Linear Space Streaming Lower Bounds for Approximating CSPs

Chi-Ning Chou*
chiningchou@g.harvard.edu
School of Engineering and Applied Sciences, Harvard University
Cambridge, Massachusetts, USA

Alexander Golovnev
alexgolovnev@gmail.com
Department of Computer Science, Georgetown University
Washington, D.C., USA

Madhu Sudan†
madhu@cs.harvard.edu
School of Engineering and Applied Sciences, Harvard University
Cambridge, Massachusetts, USA

Ameya Velingker
ameyav@google.com
Google Research
Mountain View, California, USA

Santhoshini Velusamy‡
svelusamy@g.harvard.edu
School of Engineering and Applied Sciences, Harvard University
Cambridge, Massachusetts, USA

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 MaxCut 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.

*Supported by NSF grants DMS-2134157 and CCF-1565264, DARPA grant W911NF2010021, DOE grant DE-SC0022199, and the Simons foundation.
†Supported in part by a Simons Investigator Award and NSF Awards CCF 1715187 and CCF 2152413.
‡Supported in part by a Google Ph.D. Fellowship, a Simons Investigator Award to Madhu Sudan, and NSF Awards CCF 1715187 and CCF 2152413.

STOC '22, June 20–24, 2022, Rome, Italy

ACM ISBN 978-1-4503-9264-8/22/06.
https://doi.org/10.1145/3519935.3519983

CCS CONCEPTS

• **Theory of computation → Sketching and sampling**; **Communication complexity**; **Approximation algorithms analysis**.

KEYWORDS

streaming algorithms, communication lower bound, inapproximability, constraint satisfaction problems

ACM Reference Format:
Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, Ameya Velingker, and Santhoshini Velusamy. 2022. Linear Space Streaming Lower Bounds for Approximating CSPs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC '22), June 20–24, 2022, Rome, Italy.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3519935.3519983

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 Max-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 Max-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 $\alpha \cdot \mathrm{val}_\Psi \le \eta \le \mathrm{val}_\Psi$.

In this work we consider the space complexity of approximating Max-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 Max-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 Max-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 Max-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 Max-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 Max-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 Max-CSP($\mathcal{F}$) using $o(n)$ space. We say that Max-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 $\mathcal{F}$, Max-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 also approximation resistant. We elaborate on some of these examples in Section 4. 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 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 [1–4, 7–13, 15, 16]. 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 [2, 3] — 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 [4].)

On the front of general lower bounds, Chou, Golovnev, Sudan and Velusamy [4] explored the same set of CSP problems as we do, i.e, Max-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. [4] 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 [13], who show that approximating Max Cut (which translates in our setting to Max-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 [12] showing that there exists $\varepsilon > 0$ such that $(1-\varepsilon)$-approximation for Max Cut requires linear space. Finally, Chou, Golovnev and Velusamy [7] get a tight inapproximability for Max Exact 2-SAT (corresponding to Max-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 Max-CSP($\mathcal{F}$) allowed a 1/2-approximating streaming algorithm using $o(n)$ space. Our work is the first to prove inapproximability $\alpha \le 1/2$ for any Max-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 [13] to problems beyond Max Cut. Indeed the bulk of our proof takes the tour-de-force proof in [13] 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 [13]. 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 [4, Theorem 2.9], i.e., families supporting one-wise independent distributions, but this

[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.

remains open. Indeed to get our extensions we have to formulate a new communication problem which generalizes the one in [13] and is different from the many variations considered in [5] and [4]. 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 [4] 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 [4] 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 [13] which is much closer to our work and needs more detailed comparison.

Comparison with [4]. While there is some obvious overlap in the set of problems considered in [4] 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, [4] 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 [13] 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 [4] 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 [4] due to a concept we call "folding". Folding makes our problems too restrictive to work for [4] (i.e., would prevent them for addressing every $(\gamma, \beta) - \text{Max-CSP}(\mathcal{F})$), whereas we do not know how to get our lower bounds without folding. We also note that [4] 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 [13]. We start with a quick review of the main steps of [13] and then describe our analysis and conclude with a summary of the differences/new contributions relative to [13].

Summary of [13]. [13] 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 $c_{1:t}$. [13] 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 $U(w) = U_{S,n}(w)$.

[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.

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 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 [13] 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 [13]. 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 5.17) 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 5.17 — and of course this only happens with high probability over the choice of M_t.

Incremental contribution over [13]. Given that our result closely follows [13] we now focus on some key differences, and why these contributions are conceptually significant.

(1) The analysis of [13] 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 [13]. Part of the challenge is that in the inductive proof of boundedness even the base case (which is quite simple in [13]) 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 5.17) 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 [13].
(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 [13] 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 [13]).

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.

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.,[11, 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. Proof is given in the full version [6]. □

We will also need the following lemma from [13].

Lemma 2.4 ([13] 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*

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

2.2 Concentration Inequality

We will use the following concentration inequality from [13] which is essentially an Azuma-Hoeffding style inequality for submartingales.

Lemma 2.5 ([13, Lemma 2.5])**.** *Let $X = \sum_{i \in [N]} X_i$ where X_i are Bernoulli random variables such that for every $k \in [N]$,*

$$\mathbb{E}[X_k \mid X_1, \dots, X_{k-1}] \leq p \, ,$$

for some $p \in (0, 1)$. Let $\mu = Np$. For every $\Delta > 0$, we have:

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

2.3 Fourier Analysis

In this paper, we will use Fourier analysis over $\mathbb{Z}_q$ (see, for instance, [9, 14]). 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 $\mathbb{Z}_q^n$:

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

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. Proof is given in the full version [6]. □

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. Proof is given in the full version [6]. □

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 [14, 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 $f : \mathbb{Z}_q^n \to \{a \in \mathbb{C} \mid |a| \le 1\}$ and $B = \{\mathbf{a} \in \mathbb{Z}_q^n \mid f(\mathbf{a}) \ne 0\}$ 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 $h \in \{1, \dots, 4b\}$, 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. Proof is given in the full version [6]. □

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 $e_1, \dots, e_m$ where $e_i \subseteq [n]$, $|e_i| = k$ and $e_i \cap e_j = \emptyset$ for every $i \ne j \in [m]$. We let $e_i = \{(e_i)_1, \dots, (e_i)_k\}$. The direct encoding of a matching $M = \{e_1, \dots, e_m\}$ 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 $e_1, e_2, \dots, e_m$ 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.) $A_t \in \{0,1\}^{\alpha kn\times n}$, the hypermatching matrix (see above) corresponding to a random hypermatching M_t of size αn and (b.) a vector $\mathbf{z}_t \in \mathbb{Z}_q^{\alpha kn}$ that can be generated from one of two different distributions:*

- *(Yes) $\mathbf{z}_t = A_t\mathbf{x}^* + \mathbf{b}_t$ where $\mathbf{b}_t \in \mathbb{Z}_q^{\alpha kn}$ is of the form $\mathbf{b}_t = (\mathbf{b}_{t,1}, \dots, \mathbf{b}_{t,\alpha n})$ 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 $\mathbf{b}_t \in \mathbb{Z}_q^{\alpha kn}$ is of the form $\mathbf{b}_t = (\mathbf{b}_{t,1}, \dots, \mathbf{b}_{t,\alpha n})$ 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 3.2. 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 $\{0^k, 1^k, \dots, (q-1)^k\}$ 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 3.3 (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 $\{0^k, 1^k, \dots, (q-1)^k\}$ and $\mathbb{Z}_q^k$ respectively and $n \ge n_0$ 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 $M = (e_1, \dots, e_m)$ and a sequence of centers $\mathbf{c} = (c_1, \dots, c_m)$ with $e_i = ((e_i)_1, \dots, (e_i)_k = c_i)$, 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

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

We define the folded version of the IRMD problem below (note that all the arithmetic is over $\mathbb{Z}_q$):

Definition 3.4 (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 the absolute value of*

$$\Pr_{(A_{t,\mathbf{c}_t},\mathbf{w}_t)_{t\in T}\sim \textbf{YES}}[\text{the } T\text{-th player outputs Yes}] - \Pr_{(A_{t,\mathbf{c}_t},\mathbf{w}_t)_{t\in T}\sim \textbf{NO}}[\text{the } T\text{-th player outputs Yes}].$$

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

Theorem 3.5 (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 3.5 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 3.3 will be an immediate corollary of Theorem 3.5.

Lemma 3.6. *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

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

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 3.3 using Theorem 3.5. 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 3.6 there exists a protocol for IFRMD with advantage δ using fewer than τn bits of communication, which contradicts Theorem 3.5. This completes the proof of Theorem 3.3. □

In the following section we show how Theorem 3.3 yields the claimed hardness of streaming problems. In the rest of this paper, we focus on the proof of Theorem 3.5, 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$_\Psi$. 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 [4] 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 [4].

Proposition 4.2 ([4, 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 3.3.

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 $\{\Psi \mid \mathrm{val}_\Psi \geq (1-\varepsilon/3)\omega(\mathcal{F})\}$ from instances from the set $\{\Psi \mid \mathrm{val}_\Psi \leq (1+\varepsilon/3)\rho(\mathcal{F})\}$ with success probability at least 2/3. We show how to use **ALG** to device an s-bit communication protocol for 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 Max-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 Max-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, $\Pr_{\mathbf{YES}}[\mathrm{val}_\Psi \leq (1-\varepsilon/3)\cdot\omega(\mathcal{F})] = o(1)$.
(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, $\Pr_{\mathbf{NO}}[\mathrm{val}_\Psi \geq (1+\varepsilon/3)\cdot\rho(\mathcal{F})] = o(1)$.

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 3.3 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, $\mathbb{E}[X_i \mid X_{1:i-1}] \leq q^{-(k-1)}\cdot(1+\gamma+\alpha qk)^k$. 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 $\Pr[\mathbf{x}^*|_{\mathbf{j}(i)} \in S_f \mid X_{1:i-1}] \leq |S_f|\cdot((1+\gamma+\alpha kq)/q)^k = q^{-(k-1)}\cdot(1+\gamma+\alpha qk)^k$. Setting $\gamma = \varepsilon/(100k)$ and using $\alpha \leq \varepsilon/(100k^2q)$, we conclude $\mathbb{E}[X_i \mid X_{1:i-1}] \leq q^{-(k-1)}\cdot(1+\varepsilon/(50k))^k \leq q^{-(k-1)}\cdot(1+\varepsilon/20)$. 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 $\mathbf{b} \in \mathbb{Z}_q^k$, $\Pr[\mathbf{x}^*|_{\mathbf{j}(i)} = \mathbf{b} \mid Y_{1:i-1}] \geq (1-\varepsilon/50)\cdot q^{-k}$. 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 $\Pr[\mathrm{val}_\Psi \leq (1-\varepsilon/3)\cdot\omega(\mathcal{F})] = o(1)$.

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 $\mathrm{val}_\Psi \leq (1+\varepsilon/3)\cdot\rho(\mathcal{F})$ 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. □

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 $\mathcal{D} = \mathrm{Unif}(\mathbb{Z}_q)$. 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 [5, 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 [15] who use it to show the approximation resistance of the "maximum acyclic subgraph" problem.

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

For $k = 2$ and $q \geq 2$ we define $\mathcal{F} = \{<_q\}$ where $<_q : \mathbb{Z}_q^2 \to \{0,1\}$ 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 3.5). 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, 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 3.5 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.,

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

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).

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 $n \ge n_0$ 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 $x \sim$ 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^Y_t and S^N_t 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 $S^Y_{1:t-1}$, (ii) $\Pr[\mathcal{E}_1] \ge 1 - \delta/T$, (iii) $\Pr[\overline{\mathcal{E}}_t \mid \mathcal{E}_{t-1}] \le \delta/T$ for all $t = 2, 3, \ldots, T$, and (iv) for every fixed $(A_{1:t}, \mathbf{c}_{1:t})$ and $S^Y_{1:t-1}$ satisfying $\mathcal{E}_t$, one has

$$\|S^Y_t - r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, U_t)\|_{tvd} \le \delta/T, \tag{5.2}$$

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

The proof of Lemma 5.1 will be provided in Section 5.4. Theorem 3.5 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 [11]) 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 [13], 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 [13] 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 5.6). The boundedness condition places restrictions on the Fourier spectrum of the indicator function of this set. (See Definition 5.11.) 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 5.16. 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 5.15. 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 5.17. In the rest of this section, after showing that Lemma 5.1 implies Theorem 3.5 in Section 5.1, we introduce the posterior sets and discuss their basic properties in Section 5.2, we introduce boundedness and state the

[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.

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 3.5

We now show how the lemma suffices to prove Theorem 3.5. The proof is analogous to the proof of Lemma 6.3 in [13]. 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 3.5. 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 $\Pr\left[\overline{\mathcal{E}_t} \mid \mathcal{E}_{t-1}\right] \leq \delta'/T$ 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:

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

We now apply Lemma 2.4. 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

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

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

$$\|Q^Y_{t-1} - Q^N_{t-1}\|_{tvd, \mathcal{E}_{t-1}} \leq \frac{(t-1)\delta'}{T}. \tag{5.3}$$

Next, we bound the second term on the right hand side, i.e.,

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

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

$$\begin{aligned}
&\|(Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, A_{t,\mathbf{c}_t}\mathbf{x}^*)) \\
&- (Q^Y_{t-1}, A_t, \mathbf{c}_t, r_t(Q^Y_{t-1}, A_t, \mathbf{c}_t, U_t))\|_{tvd, \mathcal{E}_t} \\
&\leq \frac{\delta'}{T}.
\end{aligned} \tag{5.4}$$

Combining all the above equations, 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., $S_t^N = r_t(A_{1:t}, \mathbf{c}_{1:t}, S_{1:t-1}^Y, U_t)$). 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 5.5 (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*

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

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 $S_1, S_2, \dots, S_t$:

Lemma 5.6 (Posterior function $\mathbf{1}_{B_{1:t}}$ captures the posterior distribution.)**.** *For every $t \in [T]$, the conditional distribution of $\mathbf{x}$ given messages $S_1, S_2, \dots, S_t$ 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 $\mathbf{x}^* \sim$ Unif$(B_{1:t})$.*

Proof. Proof is given in the full version [6]. □

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

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

Proof. By Lemma 5.6, 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 Unif$(B_{1:t})$, the distribution of the resulting string is close to the uniform distribution 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 5.8 (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 5.5 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 5.7) 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 5.9 (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 [13]. First, we introduce some notation:

$$U_{C,s}(h) := \begin{cases} 1, & h = 0 \\ \left(\frac{C\sqrt{sn}}{h}\right)^{h/2}, & 1 \le h \le s \\ \left(\frac{2q^2e^2n}{h}\right)^{h/2}, & h > s. \end{cases} \tag{5.10}$$

Definition 5.11 (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 $h \in [s]$,*

$$\sum_{\substack{\mathbf{u} \in \mathbb{Z}_q^n \\ \|\mathbf{u}\|_0 = h}} \frac{q^n}{|B|} \left|\widehat{\mathbf{1}_B}(\mathbf{u})\right| \le U_{C,s}(h). \tag{5.12}$$

Remark 5.13. 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 5.16). 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. 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 $|B| \ge q^{n-s}$, then we have that

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

for *all* $0 \le h \le n$.

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 5.14 (Reduced set)**.** *Let $n, q \in \mathbb{N}$, $0 \le s \le n$, $C > 0$, 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 $h \in \{1, \dots, s\}$ and $\mathbf{v} \in \mathbb{Z}_q^n$,*

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

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 5.15 (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.*

The proof of Lemma 5.15 is given in the full version [6]. (We note that the proof yields that $C \ge 2\zeta^2 ek^2 q^{3k}$ where ζ is the constant from Lemma 2.11.)

Recall from Corollary 5.7 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 5.16 (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

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

As a consequence, we also have

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

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

The proof of Lemma 5.16 is given in the full version [6].

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 5.17 (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 $n \ge n_0$, 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 5.17 is proved in the full version [6]. 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 5.17) and the "boundedness implies uniformity" lemma (i.e., Lemma 5.16). We set C_0 according to Lemma 5.17 and for every $\alpha \in (0, \alpha_0']$, $T \in \mathbb{N}$, and $\delta \in (0,1)$, we invoke Lemma 5.16 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 5.5. By the "posterior set is large" lemma (i.e., Lemma 5.8) and the "base case" lemma (i.e., Lemma 5.15) we have $\Pr[\overline{\mathcal{E}_1}] \leq \delta'/10 \leq \delta/T$ as desired. This satisfies condition (ii) of Lemma 5.1.

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., $|B_{1:t}| \geq q^{n-2tb}$ 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 5.8) and the "base case" lemma (i.e., Lemma 5.15), we have $|B_t| \geq q^{n-b}$ 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 5.16), $\mathbb{Z}_q^{(k-1)\alpha n}$, we have the following claim.

Claim 5.18. *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*

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

Proof. Proof is given in the full version [6]. □

Thus, by invoking the "induction step" lemma (i.e., Lemma 5.17) 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 $\Pr[\overline{\mathcal{E}_t} \mid \mathcal{E}_{t-1}] \leq 5\delta' \leq \delta/T$. 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 5.16), we know that $\|(A_{\mathbf{c}}\mathbf{x}^*) - U_t\|_{tvd} \leq \delta'$ 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 $\|S^Y_t - r_t(A_{1:t}, \mathbf{c}_{1:t}, S^Y_{1:t-1}, U_t)\|_{tvd} \leq \delta/T$ 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.

□

REFERENCES

[1] Sepehr Assadi, Sanjeev Khanna, and Yang Li. 2016. Tight bounds for single-pass streaming complexity of the set cover problem. In *STOC 2016*. 698–711. https://doi.org/10.1145/2897518.2897576

[2] Sepehr Assadi, Gillat Kol, Raghuvansh R. Saxena, and Huacheng Yu. 2020. Multi-Pass Graph Streaming Lower Bounds for Cycle Counting, MAX-CUT, Matching Size, and Other Problems. In *FOCS 2020*. 354–364. https://doi.org/10.1109/FOCS46700.2020.00041

[3] Sepehr Assadi and Vishvajeet N. 2021. Graph Streaming Lower Bounds for Parameter Estimation and Property Testing via a Streaming XOR Lemma. In *STOC 2021*. 612–-625. https://doi.org/10.1145/3406325.3451110

[4] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. 2021. Approximability of all finite CSPs with linear sketches. In *FOCS 2021*. IEEE, 1197–1208. https://doi.org/10.1109/FOCS52979.2021.00117

[5] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. 2021. Approximability of all Boolean CSPs with linear sketches. *CoRR* abs/2102.12351v3 (14th April 2021). arXiv:2102.12351 https://arxiv.org/abs/2102.12351v3

[6] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, Ameya Velingker, and Santhoshini Velusamy. 2021. Linear Space Streaming Lower Bounds for Approximating CSPs. (June 2021). arXiv:2106.13078 [cs.CC]

[7] Chi-Ning Chou, Alexander Golovnev, and Santhoshini Velusamy. 2020. Optimal Streaming Approximations for all Boolean Max-2CSPs and Max-kSAT. In *FOCS 2020*. IEEE, 330–341. https://doi.org/10.1109/FOCS46700.2020.00039

[8] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. 2009. Exponential Separation for One-Way Quantum Communication Complexity, with Applications to Cryptography. *SIAM J. Comput.* 38, 5 (2009), 1695–1708. https://doi.org/10.1137/070706550

[9] Venkatesan Guruswami and Runzhou Tao. 2019. Streaming Hardness of Unique Games. In *APPROX 2019*. LIPIcs, 5:1–5:12. https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2019.5

[10] Venkatesan Guruswami, Ameya Velingker, and Santhoshini Velusamy. 2017. Streaming Complexity of Approximating Max 2CSP and Max Acyclic Subgraph. In *APPROX 2017*. LIPIcs. https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2017.8

[11] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. 2015. Streaming Lower Bounds for Approximating MAX-CUT. In *SODA 2015*. SIAM, 1263–1282. https://doi.org/10.1137/1.9781611973730.84

[12] Michael Kapralov, Sanjeev Khanna, Madhu Sudan, and Ameya Velingker. 2017. $(1+\Omega(1))$-Approximation to MAX-CUT Requires Linear Space. In *SODA 2017*. SIAM, 1703–1722. https://doi.org/10.1137/1.9781611974782.112

[13] Michael Kapralov and Dmitry Krachun. 2019. An optimal space lower bound for approximating MAX-CUT. In *STOC 2019*. ACM, 277–288. https://doi.org/10.1145/3313276.3316364

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

[15] Noah Singer, Madhu Sudan, and Santhoshini Velusamy. 2021. Streaming approximation resistance of every ordering CSP. In *APPROX 2021*, Vol. 207. LIPIcs, 17:1–17:19. https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2021.17

[16] Elad Verbin and Wei Yu. 2011. The streaming complexity of cycle counting, sorting by reversals, and other problems. In *SODA 2011*. SIAM, 11–25. https://doi.org/10.1137/1.9781611973082.2