 (Reduced set)**.** *Let $n, q \in \mathbb{N}$, $0 \leq s \leq n$, $C > 0$, and $B \subset \mathbb{Z}_q^n$. Let M be a k-hypermatching. We say B (as well as its indicator function $\mathbf{1}_B$) is (M, C, s)-reduced if the following hold.*

- *For every $\mathbf{u} \in \mathbb{Z}_q^n$, if there exists $i \in [n]$ such that $u_i = 1$ but i is not contained in M, then $\widehat{\mathbf{1}_B}(\mathbf{u}) = 0$.*
- *For every $\mathbf{u} \in \mathbb{Z}_q^n$, if there exists a hyperedge e_i of M such that $\langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \bmod q$, then $\widehat{\mathbf{1}_B}(\mathbf{u}) = 0$.*
- *For every $h \in \{1, \ldots, s\}$ and $\mathbf{v} \in \mathbb{Z}_q^n$,*

$$\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{v}\|_0=h}} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \leq U_{C,s}(h)\,.$$

There are two key lemmas about these Fourier analytic conditions. The first lemma establishes the base case of the induction toward showing the aggregated posterior set being (C, s)-bounded (for some $C = O(1)$ and $s = \Omega(n)$). In fact, we show a stronger guarantee in which every posterior set B_t is (M_t, C, s)-reduced.

Lemma 5.16 (Base case)**.** *For every $q, k \geq 2$, $\alpha \in (0, 1/k)$, there exists a constant C such that for every k-hypermatching M on $[n]$ of size $m \leq \alpha n$, suppose $n \in \mathbb{N}$ is large enough and $0 < b \leq s \leq n/32$, then the following holds. Let $B \subseteq \mathbb{Z}_q^n$. If (i) there exists a sequence of centers $\mathbf{c}$ such that B is $(M, \mathbf{c})$-restricted and (ii) $|B| \geq q^{n-b}$, then B is (M, C, s)-reduced.*

The proof of Lemma 5.16 is postponed to Section 6.3. (We note that the proof yields that $C \geq 2\zeta^2 e k^2 q^{3k}$ where ζ is the constant from Lemma 2.11.)

Recall from Corollary 5.9 that the condition (iv) in Lemma 5.1 are implied by showing $A_{\mathbf{c}}\mathbf{x}$ is close to the uniform distribution over $\mathbb{Z}_q^{(k-1)m}$ with high probability over to choice of $A_{\mathbf{c}}$ where $\mathbf{x}$ is sampled uniformly from the posterior set $B_{1:t}$. The second key lemma shows that $A_{\mathbf{c}}\mathbf{x}^*$ is indeed close to uniform when the posterior set is bounded.

Lemma 5.17 (Boundedness implies closeness to uniformity)**.** *For every $q, k \geq 2$ and $\delta \in (0, 1/2)$, there exists $\alpha_0 = \alpha_0(k, q)$ such that for every $\alpha \in (0, \alpha_0)$, $C > 0$, there exists $\tau_0 = \tau_0(q, k, \alpha, \delta, C)$ such that the following holds for any $\tau \in (0, \tau_0)$ and sufficiently large n:*

Let $B \subset \mathbb{Z}_q^n$ be a (C, s)-bounded set with $|B| \geq q^{n-b}$, for $4\log(3/\delta) \leq b \leq s \leq \tau n$. Let M be a random k-hypermatching of size αn and $\mathbf{c}$ be a sequence of centers for M and let $A_{\mathbf{c}}$ denote the $\mathbf{c}$ centered folded encoding and M. Then, with probability at least $1 - \delta$ over the choice of M, we have that for every $\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)\alpha n}$, we have

$$1 - \delta < q^{(k-1)\alpha n} \Pr_{\mathbf{x} \sim \mathit{Unif}(B)}[A_{\mathbf{c}}\mathbf{x} = \mathbf{z}_0] < 1 + \delta\,.$$

As a consequence, we also have

1. *$\|(A_{\mathbf{c}}\mathbf{x}) - U\|_{tvd} \leq \delta$ where $\mathbf{x} \sim \mathit{Unif}(B)$ and $U \sim \mathit{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$.*
2. *For every non-negative function f over $\mathbb{Z}_q^{(k-1)\alpha n}$,*

$$(1 - \delta) \leq \frac{\mathbb{E}_{\mathbf{x} \sim \mathit{Unif}(B)}[f(A_c\mathbf{x})]}{\mathbb{E}_{\mathbf{z} \sim \mathit{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})}[f(\mathbf{z})]} \leq (1 + \delta)\,.$$

The proof of Lemma 5.17 is postponed to Section 6.4.

Our final lemma of this section asserts that if $\mathbf{1}_{B_{1:t}}$ is (C, s)-bounded, then $f_{1:t+1}$ is $(O(C), s)$-bounded with high probability.

Lemma 5.18 (Induction step)**.** *For every $q, k \in \mathbb{N}$ there exist $\alpha_0 \in (0, 1/k)$ and $C_0 > 0$ such that for every $\alpha \in (0, \alpha_0]$, $C > C_0$, and $\delta \in (0, 1/2)$, there exist $C' > 0$, $n_0 \in \mathbb{N}$, and $\tau_0 \in (0, 1)$ such that the following holds. For every $n \geq n_0$, every $0 < b, b', s < \tau_0 n$, and every $B \subset \mathbb{Z}_q^n$ that satisfies $|B| \geq q^{n-b}$ and is (C, s)-bounded, let M be a uniformly random k-hypermatching of size at most αn, with probability at least $1 - 4\delta$ over the randomness of M, for every (M, C_0, s)-reduced set $B' \subset \mathbb{Z}_q^n$ with $|B'| \geq q^{n-b'}$ and $|B \cap B'| \geq (1 - \delta) \cdot |B| \cdot |B'|/q^n \geq q^{n-s}$, we have $B \cap B'$ is (C', s)-bounded.*

Lemma 5.18 is proved in Section 6.5. In our inductive application of the lemma above, we set $B \leftarrow B_{1:t-1}$ and $B' \leftarrow B_t$ for every $t \in \{2, 3, \ldots, T\}$ to get that all the B_t's are bounded and this is the core of the proof of Lemma 5.1.

5.4 Proof of Lemma 5.1

Proof of Lemma 5.1. For every $q, k \in \mathbb{N}$, we choose α_0' to be the minimum of the α_0's from the induction step (i.e., Lemma 5.18) and the "boundedness implies uniformity" lemma (i.e., Lemma 5.17). We set C_0 according to Lemma 5.18 and for every $\alpha \in (0, \alpha_0']$, $T \in \mathbb{N}$, and $\delta \in (0, 1)$, we invoke Lemma 5.17 with $\delta' = \delta/10T$ and $C = C_0$ to get $\tau_0 = \tau_0(q, k, \alpha, \delta', C_0) > 0$ and set $s = \tau_0 n$. Let $\tau > 0$ be a small constant. (We will explicitly fix this quantity later.) Let $b = \tau n + \log_q(10/\delta')$. We choose τ such that $2Tb < s = \tau_0 n$. Let $C_1 = C_0$. We define $\mathcal{E}_1$ to be the event that $|B_1| \geq q^{n-2b}$ and B_1 is (C_1, s)-bounded, where B_1 refers to the posterior set defined in Definition 5.7. By the "posterior set is large" lemma (i.e., Lemma 5.10) and the "base case" lemma (i.e., Lemma 5.16) we

have $\Pr[\overline{\mathcal{E}_1}] \le \delta'/10 \le \delta/T$ as desired. This satisfies condition (ii) of Lemma 5.1. See Figure 1 for a pictorial overview of the proof.

Next, for each $t \in \{2, 3, \ldots, T\}$, let $\mathcal{E}_t = \mathcal{E}_1 \cap \cdots \cap \mathcal{E}_{t-1} \cap \mathcal{E}'_t$ where $\mathcal{E}'_t$ denotes the event that the aggregate posterior set $B_{1:t}$ is large, i.e., $|B_{1:t}| \ge q^{n-2tb}$ and $B_{1:t}$ is (C_t, s)-bounded, where $C_t > 0$ is a constant that will be inductively chosen later. Note that by construction $\mathcal{E}_t$ only depends on $(A_{1:t}, \mathbf{c}_{1:t})$ and $S^Y_{1:t-1}$ and hence satisfies condition (i) of the lemma. To show that $\mathcal{E}_t$ happens with high probability conditioned on $\mathcal{E}_{t-1}$, note that by the "posterior set is large" lemma (i.e., Lemma 5.10) and the "base case" lemma (i.e., Lemma 5.16), we have $|B_t| \ge q^{n-b}$ and B_t is (M_t, C_0, s)-reduced with probability at least $1 - \delta'$. Moreover, the event $\mathcal{E}_{t-1}$ implies that $B_{1:t-1}$ is (C_{t-1}, s)-bounded and hence if we set $\tau < \tau_0(q, k, \alpha, \delta', C_{t-1})$, by the "boundedness implies uniformity" lemma (i.e., Lemma 5.17), $\mathbb{Z}_q^{(k-1)\alpha n}$, we have the following claim.

Claim 5.19. *When $2Tb < s$, conditioned on $\mathcal{E}_{t-1}$, with probability at least $(1 - \delta')$ over the choice of M_t, the set $B_{1:t}$ satisfies*

$$|B_{1:t}| \ge (1 - \delta') \cdot |B_{1:t-1}| \cdot |B_t|/q^n \,.$$

Proof. By the "boundedness implies uniformity" lemma (i.e., Lemma 5.17), $\mathbb{Z}_q^{(k-1)\alpha n}$, we have

$$(1 - \delta') \le \frac{\mathbb{E}_{x \sim \mathsf{Unif}(B_{1:t-1})}\left[f(A_{t,c_t} x)\right]}{\mathbb{E}_{z \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})}\left[f(z)\right]} \le (1 + \delta') \tag{5.20}$$

for every non-negative function f. Set f to be the indicator function of $B_{r,t}$ (recall that $B_{r,t}$ is the "reduced posterior set" from Definition 5.7) and apply Eq. (5.20), we have

$$\mathbb{E}_{x \sim \mathsf{Unif}(B_{1:t-1})}\left[\mathbf{1}_{B_{r,t}}(A_{t,c_t} x)\right] = \frac{|B_{1:t}|}{|B_{1:t-1}|} \,,$$

and

$$\mathbb{E}_{z \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})}\left[\mathbf{1}_{B_{r,t}}(z)\right] = \frac{|B_{r,t}|}{q^{(k-1)\alpha n}} = \frac{|B_t|}{q^n} \,.$$

We have

$$\begin{aligned}
\frac{|B_{1:t}|}{q^n} &= \frac{|B_{1:t-1}|}{q^n} \cdot \mathbb{E}_{x \sim \mathsf{Unif}(B_{1:t-1})}\left[\mathbf{1}_{B_{r,t}}(A_{t,c_t} x)\right] \\
&\ge (1 - \delta') \frac{|B_{1:t-1}|}{q^n} \cdot \mathbb{E}_{z \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})}\left[\mathbf{1}_{B_{r,t}}(z)\right] \\
&= (1 - \delta') \frac{|B_{1:t-1}|}{q^n} \cdot \frac{|B_t|}{q^n} \,.
\end{aligned}$$

Hence $|B_{1:t}| = |B_{1:t-1} \cap B_t| \ge (1-\delta') \cdot |B_{1:t-1}| \cdot |B_t|/q^n \ge q^{n-2tb} \ge q^{n-s}$, where the last inequality is due to $2Tb < s$. □

Thus, by invoking the "induction step" lemma (i.e., Lemma 5.18) on $B_{1:t-1}$ and B_t with $C = C_{t-1}$, there exists a constant C_t such that that $B_{1:t} = B_{1:t-1} \cap B_t$ is (C_t, s)-bounded with probability at least $1 - 4\delta'$. Namely, we have $\Pr[\overline{\mathcal{E}_t} \,|\, \mathcal{E}_{t-1}] \le 5\delta' \le \delta/T$. This satisfies condition (iii).

Finally, for every $t \in [T]$, if we set $\tau < \tau_0(q, k, \alpha, \delta', C_{t-1})$, by the "boundedness implies uniformity" lemma (i.e., Lemma 5.17), we know that $\|(A_{\mathbf{c}} \mathbf{x}^*) - U_t\|_{tvd} \le \delta'$ where $\mathbf{x}^* \sim \mathsf{Unif}(B_{1:t})$ $U_t \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$. As $S^Y_{1:t} = r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, A_{t,\mathbf{c}} \mathbf{x}^*)$ where $\mathbf{x}^* \sim \mathsf{Unif}(B_{1:t})$, by the data processing inequality we have $\|S^Y_t - r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, U_t)\|_{tvd} \le \delta/T$ as desired. This satisfied condition (iv).

To conclude, we set $\tau > 0$ to be a small constant that satisfies $2Tb < \tau_0 n$ and $\tau < \tau_0(q, k, \alpha, \delta', C_{t-1})$ for all $t \in [T]$ where C_t is inductively chosen as described above. This completes the proof of Lemma 5.1.

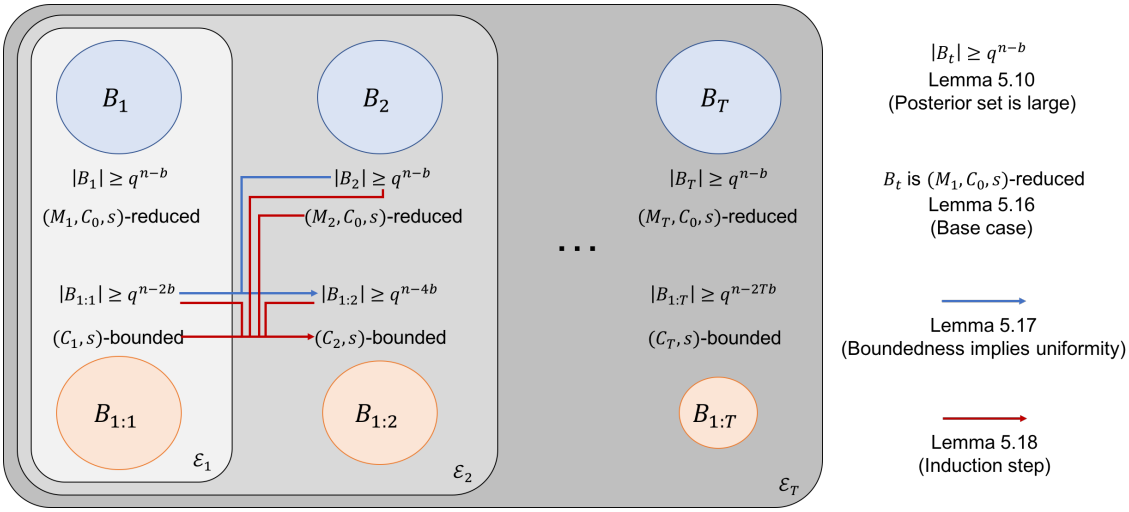

Figure 1: A pictorial overview of the proof of Lemma 5.1.

□

6 Analysis of Bounded Functions

In this section, we provide the complete proof for the three important lemmas in Section 5: the "base case" lemma (i.e., Lemma 5.16), the "boundedness implies uniformity" lemma (i.e., Lemma 5.17), and the "induction step" lemma (i.e., Lemma 5.18). We first establish useful structure on the Fourier coefficients of restricted sets (posterior set is a special case of restricted set) in Section 6.1. Next, we prove useful properties for the Fourier analytic conditions in Section 6.2. Finally, we prove the three lemmas in Section 6.4, Section 6.4, and Section 6.5 respectively.

6.1 Fourier coefficients of the posterior function

Given a k-hypermatching $M = (e_1, \ldots, e_m)$ and centers $\mathbf{c} = (c_1, \ldots, c_m)$ we say that a set $B \subseteq \mathbb{Z}_q^n$ is $(M, \mathbf{c})$-restricted if there exists a ("reduced") set $B_r \subseteq \mathbb{Z}_q^{(k-1)m}$ such that $B = \{\mathbf{x} \in \mathbb{Z}_q^n | A_{\mathbf{c}}\mathbf{x} \in B_r\}$, where $A_{\mathbf{c}}$ is the $\mathbf{c}$-centered folded encoding of M. In this section we aim to prove that large restricted sets are bounded. Recall that given a k-hypermatching $M = (e_1, \ldots, e_m)$ on vertex set $[n]$ with $m = \alpha n$ edges and sequence of centers $\mathbf{c} = (c_1, \ldots, c_m)$ with $c_i \in e_i \subseteq [n]$, the $\mathbf{c}$-centered folded representation of M was denoted $A_{\mathbf{c}} \in \mathbb{Z}_q^{(k-1)m \times n}$. We say that a set $B \subseteq \mathbb{Z}_q^n$ is $(M, \mathbf{c})$-restricted if there exists a ("reduced") set $B_r \subseteq \mathbb{Z}_q^{(k-1)m}$ such that $B = \{\mathbf{x} \in \mathbb{Z}_q^n | A_{\mathbf{c}}\mathbf{x} \in B_r\}$. For our next lemma we will also need a variant of this matrix named the $\mathbf{c}$-centered projection induced by M, which we denote $\widetilde{A}_{\mathbf{c}} \in \mathbb{Z}_q^{(k-1)m \times n}$, which is simply the matrix $A_{\mathbf{c}}$ with the columns corresponding to $c_1, \ldots, c_m$ zeroed out. (In $A_{\mathbf{c}}$ each of these columns has $(k-1)$ -1's. See Figure 2.) With this definition in place we can now relate the Fourier coefficients of the indicator of a restricted set to its image.

Recall that we use $\mathbf{e}_i \in \mathbb{Z}_q^n$ to denote the indicator vector of hyperedge e_i (see Section 3).

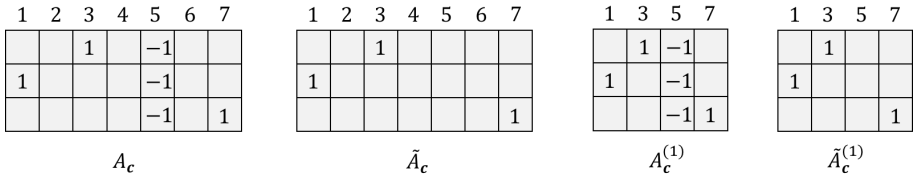

Figure 2: An example of $A_{\mathbf{c}}, \tilde{A}_{\mathbf{c}}, A_{\mathbf{c}}^{(1)}, \tilde{A}_{\mathbf{c}}^{(1)}$ with $m = 1$, $n = 7$, $k = 4$, $e_1 = \{1, 3, 5, 7\}$ and $c_1 = \{5\}$.

Lemma 6.1 (Fourier coefficients of the posterior function)**.** *Let M be a k-hypermatching of size m and $\mathbf{c}$ be a sequence of centers. Let $A_{\mathbf{c}}$ be the folded representation of M and $\widetilde{A}_{\mathbf{c}}$ be the projection induced by M. Furthermore, let $B \subseteq \mathbb{Z}_q^n$ be an $(M, \mathbf{c})$-restricted set with $B_r \in \mathbb{Z}_q^{(k-1)m}$ satisfying $B = \{\mathbf{x} \in \mathbb{Z}_q^n | A_{\mathbf{c}}\mathbf{x} \in B_r\}$. Let $\mathbf{1}_B$ denote the indicator function of B. Then for every $\mathbf{u} \in \mathbb{Z}_q^n$ we have:*

$$\widehat{\mathbf{1}_B}(\mathbf{u}) = \begin{cases} 0, & \text{if } \mathbf{u} \text{ contains a node not in } M. \\ 0, & \text{if } \exists i \in [m] \text{ such that } \langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \pmod q. \\ \widehat{\mathbf{1}}_{B_r}(\widetilde{A}_{\mathbf{c}}\mathbf{u}), & \text{otherwise.} \end{cases}$$

Proof. From the definition of the Fourier coefficient we have $\widehat{\mathbf{1}_B}(\mathbf{u}) = \frac{1}{q^n} \sum_{\mathbf{x} \in \mathbb{Z}_q^n} \mathbf{1}_B(\mathbf{x}) \omega^{\mathbf{u}^\top \mathbf{x}}$ where $\omega = e^{2\pi i/q}$ being the primitive q-th root of unity. Using the fact that B is restricted, we get

$$\widehat{\mathbf{1}_B}(\mathbf{u}) = \frac{1}{q^n} \sum_{\mathbf{x} \in \mathbb{Z}_q^n} \mathbf{1}_B(\mathbf{x}) \cdot \omega^{\mathbf{u}^\top \mathbf{x}} = \frac{1}{q^n} \sum_{\mathbf{z} \in \mathbb{Z}_q^{(k-1)m}} \sum_{\substack{\mathbf{x} \in \mathbb{Z}_q^n \\ A_{\mathbf{c}}\mathbf{x} = \mathbf{z}}} \mathbf{1}_{B_r}(\mathbf{z}) \cdot \omega^{\mathbf{u}^\top \mathbf{x}} = \frac{1}{q^n} \sum_{\mathbf{z} \in B_r} \sum_{\substack{\mathbf{x} \in \mathbb{Z}_q^n \\ A_{\mathbf{c}}\mathbf{x} = \mathbf{z}}} \omega^{\mathbf{u}^\top \mathbf{x}}.$$

We now fix $\mathbf{z} \in B_r$ and explore the final term $\sum_{\mathbf{x} \in \mathbb{Z}_q^n, A_{\mathbf{c}}\mathbf{x} = \mathbf{z}} \omega^{\mathbf{u}^\top \mathbf{x}}$. Let $\mathbf{z} = (\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(m)})$ where $\mathbf{z}^{(i)} \in \mathbb{Z}_q^{k-1}$. Also let $A_{\mathbf{c}}^{(1)}, \ldots, A_{\mathbf{c}}^{(m)} \in \mathbb{Z}_q^{(k-1) \times n}$ denote the blocks of $A_{\mathbf{c}}$ corresponding to the m edges. Now, think of $\mathbb{Z}_q^n$ as a free module over $\mathbb{Z}_q$ and consider the direct sum decomposition $\mathbb{Z}_q^n = W^{(0)} \oplus \cdots \oplus W^{(m)}$ where for $i \in [m]$, $W^{(i)}$ is the sub-module of $\mathbb{Z}_q^n$ generated by e_i, $W^{(0)}$ is the sub-module generated by $[n] - (\cup_i e_i)$, and "$\oplus$" denotes the direct sum of modules. Let us write $\mathbf{x} = \mathbf{x}^{(0)} + \cdots + \mathbf{x}^{(m)}$ where for $i \in \{0, \ldots, m\}$, $\mathbf{x}^{(i)} \in W^{(i)}$. Similarly write $\mathbf{u} = \mathbf{u}^{(0)} + \cdots + \mathbf{u}^{(m)}$. Since $(\mathbf{u}^{(i)})^\top \mathbf{x}^{(j)} = 0$ if $i \neq j$ we have $\mathbf{u}^\top \mathbf{x} = \sum_{i=0}^m (\mathbf{u}^{(i)})^\top \mathbf{x}^{(i)}$. Note also that $\mathbf{z} = A_{\mathbf{c}}\mathbf{x}$ if and only if $\mathbf{z}^{(i)} = A_{\mathbf{c}}^{(i)} \mathbf{x}^{(i)}$ for every $i \in [m]$. Using this notation, we have

$$\sum_{\substack{\mathbf{x} \in \mathbb{Z}_q^n \\ A_{\mathbf{c}}\mathbf{x} = \mathbf{z}}} \omega^{\mathbf{u}^\top \mathbf{x}} = \left(\sum_{\mathbf{x}^{(0)} \in W^{(0)}} \omega^{(\mathbf{u}^{(0)})^\top \mathbf{x}^{(0)}} \right) \cdot \prod_{i=1}^m \left(\sum_{\substack{\mathbf{x}^{(i)} \in W^{(i)} \\ A_{\mathbf{c}}^{(i)} \mathbf{x}^{(i)} = \mathbf{z}^{(i)}}} \omega^{(\mathbf{u}^{(i)})^\top \mathbf{x}^{(i)}} \right).$$

Now note that if $\mathbf{u}^{(0)} = 0$ then the first term is $|W^{(0)}| = q^{n-km}$, else it is zero. Similarly for $i \in [m]$, there are exactly q vectors $\mathbf{x}^{(i)} \in W^{(i)}$ such that $A_{\mathbf{c}}^{(i)} \mathbf{x}^{(i)} = \mathbf{z}^{(i)}$ (which are additive shifts of each other on coordinates in e_i). Concretely, these two solutions are of the form $(\tilde{A}_{\mathbf{c}}^{(i)})^\top \mathbf{z}^{(i)} + a^k$ for some $a \in \mathbb{Z}_q$. So we have

$$\sum_{\mathbf{x}^{(i)} \in W^{(i)} : A_{\mathbf{c}}^{(i)} \mathbf{x}^{(i)} = \mathbf{z}^{(i)}} \omega^{(\mathbf{u}^{(i)})^\top \mathbf{x}^{(i)}} = \sum_{a \in \mathbb{Z}_q} \omega^{(\mathbf{u}^{(i)})^\top (\tilde{A}_{\mathbf{c}}^{(i)})^\top \mathbf{z}^{(i)} + (\mathbf{u}^{(i)})^\top a^k} = \omega^{(\mathbf{u}^{(i)})^\top (\tilde{A}_{\mathbf{c}}^{(i)})^\top \mathbf{z}^{(i)}} \sum_{a \in \mathbb{Z}_q} \omega^{a \cdot \|\mathbf{u}^{(i)}\|_1}.$$

Moreover,

$$\sum_{a\in\mathbb{Z}_q} \omega^{a\cdot\|\mathbf{u}^{(i)}\|_1} = \begin{cases} 0 & \text{if } \|\mathbf{u}^{(i)}\|_1 \not\equiv 0 \bmod q \\ q & \text{otherwise.} \end{cases}.$$

Putting all the above together we get

$$\sum_{\mathbf{x}\in\mathbb{Z}_q^n A_{\mathbf{c}}\mathbf{x}=\mathbf{z}} \omega^{\mathbf{u}^\top\mathbf{x}} = \begin{cases} 0 & \text{if } \mathbf{u} \text{ contains a node not in } M. \\ 0 & \text{if } \exists i\in[m] \text{ such that } \langle \mathbf{u}, \mathbf{e}_i\rangle \not\equiv 0 \bmod q\,. \\ q^{n-(k-1)m}\cdot\omega^{(\widetilde{A}_{\mathbf{c}}\mathbf{u})^\top\mathbf{z}} & \text{otherwise.} \end{cases}$$

Summing up over all $\mathbf{z}\in\mathbb{Z}_q^{(k-1)m}$ and normalizing yields the lemma.

□

6.2 Properties of the Fourier analytic conditions

First, recall that we define a set being restricted in Definition 5.11. Observe that the restrictedness of a set is independent to the choice of centers.

Lemma 6.2 (Recentering)**.** *Let $\mathbf{c}=(c_1,\ldots,c_m)$ and $\mathbf{c}'=(c'_1,\ldots,c'_m)$ be two sequences of centers for the same matching M. Then a set $B\subseteq\mathbb{Z}_q^n$ is $(M,\mathbf{c})$-restricted if and only if it is $(M,\mathbf{c}')$-restricted.*

Proof. Let $e_i^{(t)} = ((e_i^{(t)})_1,\ldots,(e_i^{(t)})_k = c_t)$ (for $t=1,2,\ldots,m$) be the ordering of hyperedges corresponding to centering $\mathbf{c}$, and let $e'^{(t)}_i = ((e'^{(t)}_i)_1,\ldots,(e'^{(t)}_i)_k = c'_t)$.

Given a permutation $\pi:[k]\to[k]$, let P_π be a $(k-1)\times(k-1)$ matrix defined as follows: For $1\le i,j\le k-1$, let

$$(P_\pi)_{i,j} = \begin{cases} 1 & \text{if } j=\pi(i) \\ -1 & \text{if } j=\pi(k)\,. \\ 0 & \text{otherwise} \end{cases}$$

For $t=1,2,\ldots,m$, let $\pi_t:[k]\to[k]$ be the permutation defined by $(e'^{(t)}_i)_j = (e_i^{(t)})_{\pi(j)}$, and let $\pi'_t:[k]\to[k]$ be the permutation defined by $(e_i^{(t)})_j = (e'^{(t)}_i)_{\pi(j)}$. Then, it is not hard to see that $A_{\mathbf{c}'} = Q\cdot A_{\mathbf{c}}$ and $A_{\mathbf{c}} = Q'\cdot A_{\mathbf{c}'}$ where

$$Q = \begin{pmatrix} P_{\pi_1} & & & \\ & P_{\pi_2} & & \\ & & \ddots & \\ & & & P_{\pi_m} \end{pmatrix}, \qquad Q' = \begin{pmatrix} P_{\pi'_1} & & & \\ & P_{\pi'_2} & & \\ & & \ddots & \\ & & & P_{\pi'_m} \end{pmatrix},$$

and $QQ' = Q'Q = I$, the $(k-1)m\times(k-1)m$ identity matrix.

Now, suppose $B\subseteq\mathbb{F}_2^n$ is $(M,\mathbf{c})$-restricted. Let $B_r\subseteq\mathbb{Z}_q^{(k-1)m}$ be the corresponding reduced set satisfying $B=\{\mathbf{x}\in\mathbb{Z}_q^n \mid A_{\mathbf{c}}\mathbf{x}\in B_r\}$. Then, let $B'_r=\{Q\mathbf{y}\mid \mathbf{y}\in B_r\}$.

Note that if $\mathbf{x}\in B$, then $A_{\mathbf{c}'}\mathbf{x} = Q(A_{\mathbf{c}}\mathbf{x})\in B'_r$. Similarly, if $A_{\mathbf{c}'}\mathbf{x}\in B'_r$, then there is some $\mathbf{y}\in B_r$ such that $A_{\mathbf{c}'}\mathbf{x}=Q\mathbf{y}$, and so, $A_{\mathbf{c}}\mathbf{x} = Q'A_{\mathbf{c}'}\mathbf{x} = Q'Q\mathbf{y} = \mathbf{y}\in B_r$, implying that $\mathbf{x}\in B$. It follows that B is $(M,\mathbf{c}')$-restricted with reduced set B'_r.

In an analogous fashion, it follows that if B is $(M,\mathbf{c}')$-restricted, then B is also $(M,\mathbf{c})$-restricted. This completes the proof. □

Recall the notion of *bounded* sets from Definition 5.13.

Lemma 6.3. *If $C > e$, then $U_{C,s}(h)$ is monotonically increasing in $h \in [1, s]$.*

Proof. Recall that the function $f(x) = x^{1/x}$ is decreasing in the interval (e, ∞) (since $f'(x) = x^{1/x} \cdot \frac{1-\ln x}{x^2}$ is negative for $x > e$). Note that $U_{C,s}(h) = x^{\frac{C\sqrt{sn}}{2x}}$ for $x = \frac{C\sqrt{sn}}{h}$. Moreover, for $h \le s$, we have $x \ge \frac{C\sqrt{sn}}{s} \ge C > e$. Hence, it follows from monotonically decreasing property of f that $U_{C,s}(h)$ is monotonically increasing in the described interval, as desired. □

Recall the definition of a set being *reduced* from Definition 5.15. Note that if a set B is (M, C, s)-reduced then B is also (C, s)-bounded. Given these definitions, there are three tasks left to complete the proof of Lemma 5.1. First, show in Section 6.4 that if the aggregated posterior set $B_{1:T}$ is (C, s)-bounded for some $C = O(1)$ and $s = \Omega(n)$, then it will satisfy the conditions needed in Lemma 5.1 with high probability. Next, we show the posterior set $B_{1:T}$ is indeed bounded by induction. In Section 6.3 we analyze the base case and show that every posterior set B_t is (M_t, C', s)-reduced for some $C' = O(1)$. The induction step is much more involved where we show that the intersection of a (C, s)-bounded set (e.g., $B_{1:t}$) and a (M, C', s)-reduced set (e.g., B_{t+1}) is (C'', s)-bounded with high probability (over the randomness of M) for some $C'' = O(1)$. The analysis for the induction step is postponed to Section 6.5.

6.3 Proof of the "base case" lemma

In this subsection, we prove the "base case" lemma (Lemma 5.16), which shows that every posterior set B_t is (M_t, C, s)-reduced for some constant C. We include the statement again below for convenience.

Lemma 5.16 (Base case)**.** *For every $q, k \ge 2$, $\alpha \in (0, 1/k)$, there exists a constant C such that for every k-hypermatching M on $[n]$ of size $m \le \alpha n$, suppose $n \in \mathbb{N}$ is large enough and $0 < b \le s \le n/32$, then the following holds. Let $B \subseteq \mathbb{Z}_q^n$. If (i) there exists a sequence of centers $\mathbf{c}$ such that B is $(M, \mathbf{c})$-restricted and (ii) $|B| \ge q^{n-b}$, then B is (M, C, s)-reduced.*

To see how the above lemma connects to posterior sets, think of B as B_t, M as M_t, and $\mathbf{c}$ as $\mathbf{c}_t$. Note that condition (i) of Lemma 5.16 holds by the definition of B_t. As for condition (ii), it holds when the message S_t is *typical* and we know by averaging argument that this is the case with high probability (see Lemma 5.10 for more details).

To prove this lemma, we need to put together a few ingredients. First, Lemma 6.2 which tells us that being restricted is independent of the choice of centers. We also recall Parseval's lemma in this context.

Lemma 6.4. *For every $B \subseteq \mathbb{Z}_q^n$ we have $\sum_{\mathbf{v} \in \mathbb{Z}_q^n} \widehat{\mathbf{1}_B}(\mathbf{v})^2 \le |B|/q^n$.*

(Lemma 6.4 follows from Lemma 2.6 by noticing that $\sum_{\mathbf{a} \in \mathbb{Z}_q^n} \mathbf{1}_B(\mathbf{a}) = |B|$.) Recall that the (C, s)-bounded criterion bounds the sum of Fourier coefficients with a fixed weight at most s. As we also need to bound the sum of Fourier coefficients of high weight, this can be guaranteed from Parseval's inequality as shown in the following lemma.

Lemma 6.5. *Suppose $B \subset \mathbb{Z}_q^n$ satisfies $|B| \ge q^{n-b}$ for some $b \in \mathbb{N}$. Then, for every $\mathbf{v} \in \mathbb{Z}_q^n$ and $b < h \le n$, we have*

$$\sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{v}\|_0 = h}} \frac{q^n}{|B|} \left| \widehat{\mathbf{1}_B}(\mathbf{u}) \right| \le \left(\frac{2q^2 e^2 n}{h} \right)^{h/2}.$$

Proof. Note that by Lemma 6.4, we have $\sum_{\mathbf{u}\in S_h}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right|^2 \le |B|/q^n$. Using

$$\left|\{\mathbf{u}\in\mathbb{Z}_q^n \mid \|\mathbf{u}+\mathbf{v}\|_0 = h\}\right| \le (q-1)^h\cdot\binom{n}{h} \le \left(\frac{qen}{h}\right)^h$$

and the Cauchy-Schwarz inequality we get

$$\begin{aligned}
\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n\\ \|\mathbf{u}+\mathbf{v}\|_0=h}} \frac{q^n}{|B|}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| &\le \frac{q^n}{|B|}\sqrt{(|B|/q^n)\cdot(qen/h)^h} \\
&= \sqrt{(q^n/|B|)\cdot(qen/h)^h} \\
&\le \sqrt{q^b(qen/h)^h} && (\because |B|\ge q^{n-b}) \\
&\le \sqrt{q^h(qen/h)^h} && (\because h > b) \\
&= (2q^2e^2n/h)^{h/2}.
\end{aligned}$$

□

We now turn to the proof of Lemma 5.16. The overall proof follows the outline of [KK19], but we require extra care in our case, and the proof crucially depends on the ability to recenter (Lemma 6.2) and a slightly more careful probabilistic analysis.

Proof of Lemma 5.16. The first two conditions of (M,C,s)-reducedness are immediate corollaries of Lemma 6.1. So in the rest of the proof we focus on showing for every $h\in\{1,\dots,s\}$ and $\mathbf{v}\in\mathbb{Z}_q^n$,

$$\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n\\ \|\mathbf{u}+\mathbf{v}\|_0=h}} \frac{q^n}{|B|}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \le U_{C,s}(h)\,. \qquad \text{(Goal of Lemma 5.16)}$$

Fix an arbitrary $\mathbf{v}\in\mathbb{Z}_q^n$. For each $h\in\{1,\dots,s\}$, let $S_h = S_{\mathbf{v},h} = \{\mathbf{u}\mid\|\mathbf{u}+\mathbf{v}\|_0 = h\}$, i.e., the set of Fourier coefficients in the above summation.

Next, we are only interested in the subset of S_h that gives non-zero contribution in the summation. Recall by Lemma 6.1 that $\widehat{\mathbf{1}_B}(\mathbf{w}) = 0$ if $\mathsf{supp}(\mathbf{w})\not\subseteq\mathsf{supp}(M)$, or if there exists $i\in[\alpha n]$ such that $\langle\mathbf{w},\mathbf{e}_i\rangle\not\equiv 0 \bmod q$. Let

$$T_{\mathbf{v},h,M} = \{\mathbf{u}\in\mathbb{Z}_q^n \mid \|\mathbf{u}+\mathbf{v}\|_0 = h, \mathsf{supp}(\mathbf{u})\subseteq\mathsf{supp}(M), \langle\mathbf{u},\mathbf{e}_i\rangle\equiv 0 \bmod q\ \forall i\in[\alpha n]\},$$

denote this set of vectors which contains the non-zero Fourier coefficients. Roughly, our approach below is to (1) give an upper bound on the size of the set $T_{\mathbf{v},h,M}$ and (2) bound the sum of the squares of the coefficients in this set. Once we have both these bounds, we can apply Cauchy-Schwartz to conclude the desired bound. Before we undertake these steps, we make some simplifications and some refinements.

Step 0: Regular condition of v. First note that we can assume $\mathsf{supp}(\mathbf{v})\subseteq\mathsf{supp}(M)$. If this is not the case, consider the vector $\widetilde{\mathbf{v}}$ given by $\widetilde{v_i} = v_i$ if $i\in\mathsf{supp}(M)$ and $\widetilde{v_i} = 0$ otherwise. Also, let $a = |\{i\mid v_i\neq 0 \text{ and } i\notin\mathsf{supp}(M)\}|$ be the number of nodes in the support of $\mathbf{v}$ that are untouched by the matching M. Then note that $T_{\mathbf{v},h,M} = T_{\widetilde{\mathbf{v}},h-a,M}$. If we show that $(q^n/|B|)\cdot\sum_{\mathbf{u}\in T_{\widetilde{\mathbf{v}},h-a,M}}|\widehat{\mathbf{1}_B}(\mathbf{u})| \le U_{C,s}(h-a)$ then, by the monotonicity of $U_{C,s}(\cdot)$ in the interval $[1,s]$ (see Lemma 6.3), it follows that $(q^n/|B|)\cdot\sum_{\mathbf{u}\in T_{\mathbf{v},h,M}}|\widehat{\mathbf{1}_B}(\mathbf{u})| \le U_{C,s}(h)$. Thus, from now on, we assume $\mathsf{supp}(\mathbf{v})\subseteq\mathsf{supp}(M)$.

Step 1: A partition of $T_{\mathbf{v},h,M}$. We now further refine $T_{\mathbf{v},h,M}$, i.e., the set of non-zero Fourier coefficients. For an integer ℓ, let $T_{\mathbf{v},h,\ell,M} = \{\mathbf{u} \in T_{\mathbf{v},h,M} \mid \#\{i \in [\alpha n] \mid e_i \cap \mathsf{supp}(\mathbf{u}+\mathbf{v}) \neq \emptyset\} = \ell\}$ be the set $\mathbf{u} \in T_{\mathbf{v},h,\ell,M}$'s such that the support of $\mathbf{u}+\mathbf{v}$ touches exactly ℓ edges. Since $\mathbf{v}, h$ and M will be fixed in the rest of this proof, we simplify the notation and refer to this set as T_ℓ. Note that $h/k \leq \ell \leq h$. Thus, the quantity we are interested in this lemma can be upper bounded as follows.

$$\frac{q^n}{|B|}\sum_{\mathbf{u}\in S_h}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| = \frac{q^n}{|B|}\sum_{\mathbf{u}\in T_{\mathbf{v},h,M}}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| = \sum_{\ell=h/k}^{h}\frac{q^n}{|B|}\sum_{\mathbf{u}\in T_\ell}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \leq \sum_{\ell=h/k}^{h}\frac{q^n}{|B|}\sqrt{|T_\ell|\sum_{\mathbf{u}\in T_\ell}\widehat{\mathbf{1}_B}(\mathbf{u})^2}$$

where the first equality is due to Lemma 6.1, the second equality is due to the partition, and the last inequality is by Cauchy-Schwarz inequality. The reason why we partition $T_{\mathbf{v},h,M}$ into T_ℓ's is that the Fourier square-mass within T_ℓ and the cardinality of T_ℓ can be properly upper bounded respectively.

Step 2: Upper bound the Fourier square-mass within T_ℓ. To upper bound the Fourier square-mass within T_ℓ, we utilize the fact that the posterior set B is independent to the choice of center $\mathbf{c}$ (i.e., Lemma 6.2) and a hypercontractivity inequality. We remark that the fact that the exponent of b below is $h-\ell$ (as opposed to the more trivial h, or $h(k-1)/k$) is crucial to our proof and ensures we eventually get a linear space lower bound. In turn this bound is obtained by using a random center $\mathbf{c}$ and this randomization is permitted at the analysis stage by Lemma 6.2.

Claim 6.6. *There exists a constant ζ_1 such that if $|B| \geq q^{n-b}$ for some $b \in \mathbb{N}$ then for every $1 \leq \ell < h \leq b$, we have*

$$\sum_{\mathbf{u}\in T_\ell}\widehat{\mathbf{1}_B}(\mathbf{u})^2 \leq k^\ell\left(\frac{|B|}{q^n}\right)^2\left(\frac{\zeta_1\cdot b}{h-\ell}\right)^{h-\ell}.$$

Proof. The proof of this claim uses Lemma 6.1 to relate the Fourier coefficients of the function $\mathbf{1}_B$ to those of $\mathbf{1}_{B_r}$. But note that the "reduced set" B_r depends on the choice of the center. Furthermore the weight of the Fourier coefficient in the reduced space depends on how the centers overlap with $\mathsf{supp}(\mathbf{u}+\mathbf{v})$. Specifically we have that for centers $\mathbf{c}$, $\mathbf{c}$, $\|\widetilde{A}_{\mathbf{c}}(\mathbf{u}+\mathbf{v})\|_0 = \|\mathbf{u}+\mathbf{v}\|_0 - t$, where $t = |\{i \in [\alpha n] \mid c_i \in \mathbf{u}+\mathbf{v}\}|$ is the number of centers contained in $\mathbf{u}+\mathbf{v}$. Note that $t \leq \ell$ since the number centers in $\mathsf{supp}(\mathbf{u}+\mathbf{v})$ can not exceed the number of edges touching this set. The crux of the proof of this claim is that we if choose the centers randomly and then there is a positive probability that all centers (of the edges that touch $\mathsf{supp}(\mathbf{u}+\mathbf{v})$) are in $\mathsf{supp}(\mathbf{u}+\mathbf{v})$. We argue the formal details below.

For a random center $\mathbf{c}$, let $A_{\mathbf{c}}$ denote the $\mathbf{c}$-centered folded encoding of M, and let $B_{r,\mathbf{c}} = \{A_{\mathbf{c}}\mathbf{x} \mid \mathbf{x} \in B\} \subseteq \mathbb{Z}_q^{(k-1)\alpha n}$. Note that the random center $\mathbf{c}$ here is for the analysis purpose and in general might be different from the center used in the communication game. Nevertheless, due to Lemma 6.2 the choice of center does not affect the analysis. For $\mathbf{u} \in T_\ell$, let $I_{\mathbf{u}}(\mathbf{c}) = 1$ if $c_i \in \mathsf{supp}(\mathbf{u}+\mathbf{v})$ for every $i \in [\alpha n]$ with $e_i \cap \mathsf{supp}(\mathbf{u}+\mathbf{v}) \neq \emptyset$ and 0 otherwise. Note that

$\Pr_{\mathbf{c}}[I_{\mathbf{u}}(\mathbf{c}) = 1] \geq k^{-\ell}$. Now consider the following expression:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{c}}\left[\sum_{\substack{\mathbf{w}\in\mathbb{Z}_q^{(k-1)\alpha n}\\ \|\mathbf{w}+\tilde{A}_{\mathbf{c}}\cdot\mathbf{v}\|_0=h-\ell}} \widehat{\mathbf{1}_{B_{r,\mathbf{c}}}}(\mathbf{w})^2\right] &\geq \mathop{\mathbb{E}}_{\mathbf{c}}\left[\sum_{\mathbf{u}\in T_\ell} I_{\mathbf{u}}(\mathbf{c})\cdot \widehat{\mathbf{1}_B}(\mathbf{u})^2\right] && (\because \textit{Lemma 6.1})\\
&= \sum_{\mathbf{u}\in T_\ell}\left(\widehat{\mathbf{1}_B}(\mathbf{u})^2 \mathop{\mathbb{E}}_{\mathbf{c}}[I_{\mathbf{u}}(\mathbf{c})]\right) && (\because B \text{ and } T_\ell \text{ are independent to } \mathbf{c})\\
&\geq k^{-\ell}\sum_{\mathbf{u}\in T_\ell}\widehat{\mathbf{1}_B}(\mathbf{u})^2\,. && (\because \Pr_{\mathbf{c}}[I_{\mathbf{u}}(\mathbf{c})=1]\geq k^{-\ell})
\end{aligned}$$

On the other hand, note that $|B_{r\mathbf{c}}| = |B| \geq q^{n-b}$ and hence by Lemma 2.11, for every $\mathbf{c}$ we have

$$\sum_{\substack{\mathbf{w}\in\mathbb{Z}_q^{(k-1)\alpha n}\\ \|\mathbf{w}+\tilde{A}_{\mathbf{c}}\cdot\mathbf{v}\|_0=h-\ell}} \widehat{\mathbf{1}_{B_{r,\mathbf{c}}}}(\mathbf{w})^2 \leq \left(\frac{|B_{r,\mathbf{c}}|}{q^{(k-1)\alpha n}}\right)^2\left(\frac{4b}{h-\ell}\right)^{h-\ell},$$

where $\zeta_1 = \zeta$ is the constant from Lemma 2.11. Also, note that since B is $(M,\mathbf{c})$-restricted, we have $|B| = |\{\mathbf{x}\in\mathbb{Z}_q^n \mid A_{\mathbf{c}}\mathbf{x}\in B_{r,\mathbf{c}}\}| = |B_{r,\mathbf{c}}|\cdot q^{n-\mathsf{rank}(A_{\mathbf{c}})}$. As $\mathsf{rank}(A_{\mathbf{c}}) = (k-1)\alpha n$, the above inequality becomes

$$\sum_{\substack{\mathbf{w}\in\mathbb{Z}_q^{(k-1)\alpha n}\\ \|\mathbf{w}+\tilde{A}_{\mathbf{c}}\cdot\mathbf{v}\|_0=h-\ell}} \widehat{\mathbf{1}_{B_{r,\mathbf{c}}}}(\mathbf{w})^2 \leq \left(\frac{|B|}{q^n}\right)^2\left(\frac{\zeta_1\cdot b}{h-\ell}\right)^{h-\ell}.$$

Putting the two inequalities together we get

$$\sum_{\mathbf{u}\in T_\ell}\widehat{\mathbf{1}_B}(\mathbf{u})^2 \leq k^\ell\left(\frac{|B|}{q^n}\right)^2\left(\frac{\zeta_1\cdot b}{h-\ell}\right)^{h-\ell},$$

thus proving the claim. □

Step 3: Upper bound the cardinality of T_ℓ. Next, we turn to bounding the size of the set T_ℓ. To do so we explore the structure of the vectors in T_ℓ. We start with some notation. Let $E = \{i\in[\alpha n] \mid \langle\mathbf{v},\mathbf{e}_i\rangle \equiv 0 \pmod q\}$ and $O = \{i\in[\alpha n] \langle\mathbf{v},\mathbf{e}_i\rangle \not\equiv 0 \pmod q\}$. Given a vector $\mathbf{u}$, we define $W_e = W_e(\mathbf{u}) = (\mathbf{u}+\mathbf{v})\odot\left(\sum_{i\in E}\mathbf{e}_i\right)$ and $W_o = W_o(\mathbf{u}) = (\mathbf{u}+\mathbf{v})\odot\left(\sum_{i\in O}\mathbf{e}_i\right)$, where $\odot$ is used to denote the Hadamard product (entrywise product) of two vectors. Let η denote the number of edges touched by W_e and let o denote the number of edges touched by W_o. Note the following conditions hold when $\mathbf{u}\in T_\ell$.

Claim 6.7. *If $\mathbf{u}\in T_\ell$, then all the following conditions hold: (1) $|O|\leq h$, (2) $\eta + o = \ell$, and (3) $\eta \leq h/2$.*

Proof. We prove each of the individual claims below:

1. Note that $\langle\mathbf{u}+\mathbf{v},\mathbf{e}_i\rangle \not\equiv 0 \pmod q$ for every $i\in O$, since $\langle\mathbf{u},\mathbf{e}_i\rangle\equiv 0 \pmod q$ and $\langle\mathbf{v},\mathbf{e}_i\rangle\not\equiv 0 \pmod q$. Therefore, $|\mathsf{supp}(\mathbf{u}+\mathbf{v})\cap e_i| \geq 1$ for every $i\in O$, implying that $|O| \leq \sum_{i\in O}|\mathsf{supp}(\mathbf{u}+\mathbf{v})\cap e_i| \leq \|\mathbf{u}+\mathbf{v}\|_0 = h$.

2. Since $\mathbf{u}+\mathbf{v}$ touches ℓ edges, $\eta + o = \ell$.

3. Note that $\langle \mathbf{u}+\mathbf{v}, \mathbf{e}_i \rangle \equiv 0 \pmod q$ for every $i \in E$, since $\langle \mathbf{u}, \mathbf{e}_i \rangle \equiv 0 \pmod q$ and $\langle \mathbf{v}, \mathbf{e}_i \rangle \equiv 0 \pmod q$. Therefore, if $i \in E$ is touched by W_e (i.e., $W_e \odot \mathbf{e}_i \neq \mathbf{0}$), then it follows that W_e touches it in at least two points, i.e., $|\mathsf{supp}(W_e \odot \mathbf{e}_i)| \geq 2$ (see Figure 3). Combined with the fact that $|\mathsf{supp}(W_e)| \leq \|\mathbf{u}+\mathbf{v}\|_0 = h$, we obtain $\eta \leq h/2$, as desired.

□

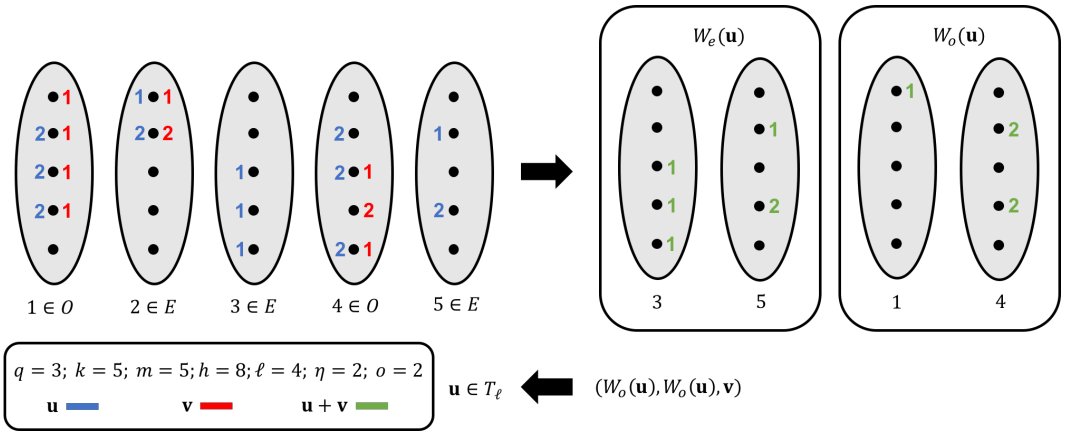

Figure 3: Upper bound the cardinality of T_ℓ. In this example, $q = 3, k = 5, m = 5$, and $\mathbf{v}$ is specified by integers in red. Note that $E = \{2, 3, 5\}$ and $O = \{1, 4\}$. Next, we consider $h = 8, \ell = 4$ and a $\mathbf{u} \in T_\ell$ specified by integers in blue. Note that by definition we have $\eta = o = 2$. In particular, the tuple $(W_e(\mathbf{u}), W_o(\mathbf{u}))$ is described on the right and $\mathbf{u}+\mathbf{v}$ is specified by integers in green. It is immediate to see that $(W_e(\mathbf{u}), W_o(\mathbf{u}), \mathbf{v})$ uniquely specifies $\mathbf{u}$ because one can subtract $W_e(\mathbf{u})$ and $W_o(\mathbf{u})$ by $\mathbf{v}$ to get the value of $\mathbf{u}$ in those coordinates. In the rest of the coordinates, $\mathbf{u}$ has the same values as $\mathbf{v}$. Moreover, observe that every hyperedges in $W_e(\mathbf{u})$ should contain at least 2 non-zero points because both $\mathbf{u}$ and $\mathbf{v}$ sum up to 0 mod q within those hyperedges.

We use these to bound $|T_\ell|$. Recall that each $\mathbf{u} \in T_\ell$ is uniquely specified by the pair (W_e, W_o) (see Figure 3) and it therefore suffices to count the number of distinct choices of (W_e, W_o). First, we see that the number of possibilities for W_o is at most $(q^k)^{|O|} \leq q^{kh}$ (by the first item of Claim 6.7). Now, having fixed W_o and o, consider the number of possibilities of W_e. We may choose W_e by picking a set $F \subseteq E$ with η edges, and then picking $|\mathsf{supp}(W_e)| = h - |\mathsf{supp}(W_o)|$ elements from the union of the edges in F, each of which is given a value in $\mathbb{Z}_q \setminus \{0\}$. Note that F can be chose in at most $\binom{|E|}{\eta} \leq \binom{\alpha n}{\eta}$ ways, after which W_e can be chosen in $\leq q^{k\eta}$ ways. Finally, note that by the second and the third item of Claim 6.7 we have $\eta \leq \min\{\ell, h/2\}$. Putting these together, we obtain

$$\begin{aligned} |T_\ell| &\leq \sum_{\eta=0}^{\min\{\ell, h/2\}} \left\{ q^{kh} \binom{\alpha n}{\eta} q^{k\eta} \right\} \\ &\leq \ell \cdot \max_{0 \leq \eta \leq \min\{\ell, h/2\}} \left\{ q^{kh} (e\alpha n/\eta)^\eta q^{k\eta} \right\} \\ &\leq \min_{\eta \in \{\ell, h/2\}} \left\{ q^{kh} (e\alpha n/\eta)^\eta q^{k\eta} \right\}. \end{aligned} \tag{6.8}$$

Step 4: Proof of Lemma 5.16. The boundedness of B now follows from some straightforward calculations. First, using the fact that $T_{\mathbf{v},h,M}$ contains all the non-zero Fourier coefficients of S_h and further partitioning it with T_ℓ, we have by Cauchy-Schwarz inequality

$$\frac{q^n}{|B|}\sum_{\mathbf{u}\in S_h}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| = \frac{q^n}{|B|}\sum_{\mathbf{u}\in T_{\mathbf{v},h,M}}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| = \sum_{\ell=h/k}^{h}\frac{q^n}{|B|}\sum_{\mathbf{u}\in T_\ell}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \le \sum_{\ell=h/k}^{h}\frac{q^n}{|B|}\sqrt{|T_\ell|\sum_{\mathbf{u}\in T_\ell}\widehat{\mathbf{1}_B}(\mathbf{u})^2}$$

By step 2 (i.e., Claim 6.6), step 3 (i.e., Equation 6.8), and the fact that $b \le s$, we further have

$$\le \sum_{\ell=h/k}^{h}\frac{q^n}{|B|}\sqrt{\min_{\eta\in\{\ell,h/2\}}\left\{q^{kh}(e\alpha n/\eta)^\eta q^{k\eta}\right\}k^\ell\left(\frac{|B|}{q^n}\right)^2\left(\frac{\zeta_1 s}{h-\ell}\right)^{h-\ell}}$$

Next, we consider two regimes of ℓ: (i) $h/k \le \ell \le h/2$ and (ii) $h/2+1 \le \ell \le h$.

$$\begin{aligned}
&\le \sum_{\ell=h/k}^{h/2}\sqrt{q^{kh}(e\alpha n/\ell)^\ell q^{k\ell}k^\ell\left(\frac{\zeta_1\cdot s}{h-\ell}\right)^{h-\ell}} + \sum_{\ell=h/2+1}^{h}\sqrt{q^{kh}(2e\alpha n/h)^{h/2}q^{kh/2}k^\ell\left(\frac{\zeta_1 s}{h/2}\right)^{h/2}}\\
&\le \sum_{\ell=h/k}^{h/2}\sqrt{q^{kh}(e\alpha n/\ell)^\ell 2^{k\ell}k^\ell\left(\frac{2\zeta_1 s}{h}\right)^{h-\ell}} + (h/2)(ns/h^2)^{h/4}(2^{3k}k^2\zeta_1^2 e)^{h/4}\\
&\le (q^{3k}ek^2)^{h/4}(ns/h^2)^{h/4}\sum_{\ell=k/h}^{h/2}\sqrt{\left(\frac{2\zeta_1 s}{n}\right)^{h/2-\ell}} + (h/2)(ns/h^2)^{h/4}(2^{3k}k^2\zeta_1^2 e)^{h/4}\\
&\le (C\sqrt{ns}/h)^{h/2}\\
&= U_{C,s}(h),
\end{aligned}$$

for $C \ge 2\zeta_1^2 ek^2q^{3k}$. Thus, we conclude that B is (M,C,s)-reduced. □

6.4 Proof: Boundedness implies near uniformity

In this section we prove Lemma 5.17 which is used to prove condition (iv) of Lemma 5.1. We restate the lemma below for convenience.

Lemma 5.17 (Boundedness implies closeness to uniformity)**.** *For every $q,k \ge 2$ and $\delta \in (0,1/2)$, there exists $\alpha_0 = \alpha_0(k,q)$ such that for every $\alpha \in (0,\alpha_0)$, $C > 0$, there exists $\tau_0 = \tau_0(q,k,\alpha,\delta,C)$ such that the following holds for any $\tau \in (0,\tau_0)$ and sufficiently large n:*

Let $B \subset \mathbb{Z}_q^n$ be a (C,s)-bounded set with $|B| \ge q^{n-b}$, for $4\log(3/\delta) \le b \le s \le \tau n$. Let M be a random k-hypermatching of size αn and $\mathbf{c}$ be a sequence of centers for M and let $A_{\mathbf{c}}$ denote the $\mathbf{c}$ centered folded encoding and M. Then, with probability at least $1-\delta$ over the choice of M, we have that for every $\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)\alpha n}$, we have

$$1-\delta < q^{(k-1)\alpha n}\Pr_{\mathbf{x}\sim \mathit{Unif}(B)}[A_{\mathbf{c}}\mathbf{x} = \mathbf{z}_0] < 1+\delta\,.$$

As a consequence, we also have

1. $\|(A_{\mathbf{c}}\mathbf{x}) - U\|_{tvd} \le \delta$ *where* $\mathbf{x} \sim \mathit{Unif}(B)$ *and* $U \sim \mathit{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$.

2. For every non-negative function f over $\mathbb{Z}_q^{(k-1)\alpha n}$,

$$(1-\delta) \leq \frac{\mathbb{E}_{\mathbf{x}\sim \mathit{Unif}(B)}\left[f(A_c\mathbf{x})\right]}{\mathbb{E}_{\mathbf{z}\sim \mathit{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})}\left[f(\mathbf{z})\right]} \leq (1+\delta)\,.$$

In the following, we denote $m = \alpha n$ for simplicity. Let us start with defining a combinatorial quantity $p(h,k,m,n)$ and showing an upper bound.

Definition 6.9. *Suppose $k, m, n > 0$ are integers. We define $p(h,k,m,n)$ to be the probability that, for a uniformly random k-hypermatching M with m edges, h fixed nodes in $[n]$, say $[h]$, are contained in M and further satisfy the condition that the intersection size of each hyperedge of M with these fixed nodes is either zero or at least two.*

Lemma 6.10 ([**?**, Lemma 6.8])**.** *For all integers n, k, $\alpha \in (0, 1/100k)$, $m = \alpha n$, and $0 \leq h \leq m$, we have*

$$p(h,k,m,n) \leq \frac{(4h^{1/2}k\alpha^{1/k})^h}{n^{h/2}}\,.$$

For completeness we include the proof from [**?**].

Proof. Let S_h denote the set $\{1, 2, \ldots, h\}$. The probability event we are interested in is captured by two conditions: (i) M containing S_h and (ii) each hyperedge of M either does not touch any vertex in S_h or touches at least 2 vertices in S_h. Thus,

$$p(h,n) \leq \Pr_M[M \text{ satisfies (i) and (ii)}]\,.$$

Now, let us denote the number of size t k-uniform hypermatchings over n vertices (assuming $tk \leq n$) as

$$M_k(n,t) = \frac{n^{\underline{tk}}}{(k!)^t \cdot t!}$$

where $n^{\underline{tk}} = n \cdot (n-1) \cdots (n-tk+1)$ denotes the downward factorial. Next, let u be the number of hyperedges that touch S_h and use u to partition the probability space. Observe that due to condition (ii), u only takes value within h/k to $h/2$. Thus, the number of M satisfying (i) and (ii) is at most

$$\sum_{u=h/k}^{h/2} \binom{n-h}{uk-h} \cdot M_k(uk,u) \cdot M_k(n-uk, \alpha n - u) \tag{6.11}$$

where $\binom{n-h}{uk-h}$ is the set of vertices outside S_h that are touched by the hyperedges that touch S_h, $M(uk,u)$ counts the number of size u k-uniform hypermatchings that touch S_h, and $M(n-uk, \alpha n - u)$ counts the number of ways to choose the remaining part (that does not touch S_h) of the k-uniform hypermatching.

Note that Equation 6.11 is an overestimation. Nevertheless, in the following we show that this

is sufficient for the desired upper bound.

$$\begin{aligned} p(h,n) &\le \frac{\sum_{u=h/k}^{h/2} \binom{n-h}{uk-h} \cdot M_k(uk,u) \cdot M_k(n-uk,\alpha n-u)}{M_k(n,\alpha n)} \\ &\le \sum_{u=h/k}^{h/2} \frac{n^{uk-h}}{(uk-h)!} \cdot \frac{\frac{(uk)!}{(k!)^u(u!)} \cdot \frac{n^{\underline{k(\alpha n-u)}}}{(k!)^{\alpha n-u}(\alpha n-u)!}}{\frac{n^{\underline{\alpha k n}}}{(k!)^{\alpha n}(\alpha n!)}} \\ &\le \sum_{u=h/k}^{h/2} \frac{(uk)!}{(uk-h)!} \cdot \frac{n^{uk-h} \cdot n^{\underline{\alpha k n-uk}}}{n^{\underline{\alpha k n}}} \cdot \frac{(\alpha n)!}{(\alpha n-u)!u!} \\ &\le \sum_{u=h/k}^{h/2} (uk)^h \cdot \left(\frac{1}{1-\alpha k}\right)^{uk} \cdot \frac{1}{n^h} \cdot \frac{(\alpha n)^u}{u!} \end{aligned}$$

As $\alpha k \le 1/100$, the equation can be further simplified as follows.

$$\begin{aligned} &\le \sum_{u=h/k}^{h/2} (2k)^{u\alpha^{1/k}k} u^{h-u} n^{u-h} \\ &\le \frac{(4h^{1/2}k\alpha^{1/k})^h}{n^{h/2}} . \end{aligned}$$

This completes the proof of Lemma 6.10 □

The following lemma is an immediate corollary of Lemma 6.10 and will be useful later in the proof of Lemma 5.17.

Lemma 6.12. *Suppose $k, q \ge 2$, $\delta \in (0, 1/2)$, and $C > 0$. Then, there exist $\alpha_0 = \alpha_0(k,q)$ and $\tau_0 = \tau_0(C,k)$ such that for every $\alpha \in (0, \alpha_0]$ and $\tau \in (0, \tau_0/\alpha^{4/k})$, it follows that for every sufficiently large n and $4\log(3/\delta) \le s \le \tau n$, $m = \alpha n$,*

$$p(h,k,m,n)U_{C,s}(h) \le \begin{cases} \delta^{2h} & , \ 1 \le h \le s \\ 2^{-h/2} & , \ s < h \le km \\ 0 & , \ h > km \end{cases} .$$

Specifically,

$$\sum_{h=2}^{n} p(h,k,m,n)U_{C,s}(h) \le \delta^2 .$$

Proof. Let us choose $\alpha_0 = 1/(8eqk)^k$ and $\tau_0 = (\delta^4/16k^2C)^2$. In particular, note that $\alpha_0 < 1/100k$, which implies that Lemma 6.10 provides an upper bound on $p(h,k,m,n)$. Thus, we can prove the first inequality for each interval using Lemma 6.10 and (5.12).

- If $1 \le h \le s$, then

$$p(h,k,m,n)U_{C,s}(h) \le \frac{(4h^{1/2}k\alpha^{1/k})^h}{n^{h/2}} \cdot \left(\frac{C\sqrt{sn}}{h}\right)^{h/2} = \left(\frac{16k^2\alpha^{2/k} \cdot C\sqrt{s}}{\sqrt{n}}\right)^{h/2} .$$

  Since $s \le \tau n$ and $\tau < \tau_0/\alpha^{4/k}$, we have $p(h,k,m,n)U_{C,s}(h) \le \delta^{2h}$, as desired.

- If $s < h \le m$, then

$$p(h,k,m,n)U_{C,s}(h) \le \frac{(4h^{1/2}k\alpha^{1/k})^h}{n^{h/2}} \cdot \left(\frac{2q^2e^2n}{h}\right)^{h/2} \le (32e^2q^2k^2\alpha^{2/k})^{h/2}.$$

  Since $\alpha \le \alpha_0 = 1/(8eqk)^k$, we have $p(h,k,m,n)U_{C,s}(h) \le 2^{-h/2}$, as desired.

- If $h > m$, we have $p(h,k,m,n) = 0$ and hence $p(h,k,m,n)U_{C,s}(h) = 0$.

Finally, we have

$$\begin{aligned}\sum_{h=2}^{n} p(h,k,m,n)U_{C,s}(h) &\le \sum_{h=2}^{s}\delta^{2h} + \sum_{h=s+1}^{km} 2^{-h/2}\\ &\le \frac{\delta^4}{1-\delta^2} + \frac{2^{-s-1/2}}{1-(1/\sqrt{2})}\\ &\le \frac{\delta^2}{2} + \frac{\delta^2}{2}\\ &= \delta^2,\end{aligned}$$

since $s \ge 4\log(3/\delta)$ and $\delta < 1/2$. □

Now, we are ready to prove the main lemma of this subsection.

Proof of Lemma 5.17. Let $q, k \ge 2$ and let $\delta \in (0, 1/2)$. Let us choose $\alpha_0 = \alpha_0(k,q)$ as in Lemma 6.12. Moreover, given any $C > 0$, let us choose $\tau_0 = \tau_0(C,k)$ as in Lemma 6.12.

Let $4\log(3/\delta) \le b \le s \le \tau n$, and let $B \subset \mathbb{Z}_q^n$ be a (C,s)-bounded set with $|B| \ge q^{n-b}$. The goal is to prove that with probability at least $1-\delta$

$$1 - \delta \le q^{(k-1)m} \Pr_{\mathbf{x}\sim\mathsf{Unif}(B)}[A_{\mathbf{c}}\mathbf{x} = -\mathbf{z}_0] \le 1 + \delta$$

for every $\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)\alpha n}$. (Note that the switch from $\mathbf{z}_0$ to $-\mathbf{z}_0$ in the event described above does not alter the statement being proved since we are proving this for every vector $\mathbf{z}_0$.)

Now, for a fixed k-hypermatching M and fixed choice of centers $\mathbf{c}$, let us expand the marginal probability as follows. Define $g_{A_{\mathbf{c}},\mathbf{z}_0}(\mathbf{x}) = \mathbf{1}_{A_{\mathbf{c}}\mathbf{x}=\mathbf{z}_0}$ and f to be the indicator function of the set B. For any fixed $\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)m}$, letting $g = g_{A_{\mathbf{c}},\mathbf{z}_0}$, we have

$$\begin{aligned}q^{(k-1)m} \Pr_{\mathbf{x}\sim\mathsf{Unif}(B)}[A_{\mathbf{c}}\mathbf{x} = -\mathbf{z}_0] &= \frac{q^{(k-1)m}}{|B|}\sum_{\mathbf{x}\in\mathbb{Z}_q^n} f(\mathbf{x})g(-\mathbf{x})\\ &= \frac{q^{(k-1)m}}{|B|}(f \star g)(0)\\ &= \frac{q^{(k-1)m}}{|B|}\sum_{\mathbf{u}\in\mathbb{Z}_q^n}\widehat{f\star g}(\mathbf{u})\\ &= \frac{q^{(k-1)m+n}}{|B|}\sum_{\mathbf{u}\in\mathbb{Z}_q^n}\widehat{f}(\mathbf{u})\widehat{g}(\mathbf{u}) \quad \text{(By Lemma 2.8)}\\ &= 1 + \frac{q^{n+(k-1)m}}{|B|}\sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n\\ \mathbf{u}\neq 0^n}}\widehat{f}(\mathbf{u})\widehat{g}(\mathbf{u}) \quad \text{(Since } q^n\widehat{f}(0) = |B| \text{ and } q^n\widehat{g}(0) = q^{n-(k-1)m}\text{).}\end{aligned} \tag{6.13}$$

We now analyze the Fourier coefficients of g and use this to bound the right hand side above.

Claim 6.14. *Let M be a k-hypermatching of size m, $\mathbf{c}$ be centers, and $z_0 \in \mathbb{Z}_q^{(k-1)m}$. Let $g(\mathbf{x}) = \mathbf{1}_{A_\mathbf{c}\mathbf{x}=\mathbf{z}_0}$. For every $\mathbf{u} \in \mathbb{Z}_q^n$, the following conditions hold:*

1. *If $\mathbf{u}$ is not perfectly matched by M, then $\widehat{g}(\mathbf{u}) = 0$.*
2. *If there exists $i \in [m]$ such that $\langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \bmod q$ where e_i is the i-th hyperedge of M, then $\widehat{g}(\mathbf{u}) = 0$.*
3. *$|\widehat{g}(\mathbf{u})| \le q^{-(k-1)m}$.*

Proof of Claim 6.14. Recall that $q^n \widehat{g}(\mathbf{u}) = \sum_\mathbf{x} g(\mathbf{x}) \omega^{\mathbf{u}^\top \mathbf{x}}$.

1. If $\mathbf{u}$ is not perfectly covered by M, then there exists $i \in [n]$ such that $u_i \neq 0$ but the i-th column of $A_\mathbf{c}$ is zero. For each $\mathbf{x} \in \mathbb{Z}_q^n$, for every $a \in \mathbb{Z}_q$ we have $g(\mathbf{x}) = g(\mathbf{x} + a\boldsymbol{\delta}_i)$, where $\boldsymbol{\delta}_i \in \mathbb{Z}_q^n$ denotes the coordinate vector in the ith direction (i.e., $\boldsymbol{\delta}_i = 0^{i-1}10^{n-i}$). Also, note that $\sum_{a \in \mathbb{Z}_q} \omega^{\mathbf{u}^\top(\mathbf{x}+a\boldsymbol{\delta}_i)} = \omega^{\mathbf{u}^\top \mathbf{x}} \sum_{a \in \mathbb{Z}_q} \omega^{u_i \cdot a} = 0$. This implies $\hat{g}(\mathbf{u}) = 0$.
2. Suppose $\langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \bmod q$. For each $\mathbf{x} \in \mathbb{Z}_q^n$ and $a \in \mathbb{Z}_q$, note that $g(\mathbf{x}) = g(\mathbf{x} + a\mathbf{e}_i)$ because a^k lies in the kernel of the folded matrix of this hyperedge. Second, since $\langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \bmod q$, we have $\sum_{a \in \mathbb{Z}_q} \omega^{\mathbf{u}^\top \mathbf{x} + a\mathbf{e}_i} = \omega^{\mathbf{u}^\top \mathbf{x}} \sum_{a \in \mathbb{Z}_q} \omega^{a \cdot \langle \mathbf{u}, \mathbf{e}_i \rangle} = 0$. This implies $\hat{g}(\mathbf{u}) = 0$.
3. By definition, we have $q^n \widehat{g}(\mathbf{u}) = \sum_\mathbf{x} \mathbf{1}_{A_\mathbf{c}\mathbf{x}=z_0} \omega^{\mathbf{u}^\top \mathbf{x}}$. Note that for fixed $M, \mathbf{c}, z_0$, there are at most $q^{n-(k-1)m}$ $\mathbf{x}$ such that $g(\mathbf{x}) = 1$. Thus, we have $|\widehat{g}(\mathbf{u})| \le q^{-(k-1)m}$ as desired.

□

Now, we can use Claim 6.14 to further upper bound Equation 6.13 as follows. Recall that $\odot$ stands for the coordinate-wise product of vectors.

$$\frac{q^{n+(k-1)m}}{|B|} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \mathbf{u} \neq 0^n}} \widehat{f}(\mathbf{u}) \widehat{g}(\mathbf{u}) \le \frac{q^n}{|B|} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \mathbf{u} \neq 0^n \\ \mathbf{u} \text{ is matched by } M \\ \|\mathbf{u} \odot \mathbf{e}_i\|_1 \equiv 0 \bmod q \ \forall i \in [m]}} |\widehat{f}(\mathbf{u})| \,.$$

One key thing here is that the above bound is independent of $\mathbf{z}_0$ an therefore holds even if we take the maximum of the left hand side over all $\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)m}$. We thus get, for every M and $\mathbf{c}$:

$$\max_{\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)m}} \left| q^{(k-1)m} \Pr_{\mathbf{x} \sim \mathsf{Unif}(B)} [A_\mathbf{c}\mathbf{x} = -\mathbf{z}_0] - 1 \right| \le \frac{q^n}{|B|} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \mathbf{u} \neq 0^n \\ \mathbf{u} \text{ is contained in } M \\ \|\mathbf{u} \odot \mathbf{e}_i\|_1 \equiv 0 \bmod q \ \forall i \in [m]}} |\widehat{f}(\mathbf{u})| \,.$$

Finally, let us take the expectation of the above quantity over the randomness of M and $\mathbf{c}$.

$$\mathop{\mathbb{E}}_{M,\mathbf{c}} \left[\max_{\mathbf{z}_0 \in \mathbb{Z}_q^{(k-1)m}} \left| q^{(k-1)m} \Pr_{\mathbf{x} \sim \mathsf{Unif}(B)} [A_\mathbf{c}\mathbf{x} = -\mathbf{z}_0] - 1 \right| \right] \le \mathop{\mathbb{E}}_{M,\mathbf{c}} \left[\frac{q^n}{|B|} \sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \mathbf{u} \neq 0^n \\ \mathbf{u} \text{ is contained in } M \\ \|\mathbf{u} \odot \mathbf{e}_i\|_1 \equiv 0 \bmod q \ \forall i \in [m]}} |\widehat{f}(\mathbf{u})| \right]$$

Next, we partition the summation according to the ℓ_0-norm of the Fourier coefficients.

$$\leq \sum_{h=1}^{n} \mathop{\mathbb{E}}_{M,\mathbf{c}} \left[\frac{q^n}{|B|} \sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0=h \\ \mathbf{u} \text{ is contained in } M \\ \|\mathbf{u}\odot\mathbf{e}_i\|_1\equiv 0 \bmod q \ \forall i\in[m]}} |\widehat{f}(\mathbf{u})| \right]$$

Observe that the event $\|\mathbf{u} \odot \mathbf{e}_i\|_1 \equiv 0 \pmod q$ implies that either $\|\mathbf{u} \odot \mathbf{e}_i\|_0 = 0$ or $\|\mathbf{u} \odot \mathbf{e}_i\|_0 \geq 2$ holds. Hence, the above summation can be replaced with a summation beginning at $h = 2$, and the equation becomes

$$\leq \sum_{h=2}^{n} \mathop{\mathbb{E}}_{M,\mathbf{c}} \left[\frac{q^n}{|B|} \sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0=h \\ \mathbf{u} \text{ is contained in } M \\ \|\mathbf{u}\odot\mathbf{e}_i\|_0=0 \text{ or } \|\mathbf{u}\odot\mathbf{e}_i\|_0\geq 2 \ \forall i\in[m]}} |\widehat{f}(\mathbf{u})| \right] \leq \sum_{h=2}^{n} p(h,k,m,n) \frac{q^n}{|B|} \sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ \|\mathbf{u}\|_0=h}} |\widehat{f}(\mathbf{u})|$$

When B is (C, s)-bounded, we can further upper bound the above quantity as follows.

$$\leq \sum_{h=2}^{n} p(h,k,m,n) \cdot U_{C,s}(h) \leq \delta^2,$$

where the last inequality is due to Lemma 6.12. Thus, when $\mathbf{x} \sim \mathsf{Unif}(B)$ and $U \sim \mathsf{Unif}(\mathbb{Z}_q^{(k-1)\alpha n})$, we have

$$\mathop{\mathbb{E}}_{M,\mathbf{c}} \left[\max_{z_0\in\mathbb{Z}_q^{(k-1)m}} \left| q^{(k-1)m} \Pr_{\mathbf{x}\sim\mathsf{Unif}(B)}[A_{\mathbf{c}}\mathbf{x} = -\mathbf{z}_0] - 1 \right| \right] \leq \delta^2 .$$

By Markov's inequality, we have

$$\max_{\mathbf{z}_0\in\mathbb{Z}_q^{(k-1)m}} \left| q^{(k-1)m} \Pr_{\mathbf{x}\sim\mathsf{Unif}(B)}[A_{\mathbf{c}}\mathbf{x} = -\mathbf{z}_0] - 1 \right| \leq \delta$$

with probability at least $1 - \delta$. This yields the main part of the lemma. The consequences follow directly from the main part (since pointwise bounds on the distance between distributions imply total variation distance as well as expectation of a non-negative weight). This completes the proof of Lemma 5.17. □

6.5 Proof of the "induction step" lemma

The goal of this section is to prove the "induction step" lemma. By Markov's inequality, it suffices to prove the following lemma which is the expectation version of Lemma 5.18. We first show that how Lemma 6.15 imply Lemma 5.18 and then focus on proving the former in the rest of this subsection.

Lemma 6.15 (Induction step in expectation)**.** *For every $q, k \in \mathbb{N}$ there exist $\alpha_0 \in (0, 1/k]$ and $C_0 > 0$ such that for every $\alpha \in (0, \alpha_0]$, $C > C_0$, $\delta \in (0, 1/2)$, there exist $C' > 0$, $n_0 \in \mathbb{N}$,*

and $\tau \in (0,1)$ such that the following holds. For every $n \geq n_0$, every $0 < b, b', s < \tau n$, and every $B \subset \mathbb{Z}_q^n$ that satisfies $|B| \geq q^{n-b}$ and is (C,s)-bounded, let M be a uniformly random k-hypermatching of size at most αn and $B' \subset \mathbb{Z}_q^n$ be (M, C_0, s)-reduced and $|B'| \geq q^{n-b'}$. If $|B \cap B'| \geq (1-\delta) \cdot |B| \cdot |B'|/q^n \geq q^{n-s}$, then for every $h \in [s]$, we have

$$\sum_{\mathbf{u} \in \mathbb{Z}_q^n} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \mathop{\mathbb{E}}_M \left[\sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')\right| \right] \leq U_{C',s}(h)\,.$$

Proof of Lemma 5.18 using Lemma 6.15. For every $h \in \{1, \ldots, s\}$, by the convolution theorem (see Lemma 2.9) for Fourier coefficients and triangle inequality, we have

$$\begin{aligned}
\sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} \frac{q^n}{|B \cap B'|} \left|\widehat{\mathbf{1}_{B \cap B'}}(\mathbf{u})\right| &\leq \frac{q^n}{|B \cap B'|} \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \left|\widehat{\mathbf{1}_B(\mathbf{u})\mathbf{1}_{B'}(\mathbf{u}')}\right| \\
&\leq \frac{|B| \cdot |B'|}{q^n \cdot |B \cap B'|} \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')\right| \\
&\leq \frac{1}{1-\delta} \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')\right| .
\end{aligned}$$

For $h \in [s]$, let $F(h)$ denote the event that the random matching M is such that

$$\frac{1}{1-\delta} \sum_{\mathbf{u} \in \mathbb{Z}_q^n} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')\right| > \frac{1}{\delta^h} \cdot U_{C',s}(h) = U_{C'/\delta^2,s}(h).$$

Let $F = \cup_{h \in [s]} F(h)$ be the union of these events. Note that if F does not hold, then $B \cap B'$ is $(C'/\delta^2, s)$-bounded. An application of Markov's inequality to Lemma 6.15 yields that

$$\Pr[F(h)] \leq \frac{U_{C',s}(h)}{(1-\delta) \cdot U_{C'/\delta^2,s}(h)} \leq \frac{\delta^h}{1-\delta}.$$

We thus get $\Pr[F] \leq \sum_{h \in [s]} F(h) \leq \frac{1}{1-\delta} \sum_h \delta^h \leq 4\delta$ where the final step uses the fact that $\delta < 1/2$. We conclude that $B \cap B'$ is $(C'/\delta^2, s)$-bounded with probability at least $1 - 4\delta$ over the randomness of M.

□

Now we prove Lemma 6.15 in the following three steps.

Proof of Lemma 6.15. At a high level, in order to prove Lemma 6.15, we split the inner sum over $\mathbf{u}'$ according to the combinatorial structure of the underlying set of vertices (viewing $\mathbf{u}'$ as an indicator vector). It turns out that the combinatorial structure nicely convolves with the boundedness parameters and helps us complete the proof.

Step 1: Partitioning the inner sum via a combinatorial structure. We start by defining the following combinatorial quantity, based on intersection properties of a random k-hypermatching.

Definition 6.16. *Let $n, q, k, u \in \mathbb{N}$ and $\alpha \in (0, 1/k)$. Let $\mathbf{u} \in (\mathbb{Z}_q \setminus \{0\})^u \times 0^{n-u}$ be a vector that is non-zero on exactly the first u coordinates. For a k-hypermatching M of size m, let $K_{\mathbf{u}}(M) := \{i \in [m] \,|\, \langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \pmod q\}$ be the set of edges with "odd intersection" (formally non-zero inner product mod q) with $\mathbf{u}$. Let $E_{\mathbf{u}}(M) := \{j \in [n] \mid u_j \neq 0, \exists i \notin K_{\mathbf{u}}(M),\ j \in e_i\}$ denote the set of vertices in the support of $\mathbf{u}$ that are in "even" edges.[5] Finally, let $O_{\mathbf{u}}(M) := \{j \in [n] \mid u_j \neq 0, \exists i \in K_{\mathbf{u}}(M),\ j \in e_i\}$ be the vertices in the support of $\mathbf{u}$ from odd edges. For $o, \eta, \kappa \in \mathbb{N}$, we define*

$$p_q(n, u, o, \eta, \kappa) := \max_{\mathbf{u} \in (\mathbb{Z}_q \setminus \{0\})^u \times 0^{n-u}} \Pr_M \left[|K_{\mathbf{u}}(M)| = \kappa, |E_{\mathbf{u}}(M)| = \eta, |O_{\mathbf{u}}(M)| = o \right], \tag{6.17}$$

where M is a uniformly random k-hypermatching of size αn. (In other words p_q is the maximum probability of a vector $\mathbf{u}$ of support size u having κ odd edges, η even vertices and o odd vertices when the matching M is drawn at random.)

Fig. 4 illustrates some of the parameters in the definition above. We remark that $p_q(\cdots)$ should not be confused with the function $p(\cdots)$ defined in Definition 6.9, which is a similar combinatorial quantity but not the same.

Note that as each edge in $K_{\mathbf{u}}(M)$ contributes at least one element to $O_{\mathbf{u}}(M)$, we have $o \geq \kappa$.

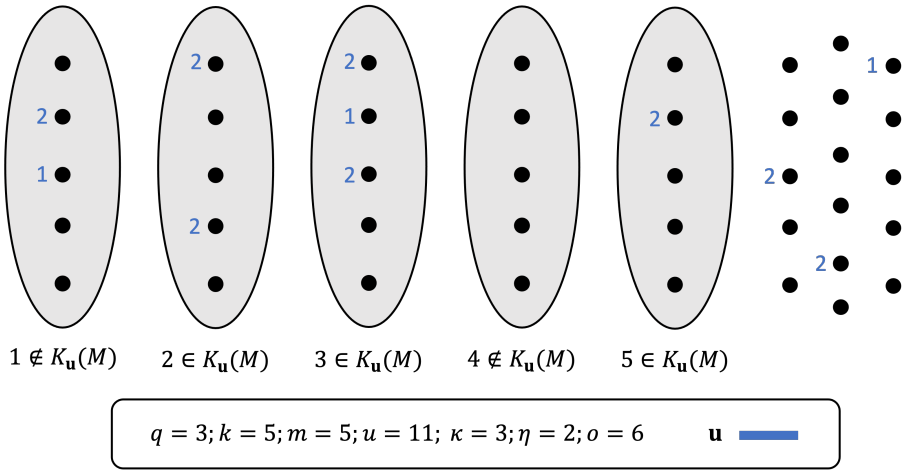

Figure 4: A graphical intuition for the parameters appeared in Definition 6.16.

We now show how to bound a certain expected value of the sum of Fourier coefficients of a fixed "level" from above in terms of the combinatorial quantity defined in Definition 6.16.

Lemma 6.18. *For every $\mathbf{u} \in \mathbb{Z}_q^n$ with $u = |\mathit{supp}(\mathbf{u})|$ and $h \in [n]$, we have*

$$\begin{aligned} &\mathop{\mathbb{E}}_M \left[\max_{\substack{B' \subset \mathbb{Z}_q^n \\ B' \text{ is } (M,C,s)\text{-reduced}}} \left\{ \sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left| \widehat{\mathbf{1}_{B'}}(\mathbf{u}') \right| \right\} \right] \\ &\leq \sum_{o,\eta,\kappa} p_q(n, u, o, \eta, \kappa) \cdot (h+1) \cdot q^{k\kappa} \cdot U_{C,s}(h + o + \eta - (u + \kappa)) \,. \end{aligned}$$

[5] Informally we refer to edges as "even" (or "odd") which would be the right terminology if $q = 2$. For $q \neq 2$ these words are formalized as having zero (or non-zero) inner product with $\mathbf{u}$.

Proof. As suggested by the right hand side, we consider the various possibilities for o, η, κ and bound the left hand side conditioned on the event defined by the probability (6.17).

Let $u = |\mathsf{supp}(\mathbf{u})|$. First, we consider a fixed matching $M = \{e_1, \dots, e_m\}$ with $m = \alpha n$ and $|K_{\mathbf{u}}(M)| = \kappa$, $|E_{\mathbf{u}}(M)| = \eta$, and $|O_{\mathbf{u}}(M)| = o$. Moreover, let $A = \mathsf{supp}(\mathbf{u}) \setminus (E_{\mathbf{u}}(M) \cup O_{\mathbf{u}}(M))$ be the set of unmatched vertices of $\mathsf{supp}(\mathbf{u})$. Furthermore, let $a = |A|$, so that $a = u - (\eta + o)$. For ease of notation, we drop the dependence on $\mathbf{u}$ and M and simply write $E = E_{\mathbf{u}}(M)$ and $O = O_{\mathbf{u}}(M)$

Let $B' \subset \mathbb{Z}_q^n$ be an (M, C, s)-reduced set. We give an upper bound on

$$\sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left| \widehat{\mathbf{1}_{B'}}(\mathbf{u}') \right|$$

in terms of the parameters o, η, κ, which will suffice to establish the lemma.

Let $U' = \mathsf{supp}(\mathbf{u}')$. First, note that $\mathbf{1}_B(\mathbf{u}') = 0$ if $U' \not\subset M$. Thus, in what follows, let us assume that $U' \subseteq M$. Let $\mathbf{u}'|_O = \mathbf{u}'|_{(\cup_{i \in O} e_i)}$ denote the restriction of $\mathbf{u}'$ to the coordinates corresponding to odd edges.

Observe that

$$\mathbf{u} + \mathbf{u}' = \mathbf{u}|_A + (\mathbf{u}|_E + \mathbf{u}'|_E) + (\mathbf{u}|_O + \mathbf{u}'|_O),$$

and the three terms on the right hand side have disjoint supports. Moreover, let $\tau(\mathbf{u}') = |\mathsf{supp}(\mathbf{u}|_O + \mathbf{u}'|_O)|$. Note that

$$\tau(\mathbf{u}') \geq |K| = \kappa, \tag{6.19}$$

since (i) $\langle \mathbf{u}', \mathbf{e}_i \rangle \equiv 0 \bmod q$ for every edge (including the odd edges) and (ii) we have $\langle \mathbf{u}, \mathbf{e}_i \rangle \not\equiv 0 \bmod q$ for every odd edge.

If $\mathbf{u}'$ satisfies $\|\mathbf{u} + \mathbf{u}'\|_0 = h$, then it follows that $\|\mathbf{u}' + (\mathbf{z} + \mathbf{u}|_E)\|_0 = h - a - \tau(\mathbf{u}')$ for some $\mathbf{z} \in \mathbb{Z}_q^n$ supported on the vertices of O (note that in particular, we can take $\mathbf{z} = \mathbf{u}'|_O$). Combining this observation with (6.19), it follows that

$$\begin{aligned}
\sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}+\mathbf{u}'\|_0 = h}} \frac{q^n}{|B'|} \left| \widehat{\mathbf{1}_{B'}}(\mathbf{w}) \right| &\leq \sum_{\tau=\kappa}^{h-a} \sum_{\substack{\mathbf{z} \in \mathbb{Z}_q^n \\ \mathsf{supp}(\mathbf{z}) \subseteq O}} \left[\sum_{\substack{\mathbf{u}' \in \mathbb{Z}_q^n \\ \|\mathbf{u}' + (\mathbf{z}+\mathbf{u}|_E)\|_0 = h-a-\tau}} \frac{q^n}{|B'|} |\widehat{\mathbf{1}_{B'}}(\mathbf{u}')| \right] \\
&\leq \sum_{\tau=\kappa}^{h-a} \sum_{\substack{\mathbf{z} \in \mathbb{Z}_q^n \\ \mathsf{supp}(\mathbf{z}) \subseteq O}} U_{C,s}(h - a - \tau) \qquad (6.20) \\
&\leq \sum_{\tau=\kappa}^{h-a} q^{k\kappa} \cdot U_{C,s}(h - a - \tau) \\
&\leq (h - a - \kappa + 1) \cdot q^{k\kappa} \cdot U_{C,s}(h - a - \kappa) \\
&\leq (h+1) \cdot q^{k\kappa} \cdot U_{C,s}(h + \eta + o - (u + \kappa)),
\end{aligned}$$

where (6.20) follows from the fact that B' is (M, C, s)-reduced (and, therefore, (C, s)-bounded) as well as Lemma 6.5 (see the remark after Definition 6.16

Recall that the above is the bound for a fixed M satisfying the conditions $K_{\mathbf{u}}(M) = \kappa$, $E_{\mathbf{u}}(M) = \eta$, and $O_{\mathbf{u}}(M) = o$. Thus, maximizing over all (M, C, s)-reduced sets B' and then summing over

all possible o, η, κ and using (6.17), we obtain

$$\begin{aligned}
&\mathop{\mathbb{E}}_{M}\left[\max_{\substack{B'\subset\mathbb{Z}_q^n\\ B' \text{ is } (M,C,s)\text{-reduced}}}\left\{\sum_{\substack{\mathbf{u}'\in\mathbb{Z}_q^n\\ \|\mathbf{u}+\mathbf{u}'\|_0=h}}\frac{q^n}{|B'|}|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')|\right\}\right]\\
&\leq \sum_{o,\eta,\kappa} p_q(n,u,o,\eta,\kappa)\cdot(h+1)\cdot q^{k\kappa}\cdot U_{C,s}(h+o+\eta-(u+\kappa))
\end{aligned}$$

as desired. □

Step 2: Useful inequalities about the boundedness parameters and the combinatorial structure. In order to quantify the upper bound in Lemma 6.18, we need to obtain an upper bound for the combinatorial quantity $p_q(n,u,o,\eta,\kappa)$.

Lemma 6.21. *For every $k \in \mathbb{N}$ there exists a constant C such that for every $\alpha \in (0, 1/k)$ and every $n, u, \kappa, o, \eta \in \mathbb{N}$ we have:*

$$p_q(n,u,o,\eta,\kappa) \leq \alpha^{(o+\eta)/k}\cdot C^u\cdot(n/\kappa)^\kappa\cdot(u/\sqrt{n\eta})^\eta\cdot(u/n)^o\,.$$

Proof. Note that p bounds the probability, over a random hypermatching, of an event involving a fixed vector $\mathbf{u}$ in $\mathbb{Z}_q^n$ with exactly u non-zero entries. By symmetry, we can equivalently fix the hypermatching M and allow $\mathbf{u}$ to be uniformly random over $S_u = \{\mathbf{u}\in\mathbb{Z}_q^n \mid |\mathsf{supp}(\mathbf{u})| = u\}$.

Therefore, without loss of generality, let M consist of $m = \alpha n$ hyperedges $e_1, e_2, \ldots, e_m$, where $e_i = (k(i-1)+1, k(i-1)+2, \ldots, k(i-1)+k)$. Moreover, let

- $K = \{i \in [m] \mid \langle \mathbf{u}, \mathbf{e}_i\rangle \not\equiv 0 \pmod q\}$.
- $O = \{j \in \mathsf{supp}(\mathbf{u}) \mid \exists i \in K, j \in e_i\}$.
- $E = \{j \in \mathsf{supp}(\mathbf{u}) \mid \exists i \in [m]\setminus K, j \in e_i\}$.

Then, we can write

$$p_q(n,u,o,\eta,\kappa) \leq \Pr_{\mathbf{u}\in S_v}[\mathcal{E}(\mathbf{u},o,\eta,\kappa)],$$

where $\mathcal{E}(\mathbf{u},o,\eta,\kappa)$ is the event that $|K| = \kappa$, $|O| = o$, and $|N| = \eta$ simultaneously hold. Thus, in order to bound $\mathcal{E}(\mathbf{u},o,\eta,\kappa)$ from above, it suffices to count the number of ways of picking $\mathbf{u}\in\mathbb{Z}_q^n$ with $|\mathsf{supp}(\mathbf{u})| = u$ such that the aforementioned conditions hold.

Let $D = \{i \in [m] : E \cap e_i \neq \emptyset\}$ and $d = |D|$. Note that each hyperedge in D contributes at least two points to E (since $\langle \mathbf{u}, \mathbf{e}_i\rangle \equiv 0 \pmod q$ for $i \in D$). Hence, $d \leq \eta/2$. Moreover, as each edge in D can contribute at most k points to E, we also have $d \geq \eta/k$. Therefore,

$$\begin{aligned}
p_q(n,v,o,\eta,\kappa) &= \sum_{d=\eta/k}^{\eta/2}\Pr_{\mathbf{u}\in S_u}[\mathcal{E}(\mathbf{u},o,\eta,\kappa) \text{ and } |D| = d]\\
&= \frac{1}{\binom{n}{u}\cdot(q-1)^u}\sum_{d=\eta/k}^{\eta/2} N_q(u,d,o,\eta,\kappa)\,. \qquad (6.22)
\end{aligned}$$

where $N_q(u,d,o,\eta,\kappa)$ is the number of $\mathbf{u}$ with support size u such that $|O| = o$, $|E| = \eta$, $|K| = \kappa$, and $|D| = d$.

Thus, it suffices to bound $N(u,d,o,\eta,\kappa)$ from above. Note that we can choose K in $\binom{m}{\kappa}$ ways, and we can then choose D in $\binom{m-\kappa}{d} \le \binom{m}{d}$ ways. Given K, we can then choose O in $\le \binom{k\kappa}{o}$ ways (since the vertices of O must be among the $k\kappa$ vertices covered by the hyperedges in K), and similarly, given D, we can choose E in $\le \binom{kd}{\eta}$ ways. Finally, we can choose the vertices in $\mathsf{supp}(\mathbf{u}) \setminus (O \cup E)$ in $\binom{n-km}{u-o-\eta}$ ways, since they can be any set of $u-o-\eta$ vertices outside the m edges in our hypermatching. Finally, for each of the u vertices chosen to be in the support of $\mathbf{u}$, we can choose any value in $\mathbb{Z}_q \setminus \{0\}$, resulting in $(q-1)^v$ choices. Therefore,

$$N_q(u,d,o,\eta,\kappa) \le \binom{\alpha n}{\kappa}\binom{\alpha n}{d}\binom{k\kappa}{o}\binom{kd}{\eta}\binom{n(1-\alpha k)}{u-o-\eta}(q-1)^u.$$

Using the bounds $\left(\frac{a}{b}\right)^b \le \binom{a}{b} \le \left(\frac{ea}{b}\right)^b$, we have that

$$\begin{aligned}\frac{N_q(u,d,o,\eta,\kappa)}{\binom{n}{u}\cdot(q-1)^u} &\le \left(\frac{e\alpha n}{\kappa}\right)^\kappa \left(\frac{e\alpha n}{d}\right)^d \left(\frac{ek\kappa}{o}\right)^o \left(\frac{ekd}{\eta}\right)^\eta \frac{(n(1-\alpha k))^{u-o-\eta}}{(u-o-\eta)!}(q-1)^u \cdot \frac{1}{\binom{n}{u}\cdot(q-1)^u} \\ &\le n^{\kappa+d-\eta-o}\kappa^{-\kappa}d^{-d}\frac{u^u}{(u-o-\eta)!}\cdot\left(\alpha^{\kappa+d}e^{\kappa+d+o+\eta}k^{o+\eta}\cdot\left(\frac{\kappa}{o}\right)^o\left(\frac{d}{\eta}\right)^\eta(1-k\alpha)^{u-o-\eta}\right).\end{aligned}$$

Recall that $\kappa \le o$, $2d \le \eta$, and $o+\eta \le u$. Hence, we have that $e^{\kappa+d+o+\eta}k^{o+\eta} \le e^{2u}\cdot k^u = (e^2k)^u$. Moreover,

$$\frac{u^{u-o-\eta}}{(u-o-\eta)!} \le 2^u.$$

Therefore, letting $C_k = 2e^2k$,

$$\frac{N_q(u,d,o,\eta,\kappa)}{\binom{n}{u}} \le C_k^u\alpha^{\kappa+d}\cdot n^{\kappa+d-\eta-o}\kappa^{-\kappa}d^{-d}u^{o+\eta} = \left(\frac{\alpha n}{d}\right)^d \cdot C_k^u\alpha^\kappa\cdot n^{\kappa-\eta-o}\kappa^{-\kappa}u^{o+\eta}.$$

Hence, by (6.22), we have that for $C_k' = 2C_k$,

$$\begin{aligned}p_q(n,u,o,\eta,\kappa) &\le C_k^u\alpha^\kappa\cdot n^{\kappa-\eta-o}\kappa^{-\kappa}u^{o+\eta}\sum_{d=\eta/k}^{\eta/2}\left(\frac{\alpha n}{d}\right)^d \\ &\le C_k^u\alpha^{\kappa+\frac{\eta}{k}}\cdot n^{\kappa-\eta-o}\kappa^{-\kappa}u^{o+\eta}\sum_{d=\eta/k}^{\eta/2}\left(\frac{n}{d}\right)^d \\ &\le C_k^u\alpha^{\frac{\eta+o}{k}}\cdot n^{\kappa-\eta-o}\kappa^{-\kappa}u^{o+\eta}\cdot\frac{\eta}{2}\left(\frac{2n}{\eta}\right)^{\eta/2} \\ &\le C_k'^u\cdot(n/\kappa)^\kappa\cdot(u/\sqrt{n\eta})^\eta\cdot(u/n)^o,\end{aligned}$$

where the second-to-last inequality follows from the fact that $n/d \ge e$ and $x^{1/x}$ is a decreasing function of x on $x \in (e,\infty)$. This completes the proof of Lemma 6.21. □

Finally, we prove an additional inequality about the boundedness parameters. This will simplify the final proof of Lemma 6.15.

Lemma 6.23. *For every $q,k \in \mathbb{N}$, there exists $\alpha_0 \in (0,1/k)$ so that the following holds. For every $\alpha \in (0,\alpha_0)$ and $C_1,C_2>0$ there exists $\varepsilon_0>0$ and $C_3>0$ such that for every $s,n,u,h,\eta,o,\kappa \in \mathbb{N}$ with $s=\varepsilon n \le \varepsilon_0 n$, we have*

$$U_{C_1,s}(u)\cdot p_q(n,u,o,\eta,\kappa)\cdot h\cdot 2^{k\kappa}\cdot U_{C_2,s}(h+\eta+o-(u+\kappa)) \le 4^{-u-2}U_{C_3,s}(h),$$

provided that $h+\eta+o-(u+\kappa)\ge 0$, $o\ge\kappa$ and $1\le h\le s\le n$.

Proof. Let C be the constant from Lemma 6.21. Let $h' = h+\eta+o-(u+\kappa)$. Let $C_4 = 4\cdot\sqrt{C_1}\cdot C\cdot 2^k$, $C_5 = 2C_2$, $C_6 = eC_5$ (where e is the base of the natural logarithm). We prove the lemma for $\alpha_0 = (\frac{1}{e^2C_4})^k$, $\varepsilon_0 = \frac{1}{(e^2C_4)^4}$ and $C_3 = (32C_6/\alpha_0^{1/k})^2$. Note in particular that this choice of α_0 depends only on k but not on C_1 and C_2 (as required).

Note that $h' \le h \le s$. If $u \le s$, then it suffices to prove

$$\begin{aligned}&\left((16C_1)^{u/2}((sn)/u^2)^{u/4}\right)\left(\alpha^{(o+\eta)/k}C^u(n/\kappa)^\kappa(u/\sqrt{n\eta})^\eta(u/n)^o\right)\left(C_2^{h'/2}(sn/h'^2)^{h'/4}\right)\cdot h\cdot 2^{k\kappa}\\ \le\ &\frac{1}{16}C_3^{h/2}\cdot(sn/h^2)^{h/4} = 4^{-2}\cdot U_{C_3,s}(h)\,,\end{aligned}\tag{6.24}$$

where the left hand side is obtained by expanding terms corresponding to $U_{C_1,s}(\cdot), p_q(\cdots)$ and $U_{C_2,s}(\cdot)$ from the left hand side in the lemma statement, and multiplying by 4^u. Similarly, if $u > s$, then it suffices to show

$$\begin{aligned}&\left((2q^2e^2)^{u/2}\cdot(n/u)^{u/2}\right)\left(\alpha^{(o+\eta)/k}C^u(n/\kappa)^\kappa(u/\sqrt{n\eta})^\eta(u/n)^o\right)\left(C_2^{h'}(sn/h'^2)^{h'/4}\right)\cdot h\cdot 2^{k\kappa}\\ \le\ &\frac{1}{16}C_3^{h/2}\cdot(sn/h^2)^{h/4}\,.\end{aligned}\tag{6.25}$$

For the case $u \le s$, we use the following sequence of inequalities:

$$\begin{aligned}&(16C_1)^{u/2}(sn/u^2)^{u/4}\cdot\alpha^{(o+\eta)/k}C^u(n/\kappa)^\kappa(u/\sqrt{n\eta})^\eta(u/n)^o\cdot C_2^{h'}(sn/h'^2)^{h'/4}\cdot h\cdot 2^{k\kappa}\cdot(h^2/(sn))^{h/4}\\
&\le\ C_4^uC_5^h\cdot(sn/u^2)^{u/4}\cdot(n/\kappa)^\kappa\cdot(u/\sqrt{n\eta})^\eta\cdot(u/n)^o\cdot(sn/h'^2)^{h'/4}\cdot(h^2/(sn))^{h/4}\\
&\qquad\text{(Using } \alpha\le\alpha_0\le 1,\ h'\le h,\ h\le 2^h,\ \kappa\le u,\ C_4 \triangleq 4\sqrt{C_1}\cdot C\cdot 2^k,\ C_5\triangleq 2C_2)\\
&\le\ C_4^uC_6^h\cdot(sn/u^2)^{u/4}\cdot(n/\kappa)^\kappa\cdot(u/\sqrt{n\eta})^\eta\cdot(u/n)^o\cdot(sn/h^2)^{-(h-h')/4}\\
&\qquad\text{(Using } (h/h')^{h'/2}\le e^h) \text{ and } C_6 = eC_5\ )\\
&=\ C_4^uC_6^h\cdot(sn/u^2)^{u/4}\cdot(n/\kappa)^\kappa\cdot(u/\sqrt{n\eta})^\eta\cdot(u/n)^o\cdot(sn/h^2)^{-(u+\kappa-(\eta+o))/4}\\
&=\ C_4^uC_6^h\cdot(h^2/u^2)^{u/4}\cdot(h^2n^3/(s\kappa^4))^{\kappa/4}\cdot(su^4/(n\eta^2h^2))^{\eta/4}\cdot(su^4/n^3h^2)^{o/4}\\
&=\ C_4^uC_6^h\cdot(h^2/u^2)^{u/4}\cdot(h^2n^2/(\varepsilon\kappa^4))^{\kappa/4}\cdot(\varepsilon u^4/(\eta^2h^2))^{\eta/4}\cdot(\varepsilon u^4/(n^2h^2))^{o/4}\\
&\le\ (\varepsilon^{1/4}C_4)^u(C_6/\varepsilon^{1/4})^h\cdot(h^2/u^2)^{u/4}\cdot(h^2n^2/(\kappa^4))^{\kappa/4}\cdot(u^4/(\eta^2h^2))^{\eta/4}\cdot(u^4/(n^2h^2))^{o/4}\\
&\qquad\text{(Collecting } \varepsilon \text{ terms and using } \eta+o-\kappa\ge u-h)\\
&\le\ (e^2\varepsilon^{1/4}C_4)^u(C_6/\varepsilon^{1/4})^h\cdot(h/u)^{u/2}\cdot(hn/(u^2))^{\kappa/2}\cdot(u/h)^{\eta/2}\cdot(u^2/(nh))^{o/2}\\
&\qquad\text{(Using } (u/\kappa)^\kappa\le e^u \text{ and } (u/\eta)^\eta\le e^u)\\
&=\ (e^2\varepsilon^{1/4}C_4)^u(C_6/\varepsilon^{1/4})^h\cdot(h/u)^{u/2}\cdot(u/h)^{\eta/2}\cdot(u^2/(nh))^{(o-\kappa)/2}\\
&\le\ (e^2\varepsilon^{1/4}C_4)^u(C_6/\varepsilon^{1/4})^h\cdot(h/u)^{u/2}\cdot(u/h)^{\eta/2}\cdot(u/h)^{(o-\kappa)/2}\\
&\qquad\text{(Using } u\le n \text{ and } o\ge\kappa)\\
&=\ (e^2\varepsilon^{1/4}C_4)^u(C_6/\varepsilon^{1/4})^h\cdot(h/u)^{(u-\eta-o+\kappa)/2}\\
&\le\ (C_6/\varepsilon^{1/4})^h\cdot(h/u)^{(u-\eta-o+\kappa)/2}\quad(\text{Using } \varepsilon\le\varepsilon_0\triangleq\tfrac{1}{(e^2C_4)^4})\\
&\le\ (C_6/\varepsilon^{1/4})^h\cdot 2^h\qquad(\text{See below})\\
&=\ (2C_6)^h(n/s)^{h/4}\qquad(\text{Using } s=\varepsilon n)\\
&\le\ \frac{1}{16}C_3^{h/2}\cdot(sn/h^2)^{h/4}\qquad(\text{Using } C_3=(32C_6/\alpha_0^{1/k})^2\ge 64C_6^2,\ h\ge 2 \text{ and } h\le s\le n)\,,\end{aligned}$$

where the inequality $(h/u)^{(u-\eta-o+\kappa)/2}\le 2^h$ (Used three lines above) holds for the following reason: We consider two cases depending on whether $h \le u$ or not. In the case where $h \le u$, we use

$u \geq \eta + o$ to conclude that $u - o - \eta + \kappa \geq 0$. Then we have that $(h/u)^{(u-\eta-o+\kappa)/2} \leq 1 \leq 2^h$. For the other case where $h > u$, we use $o \geq \kappa$ to conclude that $u \geq u - o - \eta + \kappa$. We apply $(a/b)^b \leq 2^a$ that holds for all $a, b \geq 0$, to infer that $(h/u)^{(u-\eta-o+\kappa)/2} \leq (h/u)^u \leq 2^h$. Thus we have (6.24) in the case $u \leq s$.

Next, we turn to the case $u > s$. Here we consider

$$
\begin{aligned}
&(2q^2e^2)^{u/2}(n/u)^{u/2} \cdot \alpha^{(o+\eta)/k} C^u (n/\kappa)^\kappa (u/\sqrt{n\eta})^\eta (u/n)^o \cdot C_2^{h'} (sn/h'^2)^{h'/4} \cdot h \cdot 2^{k\kappa} \cdot (h^2/(sn))^{h/4} \\
&\quad \leq \quad C_4^u C_5^h \alpha^{(o+\eta)/k} \cdot (n/u)^{u/2} \cdot (n/\kappa)^\kappa \cdot (u/\sqrt{n\eta})^\eta \cdot (u/n)^o \cdot (sn/h'^2)^{h'/4} \cdot (h^2/(sn))^{h/4} \\
&\qquad\qquad \text{(Using } h' \leq h \leq 2^h,\ \kappa \leq u,\ C_4 \triangleq \sqrt{2}e \cdot q \cdot C \cdot 2^k,\ C_5 \triangleq 2C_2) \\
&\quad \leq \quad C_4^u C_6^h \alpha^{(o+\eta)/k} \cdot (n/u)^{u/2} \cdot (n/\kappa)^\kappa \cdot (u/\sqrt{n\eta})^\eta \cdot (u/n)^o \cdot (sn/h^2)^{-(h-h')/4} \\
&\qquad\qquad \text{(Using } (h/h')^{h'/2} \leq e^h) \text{ and } C_6 = eC_5\ ) \\
&\quad = \quad C_4^u C_6^h \alpha^{(o+\eta)/k} \cdot (n/u)^{u/2} \cdot (n/\kappa)^\kappa \cdot (u/\sqrt{n\eta})^\eta \cdot (u/n)^o \cdot (sn/h^2)^{-(u+\kappa-(\eta+o))/4} \\
&\quad = \quad C_4^u C_6^h \alpha^{(o+\eta)/k} \cdot (nh^2/(su^2))^{u/4} \cdot (h^2n^3/(s\kappa^4))^{\kappa/4} \cdot (su^4/(n\eta^2h^2))^{\eta/4} \cdot (su^4/n^3h^2)^{o/4} \\
&\quad = \quad C_4^u C_6^h \alpha^{(o+\eta)/k} \cdot (h^2/(\varepsilon u^2))^{u/4} \cdot (h^2n^2/(\varepsilon\kappa^4))^{\kappa/4} \cdot (\varepsilon u^4/(\eta^2h^2))^{\eta/4} \cdot (\varepsilon u^4/(n^2h^2))^{o/4} \\
&\quad \leq \quad C_4^u C_6^h \alpha^{(o+\eta)/k} \cdot (h^2/u^2)^{u/4} \cdot (h^2n^2/(\kappa^4))^{\kappa/4} \cdot (u^4/(\eta^2h^2))^{\eta/4} \cdot (u^4/(n^2h^2))^{o/4} \\
&\qquad\qquad \text{(Collecting } \varepsilon \text{ terms } u + \kappa - (\eta + o) \leq 0 \text{ and } \varepsilon \leq 1) \\
&\quad \leq \quad (e^2C_4)^u C_6^h \alpha^{(o+\eta)/k} \cdot (h/u)^{u/2} \cdot (hn/(u^2))^{\kappa/2} \cdot (u/h)^{\eta/2} \cdot (u^2/(nh))^{o/2} \\
&\qquad\qquad \text{(Using } (u/\kappa)^\kappa \leq e^u \text{ and } (u/\eta)^\eta \leq e^u) \\
&\quad = \quad (e^2C_4)^u C_6^h \alpha^{(o+\eta)/k} \cdot (h/u)^{u/2} \cdot (u/h)^{\eta/2} \cdot (u^2/(nh))^{(o-\kappa)/2} \\
&\quad \leq \quad (e^2C_4)^u C_6^h \alpha^{(o+\eta)/k} \cdot (h/u)^{u/2} \cdot (u/h)^{\eta/2} \cdot (u/h)^{(o-\kappa)/2} \\
&\qquad\qquad \text{(Using } u \leq n \text{ and } o \geq \kappa) \\
&\quad \leq \quad (e^2C_4)^u C_6^h \alpha^{(u-h)/k} \cdot (h/u)^{(u-\eta-o+\kappa)/2} \\
&\qquad\qquad \text{(Using } \alpha \leq 1 \text{ and } u - h \leq \eta + o) \\
&\quad = \quad (\alpha^{1/k}e^2C_4)^u (C_6/\alpha^{1/k})^h (h/u)^{(u-\eta-o+\kappa)/2} \\
&\quad := \quad T_2\,.
\end{aligned}
$$

Once again, as in the $u \leq s$ case, we have $(h/u)^{(u-\eta-o+\kappa)/2} \leq 2^h$. Thus, as we chose $\alpha_0 = (1/(e^2C_4))^k$, we have $\alpha^{1/k}e^2C_4 \leq \alpha_0^{1/k}e^2C_4 \leq 1$ (which is possible since C_4 depends only on k and q, via C), and then choose $C_3 = (32C_6/\alpha_0^{1/k})^2$, thereby

$$
T_2 \leq (C_6/\alpha^{1/k})^h \cdot 2^h \leq \frac{1}{16^h} C_3^{h/2} \leq \frac{1}{16} C_3^{h/2} \cdot (sn/h^2)^{h/4}\,,
$$

where the last inequality uses $h \leq s \leq n$. This establishes (6.25) in the case $u > s$, and completes the proof of Lemma 6.23. □

Step 3: Proof of Lemma 6.15. We are now ready to combine the ingredients of the previous steps to prove Lemma 6.15. Note that

$$\begin{aligned}
&\sum_{\mathbf{u}\in\mathbb{Z}_q^n} \frac{q^n}{|B|}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \mathop{\mathbb{E}}_{M}\left[\sum_{\substack{\mathbf{u}'\in\mathbb{Z}_q^n\\ \|\mathbf{u}+\mathbf{u}'\|_0=h}} \frac{q^n}{|B'|}\left|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')\right|\right]\\
&= \sum_{u=0}^{n} \sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n\\ |\mathsf{supp}(\mathbf{u})|=u}} \frac{q^n}{|B|}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \mathop{\mathbb{E}}_{M}\left[\sum_{\substack{\mathbf{u}'\in\mathbb{Z}_q^n\\ \|\mathbf{u}+\mathbf{u}'\|_0=h}} \frac{q^n}{|B'|}\left|\widehat{\mathbf{1}_{B'}}(\mathbf{u}')\right|\right]\\
&\le \sum_{u=0}^{n} \sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n\\ |\mathsf{supp}(\mathbf{u})|=u}} \frac{q^n}{|B|}\left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \cdot \sum_{o,\eta,\kappa} p_q(n,u,o,\eta,\kappa)\cdot h\cdot 2^{k\kappa}\cdot U_{C,s}(h+o+\eta-(u+\kappa)) \qquad (6.26)\\
&\le \sum_{u=0}^{n}\sum_{o,\eta,\kappa} U_{C,s}(u)\cdot p_q(n,u,o,\eta,\kappa)\cdot h\cdot 2^{k\kappa}\cdot U_{C,s}(h+o+\eta-(u+\kappa)) \qquad (6.27)\\
&\le \sum_{u=0}^{n}\sum_{o,\eta,\kappa} 4^{-u-2}U_{C_1,s}(h) \qquad (6.28)\\
&\le \sum_{u=0}^{n}(u+1)^3\cdot 4^{-u-2}\cdot U_{C_1,s}\\
&\le U_{C_1,s}(h),
\end{aligned}$$

where (6.26) follows from Lemma 6.18, (6.27) follows from the fact that B is (C,s)-bounded, and (6.28) follows from Lemma 6.23. □

References

[AKL16] Sepehr Assadi, Sanjeev Khanna, and Yang Li. Tight bounds for single-pass streaming complexity of the set cover problem. In *STOC 2016*, pages 698–711, 2016.

[AKSY20] Sepehr Assadi, Gillat Kol, Raghuvansh R. Saxena, and Huacheng Yu. Multi-Pass Graph Streaming Lower Bounds for Cycle Counting, MAX-CUT, Matching Size, and Other Problems. In *FOCS 2020*, pages 354–364, 2020.

[AN21] Sepehr Assadi and Vishvajeet N. Graph streaming lower bounds for parameter estimation and property testing via a streaming xor lemma. In *STOC 2021*, pages 612—-625, 2021.

[CGSV21a] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Approximability of all finite csps with linear sketches. In *FOCS 2021*, pages 1197–1208. IEEE, 2021.

[CGSV21b] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Approximability of all Boolean CSPs with linear sketches. *CoRR*, abs/2102.12351v3, 14th April 2021.

[CGV20] Chi-Ning Chou, Alexander Golovnev, and Santhoshini Velusamy. Optimal streaming approximations for all Boolean Max-2CSPs and Max-kSAT. In *FOCS 2020*, pages 330–341. IEEE, 2020.

[GKK+09] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2009.

[GT19] Venkatesan Guruswami and Runzhou Tao. Streaming hardness of unique games. In *APPROX 2019*, pages 5:1–5:12. LIPIcs, 2019.

[GVV17] Venkatesan Guruswami, Ameya Velingker, and Santhoshini Velusamy. Streaming complexity of approximating Max 2CSP and Max Acyclic Subgraph. In *APPROX 2017*. LIPIcs, 2017.

[KK19] Michael Kapralov and Dmitry Krachun. An optimal space lower bound for approximating MAX-CUT. In *STOC 2019*, pages 277–288. ACM, 2019.

[KKS15] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Streaming lower bounds for approximating MAX-CUT. In *SODA 2015*, pages 1263–1282. SIAM, 2015.

[KKSV17] Michael Kapralov, Sanjeev Khanna, Madhu Sudan, and Ameya Velingker. $(1+\omega(1))$-approximation to MAX-CUT requires linear space. In *SODA 2017*, pages 1703–1722. SIAM, 2017.

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

[SSV21] Noah Singer, Madhu Sudan, and Santhoshini Velusamy. Streaming approximation resistance of every ordering CSP. In *APPROX 2021*, volume 207, pages 17:1–17:19. LIPIcs, 2021.

[VY11] Elad Verbin and Wei Yu. The streaming complexity of cycle counting, sorting by reversals, and other problems. In *SODA 2011*, pages 11–25. SIAM, 2011.

[Yao77] Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *FOCS 1977*, pages 222–227. IEEE, 1977.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il