

# Approximability of all finite CSPs in the dynamic streaming setting

Chi-Ning Chou\*    Alexander Golovnev†    Madhu Sudan‡    Santhoshini Velusamy§

## Abstract

A constraint satisfaction problem (CSP),  $\text{Max-CSP}(\mathcal{F})$ , is specified by a finite set of constraints  $\mathcal{F} \subseteq \{[q]^k \rightarrow \{0,1\}\}$  for positive integers  $q$  and  $k$ . An instance of the problem on  $n$  variables is given by  $m$  applications of constraints from  $\mathcal{F}$  to subsequences of the  $n$  variables, and the goal is to find an assignment to the variables that satisfies the maximum number of constraints. In the  $(\gamma, \beta)$ -approximation version of the problem for parameters  $0 \leq \beta < \gamma \leq 1$ , the goal is to distinguish instances where at least  $\gamma$  fraction of the constraints can be satisfied from instances where at most  $\beta$  fraction of the constraints can be satisfied.

In this work we consider the approximability of this problem in the context of streaming algorithms and give a dichotomy result in the dynamic setting, where constraints can be inserted or deleted. Specifically, for every family  $\mathcal{F}$  and every  $\beta < \gamma$ , we show that either the approximation problem is solvable with polylogarithmic space in the dynamic setting, or not solvable with  $o(\sqrt{n})$  space. We also establish tight inapproximability results for a broad subclass in the streaming insertion-only setting. Our work builds on, and significantly extends previous work by the authors who consider the special case of Boolean variables ( $q = 2$ ), singleton families ( $|\mathcal{F}| = 1$ ) and where constraints may be placed on variables or their negations. Our framework extends non-trivially the previous work allowing us to appeal to richer norm estimation algorithms to get our algorithmic results. For our negative results we introduce new variants of the communication problems studied in the previous work, build new reductions for these problems, and extend the technical parts of previous works. In particular, previous works used Fourier analysis over the Boolean cube to prove their results and the results seemed particularly tailored to functions on Boolean literals (i.e., with negations). Our techniques surprisingly allow us to get to general  $q$ -ary CSPs without negations by appealing to the same Fourier analytic starting point over Boolean hypercubes.

---

\*School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported by NSF awards CCF 1565264 and CNS 1618026. Email: [chiningchou@g.harvard.edu](mailto:chiningchou@g.harvard.edu).

†Department of Computer Science, Georgetown University. Email: [alexgolovnev@gmail.com](mailto:alexgolovnev@gmail.com).

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

§School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported in part by a Simons Investigator Award and NSF Award CCF 1715187. Email: [svelusamy@g.harvard.edu](mailto:svelusamy@g.harvard.edu).

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Results	4
1.2	Comparison with previous works	5
1.3	Technical contributions	7
<b>2</b>	<b>Results</b>	<b>9</b>
2.1	Main Notions	9
2.2	Results in the dynamic setting	10
2.2.1	Decidability of the Classification	11
2.2.2	Approximation resistance in the dynamic streaming setting	12
2.3	Results in the insertion-only setting	14
2.4	Some Examples	17
2.5	Classification of exact computability	19
<b>3</b>	<b>Preliminaries</b>	<b>20</b>
3.1	Approximate Constraint Satisfaction	20
3.2	Total variation distance	21
3.3	Concentration inequality	21
3.4	Fourier analysis	22
<b>4</b>	<b>A Streaming Approximation Algorithm for Max-CSP(<math>\mathcal{F}</math>)</b>	<b>23</b>
4.1	Analysis of the correctness of Algorithm 1	25
<b>5</b>	<b>Streaming Space Lower Bounds for Max-CSP(<math>\mathcal{F}</math>)</b>	<b>27</b>
5.1	The Signal Detection Problem and Results	28
5.2	The insertion-only lower bound	29
5.2.1	The (Padded) Streaming SD Problem	30
5.2.2	CSP value of padded-streaming-SD	31
5.2.3	Reduction from one-way $(\mathcal{D}_Y, \mathcal{D}_N)$ -SD to padded-streaming-SD	35
5.2.4	Proof of the insertion-only lower bound	37
5.3	The dynamic lower bound	37
5.3.1	The framework of [AHLW16]	38
5.3.2	$T$ -Player Simultaneous Version of SD	38
5.3.3	Proof of Theorem 5.1	40
<b>6</b>	<b>Hardness of Advice-Signal-Detection with Uniform Marginals</b>	<b>41</b>
6.1	Hardness of Advice-RMD	42
6.2	Hardness of Advice-SD	42
6.3	Proof of Theorem 6.2	43
<b>7</b>	<b>Hardness of Signal Detection</b>	<b>49</b>
7.1	Finite upper bound on the number of polarization steps	51
7.2	Reduction from single function to a family of functions	55
7.3	Putting it together	56

<b>8</b>	<b>Indistinguishability of the Polarization Step</b>	<b>57</b>
8.1	Reduce a Boolean SD problem to a non-Boolean SD problem . . . . .	57
8.2	Indistinguishability of shifting distributions . . . . .	62
8.3	Proof of Theorem 7.4 . . . . .	64
<b>9</b>	<b>Dichotomy for exact Computation</b>	<b>66</b>

# 1 Introduction

In this work we explore the approximability of general (finite) constraint satisfaction problems in the setting of streaming algorithms. We introduce the terms below.

For positive integers  $q$  and  $k$ , a  $q$ -ary *constraint satisfaction problem* (CSP) is given by a (finite) set of constraints  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ . A constraint  $C$  on  $x_1, \dots, x_n$  is given by a pair  $(f, \mathbf{j})$ , with  $f \in \mathcal{F}$  and  $\mathbf{j} = (j_1, \dots, j_k) \in [n]^k$  where the coordinates of  $\mathbf{j}$  are all distinct. An assignment  $\mathbf{b} \in [q]^n$  satisfies  $C = (f, \mathbf{j})$  if  $f(b_{j_1}, \dots, b_{j_k}) = 1$ . An instance  $\Psi$  of  $\text{Max-CSP}(\mathcal{F})$  consists of  $m$  constraints  $C_1, \dots, C_m$  applied to  $n$  variables  $x_1, x_2, \dots, x_n$  along with  $m$  non-negative integer weights  $w_1, \dots, w_m$ . The value of an assignment  $\mathbf{b} \in [q]^n$  on an instance  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$ , denoted  $\text{val}_\Psi(\mathbf{b})$ , is the fraction of weight of constraints satisfied by  $\mathbf{b}$ . The goal of the *exact* problem is to compute the maximum, over all assignments, of the value of the assignment on the input instance, i.e., to compute, given  $\Psi$ , the quantity  $\text{val}_\Psi = \max_{\mathbf{b} \in [q]^n} \{\text{val}_\Psi(\mathbf{b})\}$ . In this work we consider the complexity of the  $(\gamma, \beta)$ -*approximation version* of this problem for parameters  $0 \leq \beta < \gamma \leq 1$ , denoted  $(\gamma, \beta)$ - $\text{Max-CSP}(\mathcal{F})$ , defined as follows: Given an instance  $\Psi$  with the promise that  $\text{val}_\Psi \geq \gamma$  or  $\text{val}_\Psi \leq \beta$  determine which of the two conditions holds.

We study the complexity of  $(\gamma, \beta)$ - $\text{Max-CSP}(\mathcal{F})$  in the setting of randomized streaming algorithms in both the *insertion-only* setting and the *dynamic* setting. In the insertion-only setting an instance  $\Psi = (C_1, \dots, C_m)$  is presented as a stream  $\sigma_1, \sigma_2, \dots, \sigma_\ell$  where  $\sigma_i = (f(i), \mathbf{j}(i))$  represents adding a unit weight to the constraint given by the pair  $(f(i), \mathbf{j}(i))$ . In the dynamic setting constraints are allowed to be inserted or deleted subject to the constraint that at the end every constraint appears a non-negative number of times. Thus the input in the dynamic setting is a stream  $\sigma_1, \dots, \sigma_\ell$  where  $\sigma_i = (f(i), \mathbf{j}(i), w_i)$  with  $w_i \in \{-1, 1\}$  representing the addition of  $w_i$  to the weight of the constraint given by the pair  $(f(i), \mathbf{j}(i))$ . (We note that our negative results work even in the restrictive strict turnstile setting where the weight of every constraint is non-negative at all intermediate stages. In contrast our algorithmic results work in a more general turnstile setting which allows the weight of constraints to be negative mid-stream but must be non-negative at the end.) In both settings we study the space required to solve the  $(\gamma, \beta)$ -approximation version of  $\text{Max-CSP}(\mathcal{F})$ .

## 1.1 Results

Our main theorem is a decidable dichotomy theorem for  $(\gamma, \beta)$ - $\text{Max-CSP}(\mathcal{F})$  in the dynamic setting.

**Theorem 1.1** (Succinct version). *For every  $q, k \in \mathbb{N}$ ,  $0 \leq \beta < \gamma \leq 1$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , one of the following two conditions hold: Either  $(\gamma, \beta)$ - $\text{Max-CSP}(\mathcal{F})$  can be solved with  $O(\log^3 n)$  space in the dynamic setting, or for every  $\varepsilon > 0$ ,  $(\gamma - \varepsilon, \beta + \varepsilon)$ - $\text{Max-CSP}(\mathcal{F})$  in the dynamic setting requires  $\Omega(\sqrt{n})$ -space. Furthermore there is a polynomial space algorithm that decides which of the two conditions holds, given  $\gamma, \beta$  and  $\mathcal{F}$ .*

[Theorem 1.1](#) combines the more detailed [Theorem 2.3](#) with decidability coming from [Theorem 2.4](#). In the insertion-only setting we do not get a complete characterization, but we do get some sufficient conditions that allow us to pin down the approximability of some families  $\mathcal{F}$ . To describe our result in this setting, we need some notions.

Given a constraint family  $\mathcal{F}$  consider the quantity  $\rho_{\min}(\mathcal{F}) = \inf_{\Psi} \{\text{val}_\Psi\}$  where the infimum is taken over all instances  $\Psi$  of  $\text{Max-CSP}(\mathcal{F})$ . By definition for every  $\Psi$  we have  $\text{val}_\Psi \in [\rho_{\min}(\mathcal{F}), 1]$ , and consequently  $(\gamma, \beta)$ - $\text{Max-CSP}(\mathcal{F})$  is trivially solvable for  $\beta < \rho_{\min}(\mathcal{F}) \leq \gamma$  (since the **NO** set is

empty). Informally we say that  $\text{Max-CSP}(\mathcal{F})$  is approximation-resistant if no non-trivial algorithm (i.e., for  $\rho_{\min} < \beta < \gamma < 1$ ) exists. (See [Definition 2.11](#) for a formal definition.) Our next theorem shows that a broad class of functions is approximation-resistant.

We say that  $f$  supports *one-wise independence* if there exists a distribution  $\mathcal{D}$  supported on  $f^{-1}(1)$  whose marginals are uniform on  $[q]$ . We say that  $\mathcal{F}$  supports one-wise independence if every  $f \in \mathcal{F}$  supports one-wise independence.

**Theorem 1.2** (Informal). *If  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  supports one-wise independence then it is approximation-resistant.*

[Theorem 1.2](#) is formalized as [Theorem 2.17](#) in [Section 2.3](#). We also give theorems capturing hardness in the insertion-only setting beyond the 1-wise independent case. Stating the full theorem requires more notions (see [Section 2.3](#)), but as a consequence we get the following extension of theorems of [\[CGV20\]](#) and [\[CGSV21\]](#).

**Theorem 1.3.** *Let  $q = k = 2$ . Then, for every family  $\mathcal{F} \subseteq \{f : [q]^2 \rightarrow \{0, 1\}\}$ , and for every  $0 \leq \beta < \gamma \leq 1$ , at least one of the following always holds:*

1.  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) has a  $O(\log^3 n)$ -space dynamic streaming algorithm.
2. For every  $\varepsilon > 0$ ,  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) in the insertion-only setting requires  $\Omega(\sqrt{n})$  space. If  $\gamma = 1$ , then  $(1, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) in the insertion-only setting requires  $\Omega(\sqrt{n})$  space.

Furthermore, for every  $\ell \in \mathbb{N}$ , there is an algorithm using space  $\text{poly}(\ell)$  that decides which of the two conditions holds given the truth-tables of functions in  $\mathcal{F}$ , and  $\gamma$  and  $\beta$  as  $\ell$ -bit rationals.

[Theorem 1.3](#) is proved in [Section 2.3](#). [\[CGV20\]](#) study the setting where constraints are applied to literals,  $\mathcal{F}$  contains a single function and get a tight characterization of the approximability of  $\text{Max-CSP}(\mathcal{F})$ <sup>1</sup>. [Theorem 1.4](#) in [\[CGSV21\]](#) extends the study to classify all choices of  $\beta < \gamma$ . Our work extends both of their works by allowing constraints to be applied only to variables, and by allowing families of constraint functions.

For the sake of completeness we also give a simple characterization of the  $\text{Max-CSP}(\mathcal{F})$  problems that are solvable *exactly* in polylogarithmic space.

**Theorem 1.4** (Succinct version). *For every  $q, k \in \mathbb{N}$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , the Max-CSP( $\mathcal{F}$ ) problem is solvable exactly in deterministic logarithmic space if and only if there is a constant  $\sigma \in [q]$  such that every satisfiable function in  $\mathcal{F}$  is satisfied by the all  $\sigma$ -assignment. All remaining families  $\mathcal{F}$  require  $\Omega(n)$  space to solve exactly.*

The proof of this theorem is by elementary reductions from standard communication complexity problems and included in [Section 9](#).

## 1.2 Comparison with previous works

[Theorems 1.1](#), [1.3](#) and [2.17](#) significantly extend similar theorems by the authors [\[CGSV21\]](#) and we compare the results shortly. First we discuss work prior to theirs. Prior to [\[CGSV21\]](#), there were a few works where individual problems were explored – specifically the Max-CUT problem

---

<sup>1</sup>By approximability of  $\text{Max-CSP}(\mathcal{F})$  we refer to the quantity  $\inf_{\beta} \sup_{\gamma} \{\beta/\gamma\}$  over polylog space solvable  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) problems.

was explored in [KKS15, KKS17, KK19, CGV20], Max-DICUT in [GVV17, CGV20], and Unique Games (which is really a family of problems) was analyzed in [GT19]. The most extensive prior work is from Chou, Golovnev, and Velusamy [CGV20], which gives a classification along the lines of [Theorem 1.3](#) in the special case where  $q = 2$  and  $k = 2$ , but they only study the approximability of  $\text{Max-CSP}(\mathcal{F})$  where constraints are applied to literals and  $\mathcal{F}$  contains a single function. Note that in this case, the number of different  $\mathcal{F}$ 's is finite. Indeed their classification is fully explicit in that case! Modulo the explicitness and the exact polynomial in the space lower bound, all the previous results are subsumed by the prior work [CGSV21].

Turning to the previous work [CGSV21], the main theorems there are restrictions of [Theorems 1.1, 1.3](#) and [2.17](#) to the special case where (1)  $\mathcal{F}$  contains a single function  $f$ , (2)  $q = 2$  and (3) Constraints are placed on “literals” rather than variables. We discuss the technical differences shortly, but start by explaining why even in terms of expressiveness the difference is significant. Indeed, due to restriction (1), their work does not capture the Max-3SAT problem which requires different constraints for clauses of length 1, 2, and 3. This is a quantitatively significant restriction in that the approximability in this case is “smaller” than that of  $\text{Max-CSP}(f)$  for any of the constituent functions. So hard instances do involve a mix of constraints! The lack of expressiveness induced by the second restriction of Boolean variables is perhaps more obvious. Natural examples of CSPs that require larger alphabets are Max- $q$ -Coloring and Unique Games. Finally we turn to restriction (3) — the inability to capture CSP problems over variables. This restriction prevents previous works from capturing some very basic problems including Max-CUT and Max-DICUT. Furthermore, the notion of “literals” is natural only in the setting of Boolean variables — so overcoming this restriction seems crucial to eliminating the restriction of Booleanity of the variables.

We repeat some folklore knowledge about  $\text{Max-CSP}(\mathcal{F})$ . By expanding the study to allow multiple functions in  $\mathcal{F}$ , the class of  $\text{Max-CSP}(\mathcal{F})$ 's becomes more robust and complete. First we note that the requirement that all constraint functions in  $\mathcal{F}$  have the same arity  $k$  is not a restriction. Given a family  $\mathcal{F}'$  containing functions of varying arity, we can pad the functions with dummy variables that the functions do not depend on, to create a family  $\mathcal{F}$  of functions of the same arity. Similarly the requirement that the variables in a constraint be distinct is not a restriction. Rather it enhances the expressive power of  $\text{Max-CSP}(\mathcal{F})$ . For instance it allows the study of Max-3SAT (where clauses have length at most 3) as well as Max-E3SAT (where clauses have length exactly 3) — a difference that would not be captured without the distinctness requirement. The distinctness requirement can be dropped by changing the family  $\mathcal{F}$  to include the functions on a smaller number of distinct variables obtained by repeating variables explicitly in a constraint. Finally we note the use of “variables” only and not “literals” also enriches the study. On the one hand, one can always build in the ability to apply functions to literals by adding new functions to the class  $\mathcal{F}$  that explicitly negate chosen variables. On the other hand, this requirement to work with functions on variables, and not literals, allows for distinctions between problems such as Max-CUT and Max-2LIN. In the former the constraint requires that two variables not be equal whereas the latter allows arbitrary  $\mathbb{F}_2$ -linear equations on 2 distinct variables.

In terms of technical challenges in extending the previous work, we will get into the specifics more in [Section 1.3](#), but here we highlight some of the hurdles. We note that restriction (1) is not very significant and indeed it is easy to extend the techniques from the previous work to this case, so we focus on the other two. On the algorithmic side, a key ingredient in previous works is the use of a space efficient estimator for the  $\ell_1$  norm of a vector (in the “turnstile” model). Note that the  $\ell_1$  norm of a vector  $\mathbf{y} \in \mathbb{R}^n$  is the result of the following Boolean optimization problem

$\|\mathbf{y}\|_1 = \max_{\mathbf{x} \in \{-1,1\}^n} \sum_{i=1}^n x_i y_i$ . Using the ability to solve this optimization problem naturally helps with other Boolean optimization problems including some Max-CSP( $\mathcal{F}$ ) problems as well, but extending to the non-Boolean setting seems to need new algorithms (and of course new reductions to these algorithms). Turning to the negative results, Booleanity seems deeply embedded in the techniques of the previous work. The core machinery is Fourier analysis over the Boolean hypercube. Also central to their hardness result is the ability to negate variables. Roughly the hardness result creates “hard instances” by creating a random instance of a CSP that is satisfied by the all 1’s assignment, and then complementing a random subset of variables to hide the planted assignment. The entire methodology seems to fail when complementation is not allowed. In [Section 1.3](#) we describe how we overcome these hurdles to get our result.

### 1.3 Technical contributions

Our proof of [Theorem 1.1](#) follows the same main steps as the proof of the analogous theorem from [\[CGSV21\]](#). Given  $\mathcal{F}$ ,  $\gamma$  and  $\beta$ , we start by defining two convex sets  $K^Y$  and  $K^N$  in finite dimension that aims to determine efficient solvability of  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ). We then show that if  $K^Y$  and  $K^N$  are disjoint then the separating hyperplane (in finite dimension) can be used to design an efficient algorithm for solving  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) (for all instance lengths!). Finally we show that if the sets overlap, then a point in the intersection can be used to show hardness of the streaming problem in the dynamic setting via a sequence of communication complexity results. We note that while this framework remains the same in this work, each step has substantial differences with some non-obvious choices in each case, apart from technical hurdles. We elaborate on the details below. We remark that the presence of multiple functions in  $\mathcal{F}$  is not a serious issue and so in this overview, fix  $\mathcal{F} = \{f\}$  for simplicity.

**Convex Set Framework.** In the previous work, the convex sets live in  $k$  dimensions, and aim to capture hard “weighted instances” on  $k$  variables. More specifically, we view an instance as a distribution on constraints in a natural way. All constraints apply to the  $k$  variables  $x_1, \dots, x_k$  in order (and the only variation is which variables are negated and which ones are not). The “Yes” set  $K^Y$  represents the marginals (how often is a variable positively and how often negated) of distributions for which the all 1s assignment satisfies at least  $\gamma$  fraction of the constraints. The “No” set  $K^N$  is more subtle and represents the marginals of distributions for which no “symmetric probabilistic assignment” satisfies more than  $\beta$  fraction of the constraints. Specifically for every  $p \in [0, 1]$ , assigning the variables independently according to  $\text{Bern}(p)$  is expected to satisfy at most  $\beta$  fraction of the constraints. (Note in particular there may exist assignments satisfying many (even all!) the constraints, but they would be not remain satisfying under permutation of the variables.)

In our setting, when negations are not allowed, the above does not make sense. Using the “all 1s” assignment as the canonical one does not work since there is no way to use negations of variables to move this planted assignment. The key to our framework is to use  $kq$  variables  $x_{i,\sigma}$  for  $i \in [k]$  and  $\sigma \in [q]$  which we view as a  $k \times q$  matrix. (Note that even for  $q = 2$  this does not lead to the same convex sets as in [\[CGSV21\]](#).) Our distributions/instances involve constraints placed on  $k$ -tuples of the form  $(x_{1,a_1}, \dots, x_{k,a_k})$  - so the  $i$ th variable in a constraint is always a variable from the  $i$ th row. In our planted assignment the variable  $x_{i,\sigma}$  is assigned the value  $\sigma$ . This leads to  $K^Y \subseteq \mathbb{R}^{kq}$  representing the marginals (probability that the  $i$ th variable in a random constraint involves variable  $x_{i,\sigma}$ ) of distributions on constraints where the planted assignment satisfies at least  $\gamma$ -fraction of the constraints. Defining the set  $K^N$  turns out to be more subtle than in the previous



work, but the correct answer turns out to be:  $K^N$  consists of the marginals of distributions where no “column-symmetric probabilistic assignment” to the variables satisfies more than  $\beta$  fraction of the constraints. In other words we pick  $q$  distributions  $\mathcal{P}_\sigma$  supported on  $[q]$ , one for each  $\sigma \in [q]$ , then assigning  $x_{i,\sigma}$  according to  $\mathcal{P}_\sigma$  independently for each  $i$  and  $\sigma$ , does not satisfy more than  $\beta$  fraction of the constraints. Choosing the distribution of each variable to be completely arbitrary does not work (we can not design an algorithm in this case), while choosing the distributions to be all identical also does not work (we can not prove the lower bounds). The “column-symmetric” option turns out to be the sweet spot where both sides work!

**Approximation algorithm.** The challenge for the algorithmic result is turning the linear separability of  $K^Y$  and  $K^N$  to get an algorithm for instances of any length. [CGSV21] approach this question by using the linear separator to define, for an instance  $\Psi$  on  $n$  variables, a vector they call the “bias” vector  $\text{bias} \in \mathbb{R}^n$  and showing that the  $\ell_1$  norm of  $\text{bias}$  gives a good approximator to the maximum number of satisfiable constraints. The latter step is in hindsight natural given that both the  $\ell_1$  norm  $\text{bias}$  and  $\text{Max-CSP}(\mathcal{F})$  are optimization problems whose solution space is  $\{-1, 1\}^n$ . The algorithmic result follows since  $\ell_1$  norms can be estimated efficiently in polylog space [Ind00, KNW10].

In our setting we show how to use the linear separator to define an  $n \times q$  bias *matrix*  $\text{bias}$  and show that the  $\|\cdot\|_{1,\infty}$  norm of this matrix (i.e, the  $\ell_1$  norm of the vector whose coordinates are the  $\ell_\infty$  norms of the rows of  $\text{bias}$ ) gives a good approximator to the maximum number of satisfiable constraints. Fortunately estimation of a wide variety of norms have been explored in the streaming setting, and in particular the  $\|\cdot\|_{1,\infty}$ -norm is also known to be approximable in polylogarithmic space and this yields our upper bound. Thus the novel part here is in the determination of the right matrix and the right norm to look at, and then the rest follows naturally from previous works.

**Lower bounds.** Finally we turn to the lower bounds. We start with the negative result establishing the dichotomy in the dynamic setting. Here we are given that there are two distributions  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  on constraints that have the same one-wise marginals. The distributions on constraints can be viewed as distributions on  $[q]^k$  (or  $\{0, 1\}^k$  in the setting  $q = 2$ ). In [CGSV21] they create a communication problem called “Randomized Mask Detection (RMD)” based on the two distributions and prove it is hard for one-way communication and the streaming lower bound follows. In this RMD problem, Alice gets a vector  $x \in \{0, 1\}^n$  and Bob gets a set  $S \subseteq [n]$  and a randomly masked (XOR-ed) copy of  $x|_S$  where the mask is either derived as a concatenation of strings drawn independently from  $\mathcal{D}_Y$  or from  $\mathcal{D}_N$ , and they have to determine which of the two conditions holds. A one-way communication lower bound of  $s$  for RMD with  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  whose marginals are equal and contained in  $K^Y \cap K^N$  implies an  $\Omega(s)$  lower bound for  $(\gamma, \beta)$ - $\text{Max-CSP}(\mathcal{F})$ . [CGSV21] give an  $\Omega(\sqrt{n})$  lower bound for this problem in two steps: First they use Fourier analysis of Boolean functions to show hardness of RMD for a restricted class of distributions  $(\mathcal{D}_Y, \mathcal{D}_N)$ . They then reduce this class of hard problems to RMD for every pair. To highlight our challenges, note that negations and Booleanity occur everywhere including in the very definition of RMD (due to the XOR masking), the Boolean Fourier analysis, and even in the reduction which uses properties of the hypercube  $\{-1, 1\}^k$ . (In particular they use the simple feature that the cube decomposes into two smaller subcubes.)

We overcome these obstacles by a mix of steps. First, we work with a “harder”<sup>2</sup> communication

---

<sup>2</sup>By “harder” we mean that the communication task is harder so our task of proving lower bounds on communi-



problem called Signal Detection (SD) — where rather than getting an entire XOR mask of  $x|_S$  with  $m$  strings drawn independently according to  $\mathcal{D}_Y$  or  $\mathcal{D}_N$ , Bob gets weak signals indicating “Does  $x$  equal the  $i$ th masking string, on the appropriate coordinates?” This problem does not use XORs in the definition and so generalizes to  $q$ -ary alphabets and does not need negations either. It also continues to be reducible to the streaming problem of interest, so we still get space lower bounds from this. The main challenge now is proving hardness of the SD problem for distributions  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  over  $[q]^k$  with matching marginals. For this part we switch back to the Boolean RMD problem, and create a variant where Alice and Bob get some additional advice. We extend the Fourier-analytic proof of the previous work (which goes back to [GKK<sup>+</sup>09]) to show hardness of our Boolean advice-RMD problem. The advice in the Boolean problem now allows us to get hardness for general  $q$ -ary SD problems using combinatorial methods. While the natural combinatorial domain would have been  $[q]^k$ , we work with the more general product sets  $[q_1] \times \dots \times [q_k]$  and this set turns out to have the right decomposability for us. Putting these ingredients together gives a  $\Omega(\sqrt{n})$  lower bound on the communication complexity of SD whenever  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  have matching marginals.

From this point onwards our proof follows roughly the same outline as in [CGSV21]. To get hardness in the dynamic setting, we translate the hardness of the two player one-way communication SD problem into the hardness of a many (but constant) player simultaneous communication problem. This allows us to use the framework and powerful results of Li, Nguyen and Woodruff [LNU14] and Ai, Hu, Li and Woodruff [AHLW16] to get a space  $\sqrt{n}$  lower bound for every  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) whenever the corresponding  $K^Y$  and  $K^N$  intersect. For the hardness result in the insertion only setting, the lower bound on the simultaneous communication problem no longer suffices. So here we craft our own reduction to a  $T$ -player one-way communication problem which reduces in turn to  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) in the insertion-only streaming setting. Unfortunately this step works only in some restricted cases (for instance if  $\mathcal{D}_N$  is the uniform distribution on  $[q]^k$ ) and this yields our lower bound (Theorem 2.17) in the insertion-only setting.

## 2 Results

We let  $\mathbb{N}$  denote the set of positive integers. We let  $[n]$  denote the set  $\{1, \dots, n\}$ . For a finite set  $\Omega$ , let  $\Delta(\Omega)$  denote the space of all probability distributions over  $\Omega$ , i.e.,

$$\Delta(\Omega) = \left\{ \mathcal{D} : \Omega \rightarrow \mathbb{R}^{\geq 0} \mid \sum_{\omega \in \Omega} \mathcal{D}(\omega) = 1 \right\}.$$

We view  $\Delta(\Omega)$  as being contained in  $\mathbb{R}^{|\Omega|}$ . We use  $X \sim \mathcal{D}$  to denote a random variable drawn from the distribution  $\mathcal{D}$ .

### 2.1 Main Notions

We start with some definitions and notations crucial to describe the central objects of our study, which are distributions of constraints applied to a special set of variables. We start with the variables, and the planted assignment, before turning to notions about distributions.

---

cation complexity is easier.

In this part we consider distributions of constraints over a set of  $kq$  variables denoted  $\mathbf{x} = (x_{i,\sigma} \mid i \in [k], \sigma \in [q])$ . (We think of the variables as sitting in a  $k \times q$  matrix with  $i$  indexing the rows and  $\sigma$  indexing the columns.) For  $f \in \mathcal{F}$  and  $\mathbf{a} \in [q]^k$ , let  $\mathcal{C}(f, \mathbf{a})$  denote the constraint  $f(x_{1,a_1}, \dots, x_{k,a_k})$ . For an assignment  $\mathbf{b} = (b_{i,\sigma} \mid i \in [k], \sigma \in [q]) \in [q]^{kq}$  we use the notation  $\mathcal{C}(f, \mathbf{a})(\mathbf{b})$  to denote the value  $f(b_{1,a_1}, \dots, b_{k,a_k})$ . We let  $\mathbb{I} \in [q]^{kq}$  denote the assignment  $\mathbb{I}_{i,\sigma} = \sigma$ . (In the following section we will use  $\mathbb{I}$  as our planted assignment.)

We now turn to defining the “marginals” of distributions. For  $\mathcal{D} \in \Delta(\mathcal{F} \times [q]^k)$ , we let  $\boldsymbol{\mu}(\mathcal{D}) = (\mu_{f,i,\sigma})_{f \in \mathcal{F}, i \in [k], \sigma \in [q]}$  be given by  $\mu_{f,i,\sigma} = \Pr_{(g,\mathbf{a}) \sim \mathcal{D}}[g = f \text{ and } a_i = \sigma]$ . Thus the marginal  $\boldsymbol{\mu}(\mathcal{D})$  lies in  $\mathbb{R}^{|\mathcal{F}| \times qk}$ . We often reduce our considerations to families  $\mathcal{F}$  containing a single element. In such cases we simplify the notion of a distribution to  $\mathcal{D} \in \Delta([q]^k)$ . For  $\mathcal{D} \in \Delta([q]^k)$ , we let  $\boldsymbol{\mu}(\mathcal{D}) = (\mu_{i,\sigma})_{i \in [k], \sigma \in [q]}$  be given by  $\mu_{i,\sigma} = \Pr_{\mathbf{a} \sim \mathcal{D}}[a_i = \sigma]$ .

Next we introduce our family of distributions that capture our “Yes” and “No” instances. “Yes” instances are highly satisfied by our planted assignment, while “No” instances are not very satisfied by any “column-symmetric”, independent, probabilistic assignment. The fact that we only consider distributions on  $kq$  variables makes this a set in finite-dimensions.

**Definition 2.1** (Space of YES/NO distributions). *For  $q, k \in \mathbb{N}$ ,  $\gamma \in [0, 1]$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , we let*

$$S_\gamma^Y(\mathcal{F}) = \left\{ \mathcal{D} \in \Delta(\mathcal{F} \times [q]^k) \mid \mathbb{E}_{(f,\mathbf{a}) \sim \mathcal{D}} [\mathcal{C}(f, \mathbf{a})(\mathbb{I})] \geq \gamma \right\}.$$

For  $\beta \in [0, 1]$  we let

$$S_\beta^N(\mathcal{F}) = \left\{ \mathcal{D} \in \Delta(\mathcal{F} \times [q]^k) \mid \forall (\mathcal{P}_\sigma \in \Delta([q]))_{\sigma \in [q]}, \mathbb{E}_{(f,\mathbf{a}) \sim \mathcal{D}} \left[ \mathbb{E}_{\mathbf{b}, b_{i,\sigma} \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{b})] \right] \leq \beta \right\}.$$

By construction, for  $\beta < \gamma$ , the sets  $S_\gamma^Y(\mathcal{F})$  and  $S_\beta^N(\mathcal{F})$  are disjoint. (In particular for any  $\mathcal{D} \in S_\gamma^Y(\mathcal{F})$ ,  $\mathbb{I}$  corresponds to a (deterministic!) column symmetric assignment that satisfies  $\gamma > \beta$  fraction of constraints, so  $\mathcal{D} \notin S_\beta^N(\mathcal{F})$ .) The key to the analysis of low-space streaming algorithms is that they only seem to be able to estimate the marginals of a distribution — so we turn to exploring the marginals of the sets above.

**Definition 2.2** (Marginals of Yes/NO Distributions). *For  $\gamma, \beta \in [0, 1]$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , we let*

$$K_\gamma^Y(\mathcal{F}) = \{\boldsymbol{\mu}(\mathcal{D}) \in \mathbb{R}^{|\mathcal{F}|kq} \mid \mathcal{D} \in S_\gamma^Y(\mathcal{F})\} \text{ and } K_\beta^N(\mathcal{F}) = \{\boldsymbol{\mu}(\mathcal{D}) \in \mathbb{R}^{|\mathcal{F}|kq} \mid \mathcal{D} \in S_\beta^N(\mathcal{F})\}.$$

See [Section 2.5](#) for some examples of the sets  $S_\gamma^Y(\mathcal{F}), S_\beta^N(\mathcal{F}), K_\gamma^Y(\mathcal{F}), K_\beta^N(\mathcal{F})$ .

## 2.2 Results in the dynamic setting

The following theorem now formalizes the informal statement that low space algorithms can only capture the marginals of distributions in the dynamic setting.

**Theorem 2.3** (Dynamic Dichotomy Theorem). *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  and for every  $0 \leq \beta < \gamma \leq 1$ , the following hold:*

1. If  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ , then  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) in the dynamic setting admits a probabilistic streaming algorithm that uses  $O(\log^3 n)$  space<sup>3</sup> on instances on  $n$  variables.
2. If  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) \neq \emptyset$ , then for every  $\varepsilon > 0$ , the  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) in the dynamic setting requires  $\Omega(\sqrt{n})$  space<sup>4</sup> on instances on  $n$  variables. Furthermore, if  $\gamma = 1$ , then  $(1, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) in the dynamic setting requires  $\Omega(\sqrt{n})$  space.

Theorem 2.3 is proved in two parts: Theorem 4.1 proves Theorem 2.3, Part 1 while Theorem 5.1 proves Theorem 2.3, Part 2.

### 2.2.1 Decidability of the Classification

We now complement Theorem 2.3 by showing that the condition  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ ? can be decided in polynomial space given  $\gamma$  and  $\beta$  as ratios of  $\ell$ -bit integers and members of  $\mathcal{F}$  as truth tables. (So the input is of size  $O(\ell + |\mathcal{F}| \cdot q^k)$  and our algorithm needs space polynomial in this quantity.)

**Theorem 2.4.** *For every  $k, q \in \mathbb{N}$   $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , and  $\ell$ -bit rationals  $\beta, \gamma \in [0, 1]$  (i.e.,  $\beta$  and  $\gamma$  are expressible as the ratio of two integers in  $\{-2^\ell, \dots, 2^\ell\}$ ), the condition “ $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ ?” can be decided in space  $\text{poly}(|\mathcal{F}|, q^k, \ell)$  given truth tables of all elements of  $\mathcal{F}$  and  $\gamma$  and  $\beta$  as  $\ell$ -bit rationals.*

We prove Theorem 2.4 in this section. Our analysis mimics that of [CGSV21] in a straightforward manner, given the definition of  $K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N(\mathcal{F})$ . The following lemma states some basic properties of the sets  $S_\gamma^Y(\mathcal{F})$ ,  $S_\beta^N(\mathcal{F})$ ,  $K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N(\mathcal{F})$  and uses them to express the condition “ $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ ?” in the quantified theory of reals.

**Lemma 2.5.** *For every  $k, q \in \mathbb{N}$   $\beta, \gamma \in [0, 1]$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , the sets  $S_\gamma^Y(\mathcal{F})$ ,  $S_\beta^N(\mathcal{F})$ ,  $K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N(\mathcal{F})$  are bounded, closed, and convex. Furthermore, the condition  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$  can be expressed in the quantified theory of reals with 2 quantifier alternations,  $O(|\mathcal{F}|q^k + q^2)$  variables, and polynomials of degree at most  $k + 1$ .*

*Proof.* We start by observing that  $\Delta(\mathcal{F} \times [q]^k)$  is a bounded convex polytope in  $\mathbb{R}^{|\mathcal{F}| \times [q]^k}$ . Furthermore, viewing  $\mathcal{D}$  as a vector in  $\mathbb{R}^{|\mathcal{F}| \times [q]^k}$ , for any given  $\mathbf{b} \in [q]^k$  the quantity  $\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}}[C(f, \mathbf{a})(\mathbf{b})]$  is linear in  $\mathcal{D}$ . Thus  $S_\gamma^Y(\mathcal{F})$  is given by a single linear constraint on  $\Delta(\mathcal{F} \times [q]^k)$  making it a bounded convex polytope as well.  $S_\beta^N(\mathcal{F})$  is a bit more complex - in that there are infinitely many linear inequalities defining it (one for every distribution  $(\mathcal{P}_\sigma)_{\sigma \in [q]}$ ). Nevertheless this leaves  $S_\beta^N(\mathcal{F})$  bounded, closed (as infinite intersection of closed sets is closed), and convex (though it may no longer be a polytope). Finally since  $K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N(\mathcal{F})$  are linear projections of  $S_\gamma^Y(\mathcal{F})$  and  $S_\beta^N(\mathcal{F})$  respectively, they retain the features of being bounded, closed and convex.

Finally to get an effective algorithm for intersection detection, we express the intersection condition in the quantified theory of the reals. To get this, we note that  $(\mathcal{P}_\sigma)_{\sigma \in [q]}$  can be expressed by  $q^2$  variables, specifically using variables  $\mathcal{P}_\sigma(\tau)$  for every  $\sigma, \tau \in [q]$  where  $\mathcal{P}_\sigma(\tau)$  denotes the

<sup>3</sup>In particular, the space complexity is  $O(\log^3 n)$  bits, or  $O(\log^2 n)$  cells where each cell is  $O(\log n)$  bits long. Crucially while the constant in the  $O(\cdot)$  depends on  $k, \gamma$  and  $\beta$ , the exponent is a universal constant.

<sup>4</sup>Again, the constant hidden in the  $\Omega$  notation depends on  $k, \gamma$  and  $\beta$ .

probability of  $\tau$  in  $\mathcal{P}_\sigma$ . In terms of these variables (which will eventually be quantified over) the condition  $\mathbb{E}_{(f,\mathbf{a})\sim\mathcal{D}} [\mathbb{E}_{\mathbf{b},b_i,\sigma\sim\mathcal{P}_\sigma} [\mathcal{C}(f,\mathbf{a})(\mathbf{b})]] \leq \beta$  is a multivariate polynomial inequality in  $(\mathcal{P}_\sigma)_\sigma$  and  $\mathcal{D}$ . (Specifically we get a polynomial of total degree at most  $k$  in  $(\mathcal{P}_\sigma)_\sigma$ , and of total degree at most one in  $\mathcal{D}$ .) This allows us to use the following quantified system to express the condition  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) \neq \emptyset$ :

$$\exists \mathcal{D}_Y, \mathcal{D}_N \in \mathbb{R}^{|\mathcal{F}| \times q^k}, \forall ((\mathcal{P}_\sigma)_\sigma) \in \mathbb{R}^{q^2} \text{ s.t.} \\ \mathcal{D}_Y, \mathcal{D}_N, (\mathcal{P}_\sigma)_\sigma, \forall \sigma \in [q] \text{ are distributions,} \quad (2.6)$$

$$\forall f_0 \in \mathcal{F}, \forall i \in [k], \tau \in [q] \quad \Pr_{(f,\mathbf{a})\sim\mathcal{D}_Y} [f = f_0 \text{ and } a_i = \tau] = \Pr_{(f,\mathbf{a})\sim\mathcal{D}_N} [f = f_0 \text{ and } a_i = \tau], \quad (2.7)$$

$$\mathbb{E}_{(f,\mathbf{a})\sim\mathcal{D}_Y} [\mathcal{C}(f,\mathbf{a})(\mathbb{I})] \geq \gamma, \quad (2.8)$$

$$\mathbb{E}_{(f,\mathbf{a})\sim\mathcal{D}_N} \left[ \mathbb{E}_{\mathbf{b},b_i,\sigma\sim\mathcal{P}_\sigma} [\mathcal{C}(f,\mathbf{a})(\mathbf{b})] \right] \leq \beta. \quad (2.9)$$

Note that Eqs. (2.6) to (2.8) are just linear inequalities in the variables  $\mathcal{D}_Y, \mathcal{D}_N$ . As noticed above Eq. (2.9) is an inequality in the  $\mathcal{P}_\sigma$ s and  $\mathcal{D}_N$ , of total degree at most  $k + 1$ .

We thus get that the intersection problem can be expressed in the quantified theory of the reals by an expression with two quantifier alternations,  $2|\mathcal{F}|q^k + q^2$  variables and  $O(|\mathcal{F}|q^k + q^2)$  polynomial inequalities, with polynomials of degree at most  $k + 1$ . (Most of the inequalities are of the form  $\mathcal{D}_Y(\mathbf{b}) \geq 0$  or  $\mathcal{D}_N(\mathbf{b}) \geq 0$ . We also have  $O(|\mathcal{F}|kq)$  equalities (saying probabilities must add to one and matching the marginals of  $\mathcal{D}_Y$  and  $\mathcal{D}_N$ ). Of the two remaining, Eq. (2.8) is linear, only Eq. (2.9) is a higher-degree polynomial.  $\square$

As in [CGSV21] we now appeal to the following theorem on the complexity of the quantified theory of the reals.

**Theorem 2.10** ([BPR06, Theorem 14.11, see also Remark 13.10]). *The truth of a quantified formula with  $w$  quantifier alternations over  $K$  variables and polynomial (potentially strict) inequalities can be decided in space  $K^{O(w)}$  and time  $2^{K^{O(w)}}$ .*

Specifically, Theorem 14.11 in [BPR06] asserts the time complexity above, and Remark 13.10 yields the space complexity. We are now ready to prove Theorem 2.4.

*Proof of Theorem 2.4.* The quantified polynomial system given by Lemma 2.5 yields parameters  $K = O(|\mathcal{F}|q^k + q^2)$  for the number of variables and  $w = 2$  for the number of alternations. Applying Theorem 2.10 with these parameters yields the theorem.  $\square$

### 2.2.2 Approximation resistance in the dynamic streaming setting

We now turn to the notion of ‘‘approximation resistant’’  $\text{Max-CSP}(\mathcal{F})$  problems. We start with a discussion where  $\mathcal{F} = \{f\}$ . In the setting where constraints are applied to literals rather than variables, the notion of approximation resistance is used to refer to problems where it is hard to outperform the uniform random assignment. In other words if  $\rho(f)$  is defined to be the probability that a random assignment satisfies  $f$ , then  $\text{Max-CSP}(f)$  is defined to be approximation resistant if  $(1 - \varepsilon, \rho(f) + \varepsilon)\text{-Max-CSP}(f)$  is hard. In our setting however, where constraints are applied to

variables, this notion is a bit more nuanced. Here it may be possible to construct functions where a random assignment does poorly and yet every instance has a much higher value.<sup>5</sup> In our setting the correct notion is to consider the probability of satisfying  $f$  maximized over all i.i.d. distributions for variables. If this quantity is defined to be  $\rho(f)$ , then clearly every instance of  $\text{Max-CSP}(f)$  achieves value  $\rho(f)$ . And now approximation resistance can be defined as beating this trivial lower bound. Below we formalize the above discussion while also extending to general  $\mathcal{F}$ .

**Definition 2.11** (Approximation resistance). *For  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , we define*

$$\rho_{\min}(\mathcal{F}) = \lim_{\Psi \text{ instance of Max-CSP}(\mathcal{F})} \inf \{ \text{val}_{\Psi} \}.$$

We also define

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

We say that  $\text{Max-CSP}(\mathcal{F})$  is streaming-approximation-resistant in the dynamic setting (insertion-only setting) if for every  $\varepsilon > 0$  there exists  $\delta > 0$  such that  $(1 - \varepsilon, \rho(\mathcal{F}) + \varepsilon)$ - $\text{Max-CSP}(\mathcal{F})$  requires  $\Omega(n^{\delta})$  space in the dynamic setting (resp. insertion-only setting).

**Proposition 2.12.** *For every  $q, k \in \mathbb{N}$ ,  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  we have  $\rho_{\min}(\mathcal{F}) = \rho(\mathcal{F})$ .*

*Proof.* We start by showing  $\rho(\mathcal{F}) \leq \rho_{\min}(\mathcal{F})$ . Fix an instance  $\Psi$  of  $\text{Max-CSP}(\mathcal{F})$  and let  $\mathcal{D}_{\mathcal{F}}$  be the distribution on  $\mathcal{F}$  obtained by picking a random constraint of  $\Psi$  and looking at the function (while ignoring the variables that the constraint is applied to). By definition of  $\rho(\mathcal{F})$ , there exists a distribution  $\mathcal{D} \in \Delta([q])$  such that  $\mathbb{E}_{f \sim \mathcal{D}_{\mathcal{F}}, \mathbf{a} \sim \mathcal{D}^k} [f(\mathbf{a})] \geq \rho(\mathcal{F})$ . Now consider a random assignment to the variables of  $\Psi$  where variable  $x_j$  is assigned a value independently according to  $\mathcal{D}$ . It can be verified that  $\mathbb{E}_{\mathbf{x}}[\text{val}_{\Psi}(\mathbf{x})] \geq \rho(\mathcal{F})$  and so  $\text{val}_{\Psi} \geq \rho(\mathcal{F})$ . We thus conclude that  $\rho(\mathcal{F}) \leq \text{val}_{\Psi}$  for all  $\Psi$  and so  $\rho(\mathcal{F}) \leq \rho_{\min}(\mathcal{F})$ .

We now turn to the other direction. We prove that for every  $\varepsilon > 0$  we have  $\rho_{\min}(\mathcal{F}) \leq \rho(\mathcal{F}) + \varepsilon$  and the inequality follows by taking limits. Let  $\mathcal{D}_{\mathcal{F}}$  be the distribution achieving the minimum in the definition of  $\rho(\mathcal{F})$ . Given  $\varepsilon > 0$  let  $n$  be a sufficiently large integer and let  $m = O(n^k/\varepsilon)$ . Let  $\Psi$  be the instance of  $\text{Max-CSP}(\mathcal{F})$  on  $n$  variables with  $m$  constraints chosen as follows: For every  $\mathbf{j} \in [n]^k$  with distinct coordinates and every  $f \in \mathcal{F}$  we place  $\lfloor \mathcal{D}_{\mathcal{F}}(f)/\varepsilon \rfloor$  copies of the constraint  $(f, \mathbf{j})$ .

We claim that the  $\Psi$  generated above satisfies  $\text{val}_{\Psi} \leq \rho(\mathcal{F}) + \varepsilon/2 + O(1/n)$  and this suffices for the proposition. To see the claim, fix an assignment  $\nu \in [q]^n$  and let  $\mathcal{D} \in \Delta([q])$  be the distribution induced by sampling  $i \in [n]$  uniformly and outputting  $\nu_i$ . On the one hand we have from the definition of  $\rho(\mathcal{F})$  that  $\mathbb{E}_{f \sim \mathcal{D}_{\mathcal{F}}, \mathbf{a} \sim \mathcal{D}^k} [f(\mathbf{a})] \leq \rho(\mathcal{F})$ . On the other hand we have that the distribution obtained by sampling a random constraint  $(f, \mathbf{j})$  of  $\Psi$  and outputting  $(f, \nu|_{\mathbf{j}})$  is  $\varepsilon/2 + O(1/n)$  close in total variation distance to sampling  $f \sim \mathcal{D}_{\mathcal{F}}$  and  $\mathbf{a} \sim \mathcal{D}^k$ . (The  $\varepsilon/2$  gap comes from the rounding down of each constraint to an integral number, and the  $O(1/n)$  gap comes from the fact that  $\mathbf{j}$  is sampled from  $[n]$  without replacement.) We thus conclude that

$$\text{val}_{\Psi}(\nu) \leq \mathbb{E}_{f \sim \mathcal{D}_{\mathcal{F}}, \mathbf{a} \sim \mathcal{D}^k} [f(\mathbf{a})] + \varepsilon/2 + O(1/n) \leq \rho(\mathcal{F}) + \varepsilon/2 + O(1/n) \leq \rho(\mathcal{F}) + \varepsilon.$$

Since this holds for every  $\nu$  we conclude that this upper bounds  $\text{val}_{\Psi}$  as well thus establishing the claim, and hence the proposition.  $\square$

<sup>5</sup>Take for instance  $f(x_1) = 1$  iff  $x_1 = 1$ . The random assignment satisfies  $f$  with probability  $1/q$  while every instance is satisfiable!

As defined  $\rho_{\min}(\mathcal{F})$  is not immediately even computable, but using the equivalence to  $\rho(\mathcal{F})$  we get a polynomial space algorithm as shown next.

**Theorem 2.13.** *There is an algorithm  $A$  that, on input  $\mathcal{F} \subseteq \{[q]^k \rightarrow \{0,1\}\}$  presented as  $|\mathcal{F}|$  truth-tables and  $\tau \in \mathbb{R}$  presented as an  $\ell$ -bit rational, answers the question “Is  $\rho_{\min}(\mathcal{F}) \leq \tau$ ?” in space  $\text{poly}(|\mathcal{F}|, q^k, \ell)$ .*

*Proof.* By [Proposition 2.12](#) we have

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

Viewing  $\mathcal{D}_{\mathcal{F}} \in \mathbb{R}^{|\mathcal{F}|}$  and  $\mathcal{D} \in \mathbb{R}^q$  and noticing that the inner expectation is a degree  $k+1$  polynomial in  $\mathcal{D}_{\mathcal{F}}$  and  $\mathcal{D}$  we get, again using [Theorem 2.10](#), that there is a space  $\text{poly}(|\mathcal{F}|, q^k, \ell)$  algorithm answering the question “Is  $\rho_{\min}(\mathcal{F}) \leq \tau$ ?”.  $\square$

[Theorem 2.3](#) immediately yields a decidable characterization of approximation resistant Max-CSP( $\mathcal{F}$ ) problems in the dynamic streaming setting.

**Theorem 2.14.** *For every  $q, k \in \mathbb{N}$ , for every family  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0,1\}\}$ , Max-CSP( $\mathcal{F}$ ) is approximation resistant in the dynamic setting if and only if  $K_1^Y(\mathcal{F}) \cap K_{\rho(\mathcal{F})}^N(\mathcal{F}) \neq \emptyset$ . Furthermore, if Max-CSP( $\mathcal{F}$ ) is streaming-approximation-resistant in the dynamic setting, then for every  $\varepsilon > 0$  we have that  $(1, \rho(\mathcal{F}) + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space. If Max-CSP( $\mathcal{F}$ ) is not streaming-approximation-resistant, then there exists  $\varepsilon > 0$  such that  $(1 - \varepsilon, \rho(\mathcal{F}) + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) can be solved in polylogarithmic space in the dynamic setting. Finally, given the truth-table of the functions in  $\mathcal{F}$  there is an algorithm running in space  $\text{poly}(q^k |\mathcal{F}|)$  that decides whether or not Max-CSP( $\mathcal{F}$ ) is streaming-approximation-resistant in the dynamic setting.*

*Proof.* By [Theorem 2.3](#) we have that Max-CSP( $\mathcal{F}$ ) is approximation-resistant if and only if  $K_{1-\varepsilon}^Y(\mathcal{F}) \cap K_{\rho(\mathcal{F})+\varepsilon}^N(\mathcal{F}) \neq \emptyset$  for every small  $\varepsilon > 0$ . Taking limits as  $\varepsilon \rightarrow 0$ , this implies that Max-CSP( $\mathcal{F}$ ) is approximation resistant if and only if  $K_1^Y(\mathcal{F}) \cap K_{\rho(\mathcal{F})}^N(\mathcal{F}) \neq \emptyset$ . If  $K_1^Y(\mathcal{F}) \cap K_{\rho(\mathcal{F})}^N(\mathcal{F}) = \emptyset$ , then by the property that these sets are closed (see [Lemma 2.5](#)), we have that there must exist  $\varepsilon > 0$  such that  $K_{1-\varepsilon}^Y(\mathcal{F}) \cap K_{\rho(\mathcal{F})+\varepsilon}^N(\mathcal{F}) = \emptyset$ . In turn this implies, again by [Theorem 2.3](#), that the  $(1 - \varepsilon, \rho(\mathcal{F}) + \varepsilon)$ -approximation version of Max-CSP( $\mathcal{F}$ ) can be solved by a streaming algorithm with  $O(\log^3 n)$  space.

To get the decidability result, we combine the ingredients from the proof of [Theorems 2.4](#) and [2.13](#). (We can’t use them as blackboxes since  $\rho_{\min}(\mathcal{F})$  may not be rational.) We create a quantified system of polynomial inequalities using a new variable called  $\rho$  and expressing the conditions  $\rho = \rho(\mathcal{F})$  (with further variables for  $\mathcal{D}_{\mathcal{F}}$  and  $\mathcal{D}$  as in the proof of [Theorem 2.13](#)) and expressing the conditions  $K_1^Y(\mathcal{F}) \cap K_{\rho}^N(\mathcal{F}) \neq \emptyset$  as in the proof of [Theorem 2.4](#). The resulting expression is thus satisfiable if and only if  $\mathcal{F}$  is approximation resistant, and this satisfiability can be decided in polynomial space in the input length  $q^k |\mathcal{F}|$  by [Theorem 2.10](#).  $\square$

### 2.3 Results in the insertion-only setting

We now turn to some special classes of functions where we can prove lower bounds in the insertion only setting matching the algorithmic results in the dynamic setting. To define these classes we need some definitions.



We start by defining the notion of a “one-wise independent” distribution  $\mathcal{D} \in \Delta(\mathcal{F} \times [q]^k)$ . (We note that this is somewhat related to, but definitely not the same as the notion of a family  $\mathcal{F}$  that *supports* one-wise independence which was defined informally in [Section 1](#). We will recall that notion shortly.) We also define the notion of a “padded one-wise pair” of distributions”.

**Definition 2.15** (One-wise independence and Padded one-wise independence). *For  $\mathcal{D} \in \Delta(\mathcal{F} \times [q]^k)$  we say that  $\mathcal{D}$  is one-wise independent (or has “uniform marginals”) if its marginal  $\boldsymbol{\mu}(\mathcal{D}) = (\mu_{f,i,\sigma})_{f \in \mathcal{F}, i \in [k], \sigma \in [q]}$  satisfies  $\mu_{f,i,\sigma} = \mu_{f,i,\sigma'}$  for every  $f \in \mathcal{F}$ ,  $i \in [k]$  and  $\sigma, \sigma' \in [q]$ .*

*We say that a pair of distributions  $(\mathcal{D}_1, \mathcal{D}_2)$  form a padded one-wise pair if there exist  $\mathcal{D}_0, \mathcal{D}'_1, \mathcal{D}'_2$  and  $\tau \in [0, 1]$  such that for every  $i \in \{1, 2\}$  we have  $\mathcal{D}'_i$  is one-wise independent and  $\mathcal{D}_i = \tau \mathcal{D}_0 + (1 - \tau) \mathcal{D}'_i$ .*

Our main lower bound in the insertion-only setting asserts that if  $S_\gamma^Y(\mathcal{F}) \times S_\beta^N(\mathcal{F})$  contains a padded one-wise pair  $(\mathcal{D}_Y, \mathcal{D}_N)$  then  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$ -space.

**Theorem 2.16** (Streaming lower bound in the insertion-only setting). *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  and for every  $0 \leq \beta < \gamma \leq 1$ , if there exists a padded one-wise pair of distributions  $\mathcal{D}_Y \in S_\gamma^Y(\mathcal{F})$  and  $\mathcal{D}_N \in S_\beta^N(\mathcal{F})$  then, for every  $\varepsilon > 0$ , the  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space in the insertion-only setting. Furthermore, if  $\gamma = 1$ , then  $(1, \beta + \varepsilon)$ -Max-CSP( $f$ ) requires  $\Omega(\sqrt{n})$  space in the insertion-only setting.*

[Theorem 2.16](#) is proved in [Section 5.2.4](#). As stated above the theorem is more complex to apply than, say, [Theorem 2.3](#), owing to the fact that the condition for hardness depends on the entire distribution (and the sets  $S_\gamma^Y$  and  $S_\beta^N$ ) rather than just marginals (or the sets  $K_\gamma^Y$  and  $K_\beta^N$ ). However it can be used to derive some clean results, specifically [Theorem 2.17](#) and [Theorem 1.3](#), that do depend only on the marginals. We prove these (assuming [Theorem 2.16](#)) below.

We say that  $f : [q]^k \rightarrow \{0, 1\}$  supports *one-wise independence* if there exists a distribution  $\mathcal{D}$  supported on  $f^{-1}(1)$  whose marginals are uniform on  $[q]$ . We say that a family  $\mathcal{F}$  strongly supports one-wise independence if every function  $f \in \mathcal{F}$  supports one-wise independence. We say that a family  $\mathcal{F}$  weakly supports one-wise independence if there exists  $\mathcal{F}' \subseteq \mathcal{F}$  satisfying  $\rho(\mathcal{F}') = \rho(\mathcal{F})$  and every function  $f \in \mathcal{F}'$  supports one-wise independence.

**Theorem 2.17.** *For every  $q, k \in \mathbb{N}$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  such that  $\mathcal{F}$  weakly supports one-wise independence, Max-CSP( $\mathcal{F}$ ) is approximation resistant in the insertion-only setting. In particular, for every  $\varepsilon > 0$ ,  $(1, \rho(\mathcal{F}) + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space.*

**Remark 2.18.** We note that [Theorem 1.2](#) differs from [Theorem 2.17](#) in two ways. First our definition of approximation resistance implies hardness of  $(1, \rho(\mathcal{F}))$ -Max-CSP( $\mathcal{F}$ ) whereas the informal discussion before [Theorem 1.2](#) asserted hardness of  $(1, \rho_{\min}(\mathcal{F}))$ -Max-CSP( $\mathcal{F}$ ). But [Proposition 2.12](#) asserts  $\rho(\mathcal{F}) = \rho_{\min}(\mathcal{F})$ . A second difference is that [Theorem 1.2](#) asserted hardness for  $\mathcal{F}$  that strongly supports one-wise independence whereas [Theorem 2.17](#) asserts hardness for  $\mathcal{F}$  that weakly supports one-wise independence. Thus [Theorem 2.17](#) is stronger and implies [Theorem 1.2](#).

*Proof.* Let  $\mathcal{F}' \subseteq \mathcal{F}$  be a family satisfying  $\rho(\mathcal{F}') = \rho(\mathcal{F})$  such that every function  $f \in \mathcal{F}'$  supports one-wise independence. Furthermore let  $\mathcal{D}_{\mathcal{F}} \in \Delta(\mathcal{F}')$  minimize  $\max_{\mathcal{D} \in \Delta([q]^k)} \{\mathbb{E}_{f \sim \mathcal{D}_{\mathcal{F}}, \mathbf{a} \sim \mathcal{D}^k} [f(\mathbf{a})]\}$ . For  $f \in \mathcal{F}'$  let  $\mathcal{D}_{\mathcal{F}} \in \Delta([q]^k)$  be the distribution with uniform marginals supported on  $f^{-1}(1)$ . Now let  $\mathcal{D}_Y$  be the distribution where  $(f, \mathbf{a}) \sim \mathcal{D}_Y$  is sampled by picking  $f \in \mathcal{D}_{\mathcal{F}}$  (where  $\mathcal{D}_{\mathcal{F}}$  is being viewed as an element of  $\Delta(\mathcal{F})$ ) and then sampling  $\mathbf{a} \sim \mathcal{D}_{\mathcal{F}}$ . Now let  $\mathcal{D}_N = \mathcal{D}_{\mathcal{F}} \times \text{Unif}([q]^k)$ . Note that  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  are one-wise independent distributions with  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$ . In particular this



implies that  $(\mathcal{D}_Y, \mathcal{D}_N)$  are a padded one-wise pair. We claim that  $\mathcal{D}_Y \in S_1^Y(\mathcal{F})$  and  $\mathcal{D}_N \in S_{\rho(\mathcal{F})}^N(\mathcal{F})$ . The theorem then follows immediately from [Theorem 2.16](#).

To see the claim, first note that by definition we have that  $(f, \mathbf{a}) \sim \mathcal{D}_Y$  satisfies  $\mathcal{C}(f, \mathbf{a})(\mathbb{I}) = f(\mathbf{a}) = f_0(\mathbf{a}) = 1$  with probability 1. Thus we have  $\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}}[\mathcal{C}(f, \mathbf{a})(\mathbb{I})] = 1$  and so  $\mathcal{D}_Y \in S_1^Y(\mathcal{F})$ . Now consider  $(f, \mathbf{a}) \sim \mathcal{D}_N$ . To show  $\mathcal{D}_N \in S_{\rho(\mathcal{F})}^N(\mathcal{F})$  we need to show  $\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}}[\mathbb{E}_{\mathbf{b}, b_i, \sigma \sim \mathcal{P}_\sigma}[\mathcal{C}(f, \mathbf{a})(\mathbf{b})]] \leq \rho(\mathcal{F})$  for every family of distributions  $(\mathcal{P}_\sigma \in \Delta([q]))_{\sigma \in [q]}$ . Now let  $\mathcal{P}$  be the distribution where  $\tau \sim \mathcal{P}$  is sampled by picking  $\sigma \sim \text{Unif}([q])$  and then sampling  $\tau \sim \mathcal{P}_\sigma$ . We have

$$\begin{aligned} \mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}} \left[ \mathbb{E}_{\mathbf{b}, b_i, \sigma \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{b})] \right] &= \mathbb{E}_{f \sim \mathcal{D}_{\mathcal{F}}, \mathbf{a} \sim \text{Unif}([q]^k)} \left[ \mathbb{E}_{\mathbf{b}, b_i, \sigma \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{b})] \right] \\ &= \mathbb{E}_{f \sim \mathcal{D}_{\mathcal{F}}} \left[ \mathbb{E}_{\mathbf{a} \sim \mathcal{P}^k} [f(\mathbf{a})] \right] \\ &\leq \rho(\mathcal{F}') \\ &= \rho(\mathcal{F}). \end{aligned}$$

This proves  $\mathcal{D}_N \in S_{\rho(\mathcal{F})}^N(\mathcal{F})$  and thus proves the theorem.  $\square$

We now turn to the proof of [Theorem 1.3](#). Indeed we prove a more detailed statement along the lines of [Theorem 2.3](#) in this case. For this part we use the fact, which extends a similar statement in [\[CGSV21\]](#) that every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [2]^2)$  with matching marginals form a padded one-wise pair.

**Proposition 2.19.** *If  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [2]^2)$  satisfy  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$  then  $(\mathcal{D}_Y, \mathcal{D}_N)$  form a padded one-wise pair.*

*Proof.* We use [\[CGSV21, Proposition 2.12\]](#) which proves this proposition for the special case of  $|\mathcal{F}| = 1$ . For  $g \in \mathcal{F}$ , let  $P(g)$  denote the probability of sampling a constraint  $(f, \mathbf{j}) \sim \mathcal{D}_Y$  with function  $f = g$  and let  $P$  denote this distribution. (Note that since  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$ ,  $\mathcal{D}_N$  also samples  $g$  with the same probability.) Let  $\mathcal{D}_{Y|g}$  denote  $\mathcal{D}_Y$  conditioned on  $f = g$ . Similarly let  $\mathcal{D}_{N|g}$  denote  $\mathcal{D}_N$  conditioned on  $f = g$ . Since  $\mathcal{D}_{Y|g}$  and  $\mathcal{D}_{N|g}$  are distributions from  $\Delta(\{g\} \times [2]^2)$  with matching marginals we have from [\[CGSV21, Proposition 2.12\]](#) that there exist  $\mathcal{D}_{0|g}, \mathcal{D}'_{Y|g}$  and  $\mathcal{D}'_{N|g}$  and  $\tau_g$  such that (1)  $\mathcal{D}_{Y|g} = \tau_g \mathcal{D}_{0|g} + (1 - \tau_g) \mathcal{D}'_{Y|g}$ , (2)  $\mathcal{D}_{N|g} = \tau_g \mathcal{D}_{0|g} + (1 - \tau_g) \mathcal{D}'_{N|g}$  and (3)  $\mathcal{D}'_{Y|g}$  and  $\mathcal{D}'_{N|g}$  are one-wise independent.

Now let  $\tau = \mathbb{E}_{f \sim P}[\tau_f]$ , and  $\mathcal{D}_0 \in \Delta(\mathcal{F} \times [q]^k)$  be the distribution where  $\mathbf{a} = (f, \mathbf{b}) \in \{\mathcal{F}\} \times [2]^2$  is sampled with probability  $\frac{P(f) \cdot \tau_f \cdot \mathcal{D}_{0|f}(\mathbf{a})}{\tau}$ , where  $\mathcal{D}_{0|f}(\mathbf{a})$  is the probability of sampling  $\mathbf{a}$  from  $\mathcal{D}_{0|f}$ . Note that this is a valid probability distribution as

$$\sum_{f \in \mathcal{F}} \sum_{\mathbf{b} \in [2]^2} \frac{P(f) \cdot \tau_f \cdot \mathcal{D}_{0|f}(f, \mathbf{b})}{\tau} = \sum_{f \in \mathcal{F}} \frac{P(f) \cdot \tau_f}{\tau} \cdot \sum_{\mathbf{b} \in [2]^2} \mathcal{D}_{0|f}((f, \mathbf{b})) = 1.$$

Similarly define  $\mathcal{D}'_Y$  and  $\mathcal{D}'_N$  such that  $\mathbf{a}$  is sampled with probability  $\frac{P(f) \cdot (1 - \tau_f) \cdot \mathcal{D}'_{Y|f}(\mathbf{a})}{1 - \tau}$  and  $\frac{P(f) \cdot (1 - \tau_f) \cdot \mathcal{D}'_{N|f}(\mathbf{a})}{1 - \tau}$ , respectively. It can be verified that these choices satisfy (1)  $\mathcal{D}_Y = \tau \mathcal{D}_0 + (1 - \tau) \mathcal{D}'_Y$ , (2)  $\mathcal{D}_N = \tau \mathcal{D}_0 + (1 - \tau) \mathcal{D}'_N$  and (3)  $\mathcal{D}'_Y$  and  $\mathcal{D}'_N$  are one-wise independent. It follows that  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  form a padded one-wise pair.  $\square$

Combining [Proposition 2.19](#) and [Theorem 2.16](#) we immediately get the following theorem, which in turn implies [Theorem 1.3](#).

**Theorem 2.20.** *For every family  $\mathcal{F} \subseteq \{f : [2]^2 \rightarrow \{0, 1\}\}$ , and for every  $0 \leq \beta < \gamma \leq 1$ , the following hold:*

1. *If  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ , then  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) admits a probabilistic dynamic streaming algorithm that uses  $O(\log^3 n)$  space.*
2. *If  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) \neq \emptyset$ , then for every  $\varepsilon > 0$ , then  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) in the insertion-only setting requires  $\Omega(\sqrt{n})$  space<sup>6</sup>. Furthermore, if  $\gamma = 1$ , then  $(1, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) in the insertion-only setting requires  $\Omega(\sqrt{n})$  space.*

*Proof.* Part (1) is simply the specialization of Part (1) of [Theorem 2.3](#) to the case  $k = 2$ . For Part (2), suppose  $\mu \in K_\gamma^Y \cap K_\beta^N$ . Let  $\mathcal{D}_Y \in S_\gamma^Y$  and  $\mathcal{D}_N \in S_\beta^N$  be distributions such that  $\mu(\mathcal{D}_Y) = \mu(\mathcal{D}_N) = \mu$ . Then by [Proposition 2.19](#) we have that  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  form a padded one-wise pair, and so [Theorem 2.16](#) can be applied to get Part (2).  $\square$

## 2.4 Some Examples

We consider three basic examples of general  $q$ -CSP and illustrate how to apply [Theorem 2.16](#) to determine their approximability.

The first example is Max-DICUT described below. Since Max-DICUT does not allow constraints to be applied on negations of variables, the previous work [\[CGSV21\]](#) does not capture it.

### Example 1 (Max-DICUT).

Let  $f(x, y) : [2]^2 \rightarrow \{0, 1\}$  with  $f(x, y) = 1$  if and only if  $x = 2$  and  $y = 1$ . Note that Max-DICUT = Max-CSP( $\{f\}$ ) with  $q = k = 2$ . Observe that for every distribution  $\mathcal{D} \in \Delta([q]^k)$  with probability density vector  $\phi(\mathcal{D}) = (\phi_{22}, \phi_{21}, \phi_{12}, \phi_{11})$ , we have for every  $0 \leq \gamma, \beta \leq 1$

$$S_\gamma^Y(\mathcal{F}) = \{\mathcal{D} \mid \phi_{21} \geq \gamma\}$$

and

$$S_\beta^N(\mathcal{F}) = \left\{ \mathcal{D} \mid \max_{p, q \in [0, 1]} p(1-p) \cdot \phi_{22} + pq \cdot \phi_{21} + (1-q)(1-p) \cdot \phi_{12} + (1-q)q \cdot \phi_{11} \leq \beta \right\}.$$

Also, note that the marginal vector  $\mu(\mathcal{D}) = (\mu_{22}, \mu_{21}, \mu_{12}, \mu_{11})$  and  $\phi(\mathcal{D})$  satisfy the following relations:

$$\begin{cases} \mu_{22} = \phi_{12} + \phi_{22} \\ \mu_{21} = \phi_{11} + \phi_{21} \\ \mu_{12} = \phi_{21} + \phi_{22} \\ \mu_{11} = \phi_{11} + \phi_{12} \end{cases}$$

Note that for every  $\mathcal{D} \in \Delta([q]^k)$ , we have  $\mathcal{D} \in S_{1/4}^N$ . In particular, the uniform distribution  $\text{Unif}([2]^2) \in S_{1/4}^N$ . Since the distribution given by the density vector  $(\phi_{22} = 0, \phi_{21} =$

<sup>6</sup>The constant hidden in the  $\Omega$  notation may depend on  $k$  and  $\varepsilon$ .

$1/2, \phi_{12} = 1/2, \phi_{11} = 0)$  also has uniform marginals and belongs to  $S_{1/2}^Y$ , we have that for every  $\beta \geq 1/4$ ,  $K_{1/2}^Y \cap K_\beta^N(\mathcal{F}) \neq \emptyset$ . So it suffices to focus on the case where  $\gamma \geq 1/2$ .

Fix  $\gamma \geq 1/2$ , we want to compute the minimum  $\beta$  such that  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) \neq \emptyset$ . The kernel of the mapping from probability density  $\phi$  to the marginal vector  $\mu$  is spanned by  $(1, -1, -1, 1)$ . Then simple calculations show that the minimum  $\beta$  is achieved when  $\mu = (1 - \gamma, \gamma, \gamma, 1 - \gamma)$  with  $(0, \gamma, 1 - \gamma, 0) \in S_\gamma^Y(\mathcal{F})$  and  $(1 - \gamma, 2\gamma - 1, 0, 1 - \gamma) \in S_\beta^N(\mathcal{F})$ . Specifically,

$$\begin{aligned} \beta &= \max_{p,q \in [0,1]} (p(1-p) + q(1-q)) \cdot (1-\gamma) + pq \cdot (2\gamma - 1) \\ &= \max_{p,q \in [0,1]} \frac{(1-\gamma)^2}{3-4\gamma} - \frac{3-4\gamma}{2} \cdot \left( \left( p + \frac{1-\gamma}{4\gamma-3} \right)^2 + \left( q + \frac{1-\gamma}{4\gamma-3} \right)^2 \right) - \frac{(2\gamma-1)}{2} \cdot (p-q)^2. \end{aligned}$$

When  $\gamma \geq 3/4$ , the expression is maximized by  $p = q = 1$  and hence  $\beta = 2\gamma - 1$ . When  $1/2 \leq \gamma \leq 3/4$ , the expression is maximized by  $p = q = (1 - \gamma)/(3 - 4\gamma)$  and hence  $\beta = (1 - \gamma)^2/(3 - 4\gamma)$ . Finally, over  $\gamma \in [3/4, 1]$ ,  $\beta/\gamma$  is minimized at  $(\gamma, \beta) = (3/4, 1/2)$  and  $\beta/\gamma = 2/3$ ; over  $\gamma \in [1/2, 3/4]$ ,  $\beta/\gamma$  is minimized at  $(\gamma, \beta) = (3/5, 4/15)$  and  $\beta/\gamma = 4/9$ .

Now [Proposition 2.19](#) gives us that any pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [2]^2)$ ,  $\mathcal{D}_Y \in S_{3/5}^Y, \mathcal{D}_N \in S_{4/15}^N$  witnessing  $K_{3/5}^Y \cap K_{4/15}^N \neq \emptyset$  forms a padded one-wise pair. Finally, [Theorem 2.16](#), applied to the padded one-wise pair  $(\mathcal{D}^Y, \mathcal{D}^N)$ , implies that Max-DICUT cannot be approximated better with a factor  $(4/9 + \varepsilon)$  in space  $o(\sqrt{n})$  in the insertion-only setting, which is consistent with the findings in [\[CGV20\]](#) for the Max-DICUT problem.

Next, we consider Unique Games (UG) to be an example with larger alphabet which the previous work [\[CGSV21\]](#) does not capture.

### Example 2 (Max- $q$ UG).

Let  $k = 2$  and  $q \geq 2$ . Let  $\mathcal{F} = \{f : [q]^2 \rightarrow \{0, 1\} \mid f^{-1}(1) \text{ is a bijection}\}$ . Note that  $\text{Max-}q\text{UG} = \text{Max-CSP}(\mathcal{F})$ . We claim that the quantity  $\alpha = \inf_\beta \alpha(\beta) = 1/q$  where  $\alpha(\beta) = \sup_{\gamma \mid K_\gamma^Y \cap K_\beta^N = \emptyset} \{\beta/\gamma\}$ . First, note that  $\mathcal{D} \in S_{1/q}^N$  for every  $\mathcal{D}$  and hence implies  $\alpha \geq 1/q$ .

For simplicity we work with the alphabet  $\mathbb{Z}_q = \{0, \dots, q-1\}$  instead of  $[q]$ . For  $\tau \in \mathbb{Z}_q$  let  $f_\tau \in \mathcal{F}$  be the constraint  $f_\tau(x, y) = 1$  if and only if  $x - y = \tau \pmod{q}$ . Let  $\mathcal{D}^Y$  be the uniform distribution over  $\{(f_\tau, \sigma, \sigma + \tau) \mid \sigma, \tau \in \mathbb{Z}_q\}$ . Note that obviously we have  $\mathcal{D}^Y \in S_1^Y$ . Now let  $\mathcal{D}^N$  be the uniform distribution over  $\{f_\tau \mid \tau \in \mathbb{Z}_q\} \times \mathbb{Z}_q^2$ . Note that for any assignment to two variables  $x_{1,\sigma_1}, x_{2,\sigma_2}$  the probability over  $\tau$  that it satisfies  $f_\tau(x_{1,\sigma_1}, x_{2,\sigma_2})$  is exactly  $1/q$ . It follows that any assignment to  $(x_{i,\sigma})_{i,\sigma}$  satisfies exactly  $1/q$  fraction of the constraints in  $\mathcal{D}^N$  and so  $\mathcal{D}^N \in S_{1/q}^N$ . Observe that the marginals of  $\mathcal{D}^Y$  and  $\mathcal{D}^N$  are the same, i.e.,  $\mu(\mathcal{D}^Y) = \mu(\mathcal{D}^N) = \mu(\text{Unif}(\{f_\tau\} \times \mathbb{Z}_q^2))$ . This gives us  $\mu(\text{Unif}(\{f_\tau\} \times [q]^2)) \in K_1^Y \cap K_{1/q}^N$  so we have  $\alpha(\beta) = \beta$  for  $\beta \geq 1/q$ . Minimizing this over  $\beta$ , [Theorem 2.16](#), applied to the one-wise independent distribution  $\mathcal{D}^Y$  and  $\mathcal{D}^N$ , gives that the problem can not be approximated better than  $1/q$  in space  $o(\sqrt{n})$  in the insertion-only setting, which is consistent with the findings in [\[GT19\]](#) for the Max- $q$ UG problem.

### Example 3 (Max- $q$ Col)

Let  $k = 2$  and  $q \geq 2$ . Let  $\mathcal{F} = \{f_{\neq}\}$  where  $f_{\neq} : [q]^2 \rightarrow \{0, 1\}$  is given by  $f_{\neq}(x, y) = 1 \Leftrightarrow x \neq y$ . Note that  $\text{Max-}q\text{Col} = \text{Max-CSP}(\mathcal{F})$ . We claim that the quantity  $\alpha = \inf_{\beta} \alpha(\beta) = 1 - 1/q$  where  $\alpha(\beta) = \sup_{\gamma | K_{\gamma}^Y \cap K_{\beta}^N = \emptyset} \{\beta/\gamma\}$ . First, note that  $\mathcal{D} \in S_{1-1/q}^N$  for every  $\mathcal{D}$  and hence implies  $\alpha \geq 1 - 1/q$ . We now show this is also the upper bound by exhibiting  $\mathcal{D}^Y$  and  $\mathcal{D}^N$ . Let  $\mathcal{D}^Y$  be the uniform distribution over  $\{(f_{\neq}, \sigma, \tau) \mid \sigma \neq \tau \in [q]\}$ . Note that obviously we have  $\mathcal{D}^Y \in S_1^Y$ . Now let  $\mathcal{D}^N$  be the uniform distribution over  $\{f_{\neq}\} \times [q]^2$ . This leads to  $\beta = \max_{\mathcal{P}_{\sigma}} \{\mathbb{E}_{(f, a_1, a_2) \sim \mathcal{D}^N} [\mathbb{E}_{x \sim \mathcal{P}_{a_1}, y \sim \mathcal{P}_{a_2}} [f(x, y)]]\}$ . The independence of  $a_1$  and  $a_2$  in  $\mathcal{D}^N$  allows us to simplify this to  $\max_{\mathcal{P} \in \Delta([q])} \{\mathbb{E}_{x, y \sim \mathcal{P}} [f_{\neq}(x, y)]\}$  and the latter is easily seen to be at most  $1 - 1/q$ . Thus we conclude  $\mathcal{D}^N \in S_{1-1/q}^N$ . Since the marginals of  $\mathcal{D}^Y$  and  $\mathcal{D}^N$  are the same, i.e.,  $\mu(\mathcal{D}^Y) = \mu(\mathcal{D}^N) = \mu(\text{Unif}(\{f_{\neq}\} \times [2] \times [q]))$ , this gives us  $\mu(\text{Unif}(\{f_{\neq}\} \times [2] \times [q])) \in K_1^Y \cap K_{1-1/q}^N$  so we have  $\alpha(\beta) = \beta$  for  $\beta \geq 1 - 1/q$ . Minimizing this over  $\beta$ , [Theorem 2.16](#), applied to the one-wise independent distribution  $\mathcal{D}^Y$  and  $\mathcal{D}^N$ , gives that the problem can not be approximated better than  $1 - 1/q$  in space  $o(\sqrt{n})$  in the insertion-only setting.

Another example along the same vein is analyzed in a subsequent work by Singer, Sudan and Velusamy [\[SSV21\]](#) who show that  $(1 - 1/q, (1/2)(1 - 1/q))$ -Max-CSP( $\mathcal{F}$ ) is hard for  $\mathcal{F} = \{f_{<}\}$  where  $f_{<} : [q]^2 \rightarrow \{0, 1\}$  is given by  $f_{<}(x, y) = 1$  if and only if  $x < y$ . This analysis forms a critical step in their improved analysis of the Maximum Acyclic Subgraph Problem (which is not captured in our framework).

## 2.5 Classification of exact computability

Finally for the sake of completeness we show that all “non-trivial” CSPs are hard to solve exactly. “Trivial” families are those where all satisfiable constraints are satisfied by a constant assignment, as defined precisely below.

**Definition 2.21** (Constant satisfiable). *For  $\sigma \in [q]$  and  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  we say that  $\mathcal{F}$  is  $\sigma$ -satisfiable if for every  $f \in \mathcal{F} \setminus \{\mathbf{0}\}$  we have that  $f(\sigma^k) = 1$ . We say  $\mathcal{F}$  is constant-satisfiable if there exists  $\sigma \in [q]$  such that  $\mathcal{F}$  is  $\sigma$ -satisfiable.*

Our theorem below asserts that constant satisfiable families are the only ones that are solvable exactly. And for additive  $\varepsilon$  approximations to the maximum fraction of satisfiable constraints, they require space growing polynomially in  $\varepsilon^{-1}$ .

**Theorem 2.22.** *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  the following hold:*

1. *If  $\mathcal{F}$  is constant satisfiable, then there exists a deterministic dynamic streaming algorithm that uses  $O(\log n)$  space and solves Max-CSP( $\mathcal{F}$ ) exactly optimally.*
2. *If  $\mathcal{F}$  is not constant satisfiable, then the following hold in the insertion-only and dynamic settings:*
  - (a) *Every probabilistic algorithm solving Max-CSP( $\mathcal{F}$ ) exactly requires  $\Omega(n)$  space.*

- (b) For every  $\varepsilon = \varepsilon(n) > 0$ ,  $(1, 1 - \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\min\{n, \varepsilon^{-1}\})$ -space<sup>7</sup> on sufficiently large inputs.
- (c) For  $\rho_{\min}(\mathcal{F})$  defined in Definition 2.11, for every  $\rho_{\min}(\mathcal{F}) < \gamma < 1$  and every  $\varepsilon = \varepsilon(n) > 0$ ,  $(\gamma, \gamma - \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\min\{n, \varepsilon^{-2}\})$ -space<sup>7</sup> on sufficiently large inputs.

Theorem 2.22 is proved in Section 9.

**Organization of the rest of the paper.** In Section 3 we introduce some basic notation and review some probability theory and Fourier analytic basics. In Section 4 we describe our algorithm for approximating Max-CSP( $\mathcal{F}$ ) in the tractable cases, giving the positive part of Theorem 2.3. In Section 5 we introduce our communication problem, state a lower bound (Theorem 5.4) and prove our streaming lower bounds in the insertion-only and dynamic settings (Theorem 2.16 and the negative part of Theorem 2.3) modulo this lower bound. We prove Theorem 5.4 in Sections 6 to 8. In Section 9 we prove Theorem 2.22 giving the characterization of Max-CSP( $\mathcal{F}$ ) problems exactly solvable in logarithmic space.

### 3 Preliminaries

We will follow the convention that  $n$  denotes the number of variables in the CSP as well as the communication game,  $m$  denotes the number of constraints in the CSP, and  $k$  denotes the arity of the CSP. We use  $\mathbb{N}$  to denote the set of natural numbers  $\{1, 2, 3, \dots\}$  and use  $[n]$  to denote the set  $\{1, 2, \dots, n\}$ . By default, a Boolean variable in this paper takes value in  $\{-1, 1\}$ .

For variables of a vector form, we write them in boldface, *e.g.*,  $\mathbf{x} \in [q]^n$ , and its  $i$ -th entry is written without boldface, *e.g.*,  $x_i$ . For variable being a vector of vectors, we write it, for example, as  $\mathbf{b} = (\mathbf{b}(1), \mathbf{b}(2), \dots, \mathbf{b}(m))$  where  $\mathbf{b}(i) \in [q]^k$ . The  $j$ -th entry of the  $i$ -th vector of  $\mathbf{b}$  is then written as  $\mathbf{b}(i)_j$ . Let  $\mathbf{x}$  and  $\mathbf{y}$  be two vectors of the same length,  $\mathbf{x} \odot \mathbf{y}$  denotes the entry-wise product of them.

For every  $p \in [0, 1]$ ,  $\text{Bern}(p)$  denotes the Bernoulli distribution that takes value 1 with probability  $p$  and takes value  $-1$  with probability  $1 - p$ .

#### 3.1 Approximate Constraint Satisfaction

Max-CSP( $\mathcal{F}$ ) is specified by a family of constraints  $\mathcal{F}$ , where each constraint function  $f \in \mathcal{F}$  is such that  $f : [q]^k \rightarrow \{0, 1\}$ , for a fixed positive integer  $k$ . Given  $n$  variables  $x_1, x_2, \dots, x_n$ , an application of the constraint function  $f$  to these variables, which we term simply a *constraint*, is given by a  $k$ -tuple  $\mathbf{j} = (j_1, \dots, j_k) \in [n]^k$  where the  $j_i$ 's are distinct and represent the application of the constraint function  $f$  to the variables  $x_{j_1}, \dots, x_{j_k}$ . Specifically an assignment  $\mathbf{b} \in [q]^n$  satisfies a constraint given by  $(f, \mathbf{j})$  if  $f(b_{j_1}, \dots, b_{j_k}) = 1$ . An instance  $\Psi$  of Max-CSP( $\mathcal{F}$ ) consists of  $m$  constraints  $C_1, \dots, C_m$  with non-negative weights  $w_1, \dots, w_m$  where  $C_i = (f_i, \mathbf{j}(i))$  and  $w_i \in \mathbb{R}$  for each  $i \in [m]$ . For an assignment  $\mathbf{b} \in [q]^n$ , the value  $\text{val}_\Psi(\mathbf{b})$  of  $\mathbf{b}$  on  $\Psi$  is the fraction of weight of constraints satisfied by  $\mathbf{b}$ , *i.e.*,  $\text{val}_\Psi(\mathbf{b}) = \frac{1}{W} \sum_{i \in [m]} w_i \cdot f_i(\mathbf{b}|_{\mathbf{j}(i)})$ , where  $W = \sum_{i=1}^m w_i$ . The optimal value of  $\Psi$  is defined as  $\text{val}_\Psi = \max_{\mathbf{b} \in [q]^n} \{\text{val}_\Psi(\mathbf{b})\}$ . The approximation version of Max-CSP( $\mathcal{F}$ ) is defined as follows.

<sup>7</sup> The constant hidden in the  $\Omega$  depends on  $\mathcal{F}$ , but (obviously) not on  $\varepsilon$ .

We first describe how instances  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  are presented in the streaming setting.

**Remark 3.1.** In the insertion only setting we only consider the unweighted case. (Note this is a setting we only use for our lower bounds, so this makes our results stronger.) Here the input  $\Psi = (C_1, \dots, C_m)$  is presented as a stream  $\sigma_1, \dots, \sigma_m$  with  $\sigma_i = C_i$  where the  $C_i$ 's are distinct. In the dynamic setting, the input  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  is obtained by inserting and deleting (unweighted) constraints, possibly with repetitions and thus leading to a (integer) weighted instance. Formally  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  is presented as a stream  $\sigma_1, \dots, \sigma_\ell$  where  $\sigma_t = (C'_t, w'_t)$  and  $w'_t \in \{-1, 1\}$  such that  $w_i = \sum_{t \in [\ell]: C_i = C'_t} w'_t$ . For the algorithmic results we require that  $w_i$ 's are non-negative at the end of the stream. The lower bounds however work in the “strict turnstile” model where at all times, the weight of every constraint is non-negative, i.e., for all  $\ell' \leq \ell$  and for all constraints  $C$  we have  $\sum_{t \in [\ell']: C = C'_t} w'_t \geq 0$ . Given such representations of our instance, we now describe our streaming task.

**Definition 3.2** ( $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ )). Let  $\mathcal{F}$  be a constraint family and  $0 \leq \beta < \gamma \leq 1$ . For each  $m \in \mathbb{N}$ , let  $\Gamma_m = \{\Psi = (C_1, \dots, C_m; w_1, \dots, w_m) \mid \text{val}_\Psi \geq \gamma\}$  and  $B_m = \{\Psi = (C_1, \dots, C_m; w_1, \dots, w_m) \mid \text{val}_\Psi \leq \beta\}$ .

The task of  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) is to distinguish between instances from  $\Gamma = \cup_{m \leq \text{poly}(n)} \Gamma_m$  and instances from  $B = \cup_{m \leq \text{poly}(n)} B_m$ . Specifically we desire algorithms (in the dynamic/insertion-only setting) that output 1 w.p. at least  $2/3$  on inputs from  $\Gamma$  and output 1 w.p. at most  $1/3$  on inputs from  $B$ .

### 3.2 Total variation distance

The total variation distance between probability distributions plays an important role in our analysis.

**Definition 3.3** (Total variation distance of discrete random variables). Let  $\Omega$  be a finite probability space and  $X, Y$  be random variables with support  $\Omega$ . 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 3.4** (E.g., [KKS15, Claim 6.5]). For random variables  $X, Y$  and  $W$ :

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

### 3.3 Concentration inequality

We will use the following concentration inequality which is essentially an Azuma-Hoeffding style inequality for submartingales. The form we use is based on [KK19, Lemma 2.5], and allows for variables with different expectations. The analysis is a very slight modification of theirs.

**Lemma 3.5.** Let  $X = \sum_{i \in [N]} X_i$  where  $X_i$  are Bernoulli random variables such that for every  $k \in [N]$ ,  $\mathbb{E}[X_k | X_1, \dots, X_{k-1}] \leq p_k$  for some  $p_k \in (0, 1)$ . Let  $\mu = \sum_{k=1}^N p_k$ . For every  $\Delta > 0$ , we have:

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

*Proof.* Let  $v = \Delta/(\mu + \Delta)$  and  $u = \ln(1 + v)$ . We have

$$\mathbb{E}[e^{uX}] = \mathbb{E}\left[\prod_{k=1}^N e^{uX_k}\right] \leq (1 + p_N(e^u - 1)) \cdot \mathbb{E}\left[\prod_{k=1}^{N-1} e^{uX_k}\right] \leq \prod_{i=1}^N (1 + p_k(e^u - 1)) = \prod_{i=1}^N (1 + p_k v) \leq e^{v\mu},$$

where the final inequality uses  $1 + x \leq e^x$  for every  $x$  (and the definition of  $\mu$ ). Applying Markov's inequality to the above, we have:

$$\Pr[X \geq \mu + \Delta] = \Pr\left[e^{uX} \geq e^{u(\mu + \Delta)}\right] \leq \mathbb{E}[e^{uX}] / e^{u(\mu + \Delta)} \leq e^{v\mu - u\mu - u\Delta}.$$

From the inequality  $e^{v-v^2/2} \leq 1 + v$  we infer  $u \geq v - v^2/2$  and so the final expression above can be bounded as:

$$\Pr[X \geq \mu + \Delta] \leq e^{v\mu - u\mu - u\Delta} \leq e^{\frac{v^2}{2}(\mu + \Delta) - v\Delta} = e^{-\frac{\Delta^2}{2(\mu + \Delta)}},$$

where the final equality comes from our choice of  $v$ .  $\square$

### 3.4 Fourier analysis

We will need the following basic notions from Fourier analysis over the Boolean hypercube (see, for instance, [O'D14]). For a Boolean function  $f : \{-1, 1\}^k \rightarrow \mathbb{R}$  its Fourier coefficients are defined by  $\hat{f}(\mathbf{v}) = \mathbb{E}_{\mathbf{a} \in \{-1, 1\}^k} [f(\mathbf{a}) \cdot (-1)^{\mathbf{v}^\top \mathbf{a}}]$ , where  $\mathbf{v} \in \{0, 1\}^k$ . We need the following two important tools.

**Lemma 3.6** (Parseval's identity). For every function  $f : \{-1, 1\}^k \rightarrow \mathbb{R}$ ,

$$\|f\|_2^2 = \frac{1}{2^k} \sum_{\mathbf{a} \in \{-1, 1\}^k} f(\mathbf{a})^2 = \sum_{\mathbf{v} \in \{0, 1\}^k} \hat{f}(\mathbf{v})^2.$$

Note that for every distribution  $f$  on  $\{-1, 1\}^k$ ,  $\hat{f}(0^k) = 2^{-k}$ . For the uniform distribution  $U$  on  $\{-1, 1\}^k$ ,  $\hat{U}(\mathbf{v}) = 0$  for every  $\mathbf{v} \neq 0^k$ . Thus, by Lemma 3.6, for any distribution  $f$  on  $\{-1, 1\}^k$ :

$$\|f - U\|_2^2 = \sum_{\mathbf{v} \in \{0, 1\}^k} \left(\hat{f}(\mathbf{v}) - \hat{U}(\mathbf{v})\right)^2 = \sum_{\mathbf{v} \in \{0, 1\}^k \setminus \{0^k\}} \hat{f}(\mathbf{v})^2. \quad (3.7)$$

Next, we will use the following consequence of hypercontractivity for Boolean functions as given in [GKK<sup>+</sup>09, Lemma 6] which in turns relies on a lemma from [KKL88].

**Lemma 3.8.** Let  $f : \{-1, 1\}^n \rightarrow \{-1, 0, 1\}$  and  $A = \{\mathbf{a} \in \{-1, 1\}^n \mid f(\mathbf{a}) \neq 0\}$ . If  $|A| \geq 2^{n-c}$  for some  $c \in \mathbb{N}$ , then for every  $\ell \in \{1, \dots, 4c\}$ , we have

$$\frac{2^{2n}}{|A|^2} \sum_{\substack{\mathbf{v} \in \{0, 1\}^n \\ \|\mathbf{v}\|_1 = \ell}} \hat{f}(\mathbf{v})^2 \leq \left(\frac{4\sqrt{2}c}{\ell}\right)^\ell.$$



## 4 A Streaming Approximation Algorithm for Max-CSP( $\mathcal{F}$ )

In this section we give our main algorithmic result — a  $O(\log n)$ -space streaming algorithm in the dynamic setting for the  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) if  $K_\gamma^Y = K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N = K_\beta^N(\mathcal{F})$  are disjoint. (See [Definition 2.2.](#))

We state our main theorem of this section which simply repeats Part (1) of [Theorem 2.3.](#)

**Theorem 4.1.** *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  and for every  $0 \leq \beta < \gamma \leq 1$  if  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ , then  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ) in the dynamic setting admits a probabilistic streaming algorithm that uses  $O(\log^3 n)$  space.*

We start with a brief overview of our algorithm. Roughly, given an instance  $\Psi$  on  $n$  variables with  $m$  constraints, our streaming algorithm (implicitly) works with an  $n \times q$  bias non-negative matrix  $\mathbf{bias}$  whose  $(i, \sigma)$ th entry tries to capture how much the  $i$ th variable would like to be assigned the value  $\sigma$  (according to our approximation heuristic). Note that any such matrix is too large for our algorithm, so the algorithm does not explicitly maintain this matrix. Our heuristic ensures that  $\mathbf{bias}$  is updated linearly by every constraint and so the rich theory of norm-approximations of matrices under linear updates can be brought into play to compute any desired norm of this matrix. Given the intuition that  $\mathbf{bias}_{i,\sigma}$  represents the preference of variable  $i$  for value  $\sigma$ , a natural norm of interest to us is  $\|\mathbf{bias}\|_{1,\infty} \triangleq \sum_{i=1}^n \{\max_{\sigma \in [q]} \{\mathbf{bias}_{i,\sigma}\}\}$ . This norm, fortunately for us, is well-known to be computable using  $O(q \log^3 n)$  bits of space [[AKO11](#)] (assuming  $\mathbf{bias}$  is updated linearly) and we use this algorithm as a black box.

The question then turns to asking how  $\mathbf{bias}$  should be defined. On input a stream  $\sigma_1, \dots, \sigma_\ell$  representing an instance  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  (as in [Remark 3.1](#)) with  $\sigma_i = (C'_i = (\mathbf{j}(i), \mathbf{b}(i)), w'_i)$ , how should  $\mathbf{bias}$  be updated? Presumably the  $i$ -th update will only involve the rows  $\mathbf{j}(i)_1, \dots, \mathbf{j}(i)_k$  but how should these be updated and how should this update depend on the function  $f_i$ ? Here is where the disjointness of  $K^Y$  and  $K^N$  comes into play. (We suppress  $\mathcal{F}$  and  $\gamma$  and  $\beta$  in the notation of the sets  $S_\gamma^Y, S_\beta^N$  and  $K_\gamma^Y$  and  $K_\beta^N$  in this overview.) We show that these sets are convex and closed, and so there is a hyperplane (with margin) separating the two sets. Let  $\boldsymbol{\lambda} = (\lambda_{f,i,\sigma})_{f \in \mathcal{F}, i \in [k], \sigma \in [q]}$  be the coefficients of this separating hyperplane and let  $\tau_N < \tau_Y$  be thresholds such that  $\langle \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle \geq \tau_Y$  for  $\boldsymbol{\mu} \in K^Y$  and  $\langle \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle \leq \tau_N$  for  $\boldsymbol{\mu} \in K^N$ . It turns out that the coefficients of  $\boldsymbol{\lambda}$  give us exactly the right information to determine the update to the bias vector: Specifically given an element  $\sigma_i$  of the stream with constraint  $C'_i = (f_i, \mathbf{j}(i))$  and weight  $w'_i$  and  $\ell \in [k]$  and  $\sigma \in [q]$ , we add  $\lambda_{f_i,\ell,\sigma} \cdot w'_i$  to  $\mathbf{bias}_{\mathbf{j}(i)_\ell,\sigma}$ . We are unable to provide intuition for why these updates work but the proof that the algorithm works is nevertheless quite short!

We now turn to describing our algorithm. We note that from this point on, our writeup closely follows [[CGSV21](#)] with main change being in [Definition 4.3](#), [Theorem 4.4](#), and [Lemma 4.9](#). Recall by [Lemma 2.5](#) that the set  $S^Y, S^N, K^Y, K^N$  are all convex and closed. This implies the existence of a separating hyperplane when  $K^Y$  and  $K^N$  do not intersect. We use a mild additional property to conclude that the coefficients of this hyperplane are non-negative, and we later use this crucially in the computation of the  $\mathbf{bias}$  of the instance.

**Proposition 4.2.** *Let  $\beta, \gamma$  and  $\mathcal{F}$  be such that  $0 \leq \beta < \gamma \leq 1$  and  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) = \emptyset$ . Then there exists a non-negative vector  $\boldsymbol{\lambda} = (\lambda_{f,i,\sigma})_{f \in \mathcal{F}, i \in [k], \sigma \in [q]}$  and real numbers  $\tau_Y > \tau_N$  such that*

$$\forall \boldsymbol{\mu} \in K_\gamma^Y(\mathcal{F}), \quad \langle \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle \geq \tau_Y \quad \text{and} \quad \forall \boldsymbol{\mu} \in K_\beta^N(\mathcal{F}), \quad \langle \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle \leq \tau_N.$$

*Proof.* The existence of a separating hyperplane follows from standard convexity (see, e.g., [BV04, Exercise 2.22]). For us this implies there exists  $\lambda' \in \mathbb{R}^{|\mathcal{F}| \times kq}$  and  $\tau'_Y < \tau'_N$  such that

$$\forall \mu \in K_\gamma^Y(\mathcal{F}), \quad \langle \lambda', \mu \rangle \geq \tau'_Y \quad \text{and} \quad \forall \mu \in K_\beta^N(\mathcal{F}), \quad \langle \lambda', \mu \rangle \leq \tau'_N.$$

But  $\lambda'$  is not necessarily a positive vector. To remedy this we use the fact that  $K_\gamma^Y(\mathcal{F}) \cup K_\beta^N(\mathcal{F})$  is contained in a hyperplane whose coefficients are themselves positive. In particular we note that for every  $\mathcal{D} \in \Delta(\mathcal{F} \times [q]^k)$  we have  $\langle \mu(\mathcal{D}), \mathbf{1} \rangle = k$  where  $\mathbf{1} \in \mathbb{R}^{|\mathcal{F}| \times kq}$  is the all ones vector, as verified below:

$$\langle \mu(\mathcal{D}), \mathbf{1} \rangle = \sum_{f \in \mathcal{F}, i \in [k], \sigma \in [q]} \mu_{f,i,\sigma} = \sum_{i \in [k]} \left( \sum_{f \in \mathcal{F}, \sigma \in [q]} \mu_{f,i,\sigma} \right) = \sum_{i \in [k]} 1 = k.$$

Let  $\lambda'_{\min} = \min_{f,t,\sigma} \lambda'_{f,t,\sigma}$ . Now let  $\lambda$ ,  $\tau_Y$  and  $\tau_N$  be given by:

$$\lambda_{f,t,\sigma} = \lambda'_{f,t,\sigma} + |\lambda'_{\min}|, \quad \tau_Y = \tau'_Y + k \cdot |\lambda'_{\min}| \quad \text{and} \quad \tau_N = \tau'_N + k \cdot |\lambda'_{\min}|.$$

Observe that  $\lambda$  is a non-negative vector and  $\tau_Y > \tau_N$ . We also have:

$$\forall \mu \in K_\gamma^Y(\mathcal{F}), \quad \langle \lambda, \mu \rangle = \langle \lambda', \mu \rangle + |\lambda'_{\min}| \geq \langle \mathbf{1}, \mu \rangle \geq \tau'_Y + k|\lambda'_{\min}| = \tau_Y$$

as desired. Similarly also get  $\forall \mu \in K_\beta^N(\mathcal{F}), \quad \langle \lambda, \mu \rangle \leq \tau_N$ , concluding the proof.  $\square$

To use the vector  $\lambda$  given by [Proposition 4.2](#) we introduce the notion of the bias matrix and the bias of a Max-CSP( $\mathcal{F}$ ) instance  $\Psi$ .

**Definition 4.3** (Bias (matrix)). *For a non-negative vector  $\lambda = (\lambda_{f,i,\sigma})_{f \in \mathcal{F}, i \in [k], \sigma \in [q]} \in \mathbb{R}^{|\mathcal{F}|kq}$ , and instance  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  of Max-CSP( $\mathcal{F}$ ) where  $C_i = (f_i, \mathbf{j}(i))$ , where  $f_i \in \mathcal{F}$  and  $\mathbf{j}(i) \in [n]^k$ , we let the  $\lambda$ -bias matrix of  $\Psi$ , denoted  $\text{bias}_\lambda(\Psi)$ , be the matrix in  $\mathbb{R}^{n \times q}$  given by*

$$\text{bias}_\lambda(\Psi)_{\ell,\sigma} = \frac{1}{W} \cdot \sum_{i \in [m], t \in [k]; \mathbf{j}(i)_t = \ell} \lambda_{f_i,t,\sigma} \cdot w_i,$$

for  $\ell \in [n]$  and  $\sigma \in [q]$ , where  $W = \sum_{i \in [m]} w_i$ . The  $\lambda$ -bias of  $\Psi$ , denoted  $B_\lambda(\Psi)$ , is defined as  $B_\lambda(\Psi) = \sum_{\ell=1}^n \max_{\sigma \in [q]} \text{bias}_\lambda(\Psi)_{\ell,\sigma}$ .

Key to our algorithm for approximating Max-CSP( $\mathcal{F}$ ) is the following algorithm to compute the  $\ell_{1,\infty}$  norm of a matrix. Recall that for a matrix  $M \in \mathbb{R}^{a \times b}$  the  $\ell_{1,\infty}$  norm is the quantity  $\|M\|_{1,\infty} = \sum_{i \in [a]} \{\max_{j \in [b]} \{|M_{ij}|\}\}$ .

**Theorem 4.4** (Implied by [AKO11, Theorem 4.5]). *There exists a constant  $c$  such that the  $\ell_{1,\infty}$  norm of an  $n \times q$  matrix  $M$  can be estimated to within a multiplicative error of  $(1 + \varepsilon)$  in the turnstile streaming model with  $O(\varepsilon^{-c} \cdot q \cdot \log^2 n)$  words (or with  $O(\varepsilon^{-c} \cdot q \cdot \log^3 n)$  bits).*

We note that Theorem 4.5 in [AKO11] is much more general. [Theorem 4.4](#) is the special case corresponding to  $X = \ell_\infty$  and  $E_X$  being simply the identity function.  $\alpha(\dots)$  in this case turns out to be  $O(\log n)$  leading to the bounds above [And20].

Note that there is a slight distinction between the definitions of  $B_\lambda(\Psi)$  and  $\|\text{bias}_\lambda(\Psi)\|_{1,\infty}$ , but these quantities are equal since  $\text{bias}_\lambda$  is a non-negative matrix (which in turn follows from the fact that  $\lambda$  is non-negative). We thus get the following corollary.

**Corollary 4.5.** *There exists a constant  $c$  such that for every  $k, q, \mathcal{F}$  and  $\varepsilon > 0$ , there exists a streaming algorithm running in space  $O(\varepsilon^{-c} \cdot \log^3 n)$  that on input a stream  $\sigma_1, \dots, \sigma_\ell$  representing a Max-CSP( $\mathcal{F}$ ) instance  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  on  $n$  variables, outputs a  $(1 \pm \varepsilon)$  approximation to  $B_\lambda(\Psi)$ .*

We are now ready to describe our algorithm for  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ).

---

**Algorithm 1** A streaming algorithm for  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ )

---

**Input:** A stream  $\sigma_1, \dots, \sigma_\ell$  representing an instance  $\Psi$  of Max-CSP( $\mathcal{F}$ ).

- 1: Let  $\lambda \in \mathbb{R}^{|\mathcal{F}|kq}$ ,  $\tau_N$  and  $\tau_Y$  be as given by [Proposition 4.2](#) separating  $K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N(\mathcal{F})$ , so  $\lambda$  is non-negative and  $\tau_N < \tau_Y$ .
- 2: Let  $\varepsilon = \frac{\tau_Y - \tau_N}{2(\tau_Y + \tau_N)}$ .
- 3: Using [Corollary 4.5](#) compute a  $(1 \pm \varepsilon)$  approximation  $\tilde{B}$  to  $B_\lambda(\Psi)$ , i.e.,

$$(1 - \varepsilon)B_\lambda(\Psi) \leq \tilde{B} \leq (1 + \varepsilon)B_\lambda(\Psi) \text{ with probability at least } 2/3.$$

- 4: **if**  $\tilde{B} \leq \tau_N(1 + \varepsilon)$  **then**  
**Output:** NO.
  - 5: **else**  
**Output:** YES.
- 

Given [Corollary 4.5](#) it follows that the algorithm above uses space  $O(\log^3 n)$  on instances on  $n$  variables. In what follows we prove that the algorithm correctly solves  $(\gamma, \beta)$  - Max-CSP( $\mathcal{F}$ ).

#### 4.1 Analysis of the correctness of [Algorithm 1](#)

**Lemma 4.6.** *[Algorithm 1](#) correctly solves  $(\gamma, \beta)$ -Max-CSP( $\mathcal{F}$ ), if  $K_\gamma^Y(\mathcal{F})$  and  $K_\beta^N(\mathcal{F})$  are disjoint. Specifically, for every  $\Psi$ , let  $\tau_Y, \tau_N, \varepsilon, \lambda, \tilde{B}$  be as given in [Algorithm 1](#), we have:*

$$\begin{aligned} \text{val}_\Psi \geq \gamma &\Rightarrow B_\lambda(\Psi) \geq \tau_Y \text{ and } \tilde{B} > \tau_N(1 + \varepsilon), \\ \text{and } \text{val}_\Psi \leq \beta &\Rightarrow B_\lambda(\Psi) \leq \tau_N \text{ and } \tilde{B} \leq \tau_N(1 + \varepsilon), \end{aligned}$$

provided  $(1 - \varepsilon)B_\lambda(\Psi) \leq \tilde{B} \leq (1 + \varepsilon)B_\lambda(\Psi)$ .

In the rest of this section, we will prove [Lemma 4.6](#). The key to our analysis is a distribution  $\mathcal{D}(\Psi^{\mathbf{b}}) \in \Delta(\mathcal{F} \times [q]^k)$  that we associate with every instance  $\Psi$  and assignment  $\mathbf{b} \in [q]^n$  to the variables of  $\Psi$ . If  $\Psi$  is  $\gamma$ -satisfied by assignment  $\mathbf{b}$ , we prove that  $\mu(\mathcal{D}(\Psi^{\mathbf{b}})) \in K_\gamma^Y(\mathcal{F})$ . On the other hand, if  $\Psi$  is not  $\beta$ -satisfiable by any assignment, we prove that for every  $\mathbf{b}$ ,  $\mu(\mathcal{D}(\Psi^{\mathbf{b}})) \in K_\beta^N(\mathcal{F})$ . Finally we also show that the bias  $B_\lambda(\Psi)$  relates to  $\lambda(\mathcal{D}(\Psi^{\mathbf{b}})) \triangleq \langle \mu(\mathcal{D}(\Psi^{\mathbf{b}})), \lambda \rangle$ , where the latter quantity is exactly what needs to be computed (by [Proposition 4.2](#)) to distinguish the membership of  $\mu(\mathcal{D}(\Psi^{\mathbf{b}}))$  in  $K_\gamma^Y(\mathcal{F})$  versus the membership in  $K_\beta^N(\mathcal{F})$ .

We note that the analysis follows the same route as in [\[CGSV21\]](#). The key step is the definition of these distributions that allows the remaining steps (esp. [Lemma 4.9](#)) to be extended, which we present now.

Given an instance  $\Psi = (C_1, \dots, C_m; w_1, \dots, w_m)$  on  $n$  variables with  $C_i = (f_i, \mathbf{j}(i))$  and an assignment  $\mathbf{b} \in [q]^n$ , the distribution  $\mathcal{D}(\Psi^{\mathbf{b}}) \in \Delta(\mathcal{F} \times [q]^k)$  is sampled as follows: Sample  $i \in [m]$  with probability  $w_i/W$  where  $W = \sum_{i \in [m]} w_i$ , and output  $(f_i, \mathbf{b}|_{\mathbf{j}(i)})$ .

We start by relating the bias  $B_{\lambda(\Psi)}$  to  $\mathcal{D}(\Psi)$ .

**Lemma 4.7.** *For every vector  $\mathbf{b} \in [q]^n$ , we have  $\lambda(\mathcal{D}(\Psi^{\mathbf{b}})) = \sum_{\ell=1}^n \text{bias}_{\lambda}(\Psi)_{\ell, b_{\ell}}$ . Consequently we have  $B_{\lambda}(\Psi) = \sum_{\ell=1}^n \max_{\sigma \in [q]} \text{bias}_{\lambda}(\Psi)_{\ell, \sigma} = \max_{\mathbf{b} \in [q]^n} \{\lambda(\mathcal{D}(\Psi^{\mathbf{b}}))\}$ .*

*Proof.* We start with the first equality. Fix  $b \in [q]^n$ . Given  $f \in \mathcal{F}$ ,  $t \in [k]$ , and  $\sigma \in [q]$ , we have  $\mu(\mathcal{D}(\Psi^{\mathbf{b}}))_{f,t,\sigma} = \frac{1}{W} \sum_{i=1}^m w_i \cdot \mathbb{1}[f_i = f, b_{j(i)_t} = \sigma]$ . Hence,

$$\begin{aligned} \lambda(\mathcal{D}(\Psi^{\mathbf{b}})) &= \sum_{f \in \mathcal{F}, t \in [k], \sigma \in [q]} \mu(\mathcal{D}(\Psi^{\mathbf{b}}))_{f,t,\sigma} \cdot \lambda_{f,t,\sigma} \\ &= \frac{1}{W} \sum_{f \in \mathcal{F}, t \in [k], \sigma \in [q]} \sum_{i \in [m]} w_i \cdot \mathbb{1}[f_i = f, b_{j(i)_t} = \sigma] \cdot \lambda_{f,t,\sigma} \\ &= \frac{1}{W} \sum_{i \in [m], t \in [k], \sigma \in [q]: b_{j(i)_t} = \sigma} w_i \cdot \lambda_{f_i, t, \sigma} \\ &= \sum_{\ell=1}^n \frac{1}{W} \sum_{i \in [m], t \in [k]: j(i)_t = \ell} w_i \cdot \lambda_{f_i, t, b_{\ell}} \\ &= \sum_{\ell=1}^n \text{bias}_{\lambda}(\Psi)_{\ell, b_{\ell}}. \end{aligned}$$

For the final equality, observe that

$$B_{\lambda}(\Psi) = \sum_{\ell=1}^n \max_{\sigma \in [q]} \text{bias}_{\lambda}(\Psi)_{\ell, \sigma} = \max_{\mathbf{b} \in [q]^n} \sum_{\ell=1}^n \text{bias}_{\lambda}(\Psi)_{\ell, b_{\ell}} = \max_{\mathbf{b} \in [q]^n} \{\lambda(\mathcal{D}(\Psi^{\mathbf{b}}))\}.$$

□

The following lemmas relate  $\text{val}_{\Psi}$  to the properties of  $\mathcal{D}(\Psi^{\mathbf{a}})$ .

**Lemma 4.8.** *For every  $\Psi \in \text{Max-CSP}(\mathcal{F})$  and  $\mathbf{b} \in [q]^n$ , if  $\text{val}_{\Psi}(\mathbf{b}) \geq \gamma$  then  $\mathcal{D}(\Psi^{\mathbf{b}}) \in S_{\gamma}^Y(\mathcal{F})$ .*

*Proof.* Follows from the fact that

$$\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}(\Psi^{\mathbf{b}})} [C(f, \mathbf{a})(\mathbb{I})] = \frac{1}{W} \sum_{i=1}^m w_i \cdot f_i(b_{j(i)}) = \text{val}_{\Psi}(\mathbf{b}) \geq \gamma,$$

implying  $\mathcal{D}(\Psi^{\mathbf{b}}) \in S_{\gamma}^Y(\mathcal{F})$ . □

**Lemma 4.9.** *For every  $\Psi \in \text{Max-CSP}(\mathcal{F})$ , if  $\text{val}_{\Psi} \leq \beta$ , then for all  $\mathbf{b} \in [q]^n$ , we have  $\mathcal{D}(\Psi^{\mathbf{b}}) \in S_{\beta}^N(\mathcal{F})$ .*

*Proof.* We prove the contrapositive. We assume that  $\exists \mathbf{b} \in [q]^n$  such that  $\mathcal{D}(\Psi^{\mathbf{b}}) \notin S_{\beta}^N(\mathcal{F})$  and show this implies  $\text{val}_{\Psi} > \beta$ . Then there exists  $(P_{\sigma} \in \Delta([q]))_{\sigma \in [q]}$  such that  $\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}(\Psi^{\mathbf{b}})} [\mathbb{E}_{\mathbf{c}, c_i, \sigma \sim P_{\sigma}} [C(f, \mathbf{a})(\mathbf{c})]] > \beta$ .

We thus have

$$\begin{aligned}
\beta &< \mathbb{E}_{(f,a) \sim \mathcal{D}(\Psi^{\mathbf{b}})} \left[ \mathbb{E}_{\mathbf{c}, c_{i,\sigma} \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{c})] \right] \\
&= \mathbb{E}_{\mathbf{c}, c_{i,\sigma} \sim \mathcal{P}_\sigma} \left[ \mathbb{E}_{(f,a) \sim \mathcal{D}(\Psi^{\mathbf{b}})} [\mathcal{C}(f, \mathbf{a})(\mathbf{c})] \right] \\
&= \mathbb{E}_{\mathbf{c}, c_{i,\sigma} \sim \mathcal{P}_\sigma} \left[ \frac{1}{W} \sum_{i=1}^m w_i \cdot f_i((c_{t,b_{j(i)t}})_{t \in [k]}) \right] \\
&= \frac{1}{W} \sum_{i=1}^m w_i \cdot \mathbb{E}_{\mathbf{c}, c_{i,\sigma} \sim \mathcal{P}_\sigma} [f_i((c_{t,b_{j(i)t}})_{t \in [k]})] \\
&= \frac{1}{W} \sum_{i=1}^m w_i \cdot \mathbb{E}_{\mathbf{x}, x_\ell \sim \mathcal{P}_{b_\ell}} [f_i((x_{j(i)t})_{t \in [k]})] \\
&= \mathbb{E}_{\mathbf{x}, x_\ell \sim \mathcal{P}_{b_\ell}} \left[ \frac{1}{W} \sum_{i=1}^m w_i \cdot f_i((x_{j(i)t})_{t \in [k]}) \right] \\
&= \mathbb{E}_{\mathbf{x}, x_\ell \sim \mathcal{P}_{b_\ell}} [\text{val}_\Psi(\mathbf{x})] \\
&\leq \max_{\mathbf{x} \in [q]^n} \text{val}_\Psi(\mathbf{x}) \\
&= \text{val}_\Psi
\end{aligned}$$

which contradicts the assumption that  $\text{val}_\Psi \leq \beta$ . This concludes the proof of the claim and hence the lemma.  $\square$

As in [CGSV21], the key step among the equalities above is the one asserting  $\frac{1}{W} \sum_{i=1}^m w_i \cdot \mathbb{E}_{\mathbf{c}, c_{i,\sigma} \sim \mathcal{P}_\sigma} [f_i((c_{t,b_{j(i)t}})_{t \in [k]})] = \frac{1}{W} \sum_{i=1}^m w_i \cdot \mathbb{E}_{\mathbf{x}, x_\ell \sim \mathcal{P}_{b_\ell}} [f_i((x_{j(i)t})_{t \in [k]})]$  which relies crucially on column symmetry of the distributions used in the definition of  $S_\beta^N(\mathcal{F})$  in Definition 2.1. Without this restriction, or even more stringent ones, this step of the rounding would fail. And the reason we can't use a more stringent restriction will become clear in the proof of Theorem 2.16 (and is specifically used in the proof of Lemma 5.8). We also note that this key equality relies on the assumption that the variables in a single constraint are *distinct*. In particular the left hand side assumes  $c_{i,\sigma}$ s are drawn independently whereas the right side allows this only for the distinct variables  $x_\ell$  in a constraint.

## 5 Streaming Space Lower Bounds for Max-CSP( $\mathcal{F}$ )

In this section, we prove our two lower bound results, modulo a communication complexity lower bound which is proved in Sections 6 to 8. We start by restating the results to be proved. Recall (from Definition 2.15) the notion of a padded one-wise pair of distributions:  $(\mathcal{D}_1, \mathcal{D}_2)$  is a padded one-wise pair if there exist  $\mathcal{D}_0, \mathcal{D}'_1, \mathcal{D}'_2$  and  $\tau \in [0, 1]$  such that for every  $i \in \{1, 2\}$ ,  $\mathcal{D}'_i$  is one-wise independent, and  $\mathcal{D}_i = \tau \mathcal{D}_0 + (1 - \tau) \mathcal{D}'_i$ .

The first theorem we prove is the lower bound in the insertion-only setting for padded one-wise pairs of distributions. We restate the theorem below for convenience.

**Theorem 2.16** (Streaming lower bound in the insertion-only setting). *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  and for every  $0 \leq \beta < \gamma \leq 1$ , if there exists a padded one-wise pair of distributions  $\mathcal{D}_Y \in S_\gamma^Y(\mathcal{F})$  and  $\mathcal{D}_N \in S_\beta^N(\mathcal{F})$  then, for every  $\varepsilon > 0$ , the  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space in the insertion-only setting. Furthermore, if  $\gamma = 1$ , then  $(1, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space in the insertion-only setting.*

We also restate the lower bound in the dynamic setting from [Theorem 2.3](#) as a separate theorem below.

**Theorem 5.1** (Streaming lower bound in the dynamic setting). *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  and for every  $0 \leq \beta < \gamma \leq 1$ , if  $K_\gamma^Y(\mathcal{F}) \cap K_\beta^N(\mathcal{F}) \neq \emptyset$ , then for every  $\varepsilon > 0$ , the  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) problem requires  $\Omega(\sqrt{n})$  space in the dynamic setting. Furthermore, if  $\gamma = 1$ , then  $(1, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space in the dynamic setting.*

To prove both theorems, we introduce a new communication game we call the *Signal Detection (SD)* in [Section 5.1](#). In [Theorem 5.4](#) we state a lower bound on the communication complexity of this problem. This lower bound is established in [Sections 6 to 8](#). We then use this lower bound to prove [Theorem 2.16](#) in [Section 5.2](#) and to prove [Theorem 5.1](#) in [Section 5.3](#).

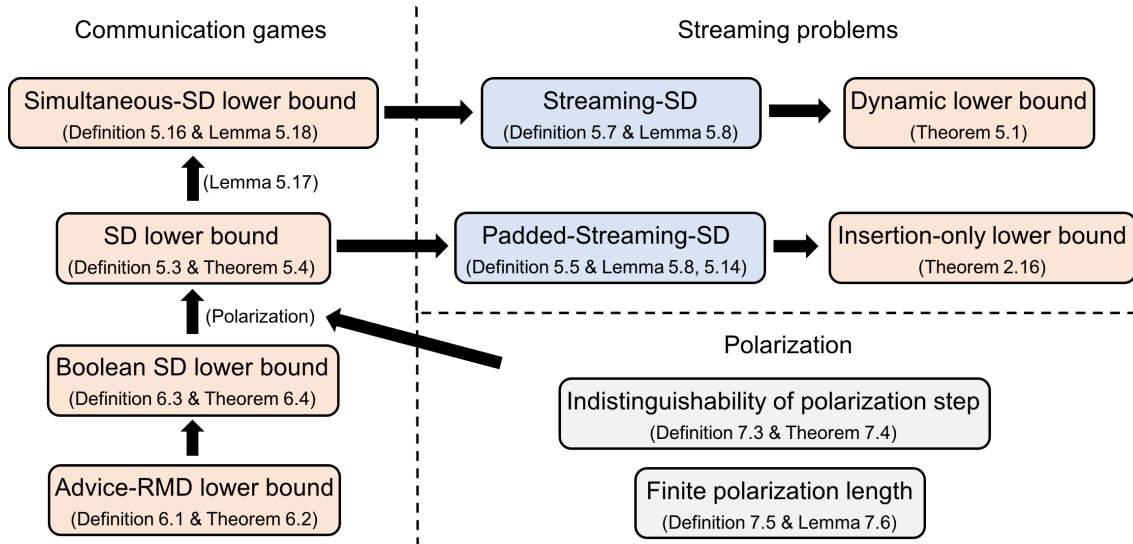


Figure 1: The roadmap of our lower bounds. The top two rows describe the results of this section, while the remaining rows describe notions and results from [Sections 6 to 8](#).

## 5.1 The Signal Detection Problem and Results

In this section we introduce our communication game and state the lower bound for this game. We start with the definition of a general one-way communication game.

**Definition 5.2** (One-way communication game). *Given two distributions  $\mathcal{Y}$  and  $\mathcal{N}$ , an instance of the two-player one-way communication game is a pair  $(X, Y)$  either drawn from  $\mathcal{Y}$  or from  $\mathcal{N}$ . Two computationally unbounded parties, Alice and Bob, receive  $X$  and  $Y$ , respectively. A protocol  $\Pi = (\Pi_A, \Pi_B)$  is a pair of functions with  $\Pi_A(X) \in \{0, 1\}^c$  denoting Alice's message to Bob, and*

$\Pi_B(\Pi_A(X), Y) \in \{\mathbf{YES}, \mathbf{NO}\}$  denoting the protocol's output. We denote this output by  $\Pi(X, Y)$ . The complexity of this protocol is the parameter  $c$  specifying the maximum length of Alice's message  $\Pi_A(X)$ . The advantage of the protocol  $\Pi$  is the quantity

$$\left| \Pr_{(X,Y) \sim \mathcal{Y}}[\Pi(X, Y) = \mathbf{YES}] - \Pr_{(X,Y) \sim \mathcal{N}}[\Pi(X, Y) = \mathbf{YES}] \right|.$$

We now define the specific game we are interested in.

**Definition 5.3** (Signal Detection (SD) Problem). *Let  $n, k, q \in \mathbb{N}, \alpha \in (0, 1)$ , where  $k, q$  and  $\alpha$  are constants with respect to  $n$ , and  $\alpha n$  is an integer less than  $n/k$ . Let  $\mathcal{F}$  be a finite set. For a pair  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  of distributions over  $\mathcal{F} \times [q]^k$ , we consider the following two-player one-way communication problem  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD.*

- The generator samples the following objects:

1.  $\mathbf{x}^* \sim \text{Unif}([q]^n)$ .
2.  $M \in \{0, 1\}^{k\alpha n \times n}$  is chosen uniformly among all matrices with exactly one 1 in each row and at most one 1 in each column. We let  $M = (M_1, \dots, M_{\alpha n})$  where  $M_i \in \{0, 1\}^{k \times n}$  is the  $i$ -th block of rows of  $M$ , where each block has exactly  $k$  rows.
3.  $\mathbf{b} = (\mathbf{b}(1), \dots, \mathbf{b}(\alpha n))$  is sampled from one of the following distributions:
  - (**YES**) each  $\mathbf{b}(i) = (f_i, \tilde{\mathbf{b}}(i)) \in \mathcal{F} \times [q]^k$  is sampled according to  $\mathcal{D}_Y$ .
  - (**NO**) each  $\mathbf{b}(i) = (f_i, \tilde{\mathbf{b}}(i)) \in \mathcal{F} \times [q]^k$  is sampled according to  $\mathcal{D}_N$ .
4.  $\mathbf{z} = (\mathbf{z}(1), \dots, \mathbf{z}(\alpha n))$  is determined from  $M, \mathbf{x}^*$  and  $\mathbf{b} = (\mathbf{b}(1), \dots, \mathbf{b}(\alpha n))$  as follows. Recall that  $\mathbf{b}(i) = (f_i, \tilde{\mathbf{b}}(i))$ . We let  $\mathbf{z}(i) = (f_i, \tilde{z}_i) \in \mathcal{F} \times \{0, 1\}$  where  $\tilde{z}_i = 1$  iff  $M_i \mathbf{x}^* = \tilde{\mathbf{b}}(i)$ .

- Alice receives  $\mathbf{x}^*$  as input.
- Bob receives  $M$  and  $\mathbf{z}$  as input.

In the special case when the set  $\mathcal{F}$  contains just one element,  $|\mathcal{F}| = 1$ , we call the corresponding communication problem  $(\mathcal{D}_Y, \mathcal{D}_N)$ -SD.

**Theorem 5.4** (Communication lower bound for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD). *For every  $k, q$ , every finite set  $\mathcal{F}$ , every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$  with  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$  there exists  $\alpha_0 > 0$  such that for every  $0 < \alpha \leq \alpha_0$  and  $\delta > 0$  there exists  $\tau > 0$  such that the following holds: Every protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD achieving advantage  $\delta$  on instances of length  $n$  requires  $\tau\sqrt{n}$  bits of communication.*

Sections 6 to 8 are devoted to proving Theorem 5.4. The specific proof can be found in Section 7.3. In the rest of this section we assume this theorem to prove Theorems 5.1 and 2.16.

## 5.2 The insertion-only lower bound

In this section we introduce two streaming problems, the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -streaming-SD problem and the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD problem in Definitions 5.5 and 5.7, where the new parameter  $T$  is an integer (a large constant) that represents the number of “parts” in a stream. We then show how to reduce these problems to  $\text{Max-CSP}(\mathcal{F})$  problems so that membership of



$\mathcal{D}_Y \in S_\gamma^Y(\mathcal{F})$  and  $\mathcal{D}_N \in S_\beta^N(\mathcal{F})$  leads to a gap in the value of the instances produced by this reduction for large  $T$ . (See [Lemma 5.8](#).) We then show that the lower bounds on the one-way communication complexity of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD translate into lower bounds on the space complexity of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -streaming-SD. (See [Lemma 5.12](#).) Combining these lemmas leads immediately to a proof of [Theorem 2.16](#) in [Section 5.2.4](#).

### 5.2.1 The (Padded) Streaming SD Problem

**Definition 5.5** ( $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -streaming-SD). For  $k, q, T \in \mathbb{N}$ ,  $\alpha \in (0, 1/k]$ , a finite set  $\mathcal{F}$  and distributions  $\mathcal{D}_Y, \mathcal{D}_N$  over  $\mathcal{F} \times [q]^k$ , the streaming problem  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T; \alpha, k, q)$ -streaming-SD is the task of distinguishing, for every  $n$ ,  $\sigma \sim \mathcal{Y}_{stream, n}$  from  $\sigma \sim \mathcal{N}_{stream, n}$  where for a given length parameter  $n$ , the distributions  $\mathcal{Y}_{stream} = \mathcal{Y}_{stream, n}$  and  $\mathcal{N}_{stream} = \mathcal{N}_{stream, n}$  are defined as follows:

- Let  $\mathcal{Y}$  be the distribution over instances of length  $n$ , i.e., triples  $(\mathbf{x}^*, M, \mathbf{z})$ , from the definition of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD. For  $\mathbf{x} \in [q]^n$ , let  $\mathcal{Y}|_{\mathbf{x}}$  denote the distribution  $\mathcal{Y}$  conditioned on  $\mathbf{x}^* = \mathbf{x}$ . The stream  $\sigma \sim \mathcal{Y}_{stream}$  is sampled as follows: Sample  $\mathbf{x}^*$  uniformly from  $[q]^n$ . Let  $(M^{(1)}, \mathbf{z}^{(1)}), \dots, (M^{(T)}, \mathbf{z}^{(T)})$  be sampled independently according to  $\mathcal{Y}|_{\mathbf{x}^*}$ . Let  $\sigma^{(t)}$  be the pair  $(M^{(t)}, \mathbf{z}^{(t)})$  presented as a stream of edges with labels in  $\mathcal{F} \times \{0, 1\}$ , i.e.,  $\mathbf{z}^{(t)} = (f_i, \tilde{z}_i)$ . Specifically for  $t \in [T]$  and  $i \in [\alpha n]$ , let  $\sigma^{(t)}(i) = (e^{(t)}(i), \mathbf{z}^{(t)}(i))$  where  $e^{(t)}(i)$  is the  $i$ -th hyperedge of  $M^{(t)}$ , i.e.,  $e^{(t)}(i) = (j^{(t)}(k(i-1)+1), \dots, j^{(t)}(k(i-1)+k))$  and  $j^{(t)}(\ell)$  is the unique index  $j$  such that  $M_{j, \ell}^{(t)} = 1$ . Finally we let  $\sigma = \sigma^{(1)} \circ \dots \circ \sigma^{(T)}$  be the concatenation of the  $\sigma^{(t)}$ s.
- $\sigma \sim \mathcal{N}_{stream}$  is sampled similarly except we now sample  $(M^{(1)}, \mathbf{z}^{(1)}), \dots, (M^{(T)}, \mathbf{z}^{(T)})$  independently according to  $\mathcal{N}|_{\mathbf{x}^*}$  where  $\mathcal{N}|_{\mathbf{x}}$  is the distribution  $\mathcal{N}$  condition on  $\mathbf{x}^* = \mathbf{x}$ .

Again when  $\alpha, k, q$  are clear from context we suppress them and simply refer to the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -streaming-SD problem.

**Remark 5.6.** We note that when  $\mathcal{D}_N = \mathcal{D}_{\mathcal{F}} \times \text{Unif}([q]^k)$  for some  $\mathcal{D}_{\mathcal{F}} \in \Delta(\mathcal{F})$ , then the distributions  $\mathcal{N}|_{\mathbf{x}^*}$  are identical for all  $\mathbf{x}^*$  (and the variables  $\mathbf{z}^{(t)}(i)$  are distributed as  $\mathcal{D}_{\mathcal{F}} \times \text{Bern}(q^{-k})$  independently for every  $t, i$ ).

For technical reasons, we need the following *padded* version of streaming-SD to extend our lower bound techniques in the insertion-only setting beyond the setting of one-wise independent distributions.

**Definition 5.7** ( $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD). For  $k, q, T \in \mathbb{N}$ ,  $\alpha \in (0, 1/k]$ ,  $\tau \in [0, 1)$ , a finite set  $\mathcal{F}$ , and distributions  $\mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0$  over  $\mathcal{F} \times [q]^k$ , the streaming problem  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau; \alpha, k, q)$ -padded-streaming-SD is the task of distinguishing, for every  $n$ ,  $\sigma \sim \mathcal{Y}_{pad-stream, n}$  from  $\sigma \sim \mathcal{N}_{pad-stream, n}$  where for a given length parameter  $n$ , the distributions  $\mathcal{Y}_{pad-stream} = \mathcal{Y}_{pad-stream, n}$  and  $\mathcal{N}_{pad-stream} = \mathcal{N}_{pad-stream, n}$  are defined as follows: Sample  $\mathbf{x}^*$  from  $[q]^n$  uniformly. For each  $i \in [\frac{\tau}{1-\tau} \alpha n T]$ , uniformly sample a tuple  $e^{(0)}(i) = (i_1, \dots, i_k) \in \binom{[n]}{k}$  and  $(f_i, \mathbf{b}^{(0)}(i)) \sim \mathcal{D}_0$ , let  $\sigma^{(0)}(i) = (e^{(0)}(i), (f_i, \mathbf{1}_{\mathbf{b}^{(0)}(i)=\mathbf{x}^*|_{e^{(0)}(i)}}))$ . Next, sample  $\sigma^{(1)}, \dots, \sigma^{(T)}$  according to the Yes and No distribution of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -streaming-SD respectively. Finally, let  $\sigma = \sigma^{(0)} \circ \dots \circ \sigma^{(T)}$  be the concatenation of the  $\sigma^{(t)}$ s.

Again when  $\alpha, k, q$  are clear from context we suppress them and simply refer to the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD problem. Note that when  $\tau = 0$ ,  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD is the same as  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -streaming-SD.

### 5.2.2 CSP value of padded-streaming-SD

There is a natural way to convert instances of padded-streaming-SD to instances of a Max-CSP( $\mathcal{F}$ ) problem. In this section we make this conversion explicit and show to use properties of the underlying distributions  $\mathcal{D}_0, \mathcal{D}_Y, \mathcal{D}_N$  to get bounds on the value of the instances produced.

Note that an instance  $\sigma$  of padded-streaming-SD is simply a sequence  $(\sigma(1), \dots, \sigma(\ell))$  where each  $\sigma(i) = (\mathbf{j}(i), \mathbf{z}(i))$  with  $\mathbf{j}(i) \in [n]^k$  and  $\mathbf{z}(i) = (f_i, \tilde{z}_i) \in \mathcal{F} \times \{0, 1\}$ . This sequence is already syntactically very close to the description of a Max-CSP( $\mathcal{F}$ ) instance. Formally, we define an instance  $\Psi(\sigma)$  of Max-CSP( $\mathcal{F}$ ) as follows. For each  $\sigma_i = (\mathbf{j}(i), \mathbf{z}(i))$  with  $\mathbf{z}(i) = (f_i, \tilde{z}_i)$ , if  $\tilde{z}_i = 1$  we add the constraint  $f_i(\mathbf{x}_{|\mathbf{j}(i)})$  to  $\Psi(\sigma)$ ; otherwise, we do not add any constraint to the formula.

In what follows we show that if  $\mathcal{D}_Y \in S_\gamma^Y$  then for all sufficiently large constant  $T$ , and sufficiently large  $n$ , if we draw  $\sigma \sim \mathcal{Y}_{\text{pad-stream}, n}$ , then with high probability,  $\Psi(\sigma)$  has value at least  $\gamma - o(1)$ . Conversely if  $\mathcal{D}_N \in S_\beta^N$ , then for all sufficiently large  $n$ , if we draw  $\sigma \sim \mathcal{N}_{\text{pad-stream}, n}$ , then with high probability  $\Psi(\sigma)$  has value at most  $\beta + o(1)$ .

**Lemma 5.8** (CSP value of padded-streaming-SD). *For every  $q, k \in \mathbb{N}$ ,  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ ,  $0 \leq \beta < \gamma \leq 1$ ,  $\varepsilon > 0$ ,  $\tau = [0, 1)$ , and distributions  $\mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0 \in \Delta(\{-1, 1\}^k)$  there exists  $\alpha_0$  such that for every  $\alpha \in (0, \alpha_0]$  the following hold for every sufficiently large  $T$ :*

1. *If  $\tau\mathcal{D}_0 + (1-\tau)\mathcal{D}_Y \in S_\gamma^Y$ , then for every sufficiently large  $n$ , the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD **YES** instance  $\sigma \sim \mathcal{Y}_{\text{pad-stream}, n}$  satisfies  $\Pr[\text{val}_{\Psi(\sigma)} < (\gamma - \varepsilon)] \leq \exp(-n)$ .<sup>8</sup>*
2. *If  $\tau\mathcal{D}_0 + (1-\tau)\mathcal{D}_N \in S_\beta^N$ , then for every sufficiently large  $n$ , the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD **NO** instance  $\sigma \sim \mathcal{N}_{\text{pad-stream}, n}$  satisfies  $\Pr[\text{val}_{\Psi(\sigma)} > (\beta + \varepsilon)] \leq \exp(-n)$ .*

Furthermore, if  $\gamma = 1$  then  $\Pr_{\sigma \sim \mathcal{Y}_{\text{pad-stream}, n}} [\text{val}_{\Psi(\sigma)} = 1] = 1$ .

*Proof.* We assume  $\varepsilon \leq 1/2$  (and if not we prove the lemma for  $\varepsilon' = \frac{1}{2}$  and this implies the lemma also for  $\varepsilon$ ). We prove the lemma for  $\alpha_0 = \frac{\varepsilon}{20kq^k}$  and  $T_0 = 1000/(\varepsilon^2\alpha)$ . In what follows we set  $\eta = \frac{\varepsilon}{20kq^k}$ .

In what follows we let  $N_0 = \frac{\tau\alpha n T}{1-\tau}$ ,  $N_t = \alpha n$  for  $t \in [T]$  and  $N = N_0 + TN_1$ . Recall that an instance of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD consists of a stream  $\sigma = \sigma^{(0)} \circ \dots \circ \sigma^{(T)}$  where  $\sigma^{(t)} = (\sigma^{(t)}(i) | i \in [N_t])$  and  $\sigma^{(t)}(i) = (e^{(t)}(i), (f_i^{(t)}, \tilde{z}^{(t)}(i)))$  where  $e^{(t)}(i)$  denotes a  $k$ -uniform hyperedge on  $[n]$  and  $f_i^{(t)} \in \mathcal{F}$  and  $\tilde{z}^{(t)}(i) \in \{0, 1\}$ . Finally recall that  $\sigma^{(t)} \sim \mathcal{Y}_{|\mathbf{x}^*}$  in the **YES** case and  $\sigma^{(t)} \sim \mathcal{N}_{|\mathbf{x}^*}$  in the **NO** case independently for each  $t$ , where  $\mathbf{x}^* \sim \text{Unif}([q]^n)$  is common across all  $t$ . We use  $\mathcal{I} = (\{0\} \times [T_0]) \cup ([T] \times [T_1])$  to denote the set of legal pairs of indices  $(t, i)$ . We let  $m$  denote the total number of constraints in  $\Psi(\sigma)$  with  $m_t$  denoting the number of constraints from  $\sigma^{(t)}$  for  $0 \leq t \leq T$ . (Note that  $m$  and the  $m_t$ 's are random variables.)

For  $\eta > 0$ , define  $\mathbf{x}^*$  to be  $\eta$ -good if for every  $\sigma \in [q]$ , we have  $|\{i \in [n] \mid \mathbf{x}_i^* = \sigma\}| \in [(1-\eta) \cdot \frac{n}{q}, (1+\eta) \cdot \frac{n}{q}]$ . A straightforward application of Chernoff bounds shows that for every  $\eta > 0$  the vector  $\mathbf{x}^*$  is  $\eta$ -good with probability  $1 - \exp(-n)$ .

Below we condition on a good  $\mathbf{x}^*$  and prove the following: (1) We show the expected value of  $m$  is roughly  $q^{-k} \cdot N$  and furthermore  $m$  is sharply concentrated around its expectation. (2) In the **YES** case we prove that the expected number of constraints satisfied by  $\mathbf{x}^*$  is roughly at least  $\gamma \cdot q^{-k} \cdot N$  and again this variable is sharply concentrated around its expectation. (3) In the **NO**

<sup>8</sup>In this lemma and proof we use  $\exp(-n)$  to denote functions of the form  $c^{-n}$  for some  $c > 1$  that does not depend on  $n$  or  $T$ , but may depend on all other parameters including  $q, k, \mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0, \beta, \gamma, \varepsilon$ .

case we prove that the expected number of constraints satisfied by any assignment is roughly at most  $\beta \cdot q^{-k} \cdot N$  and again this variable is sharply concentrated around its expectation. We note that the sharp concentration part is essentially the same in all cases and it is bounding the expectations that is different in each case. That being said the analysis of the **NO** case does require sharper concentration since we need to take a union bound over all possible assignments.

**Bounding the number of constraints.** We start with step (1). Fix an  $\eta$ -good  $\mathbf{x}^*$ . Note that  $m_t = \sum_{i \in [N_t]} \tilde{z}^{(t)}(i)$  for every  $0 \leq t \leq T$ . We divide the analysis into two subparts. In step (1a) we bound  $\mu \triangleq \mathbb{E}[\tilde{z}^{(t)}(i)]$  (in particular this expectation does not depend on  $i$  or  $t$ ). Note that  $m = \sum_{t=0}^T \sum_{i \in [N_t]} \tilde{z}^{(t)}(i)$  and so bounding  $\mu$  bounds  $\mathbb{E}[m] = \mu \cdot N$ . Then in step (1b) we show that  $m$  is concentrated around its expected value.

For step (1a), let  $p_\sigma$  denote the fraction of occurrences of the letter  $\sigma$  in  $\mathbf{x}^*$ , i.e.,  $p_\sigma = \frac{1}{n} |\{i \in [n] | \mathbf{x}_i^* = \sigma\}|$ . Note that given a sequence  $\tilde{\mathbf{b}}^{(t)}(i) = \mathbf{a} \in [q]^k$ , the probability that  $\tilde{z}^{(t)}(i) = 1$  over a random choice of  $e^{(t)}(i)$  depends on  $\mathbf{a}$  as well as the  $p_\sigma$ 's. (Specifically this probability is  $\prod_{j=1}^k p_{\mathbf{a}_j} \pm O(k^2/n)$ , where the additive correction term accounts for the sampling without replacement in the choice of  $e^{(t)}(i)$ .) However if the vector  $\mathbf{x}^*$  is good, this dependence has little quantitative effect. In particular, if  $\mathbf{x}^*$  is  $\eta$ -good, we have  $\mu \in (\frac{1}{q} \pm \eta)^k \pm O(k^2/n)$  and thus we get  $q^{-k} - 2k\eta \leq \mu \leq q^{-k} + 2k\eta$  provided  $\eta \leq 1/(4kq)$  and  $n$  is sufficiently large. This simplifies further to  $\mu \in (1 \pm \frac{\varepsilon}{10})q^{-k}$  using  $\eta \leq q^{-k}\varepsilon/(20k)$ . Summing up over  $(t, i) \in \mathcal{I}$  we get  $\mathbb{E}[m] \in (1 \pm \frac{\varepsilon}{10})q^{-k}N$ .

We now turn to step (1b), i.e., proving that  $m$  is concentrated around its expectation. (In this part we work a little harder than necessary to prove that the failure probability is  $\exp(-nT)$  rather than  $\exp(-n)$ . This is not necessary, but will be needed for the similar step in step (3).) Let  $\tilde{Z}$  denote the set of random variables  $\{\tilde{z}^{(t)}(i) | (t, i) \in \mathcal{I}\}$  and for  $(t, i) \in \mathcal{I}$ , let  $\tilde{Z}_{-(t,i)} = \tilde{Z} \setminus \{\tilde{z}^{(t)}(i)\}$ . We first show that for every  $(t, i) \in \mathcal{I}$  we have  $\mathbb{E}[\tilde{z}^{(t)}(i) | \tilde{Z}_{-(t,i)}] \in (1 \pm \frac{\varepsilon}{10}) \mathbb{E}[\tilde{z}^{(t)}(i)]$ . Let  $B_t$  denote the  $t$ -th block of variables, i.e.,  $B_t = \{\tilde{z}^{(t)}(i) | i \in [N_t]\}$ . Now note that the only dependence among the  $\tilde{z}^{(t)}(i)$ 's is among the variables within a block while the blocks themselves are independent. Furthermore the variables in the block  $B_0$  are independent of each other. Thus for  $i \in [N_0]$  we have  $\mathbb{E}[\tilde{z}^{(0)}(i) | Z_{-(0,i)}] = \mathbb{E}[\tilde{z}^{(0)}(i)]$ . For  $t > 0$ , we have the variables from block  $B_t$  may depend on each other due to the constraint that the underlying set of hyperedges are vertex disjoint. Fix  $(t, i) \in \mathcal{I}$  with  $t > 0$  and let  $S$  be the set of variables touched by the hyperedges from block  $B_t$ , excluding  $e^{(t)}(i)$ . Now consider picking a hyperedge uniformly from  $[n]$  and let  $\psi$  be the probability that this hyperedge touches  $S$ . We clearly have  $\psi \leq k|S|/n \leq k\alpha$ . On the other hand,  $\psi$  also upper bounds the difference between  $\mathbb{E}[\tilde{z}^{(t)}(i) | \tilde{Z}_{-(t,i)}]$  and  $\mathbb{E}[\tilde{z}^{(t)}(i)]$ , so we have:

$$|\mathbb{E}[\tilde{z}^{(t)}(i) | \tilde{Z}_{-(t,i)}] - \mathbb{E}[\tilde{z}^{(t)}(i)]| \leq \psi \leq k\alpha \leq \frac{\varepsilon q^{-k}}{20} \leq \frac{\varepsilon}{10} \mathbb{E}[\tilde{z}^{(t)}(i)].$$

Applying [Lemma 3.5](#) to the variables of  $\tilde{Z}$  (arranged in some arbitrary order) we have  $\Pr[m \notin ((q^{-k} \cdot (1 \pm \varepsilon/10))^3) \leq \exp(-nT)$ . Using  $(1 \pm \varepsilon/10)^3 \subseteq (1 \pm \varepsilon/2)$  for  $\varepsilon < 1$  we get:

$$\Pr[m \notin (1 \pm \varepsilon/2) \cdot q^{-k}N] \leq \exp(-nT) \tag{5.9}$$

**Lower bounding the number of satisfied constraints in the YES case.** Let  $Z^{(t)}(i)$  be the indicator variable for the event that the  $i$ -th element of  $\boldsymbol{\sigma}^{(t)}$  produces a constraint that is satisfied by  $\mathbf{x}^*$ , i.e.,  $Z^{(t)}(i) = \tilde{z}^{(t)}(i) \cdot f_i(\mathbf{x}^*|_{\mathbf{j}^{(t)}(i)})$ . Note that the number of constraints satisfied by  $\mathbf{x}^*$  is

$\sum_{(t,i) \in \mathcal{I}} Z^{(t)}(i)$ . Note further that  $Z^{(0)}(i)$ 's are identically distributed across  $i \in [N_0]$ , and  $Z^{(t)}(i)$  are also identically distributed across  $t \in [T]$  and  $i \in [N_0]$ . By construction (see [Definition 5.7](#)) we have  $\mathbb{E}[Z^{(0)}(i)] = \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_0}[f(\mathbf{b}) \cdot \mathbb{E}_{\mathbf{j}}[\mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})]]$ . By the  $\eta$ -goodness of  $\mathbf{x}^*$ , we have that for every  $\mathbf{b} \in [q]^k$ ,  $\mathbb{E}_{\mathbf{j}}[\mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})] \geq (1 - \frac{\varepsilon}{10})q^{-k}$ . Thus we get  $\mathbb{E}[Z^{(0)}(i)] \geq (1 - \frac{\varepsilon}{10})q^{-k} \cdot \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_0}[f(\mathbf{b})]$ . Similarly for  $t > 0$  we have  $\mathbb{E}[Z^{(t)}(i)] = \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_Y}[f(\mathbf{b}) \cdot \mathbb{E}_{\mathbf{j}}[\mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})]] \geq (1 - \frac{\varepsilon}{10})q^{-k} \cdot \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_Y}[f(\mathbf{b})]$ . Using linearity of expectations we now get

$$\begin{aligned} \mathbb{E} \left[ \sum_{(t,i) \in \mathcal{I}} Z^{(t)}(i) \right] &= N_0 \mathbb{E}[Z^{(0)}(1)] + TN_T \mathbb{E}[Z^{(1)}(1)] \\ &= N(\tau \mathbb{E}[Z^{(0)}(1)] + (1 - \tau) \mathbb{E}[Z^{(1)}(1)]) \\ &\geq \left(1 - \frac{\varepsilon}{10}\right) q^{-k} N \cdot \left(\tau \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_0}[f(\mathbf{b})] + (1 - \tau) \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_Y}[f(\mathbf{b})]\right) \\ &= \left(1 - \frac{\varepsilon}{10}\right) q^{-k} N \cdot \mathbb{E}_{(f,\mathbf{b}) \sim \tau \mathcal{D}_0 + (1-\tau) \mathcal{D}_Y}[f(\mathbf{b})] \\ &\geq \gamma \cdot \left(1 - \frac{\varepsilon}{10}\right) q^{-k} N, \end{aligned}$$

where the final inequality uses  $\tau \mathcal{D}_0 + (1 - \tau) \mathcal{D}_Y \in S_\gamma^Y(\mathcal{F})$ . The concentration can be analyzed exactly as in step (1b). In particular if we let  $Z$  denote all variables  $Z^{(t)}(i)$ 's, then we have  $\mathbb{E}[Z^{(t)}(i) | Z \setminus \{Z^{(t)}(i)\}] \geq \mathbb{E}[Z^{(t)}(i)] - \frac{\varepsilon}{10} q^{-k}$ .

$$\Pr \left[ \sum_{(t,i) \in \mathcal{I}} Z^{(t)}(i) \leq \left(\gamma - \frac{3\varepsilon}{10}\right) \cdot q^{-k} N \leq \gamma \cdot \left(1 - \frac{\varepsilon}{10}\right) q^{-k} N - \frac{\varepsilon}{5} q^{-k} N \right] \leq \exp(-nT). \quad (5.10)$$

**Upper bounding the number of satisfiable constraints in the NO case.** Fix an assignment  $\nu \in [q]^k$  and consider the expected number of constraints satisfied by  $\nu$ . (We will later take a union bound over all  $\nu$ .) Let  $W^{(t)}(i)$  be the indicator variable for the event that the  $i$ -th element of  $\sigma^{(t)}$  produces a constraint that is satisfied by  $\nu$ , i.e.,  $W^{(t)}(i) = \tilde{z}^{(t)}(i) \cdot f_i(\nu|_{\mathbf{j}^{(t)}(i)})$ . Note once again that  $W^{(0)}(i)$ 's are identically distributed across  $i$  and  $W^{(t)}(i)$  are identical across  $t > 0$  and  $i$ . Let  $\mu_0 = \mathbb{E}[W^{(0)}(1)]$  and  $\mu_N = \mathbb{E}[W^{(1)}(1)]$ . Note that the expected number of satisfied constraints is  $\mathbb{E}[\sum_{(t,i) \in \mathcal{I}} W^{(t)}(i)] = N \cdot (\tau \mu_0 + (1 - \tau) \mu_N)$ , so we bound  $\mu_0$  and  $\mu_N$ . By construction we have

$$\mu_0 = \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_{0,\mathbf{j}}}[f(\nu|_{\mathbf{j}}) \cdot \mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})] = \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_{0,\mathbf{j}}}[\mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})] \cdot \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_{0,\mathbf{j}}}[f(\nu|_{\mathbf{j}}) | \mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})]$$

where  $\mathbf{j}$  is a uniform random sequence of  $k$  distinct elements of  $[n]$ . As argued earlier for every  $\mathbf{b}$  we have  $\mathbb{E}_{\mathbf{j}}[\mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})] \leq (1 + \frac{\varepsilon}{10})q^{-k}$  for  $\eta$ -good  $\mathbf{x}^*$ . So we turn to bounding the second term.

For  $\sigma, \rho \in [q]$  let  $\mathcal{P}_\sigma(\rho)$  be the fraction of coordinates in  $\nu$  that take the value  $\rho$  among those coordinates where  $\mathbf{x}^*$  is  $\sigma$ , i.e.,  $\mathcal{P}_\sigma(\rho) = \frac{|\{i \in [n] | \nu_i = \rho \ \& \ \mathbf{x}_i^* = \sigma\}|}{|\{i \in [n] | \mathbf{x}_i^* = \sigma\}|}$ . Note that for every  $\sigma$ ,  $\mathcal{P}_\sigma$  is a probability distribution in  $\Delta(q)$ . Furthermore, conditioning on  $\mathbf{x}^*|_{\mathbf{j}(\ell)} = \mathbf{b}(\ell)$ , the distribution of  $\nu|_{\mathbf{j}(\ell)}$  is given by  $\mathcal{P}_{\mathbf{b}(\ell)}$ . Thus the joint distribution of  $\nu|_{\mathbf{j}}$  is  $O(k^2/n)$ -close in total variation distance to  $\mathcal{P}_{\mathbf{b}(1)} \times \cdots \times \mathcal{P}_{\mathbf{b}(k)}$ . We thus have

$$\begin{aligned} \mathbb{E}_{(f,\mathbf{b}) \sim \mathcal{D}_{0,\mathbf{j}}}[f(\nu|_{\mathbf{j}}) | \mathbb{1}(\mathbf{x}^*|_{\mathbf{j}} = \mathbf{b})] &\leq \mathbb{E}_{(f,\mathbf{a}) \sim \mathcal{D}_0} \left[ \mathbb{E}_{\mathbf{c}, \mathbf{c}_\ell \sim \mathcal{P}_{\mathbf{a}_\ell}}[f(\mathbf{c})] \right] + O(k^2/n) \\ &= \mathbb{E}_{(f,\mathbf{a}) \sim \mathcal{D}_0} \left[ \mathbb{E}_{\mathbf{d}, \mathbf{d}_\ell, \sigma \sim \mathcal{P}_\sigma}[\mathcal{C}(f, \mathbf{a})(\mathbf{d})] \right] + O(k^2/n), \end{aligned}$$

where  $\mathbf{c} \in [q]^k$  and  $\mathbf{d} \in [q]^{k \times q}$ . Note that the final expression is simply a change of notation applied to the middle expression above to make the expression syntactically closer to the notation in the definition of  $S_\beta^N(\mathcal{F})$ . Combining with the bound on  $\mathbb{E}_j[\mathbb{1}(\mathbf{x}^*|\mathbf{j} = b)]$  above we get

$$\mu_0 \leq \left(1 + \frac{\varepsilon}{10}\right)q^{-k} \cdot \left(\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}_0} \left[ \mathbb{E}_{\mathbf{d}, d_{\ell, \sigma} \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{d})] \right]\right) + O(k/n).$$

Similarly we get

$$\mu_N \leq \left(1 + \frac{\varepsilon}{10}\right)q^{-k} \cdot \left(\mathbb{E}_{(f, \mathbf{a}) \sim \mathcal{D}_N} \left[ \mathbb{E}_{\mathbf{d}, d_{\ell, \sigma} \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{d})] \right]\right) + O(k/n).$$

Now combining the two conditions above we get

$$\begin{aligned} (\tau\mu_0 + (1 - \tau)\mu_N) &\leq \left(1 + \frac{\varepsilon}{10}\right)q^{-k} \cdot \left(\mathbb{E}_{(f, \mathbf{a}) \sim \tau\mathcal{D}_0 + (1 - \tau)\mathcal{D}_N} \left[ \mathbb{E}_{\mathbf{d}, d_{\ell, \sigma} \sim \mathcal{P}_\sigma} [\mathcal{C}(f, \mathbf{a})(\mathbf{d})] \right]\right) + O(k^2/n) \\ &\leq \beta \cdot \left(1 + \frac{\varepsilon}{10}\right)q^{-k} + O(k^2/n) \\ &\leq \beta \cdot \left(1 + \frac{\varepsilon}{9}\right)q^{-k}, \end{aligned}$$

where the final inequality uses the fact that  $n$  is sufficiently large. We thus conclude the the expected number of constraints satisfied by  $\nu$  is at most  $\beta \cdot \left(1 + \frac{\varepsilon}{9}\right)q^{-k}N$ . Concentration around the mean is now similar to before. In particular if we let  $W$  denote the set of all  $W^{(t)}(i)$ 's then we still have If we  $\mathbb{E}[W^{(t)}(i)|W \setminus \{W^{(t)}(i)\}] \leq \mathbb{E}[W^{(t)}(i)] + k\alpha \leq \mathbb{E}[W^{(t)}(i)] + \frac{\varepsilon}{10}q^{-k}N$ , and so by [Lemma 3.5](#) we get

$$\Pr \left[ \sum_{(t, i) \in \mathcal{I}} W^{(t)}(i) \geq \left(\beta + \frac{2\varepsilon}{9}\right) \cdot q^{-k}N \geq \beta(1 + \varepsilon/9)q^{-k}N + \frac{\varepsilon}{9}q^{-k}N \right] \leq \exp(-nT).$$

In particular by using  $T$  sufficiently large, we get that the probability that more than  $(\beta + \frac{2\varepsilon}{9}) \cdot q^{-k}N$  constraints are satisfied by  $\nu$  is at most  $c^{-n}$  for some  $c > q$ . So by a union bound over all possible  $\nu$ 's we get the following:

$$\Pr \left[ \exists \nu \in [q]^k \text{ s.t. } \nu \text{ satisfies more than } \left(\beta + \frac{2\varepsilon}{9}\right) \cdot q^{-k}N \text{ constraints} \right] \leq \exp(-nT). \quad (5.11)$$

**Putting it together.** Putting the above together we get that in the **YES** case with probability  $1 - \exp(-n)$  we have that  $\mathbf{x}^*$  is good and the number of constraints is at most  $(1 + \frac{\varepsilon}{2})q^{-k}N$  (by [Eq. \(5.9\)](#)) while the number of satisfied constraints is at least  $(\gamma - \frac{3\varepsilon}{10}) \cdot q^{-k}N$  (by [Eq. \(5.10\)](#)). Taking ratios we get

$$\text{val}_{\Psi(\sigma)} \geq \frac{\gamma - \frac{3\varepsilon}{10}}{1 + \frac{\varepsilon}{2}} \geq \gamma - \varepsilon.$$

Similarly in the **NO** case we have with probability at least  $1 - \exp(-n)$  we have that  $\mathbf{x}^*$  is good, and the number of constraints is at least  $(1 - \frac{\varepsilon}{2})q^{-k}N$  (by [Eq. \(5.9\)](#)) while the number of satisfied constraints is at most  $(\beta + \frac{2\varepsilon}{9}) \cdot q^{-k}N$  (by [Eq. \(5.11\)](#)). Taking ratios we get

$$\text{val}_{\Psi(\sigma)} \leq \frac{\beta + \frac{2\varepsilon}{9}}{1 - \frac{\varepsilon}{2}} \leq \beta + \varepsilon.$$

This proves the main part of the lemma.

The furthermore part follows from the fact that if  $\gamma = 1$  then every constraint in the **YES** case is satisfied by  $\mathbf{x}^*$ .  $\square$

### 5.2.3 Reduction from one-way $(\mathcal{D}_Y, \mathcal{D}_N)$ -SD to padded-streaming-SD

We start by reducing SD to padded-streaming-SD in the special case where  $\mathcal{D}_N$  is “uniform on the variables” in the sense defined next. We say a distribution  $\mathcal{D} \in \Delta(\mathcal{D} \times [q]^k)$  is *uniform on the variables* if there exists a distribution  $\mathcal{D}_f \in \Delta(\mathcal{F})$  such that  $\mathcal{D} = \mathcal{D}_f \times \text{Unif}([q]^k)$ . The following lemma implies that in this special case padded-streaming-SD is hard. Since this holds for all one-wise independent distributions  $\mathcal{D}_Y$ , by applying the lemma twice we get that padded-streaming-SD is hard for all one-wise independent  $\mathcal{D}_Y$  and  $\mathcal{D}_N$ .

**Lemma 5.12.** *Let  $\mathcal{F}$  be a finite set,  $T, q, k \in \mathbb{N}$ ,  $\alpha \in (0, \alpha_0(k)]$ ,  $\tau \in [0, 1)$ , and  $\mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0 \in \Delta(\mathcal{F} \times [q]^k)$  with  $\mathcal{D}_Y$  being one-wise independent and  $\mathcal{D}_N = \mathcal{D}_f \times \text{Unif}([q]^k)$  for some  $\mathcal{D}_f \in \Delta(\mathcal{F})$  and  $\mu(\mathcal{D}_Y) = \mu(\mathcal{D}_N)$ . Suppose there is a streaming algorithm **ALG** that solves  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD on instances of length  $n$  with advantage  $\Delta$  and space  $s$ , then there is a one-way protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD on instances of length  $n$  using at most  $sT$  bits of communication achieving advantage at least  $\Delta/T$ .*

Our proof is essentially the same as that of [CGSV21] with some minor modifications. We note their proof was based on the proof of [CGV20, Lemma 4.11] and the proof of [KKS15, Lemma 6.3].

*Proof of Lemma 5.12.* Note that since we are interested in distributional advantage, we can fix the randomness in **ALG** so that it becomes a deterministic algorithm. By an averaging argument the randomness can be chosen to ensure the advantage does not decrease. Let  $\Gamma$  denote the evolution of function of **ALG** as it processes a block of edges. That is, if the algorithm is in state  $s$  and receives a stream  $\sigma$  then it ends in state  $\Gamma(s, \sigma)$ . Let  $s_0$  denote its initial state.

We consider the following collection of (jointly distributed) random variables: Let  $\mathbf{x}^* \sim \text{Unif}(\{-1, 1\}^n)$ . Denote  $\mathcal{Y} = \mathcal{Y}_{\text{pad-stream}, n}$  and  $\mathcal{N} = \mathcal{N}_{\text{pad-stream}, n}$ . Let  $(\sigma_Y^{(0)}, \sigma_Y^{(1)}, \dots, \sigma_Y^{(T)}) \sim \mathcal{Y}|_{\mathbf{x}^*}$ . Similarly, let  $(\sigma_N^{(0)}, \sigma_N^{(1)}, \dots, \sigma_N^{(T)}) \sim \mathcal{N}|_{\mathbf{x}^*}$ . Recall by Remark 5.6 that since  $\mathcal{D}_N = \mathcal{D}_f \times \text{Unif}([q]^k)$ , we have  $\mathcal{N}|_{\mathbf{x}^*}$  is independent of  $\mathbf{x}^*$ , a feature that will be crucial to this proof.

Let  $S_t^Y$  denote the state of **ALG** after processing  $\sigma_Y^{(0)}, \dots, \sigma_Y^{(t)}$ , i.e.,  $S_0^Y = \Gamma(s_0, \sigma_Y^{(0)})$  and  $S_t^Y = \Gamma(S_{t-1}^Y, \sigma_Y^{(t)})$  where  $s_0$  is the fixed initial state (recall that **ALG** is deterministic). Similarly let  $S_t^N$  denote the state of **ALG** after processing  $\sigma_N^{(0)}, \dots, \sigma_N^{(t)}$ . Note that since  $\sigma_Y^{(0)}$  has the same distribution (conditioned on the same  $\mathbf{x}^*$ ) as  $\sigma_N^{(0)}$  by definition, we have  $\|S_0^Y - S_0^N\|_{\text{tvd}} = 0$ .

Let  $S_{a:b}^Y$  denote the sequence of states  $(S_a^Y, \dots, S_b^Y)$  and similarly for  $S_{a:b}^N$ . Now let  $\Delta_t = \|S_{0:t}^Y - S_{0:t}^N\|_{\text{tvd}}$ . Observe that  $\Delta_0 = 0$  while  $\Delta_T \geq \Delta$ . (The latter is based on the fact that **ALG** distinguishes the two distributions with advantage  $\Delta$ .) Thus  $\Delta \leq \Delta_T - \Delta_0 = \sum_{t=0}^{T-1} (\Delta_{t+1} - \Delta_t)$  and so there exists  $t^* \in \{0, 1, \dots, T-1\}$  such that

$$\Delta_{t^*+1} - \Delta_{t^*} = \|S_{0:t^*+1}^Y - S_{0:t^*+1}^N\|_{\text{tvd}} - \|S_{0:t^*}^Y - S_{0:t^*}^N\|_{\text{tvd}} \geq \frac{\Delta}{T}.$$

Now consider the random variable  $\tilde{S} = \Gamma(S_{t^*}^Y, \sigma_N^{(t^*+1)})$  (so the previous state is from the **YES** distribution and the input is from the **NO** distribution). We claim below that  $\|S_{t^*+1}^Y - \tilde{S}\|_{\text{tvd}} =$



$\mathbb{E}_{A \sim_d S_{0:t^*}^Y} [\|S_{t^*+1}^Y |_{S_{0:t^*}^Y=A} - \tilde{S} |_{S_{0:t^*}^Y=A}\|_{tvd}] \geq \Delta_{t^*+1} - \Delta_{t^*}$ . Once we have the claim, we show how to get a space  $T \cdot s$  protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD with advantage  $\Delta_{t^*+1} - \Delta_{t^*}$  concluding the proof of the lemma.

**Claim 5.13.**  $\|S_{t^*+1}^Y - \tilde{S}\|_{tvd} \geq \Delta_{t^*+1} - \Delta_{t^*}$ .

*Proof.* First, by triangle inequality for the total variation distance, we have

$$\|S_{t^*+1}^Y - \tilde{S}\|_{tvd} \geq \|S_{t^*+1}^Y - S_{t^*+1}^N\|_{tvd} - \|\tilde{S} - S_{t^*+1}^N\|_{tvd}.$$

Recall that  $\tilde{S} = \Gamma(S_{t^*}^Y, \boldsymbol{\sigma}_N^{(t^*+1)})$  and  $S_{t^*+1}^N = \Gamma(S_{t^*}^N, \boldsymbol{\sigma}_N^{(t^*+1)})$ . Also, note that  $\boldsymbol{\sigma}_N^{(t^*+1)}$  follows the product distribution  $(\mathcal{D}_f \times \text{Bern}(q^{-k}))^{\alpha n}$  and in particular is independent of  $S_{t^*}^Y$  and  $S_{t^*}^N$ . (This is where we rely crucially on the property  $\mathcal{D}_N = \mathcal{D}_f \times \text{Unif}([q]^k)$ .) Furthermore  $\Gamma$  is a deterministic function, and so we can apply the data processing inequality (Item (2) of [Proposition 3.4](#) with  $X = S_{t^*}^Y$ ,  $Y = S_{t^*}^N$ ,  $W = \boldsymbol{\sigma}_N^{(t^*+1)}$ , and  $f = \Gamma$ ) to conclude

$$\|\tilde{S} - S_{t^*+1}^N\|_{tvd} = \|\Gamma(S_{t^*}^Y, \boldsymbol{\sigma}_N^{(t^*+1)}) - \Gamma(S_{t^*}^N, \boldsymbol{\sigma}_N^{(t^*+1)})\|_{tvd} \leq \|S_{t^*}^Y - S_{t^*}^N\|_{tvd}.$$

Combining the two inequalities above we get

$$\|S_{t^*+1}^Y - \tilde{S}\|_{tvd} \geq \|S_{t^*+1}^Y - S_{t^*+1}^N\|_{tvd} - \|S_{t^*}^Y - S_{t^*}^N\|_{tvd} = \Delta_{t^*+1} - \Delta_{t^*}$$

as desired. □

We now show how a protocol can be designed for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD that achieves advantage at least  $\theta = \mathbb{E}_{A \sim_d S_{0:t^*}^Y} [\|S_{t^*+1}^Y |_{S_{0:t^*}^Y=A} - \tilde{S} |_{S_{0:t^*}^Y=A}\|_{tvd}] \geq \Delta_{t^*+1} - \Delta_{t^*}$  concluding the proof of the lemma. The protocol uses the distinguisher  $T_A : \{0, 1\}^s \rightarrow \{0, 1\}$  such that  $\mathbb{E}_{A, S_{t^*+1}^Y, \tilde{S}} [T_A(S_{t^*+1}^Y)] - \mathbb{E}[T_A(\tilde{S})] \geq \theta$  which is guaranteed to exist by the definition of total variation distance.

Our protocol works as follows: Let Alice receive input  $\mathbf{x}^*$  and Bob receive inputs  $(M, \mathbf{z})$  sampled from either  $\mathcal{Y}_{\text{SD}} |_{\mathbf{x}^*}$  or  $\mathcal{N}_{\text{SD}} |_{\mathbf{x}^*}$  where  $\mathcal{Y}_{\text{SD}}$  and  $\mathcal{N}_{\text{SD}}$  are the Yes and No distribution of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD respectively.

1. Alice samples  $(\boldsymbol{\sigma}^{(0)}, \boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(T)}) \sim \mathcal{Y} |_{\mathbf{x}^*}$  and computes  $A = S_{0:t^*}^Y \in \{0, 1\}^{(t^*+1)s}$  and sends  $A$  to Bob.
2. Bob extracts  $S_{t^*}^Y$  from  $A$ , computes  $\hat{S} = \Gamma(S_{t^*}^Y, \boldsymbol{\sigma})$ , where  $\boldsymbol{\sigma}$  is the encoding of  $(M, \mathbf{z})$  as a stream, and outputs **YES** if  $T_A(\hat{S}) = 1$  and **NO** otherwise.

Note that if  $(M, \mathbf{z}) \sim \mathcal{Y}_{\text{SD}} |_{\mathbf{x}^*}$  then  $\hat{S} \sim_d S_{t^*+1}^Y |_{S_{0:t^*}^Y=A}$  while if  $(M, \mathbf{z}) \sim \mathcal{N}_{\text{SD}} |_{\mathbf{x}^*}$  then  $\hat{S} \sim \tilde{S} |_{S_{0:t^*}^Y=A}$ . It follows that the advantage of the protocol above exactly equals  $\mathbb{E}_A [T_A(S_{t^*+1}^Y)] - \mathbb{E}_A [T_A(\tilde{S})] \geq \theta \geq \Delta_{t^*+1} - \Delta_{t^*} \geq \Delta/T$ . This concludes the proof of the lemma. □

By combining [Lemma 5.12](#) with [Theorem 5.4](#), we immediately have the following consequence.

**Lemma 5.14.** *For  $k \in \mathbb{N}$  let  $\alpha_0(k)$  be as given by [Theorem 5.4](#). Let  $T \in \mathbb{N}$ ,  $\alpha \in (0, \alpha_0(k))$ ,  $\tau \in [0, 1)$ , and  $\mathcal{D}_0, \mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$  where  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  are one-wise independent distributions with  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$ . Then every streaming algorithm **ALG** solving  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD in the insertion-only setting with advantage  $1/8$  for all lengths  $n$  uses space  $\Omega(\sqrt{n})$ .*



*Proof.* Let **ALG** be an algorithm using space  $s$  solving  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD with advantage  $1/8$ . For  $g \in \mathcal{F}$ , let  $p_g = \Pr_{(f, \sigma) \sim \mathcal{D}_Y}[f = g]$  and let  $\mathcal{D}_f$  be the distribution given by  $\mathcal{D}_f(g) = p_g$ . Let  $\mathcal{D}_M = \mathcal{D}_f \times \text{Unif}([q]^k)$ . Note that  $\mathcal{D}_M$  is uniform on the variables and satisfies  $\mu(\mathcal{D}_M) = \mu(\mathcal{D}_Y) = \mu(\mathcal{D}_N)$ . Then by the triangle inequality **ALG** solves either the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_M, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD with advantage  $1/16$  or it solves the  $(\mathcal{F}, \mathcal{D}_N, \mathcal{D}_M, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD with advantage  $1/16$ . Assume without loss of generality it is the former. Then by [Lemma 5.12](#), there exists a one-way protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_M)$ -SD using at most  $sT$  bits of communication with advantage at least  $1/(16T)$ . Applying [Theorem 5.4](#) with  $\delta = 1/(16T) > 0$ , we now get that  $s = \Omega(\sqrt{n})$ .  $\square$

### 5.2.4 Proof of the insertion-only lower bound

We are now ready to prove [Theorem 2.16](#).

*Proof of Theorem 2.16.* We combine [Theorem 5.4](#), [Lemma 5.14](#) and [Lemma 5.8](#). So in particular we set our parameters  $\alpha$  and  $T$  so that the conditions of these statements are satisfied. Specifically  $k$  and  $\varepsilon > 0$ , let  $\alpha_0^{(1)}$  be the constant from [Theorem 5.4](#) and let  $\alpha_0^{(2)}$  be the constant from [Lemma 5.8](#). Let  $\alpha_0 = \min\{\alpha_0^{(1)}, \alpha_0^{(2)}\}$ . Given  $\alpha \in (0, \alpha_0)$  let  $T_0$  be the constant from [Lemma 5.8](#) and let  $T = T_0$ . (Note that these choices allow for both [Theorem 5.4](#) and [Lemma 5.8](#) to hold.) Suppose there exists a streaming algorithm **ALG** that solves  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ). Let  $\tau \in [0, 1)$  and  $\mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0$  be distributions such that (i)  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  are one-wise independent, (ii)  $\tau\mathcal{D}_0 + (1 - \tau)\mathcal{D}_Y \in S_\gamma^Y(\mathcal{F})$ , and (iii)  $\tau\mathcal{D}_0 + (1 - \tau)\mathcal{D}_N \in S_\beta^N(\mathcal{F})$ . Let  $n$  be sufficiently large and let  $\mathcal{Y}_{\text{stream}, n}$  and  $\mathcal{N}_{\text{stream}, n}$  denote the distributions of **YES** and **NO** instances of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau)$ -padded-streaming-SD of length  $n$ . Since  $\alpha$  and  $T$  satisfy the conditions of [Lemma 5.8](#), we have for every sufficiently large  $n$

$$\Pr_{\sigma \sim \mathcal{Y}_{\text{stream}, n}} [\text{val}_\Psi(\sigma) < (\gamma - \varepsilon)] = o(1) \quad \text{and} \quad \Pr_{\sigma \sim \mathcal{N}_{\text{stream}, n}} [\text{val}_\Psi(\sigma) > (\beta + \varepsilon)] = o(1).$$

We conclude that **ALG** can distinguish **YES** instances of Max-CSP( $\mathcal{F}$ ) from **NO** instances with advantage at least  $1/4 - o(1) \geq 1/8$ . However, since  $\mathcal{D}_Y, \mathcal{D}_N$  and  $\alpha$  satisfy the conditions of [Lemma 5.14](#) (in particular  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  are one-wise independent and  $\alpha \in (0, \alpha_0(k))$ ) such an algorithm requires space at least  $\Omega(\sqrt{n})$ . Thus, we conclude that any streaming algorithm that solves  $(\gamma - \varepsilon, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$  space.

Finally, note that if  $\gamma = 1$  then in [Lemma 5.8](#), we have  $\text{val}_\Psi = 1$  with probability one. Repeating the above reasoning with this information, shows that  $(1, \beta + \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\sqrt{n})$ -space.  $\square$

### 5.3 The dynamic lower bound

As in [\[CGSV21\]](#) we use the work of Ai, Hu, Li and Woodruff [\[AHLW16\]](#) (building up on Li, Nguyen and Woodruff [\[LNW14\]](#)) to use the hardness of SD to show the hardness of Max-CSP( $\mathcal{F}$ ) in the general case, in the dynamic streaming setting. To do so, we introduce a family of  $T$ -player simultaneous communication games, which we call  $(\mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD, arising from applying their framework to our problems. (See [Definition 5.16](#).) We then show a simple reduction from  $(\mathcal{D}_Y, \mathcal{D}_N)$ -SD to  $(\mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD. Combining this reduction with our

lower bounds on SD and the reduction from  $(\mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD to streaming complexity leads to the proof of [Theorem 5.1](#).

### 5.3.1 The framework of [\[AHLW16\]](#)

The work [\[AHLW16\]](#) considers problems and algorithms in the “strict turnstile” model which we define below. (The text below is reproduced from [\[CGSV21\]](#).)

In the turnstile streaming model, a problem is given by a pair of disjoint sets  $(\mathcal{A}, \mathcal{B})$ ,  $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}^N$ . An input to the problem is a vector  $\mathbf{x} \in \mathbb{Z}^N$  represented as a data stream  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_L)$  such that  $\mathbf{x} = \sum_{i=1}^L \sigma_i$  where  $\sigma_i \in \{\pm \mathbf{e}_1, \dots, \pm \mathbf{e}_N\} \subseteq \mathbb{Z}^N$  and  $\mathbf{e}_i$  denotes the indicator vector for the  $i$ -th coordinate. The goal of the problem is to distinguish  $\mathbf{x} \in \mathcal{A}$  from  $\mathbf{x} \in \mathcal{B}$ . We say that an algorithm **ALG** solves the problem  $(\mathcal{A}, \mathcal{B})$  with advantage  $\Delta$  if  $\min_{a \in \mathcal{A}} \Pr[\mathbf{ALG}(a) = 1] - \max_{b \in \mathcal{B}} \Pr[\mathbf{ALG}(b) = 1] \geq \Delta$ .

In the *strict* turnstile streaming model, we require the underlying vector to be non-negative at all points of the input data stream, i.e., for every  $i \in [L]$ ,  $\sum_{i' \leq i} \sigma_{i'} \geq 0^N$  (where the final inequality is interpreted coordinatewise). For every  $m \in \mathbb{N}$ , let  $\Lambda_m^* = \{\boldsymbol{\sigma} \mid \forall i \in \mathbb{N}, \sum_{i' \leq i} \sigma_{i'} \geq 0 \text{ \& } \|\sum_i \sigma_i\|_\infty \leq m\}$  be the set of strict turnstile data stream with magnitude at most  $m$ . The results of [\[AHLW16\]](#) hold even when inputs are restricted to the  $\Lambda_m^*$ .

To state their result we need one more notion — which of a *linear sketching algorithm*. We say that a positive vector  $q = (q_1, \dots, q_r) \in \mathbb{Z}^r$  is an  $s$ -bit vector if  $\sum_{i=1}^r \log_2 q_i \leq s$ . We say that an algorithm **ALG** is a space  $s$  randomized linear sketching algorithm if there exists an  $s$ -bit integer vector  $q \in \mathbb{Z}^r$  and a distribution over matrices  $A \in \mathbb{Z}^{r \times n}$ , and a function  $P : \mathbb{Z}^r \rightarrow \{0, 1\}$  such that for all  $\mathbf{x} \in \mathbb{Z}^n$  and for all streams  $\boldsymbol{\sigma}$  representing  $\mathbf{x}$  we have  $\mathbf{ALG}(\mathbf{x}) \sim_d P(A\mathbf{x} \pmod{q})$ . In other words **ALG** picks a random  $A$  and then produces its output by maintaining  $A \cdot (\sum_{i' \leq i} \sigma_{i'} \pmod{q})$  and outputting  $P(A \cdot \mathbf{x} \pmod{q})$  at the end.

Now, we are able to restate [Theorem 4.1](#) in [\[AHLW16\]](#) as follows.

**Theorem 5.15** ([\[AHLW16\]](#), [Theorem 4.1](#)). *Suppose there is a 1-pass randomized streaming algorithm that solves a problem  $(\mathcal{A}, \mathcal{B})$  with  $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}^N$  on  $\Lambda_m^*$  with space complexity  $s$  only depending on  $N$  and with advantage at least  $1 - \delta$ . Then there exists a space  $O(s)$  linear sketching algorithm that solves  $(\mathcal{A}, \mathcal{B})$  on  $\Lambda_m^*$  with advantage at least  $1 - 12\delta$ .*

### 5.3.2 $T$ -Player Simultaneous Version of SD

In this section, we consider the complexity of  $T$ -player number-in-hand simultaneous message passing communication games (abbrev.  $T$ -player simultaneous communication games). Such games are described by two distributions  $\mathcal{Y}$  and  $\mathcal{N}$ . An instance of the game is a  $T$ -tuple  $(X^{(1)}, \dots, X^{(T)})$  either drawn from  $\mathcal{Y}$  or from  $\mathcal{N}$  and  $X^{(t)}$  is given as input to the  $t$ -th player. A (simultaneous communication) protocol  $\Pi = (\Pi^{(1)}, \dots, \Pi^{(T)}, \Pi_{\text{ref}})$  is a  $(T + 1)$ -tuple of functions with  $\Pi^{(t)}(X^{(t)}) \in \{0, 1\}^c$  denoting the  $t$ -th player’s message to the referee, and  $\Pi_{\text{ref}}(\Pi^{(1)}(X^{(1)}), \dots, \Pi^{(T)}(X^{(T)})) \in \{\mathbf{YES}, \mathbf{NO}\}$  denoting the protocol’s output. We denote this output by  $\Pi(X^{(1)}, \dots, X^{(T)})$ . The complexity of this protocol is the parameter  $c$  specifying the maximum length of  $\Pi^{(1)}(X^{(1)}), \dots, \Pi^{(T)}(X^{(T)})$  (maximized over all  $X$ ). The advantage of the protocol  $\Pi$  is the quantity

$$\left| \Pr_{(X^{(1)}, \dots, X^{(T)}) \sim \mathcal{Y}} [\Pi(X^{(1)}, \dots, X^{(T)}) = \mathbf{YES}] - \Pr_{(X^{(1)}, \dots, X^{(T)}) \sim \mathcal{N}} [\Pi(X^{(1)}, \dots, X^{(T)}) = \mathbf{YES}] \right|.$$

**Definition 5.16** ( $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD). For  $k, T \in \mathbb{N}$ ,  $\alpha \in (0, 1/k]$ , a finite set  $\mathcal{F}$ , distributions  $\mathcal{D}_Y, \mathcal{D}_N$  over  $\mathcal{F} \times [q]^k$ , the  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD is a  $T$ -player communication game given by a family of instances  $(\mathcal{Y}_{\text{simul}, n}, \mathcal{N}_{\text{simul}, n})_{n \in \mathbb{N}, n \geq 1/\alpha}$  where for a given  $n$ ,  $\mathcal{Y} = \mathcal{Y}_{\text{simul}, n}$  and  $\mathcal{N} = \mathcal{N}_{\text{simul}, n}$  are as follows: Both  $\mathcal{Y}$  and  $\mathcal{N}$  are supported on tuples  $(\mathbf{x}^*, M^{(1)}, \dots, M^{(T)}, \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(T)})$  where  $\mathbf{x}^* \in [q]^n$ ,  $M^{(t)} \in \{0, 1\}^{kn \times n}$ , and  $\mathbf{z}^{(t)} \in (\mathcal{F} \times \{0, 1\})^{kn}$ , where the pair  $(M^{(t)}, \mathbf{z}^{(t)})$  are the  $t$ -th player's inputs for all  $t \in [T]$ . We now specify the distributions of  $\mathbf{x}^*$ ,  $M^{(t)}$ , and  $\mathbf{z}^{(t)}$  in  $\mathcal{Y}$  and  $\mathcal{N}$ :

- In both  $\mathcal{Y}$  and  $\mathcal{N}$ ,  $\mathbf{x}^*$  is distributed uniformly over  $[q]^n$ .
- In both  $\mathcal{Y}$  and  $\mathcal{N}$  the matrix  $M^{(t)} \in \{0, 1\}^{kn \times n}$  is chosen uniformly (and independently of  $\mathbf{x}^*$ ) among matrices with exactly one 1 per row and at most one 1 per column.
- The vector  $\mathbf{z}^{(t)}$  is determined from  $M^{(t)}$  and  $\mathbf{x}^*$  as follows. Sample a random vector  $\mathbf{b}^{(t)} \in (\mathcal{F} \times [q]^k)^{kn}$  whose distribution differs in  $\mathcal{Y}$  and  $\mathcal{N}$ . Specifically, let  $\mathbf{b}^{(t)} = (\mathbf{b}^{(t)}(1), \dots, \mathbf{b}^{(t)}(kn))$  be sampled from one of the following distributions (independent of  $\mathbf{x}^*$  and  $M$ ):
  - $\mathcal{Y}$ : Each  $\mathbf{b}^{(t)}(i) = (f_i, \tilde{\mathbf{b}}(i)) \in \mathcal{F} \times [q]^k$  is sampled independently according to  $\mathcal{D}_Y$ .
  - $\mathcal{N}$ : Each  $\mathbf{b}^{(t)}(i) = (f_i, \tilde{\mathbf{b}}(i)) \in \mathcal{F} \times [q]^k$  is sampled independently according to  $\mathcal{D}_N$ .

We now set  $\mathbf{z}^{(t)} = (f_i, \tilde{z}_i)$  where  $\tilde{z}_i = 1$  iff  $(M^{(t)} \mathbf{x}^*)_i = \tilde{\mathbf{b}}^{(t)}(i)$ .

If  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ , then given an instance  $\sigma = (\mathbf{x}^*, M^{(1)}, \dots, M^{(T)}, \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(T)})$ , we will let  $\Psi(\sigma)$  represent the associated instance of  $\text{Max-CSP}(\mathcal{F})$  as described in [Section 5.2.2](#).

Note that the instance  $\Psi(\sigma)$  obtained in the **YES** and **NO** cases of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD are distributed exactly according to instances derived in the **YES** and **NO** cases of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T, \mathcal{D}_0, \tau = 0)$ -padded-streaming-SD and thus [Lemma 5.8](#) can still be applied to conclude that **YES** instances usually satisfy  $\text{val}_{\Psi(\sigma)} \geq \gamma - o(1)$  and **NO** instances usually satisfy  $\text{val}_{\Psi(\sigma)} \leq \beta - o(1)$ . We will use this property when proving [Theorem 5.1](#).

We start by showing the simultaneous-SD problems above do not have low-communication protocols when the marginals of  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  match.

**Lemma 5.17.** Let  $\mathcal{F}$  be a finite set,  $k, q, T \in \mathbb{N}$ ,  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$ , and let  $\alpha \in (0, 1/k]$ . Suppose there is a protocol  $\Pi$  that solves  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD on instances of length  $n$  with advantage  $\Delta$  and space  $s$ , then there is a one-way protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD on instances of length  $n$  using at most  $s(T - 1)$  bits of communication and achieving advantage at least  $\Delta/T$ .

*Proof.* Let us first fix the randomness in  $\Pi$  so that it becomes a deterministic protocol. Note that by an averaging argument the advantage of  $\Pi$  does not decrease. Recall that  $\mathcal{Y}$  and  $\mathcal{N}$  are Yes and No input distribution of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD and we have

$$\Pr_{X \sim \mathcal{Y}}[\Pi(X) = \mathbf{YES}] - \Pr_{X \sim \mathcal{N}}[\Pi(X) = \mathbf{YES}] \geq \Delta.$$

Now, we define the following distributions  $\mathcal{D}_0, \dots, \mathcal{D}_T$ . Let  $\mathcal{D}_0 = \mathcal{Y}$  and  $\mathcal{D}_T = \mathcal{N}$ . For each  $t \in [T - 1]$ , we define  $\mathcal{D}_t$  to be the distribution of input instances of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD by sampling  $\mathbf{b}^{(t')}(i)$  independently according to  $\mathcal{D}_Y$  (resp.  $\mathcal{D}_N$ ) for all  $t' \leq t$  (resp.  $t' > t$ ) and  $i$  (see [Definition 5.16](#) to recall the definition). Next, for each  $t \in [T]$ , let

$$\Delta_t = \Pr_{X \sim \mathcal{D}_t}[\Pi(X) = \mathbf{YES}] - \Pr_{X \sim \mathcal{D}_{t-1}}[\Pi(X) = \mathbf{YES}].$$

Observe that  $\sum_{t \in [T]} \Delta_t = \Delta$  and hence there exists  $t^* \in [T]$  such that  $\Delta_{t^*} \geq \Delta/T$ .

Now, we describe a protocol  $\Pi'$  for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD as follows. On input  $(\mathbf{x}^*, M, \mathbf{z})$ , Alice receives  $\mathbf{x}^*$  and Bob receives  $(M, \mathbf{z})$ . Alice first samples matrices  $M^{(1)}, \dots, M^{(t^*-1)}, M^{(t^*+1)}, \dots, M^{(T)}$  as the second item in [Definition 5.16](#). Next, Alice samples  $\mathbf{b}^{(t')}(i) = (f_i, \tilde{\mathbf{b}}^{(t')}(i))$  according to  $\mathcal{D}_Y$  (resp.  $\mathcal{D}_N$ ) for all  $t' < t^*$  (resp.  $t' > t^*$ ) and  $i \in [\alpha n T]$  and sets  $\mathbf{z}^{(t')}(i) = (f_i, \tilde{z}_i)$  as the third item in [Definition 5.16](#). Note that this is doable for Alice because she possesses  $\mathbf{x}^*$ . Finally, Alice sends  $\{\Pi^{(t')}(M^{(t')}, \mathbf{z}^{(t')})\}_{t' \in [T] \setminus \{t^*\}}$  to Bob. After receiving Alice's message  $(X^{(1)}, \dots, X^{(t^*-1)}, X^{(t^*+1)}, \dots, X^{(T)})$ , Bob computes  $\Pi^{(t^*)}(M, \mathbf{z})$  and outputs  $\Pi'(M, \mathbf{z}) = \Pi_{\text{ref}}(X^{(1)}, \dots, X^{(t^*-1)}, \Pi^{(t^*)}(M, \mathbf{z}), X^{(t^*+1)}, \dots, X^{(T)})$ .

It is clear from the construction that the protocol  $\Pi'$  uses at most  $s(T-1)$  bits of communication. To see  $\Pi'$  has advantage at least  $\Delta/T$ , note that if  $(\mathbf{x}^*, M, \mathbf{z})$  is sampled from the Yes distribution  $\mathcal{Y}_{\text{SD}}$  of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD, then  $((M^{(1)}, \mathbf{z}^{(1)}), \dots, (M^{(t^*-1)}, \mathbf{z}^{(t^*-1)}), (M, \mathbf{z}), (M^{(t^*+1)}, \mathbf{z}^{(t^*+1)}), \dots, (M^{(T)}, \mathbf{z}^{(T)}))$  follows the distribution  $\mathcal{D}_{t^*}$ . Similarly, if  $(\mathbf{x}^*, M, \mathbf{z})$  is sampled from the No distribution  $\mathcal{N}_{\text{SD}}$  of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD, then  $((M^{(1)}, \mathbf{z}^{(1)}), \dots, (M^{(t^*-1)}, \mathbf{z}^{(t^*-1)}), (M, \mathbf{z}), (M^{(t^*+1)}, \mathbf{z}^{(t^*+1)}), \dots, (M^{(T)}, \mathbf{z}^{(T)}))$  follows the distribution  $\mathcal{D}_{t^*-1}$ . Thus, the advantage of  $\Pi'$  is at least

$$\begin{aligned} & \Pr_{(M, \mathbf{z}) \sim \mathcal{Y}_{\text{SD}}, \Pi'} [\Pi'(M, \mathbf{z}) = \mathbf{YES}] - \Pr_{(M, \mathbf{z}) \sim \mathcal{N}_{\text{SD}}, \Pi'} [\Pi'(M, \mathbf{z}) = \mathbf{YES}] \\ &= \Pr_{X \sim \mathcal{D}_{t^*}} [\Pi(X) = \mathbf{YES}] - \Pr_{X \sim \mathcal{D}_{t^*-1}} [\Pi(X) = \mathbf{YES}] = \Delta_{t^*} \geq \Delta/T. \end{aligned}$$

We conclude that there is a one-way protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD using at most  $s(T-1)$  bits of communication achieving advantage at least  $\Delta/T$ .  $\square$

As an immediate consequence of [Theorem 5.4](#) and [Lemma 5.17](#) we get that  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD requires  $\Omega(\sqrt{n})$  bits of communication when the marginals of  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  match.

**Lemma 5.18.** *For every  $k, q \in \mathbb{N}$ , there exists  $\alpha_0 > 0$  such that for every  $\alpha \in (0, \alpha_0)$  and  $\delta > 0$  the following holds: For every finite set  $\mathcal{F}$  and  $T \in \mathbb{N}$  and every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$  with  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$ , there exists  $\tau > 0$  and  $n_0$  such that for every  $n \geq n_0$ , every protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD achieving advantage  $\delta$  on instances of length  $n$  requires  $\tau\sqrt{n}$  bits of communication.*

We are now ready to prove [Theorem 5.1](#).

### 5.3.3 Proof of [Theorem 5.1](#)

*Proof of [Theorem 5.1](#).* The proof is a straightforward combination of [Theorem 5.15](#), [Lemma 5.8](#), and [Lemma 5.18](#) and so we pick parameters so that all these are applicable. Given  $\varepsilon$  and  $k$ , let  $\alpha_0^{(1)}$  be as given by [Lemma 5.8](#) and let  $\alpha_0^{(2)}$  be as given by [Lemma 5.18](#). Let  $\alpha = \min\{\alpha_0^{(1)}, \alpha_0^{(2)}\}$ . Given this choice of  $\alpha$ , let  $T_0$  be as given by [Lemma 5.8](#). We set  $T = T_0$  below. Let  $n$  be sufficiently large.

Throughout this proof we will be considering integer weighted instances of  $\text{Max-CSP}(\mathcal{F})$  on  $n$  variables with constraints. Note that such an instance  $\Psi$  can be viewed as a vector in  $\mathbb{Z}^N$  where  $N = O(|\mathcal{F}| \times n^k)$  represents the number of possibly distinct constraints applications on  $n$  variables. Let  $\Gamma = \{\Psi \mid \text{val}_\Psi \geq \gamma - \varepsilon\}$ . Let  $B = \{\Psi \mid \text{val}_\Psi \leq \beta + \varepsilon\}$ . Suppose there exists a dynamic streaming algorithm  $\mathbf{ALG}_1$  in the strict turnstile model that solves  $(\gamma - \varepsilon, \beta + \varepsilon)$ - $\text{Max-CSP}(\mathcal{F})$  using at most  $s(n)$  bits of space. Note that  $\mathbf{ALG}_1$  must achieve advantage at least  $1/3$  on the problem  $(\Gamma, B)$ .

By running several independent copies of **ALG**<sub>1</sub> and thresholding appropriately, we can get an algorithm **ALG**<sub>2</sub> with space  $O(s)$  and advantage  $1 - \frac{1}{100}$  solving  $(\Gamma, B)$ .

Now, by [Theorem 5.15](#) we get that there exists a linear sketching algorithm **ALG**<sub>3</sub> to solve to solve the  $(\Gamma, B)$  distinguishing problem with advantage at least  $1 - 12/100$ . Let  $q$ ,  $A$  and  $P$  be as given by this linear sketching algorithm. We use these to design a protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD as follows.

Let  $(M^{(t)}, \mathbf{z}^{(t)})$  denote the input to the  $t$ -th player in  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD. Each player turn his/her inputs into  $\Psi^{(t)} = (C_1^{(t)}, \dots, C_{m_t}^{(t)})$  where  $C_i^{(t)}$  corresponds to the constraint  $(\mathbf{j}^{(t)}(i), f_i^{(t)})$  with  $\mathbf{j}_i^{(t)} \in [n]^k$  the indicator vector for the  $i$ -th hyperedge of  $M^{(t)}$ . Next, the players use the shared randomness to sample a linear sketch matrix  $A$  and send  $A \sum_i C_i^{(t)} \pmod{q}$  to the referee. Finally, the referee outputs  $P(\sum_t A \sum_i C_i^{(t)} \pmod{q})$ .

To analyze the above, note that the communication is  $O(s)$ . Next, by the advantage of the linear sketching algorithm, we have that

$$\min_{\Psi \in \Gamma} [\mathbf{ALG}_3(\Psi) = 1] - \max_{\Psi \in B} [\mathbf{ALG}_3(\Psi) = 1] \geq 1 - 12/100. \quad (5.19)$$

Now we consider what happens when  $\Psi \sim \mathcal{Y}_{\text{simul}, n}$  and  $\Psi \sim \mathcal{N}_{\text{simul}, n}$ . By [Lemma 5.8](#) we have that  $\Pr_{\Psi \sim \mathcal{Y}_{\text{simul}, n}}[\Psi \in \Gamma] \geq 1 - o(1)$  and  $\Pr_{\Psi \sim \mathcal{N}_{\text{simul}, n}}[\Psi \in B] \geq 1 - o(1)$ . Combining with [Eq. \(5.19\)](#) we thus get

$$\Pr_{\Psi \sim \mathcal{Y}_{\text{simul}, n}} [\mathbf{ALG}_3(\Psi) = 1] - \Pr_{\Psi \sim \mathcal{N}_{\text{simul}, n}} [\mathbf{ALG}_3(\Psi) = 1] \geq 1 - 12/100 - o(1) \geq 1/2,$$

We thus get that there is a  $O(s)$  simultaneous communication protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N, T)$ -simultaneous-SD with at least advantage  $1/2$ .

Now we conclude by applying [Lemma 5.18](#) with  $\delta = 1/2$  to get that  $s = \Omega(\sqrt{n})/T = \Omega(\sqrt{n})$ , thus yielding the theorem.  $\square$

## 6 Hardness of Advice-Signal-Detection with Uniform Marginals

The goal of this section is to prove a variant of [Theorem 5.4](#) that will be used in [Section 7](#) and [Section 8](#) for a proof of the general case of [Theorem 5.4](#). Recall that in the  $(\mathcal{D}_Y, \mathcal{D}_N)$ -SD problem  $|\mathcal{F}| = 1$ , so we omit  $\mathcal{F}$ . The main result of this section, presented in [Theorem 6.4](#), gives an  $\Omega(\sqrt{n})$  lower bound on the communication complexity of  $(\mathcal{D}_Y, \mathcal{D}_N)$ -SD for distributions with matching marginals  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$  for the case when (i) the alphabet is Boolean  $\{-1, 1\}$ <sup>9</sup>, (ii) the marginals are uniform  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N) = 0^k$ , but (iii) both players also receive a specific advice vector  $\mathbf{a}$ . We define the corresponding Advice-SD communication game below.

In order to prove the hardness of Advice-SD, we first recall the Randomized Mask Detection (RMD) communication game [[CGSV21](#)], define its version with advice, Advice-RMD, and prove an  $\Omega(\sqrt{n})$  lower bound on the communication complexity of this game in [Theorem 6.2](#). The proof of the main result of this section, [Theorem 6.4](#), will then follow from the corresponding lower bounds for Advice-RMD in [Theorem 6.2](#).

<sup>9</sup>Throughout this section we use  $\{-1, 1\}$  to denote the Boolean domain.

## 6.1 Hardness of Advice-RMD

In this section we state a theorem that establishes hardness of RMD in the Boolean setting and with uniform marginals while allowing for advice. The proof of this theorem is postponed to [Section 6.3](#). First we define the Advice-RMD one-way communication game.

**Definition 6.1** (Advice-RMD). *Let  $n, k \in \mathbb{N}, \alpha \in (0, 1)$ , where  $k$  and  $\alpha$  are constants with respect to  $n$ , and  $\alpha n$  is an integer less than  $n/k$ . For a pair  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  of distributions over  $\{-1, 1\}^k$ , we consider the following two-player one-way communication problem  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-RMD.*

- The generator samples the following objects:
  1.  $\mathbf{x}^* \sim \text{Unif}(\{-1, 1\}^n)$ .
  2.  $\Gamma \in S_n$  is chosen uniformly among all permutations of  $n$  elements.
  3. We let  $M \in \{0, 1\}^{k\alpha n \times n}$  be a partial permutation matrix capturing  $\Gamma^{-1}(j)$  for  $j \in [k\alpha n]$ . Specifically,  $M_{ij} = 1$  if and only if  $j = \Gamma(i)$ . We view  $M = (M_1, \dots, M_{\alpha n})$  where each  $M_i \in \{0, 1\}^{k \times n}$  is a block of  $k$  successive rows of  $M$ .
  4.  $\mathbf{b} = (\mathbf{b}(1), \dots, \mathbf{b}(\alpha n))$  is sampled from one of the following distributions:
    - (YES) each  $\mathbf{b}(i) \in \{-1, 1\}^k$  is sampled according to  $\mathcal{D}_Y$ .
    - (NO) each  $\mathbf{b}(i) \in \{-1, 1\}^k$  is sampled according to  $\mathcal{D}_N$ .
  5.  $\mathbf{z} = M\mathbf{x}^* \odot \mathbf{b}$ , where  $\odot$  denotes the coordinate-wise product of the elements.
  6. Define a vector  $\mathbf{a} \in [k]^n$  as  $a_j = i$  where  $i = \Gamma^{-1}(j) \pmod{k}$  for every  $j \in [n]$ .
- Alice receives  $\mathbf{x}^*$  and  $\mathbf{a}$  as input.
- Bob receives  $M$ ,  $\mathbf{z}$ , and  $\mathbf{a}$  as input.

We follow the approach of [[GKK<sup>+</sup>09](#), [CGSV21](#)] to prove the following theorem showing a  $\Omega(\sqrt{n})$  communication lower bound for Boolean Advice-RMD. We postpone the proof to [Section 6.3](#).

**Theorem 6.2** (Communication lower bound for Boolean Advice-RMD). *For every  $k \in \mathbb{N}$ , and every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\{-1, 1\}^k)$  with uniform marginals  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N) = 0^k$  there exists  $\alpha_0 > 0$  such that for every  $\alpha \leq \alpha_0$  and  $\delta > 0$  there exists  $\tau > 0$  such that every protocol for  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-RMD achieving advantage  $\delta$  requires  $\tau\sqrt{n}$  bits of communication on instances of length  $n$ .*

## 6.2 Hardness of Advice-SD

Let us first extend the definition of the Signal Detection (SD) problem to the following Advice-SD one-way communication game.

**Definition 6.3** (Advice-SD). *Let  $n, k, q \in \mathbb{N}, \alpha \in (0, 1)$ , where  $k, q$  and  $\alpha$  are constants with respect to  $n$ , and  $\alpha n/k$  is an integer less than  $n$ . For a pair  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  of distributions over  $[q]^k$ , we consider the following two-player one-way communication problem  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-SD.*

- The generator samples the following objects:
  1.  $\mathbf{x}^* \sim \text{Unif}([q]^n)$ .
  2.  $\Gamma \in S_n$  is chosen uniformly among all permutations of  $n$  elements.



3. We let  $M \in \{0, 1\}^{k\alpha n \times n}$  be a partial permutation matrix capturing  $\Gamma^{-1}(j)$  for  $j \in [k\alpha n]$ . Specifically,  $M_{ij} = 1$  if and only if  $j = \Gamma(i)$ . We view  $M = (M_1, \dots, M_{\alpha n})$  where each  $M_i \in \{0, 1\}^{k \times n}$  is a block of  $k$  successive rows of  $M$ .
4.  $\mathbf{b} = (\mathbf{b}(1), \dots, \mathbf{b}(\alpha n))$  is sampled from one of the following distributions:
  - (YES) each  $\mathbf{b}(i) \in [q]^k$  is sampled according to  $\mathcal{D}_Y$ .
  - (NO) each  $\mathbf{b}(i) \in [q]^k$  is sampled according to  $\mathcal{D}_N$ .
5.  $\mathbf{z} = (z_1, \dots, z_{\alpha n}) \in \{0, 1\}^{\alpha n}$  is determined from  $M$ ,  $\mathbf{x}^*$  and  $\mathbf{b}$  as follows. We let  $z_i = 1$  if  $M_i \mathbf{x}^* = \mathbf{b}(i)$ , and  $z_i = 0$  otherwise.
6. Define a vector  $\mathbf{a} \in [k]^n$  as  $a_j = i$  where  $i = \Gamma^{-1}(j) \pmod{k}$  for every  $j \in [n]$ .

- Alice receives  $\mathbf{x}^*$  and  $\mathbf{a}$  as input.
- Bob receives  $M$ ,  $\mathbf{z}$ , and  $\mathbf{a}$  as input.

Almost immediately we get the following corollary for the Advice-SD problem from [Theorem 6.2](#).

**Theorem 6.4** (Communication lower bound for Boolean Advice-SD). *For every  $k \in \mathbb{N}$ , and every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\{-1, 1\}^k)$  with uniform marginals  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N) = 0^k$  there exists  $\alpha_0 > 0$  such that for every  $0 < \alpha \leq \alpha_0$  and  $\delta > 0$  there exists  $\tau > 0$ , such that every protocol for  $(\mathcal{D}_Y, \mathcal{D}_N)$ -advice-SD achieving advantage  $\delta$  requires  $\tau\sqrt{n}$  bits of communication on instances of length  $n$ .*

*Proof.* We show that a protocol achieving advantage  $\delta$  in the  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-SD game with  $s$  bits of communication implies a protocol achieving advantage  $\delta$  for the  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-RMD game with  $s$  bits of communication. Then the lower bounds of [Theorem 6.2](#) for distributions with matching marginals will finish the proof.

Assume that there exists Bob's algorithm  $\mathcal{B}(M, \mathbf{z}, \mathbf{a}, \text{Alice's message})$  that distinguishes  $\mathbf{b}_i \sim \mathcal{D}_Y$  and  $\mathbf{b}_i \sim \mathcal{D}_N$  with advantage  $\delta$  in the Advice-SD game. For the Advice-RMD game, we keep the same algorithm for Alice, and modify Bob's algorithm as follows. Bob receives  $M \in \{0, 1\}^{k\alpha n \times n}$ ,  $\mathbf{z} \in \{-1, 1\}^{k\alpha n}$ ,  $\mathbf{a}$ , and Alice's message, and partitions  $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_{\alpha n})$  where  $\mathbf{z}_i \in \{-1, 1\}^k$ . For each  $i \in [\alpha n]$ , Bob computes  $\tilde{z}_i \in \{0, 1\}$  as follows:  $\tilde{z}_i = 1$  if and only if  $\mathbf{z}_i = 1^k$ . Now Bob sets  $\mathbf{z}' = (\tilde{z}_1, \dots, \tilde{z}_{\alpha n}) \in \{0, 1\}^{\alpha n}$ , and outputs  $\mathcal{B}(M, \mathbf{z}', \mathbf{a}, \text{Alice's message})$ . It is easy to see that in both YES and NO cases, the distribution of the vectors  $\mathbf{z}'$  computed by Bob is the distribution of vectors  $\mathbf{z}$  sampled in the  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-SD game. Thus, the protocol achieves advantage  $\delta$  for the  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-SD game using  $s$  bits of communication as desired.  $\square$

### 6.3 Proof of Theorem 6.2

In order to prove [Theorem 6.2](#) we follow the approach of [[GKK+09](#), [CGSV21](#)] with the main difference in the analysis of a certain combinatorial problem in [Lemma 6.9](#).

Without loss of generality in the following we assume that  $n$  is a multiple of  $k$ . A vector  $\mathbf{a} \in [k]^n$  is called an advice vector if for every  $i \in [k]$ ,  $|\{j : a_j = i\}| = n/k$ . For an advice vector  $\mathbf{a} \in [k]^n$ , we say that a partial permutation matrix  $M \in \{0, 1\}^{k\alpha n \times n}$  of a permutation  $\Gamma$  is  $\mathbf{a}$ -respecting if for every  $i \in [k\alpha n]$  and  $j \in [n]$ ,  $M_{ij} = 1$  if and only if  $a_j = i \pmod{k}$ . Intuitively,  $\mathbf{a}$  is the advice vector that tells you which congruence class  $\Gamma(j)$  lies in.



For each advice vector  $\mathbf{a} \in [k]^n$ , each  $\mathbf{a}$ -respecting partial permutation matrix  $M \in \{0, 1\}^{k\alpha n \times n}$ , distribution  $\mathcal{D}$  over  $\{-1, 1\}^k$ , and a fixed Alice's message, the posterior distribution function  $p_{M, \mathcal{D}, \mathbf{a}} : \{-1, 1\}^{k\alpha n} \rightarrow [0, 1]$  is defined as follows. For each  $\mathbf{z} \in \{-1, 1\}^{k\alpha n}$ , let

$$p_{M, \mathcal{D}, \mathbf{a}}(\mathbf{z}) := \Pr_{\substack{\mathbf{x}^* \in \{-1, 1\}^n \\ \mathbf{b} \sim \mathcal{D}^{\alpha n}}} [\mathbf{z} = (M\mathbf{x}^*) \odot \mathbf{b} \mid M, \mathbf{a}, \text{ Alice's message}] = \mathbb{E}_{\mathbf{x}^* \in A} \mathbb{E}_{\mathbf{b} \sim \mathcal{D}^{\alpha n}} [\mathbf{1}_{\mathbf{z} = (M\mathbf{x}^*) \odot \mathbf{b}}],$$

where  $A \subset \{-1, 1\}^n$  is the set of Alice's inputs that correspond to the message.

Let  $U$  be the uniform distribution over  $\{-1, 1\}^k$ . Lemma 6.4 in [CGSV21] bounds from above the expression  $\mathbb{E}_M, M \text{ is } \mathbf{a}\text{-resp.} [\|p_{M, \mathcal{D}, \mathbf{a}} - \text{Unif}(\{-1, 1\}^{k\alpha n})\|_{\text{tvd}}^2]$  for the case without the advice vector  $\mathbf{a}$  using a combinatorial quantity. An inspection of the proof of Lemma 6.4 in [CGSV21], reveals that an analogous statement holds in the case when the advice vector  $\mathbf{a}$  is present. We provide a proof of this bound for completeness in Lemma 6.5 below.

**Lemma 6.5** (Lemma 6.4 in [CGSV21]). *Let  $\mathbf{a} \in [k]^n$ ,  $A \subseteq \{-1, 1\}^n$ , and  $f : \{-1, 1\}^n \rightarrow \{0, 1\}$  be the indicator function of  $A$ . Let  $k \in \mathbb{N}$  and  $\alpha \in (0, 1/100k)$ . Let  $\mathcal{D}$  be a distribution over  $\{-1, 1\}^k$  such that  $\mathbb{E}_{\mathbf{a} \sim \mathcal{D}}[a_j] = 0$  for all  $j \in [k]$ .*

$$\mathbb{E}_M, M \text{ is } \mathbf{a}\text{-resp.} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2] \leq \frac{2^{2n}}{|A|^2} \sum_{\ell \geq 2} h(\ell) \cdot \sum_{\substack{\mathbf{v} \in \{0, 1\}^n \\ |\mathbf{v}| = \ell}} \widehat{f}(\mathbf{v})^2,$$

where  $U \sim \text{Unif}(\{-1, 1\}^{k\alpha n})$  and for each  $\ell \in [n]$ ,

$$h(\ell) = \max_{\substack{\mathbf{v}_\ell \in \{0, 1\}^n \\ |\mathbf{v}_\ell| = \ell}} \Pr_M, M \text{ is } \mathbf{a}\text{-resp.} [\exists \mathbf{s} \in \{0, 1\}^{k\alpha n} \setminus \{0^{k\alpha n}\}, |\mathbf{s}(i)| \neq 1 \forall i, M^\top \mathbf{s} = \mathbf{v}_\ell].$$

Here for a vector  $\mathbf{s} \in \{0, 1\}^{k\alpha n}$  and integer  $i \in [\alpha n]$ ,  $\mathbf{s}(i) \in \{0, 1\}^k$  denotes the  $i$ -th group of  $k$  coordinates of  $\mathbf{s}$ .

*Proof.* Observe that

$$\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_2^2 = \sum_{\mathbf{s} \in \{0, 1\}^{k\alpha n}} \left( \widehat{p}_{M, \mathcal{D}, \mathbf{a}}(\mathbf{s}) - \widehat{U}(\mathbf{s}) \right)^2 = \sum_{\mathbf{s} \in \{0, 1\}^{k\alpha n} \setminus \{0^{k\alpha n}\}} \widehat{p}_{M, \mathcal{D}, \mathbf{a}}(\mathbf{s})^2.$$

Now by the Cauchy–Schwarz inequality we have that

$$\begin{aligned} \mathbb{E}_M, M \text{ is } \mathbf{a}\text{-resp.} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2] &\leq 2^{2k\alpha n} \mathbb{E}_M, M \text{ is } \mathbf{a}\text{-resp.} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_2^2] \\ &= 2^{2k\alpha n} \mathbb{E}_M, M \text{ is } \mathbf{a}\text{-resp.} \left[ \sum_{\mathbf{s} \in \{0, 1\}^{k\alpha n} \setminus \{0^{k\alpha n}\}} \widehat{p}_{M, \mathcal{D}, \mathbf{a}}(\mathbf{s})^2 \right]. \end{aligned} \quad (6.6)$$

The following claim shows that the Fourier coefficients of the posterior distribution  $p_{M, \mathcal{D}, \mathbf{a}}$  can be bounded from above by a certain Fourier coefficient of the indicator function  $f$ . Let's define  $\text{GOOD} := \{\mathbf{s} \in \{0, 1\}^{k\alpha n} \mid |\mathbf{s}(i)| \neq 1 \forall i\}$ .

**Claim 6.7.**

$$\mathbb{E}_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} [\|p_{M,\mathcal{D},\mathbf{a}} - U\|_{\text{tvd}}^2] \leq \frac{2^{2n}}{|A|^2} \sum_{\mathbf{s} \in \text{GOOD} \setminus \{0^{k\alpha n}\}} \mathbb{E}_M [\widehat{f}(M^\top \mathbf{s})^2].$$

*Proof.* Observe that

$$\widehat{p_{M,\mathcal{D},\mathbf{a}}}(\mathbf{s}) = \frac{1}{2^{k\alpha n}} \sum_{\mathbf{z} \in \{-1,1\}^{k\alpha n}} p_{M,\mathcal{D},\mathbf{a}}(\mathbf{z}) \prod_{\substack{i \in [\alpha n], j \in [k] \\ s(i)_j = 1}} z(i)_j.$$

Recall that  $p_{M,\mathcal{D},\mathbf{a}}(\mathbf{z}) = \mathbb{E}_{\mathbf{x}^* \in A} \mathbb{E}_{\mathbf{b} \sim \mathcal{D}^{\alpha n}} [\mathbf{1}_{\mathbf{z} = M\mathbf{x}^* \odot \mathbf{b}}]$ , the equation becomes

$$= \frac{1}{2^{k\alpha n}} \cdot \mathbb{E}_{\mathbf{x}^* \in A} \left[ \prod_{\substack{i \in [\alpha n], j \in [k] \\ s(i)_j = 1}} (M\mathbf{x}^*)_{i,j} \right] \mathbb{E}_{\mathbf{b} \sim \mathcal{D}^{\alpha n}} \left[ \prod_{\substack{i \in [\alpha n], j \in [k] \\ s(i)_j = 1}} b(i)_j \right].$$

Since  $\mathbb{E}_{\mathbf{a} \sim \mathcal{D}}[a_j] = 0$  for all  $j \in [k]$ , the right most sum is 0 if there exists  $i$  such that  $|\mathbf{s}(i)| = 1$ . This equation becomes

$$\leq \frac{1}{2^{k\alpha n}} \cdot \left| \mathbb{E}_{\mathbf{x}^* \in A} \left[ \prod_{\substack{i \in [\alpha n], j \in [k] \\ s(i)_j = 1}} (M\mathbf{x}^*)_{i,j} \right] \right| \cdot \mathbf{1}_{\mathbf{s} \in \text{GOOD}}.$$

Note that as each row and column of  $M$  has at most 1 non-zero entry, we have

$$= \frac{1}{2^{k\alpha n}} \cdot \left| \mathbb{E}_{\mathbf{x}^* \in A} \left[ \prod_{\substack{i \in [n] \\ (M^\top \mathbf{s})_i = 1}} \mathbf{x}_i^* \right] \right| \cdot \mathbf{1}_{\mathbf{s} \in \text{GOOD}}$$

Now we relate the above quantity to the Fourier coefficients of  $f$ . Recall that  $f$  is the indicator function of the set  $A$  and hence for each  $\mathbf{v} \in \{0,1\}^n$ , we have

$$\widehat{f}(\mathbf{v}) = \frac{1}{2^n} \sum_{\mathbf{x}^*} f(\mathbf{x}^*) \prod_{i \in [n]: v_i = 1} \mathbf{x}_i^* = \frac{1}{2^n} \sum_{\mathbf{x}^* \in A} \prod_{i \in [n]: v_i = 1} \mathbf{x}_i^*.$$

Thus, the Fourier coefficient of  $p_M$  can be bounded as follows.

$$\widehat{p_{M,\mathcal{D},\mathbf{a}}}(\mathbf{s}) \leq \frac{1}{2^{\alpha kn}} \cdot \frac{2^n}{|A|} \widehat{f}(M^\top \mathbf{s}) \cdot \mathbf{1}_{\mathbf{s} \in \text{GOOD}}. \quad (6.8)$$

By plugging Eq. (6.8) into Eq. (6.6), we have the desired bound and complete the proof of Claim 6.7.  $\square$

Next, by [Claim 6.7](#), we have

$$\mathbb{E}_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2] \leq \frac{2^{2n}}{|A|^2} \sum_{\mathbf{s} \in \text{GOOD} \setminus \{0^{\alpha kn}\}} \mathbb{E}_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} \left[ \widehat{f}(M^\top \mathbf{s})^2 \right].$$

Since for a fixed  $M$ , the map  $M^\top$  is injective, the right hand side of the above inequality has the following combinatorial form.

$$= \frac{2^{2n}}{|A|^2} \sum_{\mathbf{v} \in \{0,1\}^n \setminus \{0^n\}} \Pr_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} \left[ \exists \mathbf{s} \in \text{GOOD} \setminus \{0^{\alpha kn}\}, M^\top \mathbf{s} = \mathbf{v} \right] \widehat{f}(\mathbf{v})^2.$$

By symmetry, the above probability term will be the same for  $\mathbf{v}$  and  $\mathbf{v}'$  having the same Hamming weight. Recall that

$$h(\ell) = \max_{\substack{\mathbf{v}_\ell \in \{0,1\}^n \\ |\mathbf{v}_\ell| = \ell}} \Pr_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} \left[ \exists \mathbf{s} \in \text{GOOD} \setminus \{0^{\alpha kn}\}, M^\top \mathbf{s} = \mathbf{v}_\ell \right],$$

this equation becomes

$$\leq \frac{2^{2n}}{|A|^2} \sum_{\ell \geq 1} h(\ell) \cdot \sum_{\substack{\mathbf{v} \in \{0,1\}^n \\ |\mathbf{v}| = \ell}} \widehat{f}(\mathbf{v})^2.$$

Note that for  $\ell = 1$  and every  $\ell > \alpha kn$ ,  $h(\ell) = 0$  by definition. Thus, this expression simplifies to the following.

$$= \frac{2^{2n}}{|A|^2} \sum_{\ell \geq 2} h(\ell) \cdot \sum_{\substack{\mathbf{v} \in \{0,1\}^n \\ |\mathbf{v}| = \ell}} \widehat{f}(\mathbf{v})^2.$$

This completes the proof of [Lemma 6.5](#). □

Now we bound from above the combinatorial quantity  $h(\ell)$  from [Lemma 6.5](#).

**Lemma 6.9.** *For every  $0 < \alpha \in (0, 1/100k^2)$  and  $\ell \in [k\alpha n]$ , we have*

$$h(\ell) = \max_{\substack{\mathbf{v}_\ell \in \{0,1\}^n \\ |\mathbf{v}_\ell| = \ell}} \Pr_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} \left[ \exists \mathbf{s} \neq 0, |\mathbf{s}(i)| \neq 1 \forall i, M^\top \mathbf{s} = \mathbf{v}_\ell \right] \leq \left( \frac{\ell}{n} \right)^{\ell/2} (e^3 \alpha k^5)^{\ell/2}.$$

*Proof.* By symmetry, without loss of generality we can fix the advice vector  $\mathbf{a} = (1^{n/k} 2^{n/k} \dots k^{n/k})$ . For non-negative integers  $\ell_1, \dots, \ell_k$ , we say that  $\mathbf{v}_\ell \in \{0,1\}^n$  is an  $(\ell_1, \dots, \ell_k)$ -vector if for every  $i \in [k]$ ,  $\mathbf{v}$  has exactly  $\ell_i$  entries equal 1 in the  $i$ th group of  $n/k$  coordinates. For fixed values of  $\ell_i$ , let us define

$$h(\ell_1, \dots, \ell_k) = \Pr_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} \left[ \exists \mathbf{s} \neq 0, |\mathbf{s}(i)| \neq 1 \forall i, M^\top \mathbf{s} \text{ is a } (\ell_1, \dots, \ell_k)\text{-vector} \right].$$

We note that

$$h(\ell) = \max_{\substack{\ell_1, \dots, \ell_k \geq 0 \\ \sum_i \ell_i = \ell}} h(\ell_1, \dots, \ell_k). \quad (6.10)$$

An equivalent way to compute the probability  $h(\ell_1, \dots, \ell_k)$  is to fix the matching  $M = \{(i, n/k + i), \dots, (k-1)n/k + i) | i \in [\alpha n]\}$ , and to let  $\mathbf{v}$  be a random  $(\ell_1, \dots, \ell_k)$ -vector. Then

$$h(\ell_1, \dots, \ell_k) = \Pr_{\substack{\mathbf{v} \\ \mathbf{v} \text{ is } (\ell_1, \dots, \ell_k)}} \left[ \exists \mathbf{s} \neq \mathbf{0}, |\mathbf{s}(i)| \neq 1 \forall i, M^\top \mathbf{s} = \mathbf{v} \right] = \frac{|U|}{|V|}, \quad (6.11)$$

where  $V \subseteq \{0, 1\}^n$  is the set of all  $(\ell_1, \dots, \ell_k)$ -vectors, and  $U = \{\mathbf{u} \in V : \exists \mathbf{s} \neq \mathbf{0}, |\mathbf{s}(i)| \neq 1 \forall i, M^\top \mathbf{s} = \mathbf{u}\}$ . From  $\ell_1 + \dots + \ell_k = \ell$ , the number of  $(\ell_1, \dots, \ell_k)$ -vectors is

$$|V| = \prod_{i=1}^k \binom{n/k}{\ell_i} \geq \binom{n/k}{\sum_{i=1}^k \ell_i} = \binom{n/k}{\ell} \geq \left(\frac{n}{k\ell}\right)^\ell, \quad (6.12)$$

where the first inequality uses that  $n/k \geq k\alpha n \geq \ell$  for  $\alpha \leq 1/k^2$ .

For a vector  $\mathbf{s} \in \{0, 1\}^{k\alpha n}$ , let  $T_{\mathbf{s}} = \{i : |\mathbf{s}(i)| > 0\}$  be the set of indices of non-zero blocks of  $\mathbf{s}$ . In order to give an upper bound on the size of  $U$ , first we pick a set  $T_{\mathbf{s}}$ , and then we choose a vector  $\mathbf{u}$  such that  $M^\top \mathbf{s} = \mathbf{u}$  for some  $\mathbf{s}$  corresponding to the set  $T_{\mathbf{s}}$ . Note that since for each  $i \in T$ ,  $\mathbf{s}(i) > 0$  and  $\mathbf{s}(i) \neq 1$  by the definition of  $h(\ell)$ , the size of  $t = |T| \leq k/2$ . For every  $t$ , the number of ways to choose  $T_{\mathbf{s}}$  is  $\binom{\alpha n}{t}$ . For a fixed  $T_{\mathbf{s}}$ , it remains to choose the  $\ell$  coordinates of  $\mathbf{u}$  among at most  $kt$  non-zero coordinates of  $\mathbf{s}$ . For a vector  $\mathbf{s} \in \{0, 1\}^{k\alpha n}$ , let  $T_{\mathbf{s}} = \{i \in [\alpha n] : |\mathbf{s}(i)| > 0\}$  be the set of indices of non-zero blocks of  $\mathbf{s}$ . In order to give an upper bound on the size of  $U$ , first we pick a set  $T$ , and then we choose a vector  $\mathbf{u}$  such that  $M^\top \mathbf{s} = \mathbf{u}$  for some  $\mathbf{s}$  with (i)  $|\mathbf{s}(i)| \neq 1$  for all  $i$  and (ii)  $T_{\mathbf{s}} = T$ . Note that since for each  $i \in T$ ,  $\mathbf{s}(i) > 0$  and  $|\mathbf{s}(i)| \neq 1$ , the size of  $t = |T| \leq \ell/2$ . For every  $t$ , the number of ways to choose  $T$  is  $\binom{\alpha n}{t}$ . For a fixed  $T$ , it remains to choose the  $\ell$  coordinates of  $\mathbf{u}$  among at most  $kt$  non-zero coordinates of  $\mathbf{s}$ . This gives us the following upper bound on the size of  $|U|$ .

$$|U| \leq \max_{t \leq \ell/2} \binom{\alpha n}{t} \binom{kt}{\ell}. \quad (6.13)$$

The second term of the upper bound in Eq. (6.13) can be bounded from above by

$$\binom{kt}{\ell} \leq \left(\frac{ekt}{\ell}\right)^\ell \leq \left(\frac{ek\ell/2}{\ell}\right)^\ell = \left(\frac{ek}{2}\right)^\ell.$$

Now we'll show that the first term of the upper bound in Eq. (6.13) can be bounded from above by  $\left(\frac{2ek\alpha n}{\ell}\right)^{\ell/2}$ . If  $\ell \geq 2\alpha n$ , then

$$\binom{\alpha n}{t} \leq 2^{\alpha n} \leq 2^{\ell/2} \leq \left(\frac{2ek\alpha n}{\ell}\right)^{\ell/2},$$

where in the last inequality we use  $\ell \leq k\alpha n$ . If  $\ell < 2\alpha n$ , then  $t \leq \ell/2 < \alpha n$ , and

$$\binom{\alpha n}{t} \leq \left(\frac{e\alpha n}{t}\right)^t \leq \left(\frac{2e\alpha n}{\ell}\right)^{\ell/2} < \left(\frac{2ek\alpha n}{\ell}\right)^{\ell/2}.$$

The above implies that

$$|U| \leq \max_{t \leq \min\{\alpha n, \ell/2\}} \binom{\alpha n}{t} \binom{kt}{\ell} \leq \left(\frac{ek}{2}\right)^\ell \left(\frac{2ek\alpha n}{\ell}\right)^{\ell/2} \leq \left(\frac{n}{\ell}\right)^{\ell/2} (e^3 \alpha k^3)^{\ell/2}. \quad (6.14)$$

Finally, from Eqs. (6.10) to (6.12) and (6.14),

$$h(\ell) = \max_{\substack{\ell_1, \dots, \ell_k \geq 0 \\ \sum_i \ell_i = \ell}} h(\ell_1, \dots, \ell_k) = \frac{|U|}{|V|} \leq \left(\frac{k\ell}{n}\right)^\ell \cdot \left(\frac{n}{\ell}\right)^{\ell/2} (e^3 \alpha k^3)^{\ell/2} \leq \left(\frac{\ell}{n}\right)^{\ell/2} (e^3 \alpha k^5)^{\ell/2}.$$

□

In Lemma 6.15 below we give the final ingredient needed for the proof of Theorem 6.2. If  $U$  is the uniform distribution over  $\{-1, 1\}^k$ , then we show that for every large set  $A \subseteq \{0, 1\}^n$  of inputs  $x$  corresponding to a fixed Alice's message (and a fixed advice  $\mathbf{a}$ ),  $\mathbb{E}_{M, \text{Mis}} \mathbf{a}\text{-resp.} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2]$  is small.

**Lemma 6.15.** *For every  $k \in \mathbb{N}$  there exists  $\alpha_0 > 0$  such that for every  $0 < \alpha \leq \alpha_0, \delta \in (0, 1)$ , and  $c \leq \frac{\delta\sqrt{n}}{100\sqrt{\alpha k^5}}$  the following holds for all large enough  $n$ . If  $\mathcal{D}$  is a distribution over  $\{-1, 1\}^k$  such that for all  $j \in [k]$ ,  $\mathbb{E}_{\mathbf{a} \sim \mathcal{D}}[a_j] = 0$ , and  $A \subseteq \{0, 1\}^n$  is of size  $|A| \geq 2^{n-c}$ , then*

$$\mathbb{E}_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2] \leq \frac{\delta^2}{16}.$$

where  $U \sim \text{Unif}(\{-1, 1\}^{k\alpha n})$ .

*Proof.* Lemma 6.5 and Lemma 6.9 imply that for every  $A$  of size  $|A| \geq 2^{n-c}$ ,

$$\mathbb{E}_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2] \leq \frac{2^{2n}}{|A|^2} \cdot \sum_{\ell \geq 2} \left(\frac{\ell}{n}\right)^{\ell/2} (e^3 \alpha k^5)^{\ell/2} \sum_{\substack{\mathbf{v} \in \{0, 1\}^n \\ |\mathbf{v}| = \ell}} \widehat{f}(\mathbf{v})^2.$$

For every  $\ell \in [4c]$ , Lemma 3.8 implies that

$$\frac{2^{2n}}{|A|^2} \sum_{\substack{\mathbf{v} \in \{0, 1\}^n \\ |\mathbf{v}| = \ell}} \widehat{f}(\mathbf{v})^2 \leq \left(\frac{4\sqrt{2}c}{\ell}\right)^\ell.$$

By the Parseval identity,  $\sum_{\mathbf{v}} \widehat{f}(\mathbf{v})^2 \leq 1$ . This gives us that

$$\mathbb{E}_{\substack{M \\ \text{Mis } \mathbf{a}\text{-resp.}}} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{\text{tvd}}^2] \leq \sum_{\ell \geq 2} \left(\frac{\ell}{n}\right)^{\ell/2} (e^3 \alpha k^5)^{\ell/2} \cdot \left(\frac{4\sqrt{2}c}{\ell}\right)^\ell + \frac{2^{2n}}{|A|^2} \cdot \max_{4c < \ell \leq k\alpha n} \left\{ \left(\frac{\ell}{n}\right)^{\ell/2} (e^3 \alpha k^5)^{\ell/2} \right\}.$$

Recall that  $c \leq \frac{\delta\sqrt{n}}{100\sqrt{\alpha k^5}}$ . Let  $\alpha_0 = \frac{1}{2e^3 k^5}$ . Then for every  $\alpha \leq \alpha_0$ , the max term on the right hand side is maximized by  $\ell = 4c + 1$  for all large enough  $n$ ,

$$\begin{aligned} &\leq \sum_{\ell \geq 2}^{4c} \left( \frac{32e^3 \alpha k^5 c^2}{n\ell} \right)^{\ell/2} + \left( \frac{8e^3 c \alpha k^5}{n} \right)^{2c} \\ &\leq \sum_{\ell \geq 2}^{4c} \left( \frac{\delta^2}{30} \right)^{\ell/2} + \left( \frac{8e^3 \delta \sqrt{\alpha}}{100\sqrt{k^3} \sqrt{n}} \right)^{2c} \\ &< \frac{\delta^2}{16}. \end{aligned}$$

□

We are ready to finish the proof of [Theorem 6.2](#).

*Proof of Theorem 6.2.* Let us set  $\tau = \frac{\delta}{200\sqrt{\alpha k^5}}$ , and let  $\alpha_0$  be as set in [Lemma 6.15](#). Suppose that there exists a one-way communication protocol for  $(\mathcal{D}_Y, \mathcal{D}_N)$ -Advice-RMD that uses  $s = \tau\sqrt{n}$  bits of communication and has advantage at least  $\delta$ . By the triangle inequality there must exist a protocol with advantage  $\delta/2$  and  $s$  bits of communication for either the  $(\mathcal{D}_Y, \mathcal{D}_{unif})$ -Advice-RMD or the  $(\mathcal{D}_N, \mathcal{D}_{unif})$ -Advice-RMD problem. Without loss of generality, we assume that  $(\mathcal{D}_Y, \mathcal{D}_{unif})$ -Advice-RMD can be solved with advantage  $\delta/2$ . Then,

$$\|p_{M, \mathcal{D}_Y, \mathbf{a}} - p_{M, \mathcal{D}_{unif}, \mathbf{a}}\|_{tvd} \geq \frac{\delta}{2}.$$

Without loss of generality, we can assume that Alice's protocol is deterministic. In other words, for every  $\mathbf{a}$ , Alice's  $s$ -bit communication protocol partitions the set of  $\{-1, 1\}^n$  of inputs  $x$  into  $2^s$  sets  $A_1, \dots, A_{2^s} \subseteq \{-1, 1\}^n$  according to the message sent by Alice. Therefore, at least  $(1 - \delta/4)$ -fraction of inputs  $x \in \{-1, 1\}^n$  belongs to sets  $A_i$  of size  $|A_i| \geq \frac{\delta}{4} \cdot 2^{n-s} \geq 2^{n-c}$  for  $c = s + 1 - \log \delta$ . By [Lemma 6.15](#), for every  $A_i$  of size  $|A_i| \geq 2^{n-c}$ ,

$$\|p_{M, \mathcal{D}_Y, \mathbf{a}} - p_{M, \mathcal{D}_{unif}, \mathbf{a}}\|_{tvd} |_{\mathbf{x}^* \in A_i} = \underset{\text{Mis } \mathbf{a}\text{-resp.}}{\mathbb{E}_M} [\|p_{M, \mathcal{D}, \mathbf{a}} - U\|_{tvd} |_{\mathbf{x}^* \in A_i}] \leq \delta/4.$$

Finally,

$$\begin{aligned} \|p_{M, \mathcal{D}_Y, \mathbf{a}} - p_{M, \mathcal{D}_{unif}, \mathbf{a}}\|_{tvd} &\leq \Pr[x \in A_i : |A_i| < 2^{n-c}] \\ &\quad + \Pr[x \in A_i : |A_i| \geq 2^{n-c}] \cdot \|p_{M, \mathcal{D}_Y, \mathbf{a}} - p_{M, \mathcal{D}_{unif}, \mathbf{a}}\|_{tvd} |_{\mathbf{x}^* \in A_i} \\ &\leq \delta/4 + (1 - \delta/4) \cdot \delta/4 \\ &< \delta/2. \end{aligned}$$

□

## 7 Hardness of Signal Detection

In this section we extend the hardness result of the SD problems for the special distributions described in [Section 6](#) to the fully general setting, thus proving the following theorem.

**Theorem 5.4** (Communication lower bound for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD). *For every  $k, q$ , every finite set  $\mathcal{F}$ , every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$  with  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$  there exists  $\alpha_0 > 0$  such that for every  $0 < \alpha \leq \alpha_0$  and  $\delta > 0$  there exists  $\tau > 0$  such that the following holds: Every protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD achieving advantage  $\delta$  on instances of length  $n$  requires  $\tau\sqrt{n}$  bits of communication.*

The bulk of this section is devoted to proving that for every pair of distributions  $\mathcal{D}_Y$  and  $\mathcal{D}_N$ , we can find a path (a sequence) of intermediate distributions  $\mathcal{D}_Y = \mathcal{D}_0, \mathcal{D}_1, \dots, \mathcal{D}_L = \mathcal{D}_N$  such that adjacent pairs in this sequence are indistinguishable by a “basic” argument, where a basic argument is a combination of an indistinguishability result from [Theorem 7.4](#) and a shifting argument. We note here that the ideas and proofs in this section are based directly on [\[CGSV21, Section 7\]](#). The main change is to move from working with “cubes” (i.e.,  $\{-1, 1\}^k$  for various  $k$ ) to working with general product sets (i.e.,  $[q_1] \times \dots \times [q_k]$  for various  $k, q_1, \dots, q_k$ ).

Our proof comes in the following steps:

1. For every marginal vector  $\boldsymbol{\mu}$ , we identify a *canonical* distribution  $\mathcal{D}_\boldsymbol{\mu}$  that we use as the endpoint of the path. So it suffices to prove that for all  $\mathcal{D}$ ,  $\mathcal{D}$  is indistinguishable from  $\mathcal{D}_{\boldsymbol{\mu}(\mathcal{D})}$ , i.e., there is a path of finite length from  $\mathcal{D}$  to  $\mathcal{D}_{\boldsymbol{\mu}(\mathcal{D})}$ .
2. We give a combinatorial proof that there is a path of finite length (some function of  $k$ ) that takes us from an arbitrary distribution to the canonical one.

Putting these ingredients together, along with a proof that a “basic step” is indistinguishable gives us the final theorem.

Let  $\mathcal{Q} = [q_1] \times \dots \times [q_k]$  where  $\forall i, q_i \in \mathbb{N}$ . We start with the definition of the chain and the canonical distribution. For a distribution  $\mathcal{D} \in \Delta(\mathcal{Q})$ , its support is the set  $\text{supp}(\mathcal{D}) = \{\mathbf{a} \in \mathcal{Q} \mid \mathcal{D}(\mathbf{a}) > 0\}$ . For  $\mathcal{D} \in \mathcal{Q}$ , we define the marginal vector  $\boldsymbol{\mu}(\mathcal{D}) = (\mu_{i,\sigma})_{i \in [k], \sigma \in [q_i]}$  as  $\mu_{i,\sigma} = \Pr_{\mathbf{a} \sim \mathcal{D}}[a_i = \sigma]$ . Next, we consider the following partial order on  $\mathcal{Q}$ . For vectors  $\mathbf{a}, \mathbf{b} \in \mathcal{Q}$  we use the notation  $\mathbf{a} \leq \mathbf{b}$  if  $a_i \leq b_i$  for every  $i \in [k]$ . Further we use  $\mathbf{a} < \mathbf{b}$  if  $\mathbf{a} \leq \mathbf{b}$  and  $\mathbf{a} \neq \mathbf{b}$ .

**Definition 7.1** (Chain). *We refer to a sequence  $\mathbf{a}(0) < \mathbf{a}(1) < \dots < \mathbf{a}(\ell)$ ,  $\mathbf{a}(i) \in \mathcal{Q}$  for every  $i \in \{0, \dots, \ell\}$ , as a chain of length  $\ell$ . Note that chains in  $\mathcal{Q}$  have length at most  $\sum_{i=1}^k (q_i - 1)$ .*

**Lemma 7.2** (Canonical distribution). *Given a vector of marginals  $\boldsymbol{\mu} = (\mu_{i,\sigma})_{i \in [k], \sigma \in [q_i]}$ , there exists a unique distribution  $\mathcal{D}$  with matching marginals ( $\boldsymbol{\mu}(\mathcal{D}) = \boldsymbol{\mu}$ ) such that the support of  $\mathcal{D}$  is a chain. We call this the canonical distribution  $\mathcal{D}_\boldsymbol{\mu}$  associated with  $\boldsymbol{\mu}$ .*

*Proof.* We will prove the proposition by applying induction on  $\sum_{i=1}^k q_i$ . In the base case when  $\sum_{i=1}^k q_i = k$ , there is only one point in the support of the distribution and the claim holds trivially. For  $\sum_{i=1}^k q_i > k$ , define  $h = \arg \min_{i \in [k]} \mu_{i,q_i}$  and  $\tau = \mu_{h,q_h}$ . Let  $\tilde{q}_h = q_h - 1$  and  $\tilde{q}_i = q_i$ , for  $i \neq h$ . Define a vector of marginals  $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_{i,\sigma})_{i \in [k], \sigma \in [\tilde{q}_i]}$  as follows:  $\tilde{\mu}_{i,\sigma} = (\mu_{i,\sigma} - \tau)/(1 - \tau)$  if  $i \neq h$  and  $\sigma = q_i$ , and  $\tilde{\mu}_{i,\sigma} = \mu_{i,\sigma}/(1 - \tau)$  otherwise. By the induction hypothesis, there exists a unique distribution  $\tilde{\mathcal{D}}$  supported on a chain such that  $\boldsymbol{\mu}(\tilde{\mathcal{D}}) = \tilde{\boldsymbol{\mu}}$ . Observe that the distribution  $\mathcal{D} = (1 - \tau)\tilde{\mathcal{D}} + \tau\{(q_1, \dots, q_k)\}$  has marginal  $\boldsymbol{\mu}$  and is supported on a chain. We will now show that  $\mathcal{D}$  is the unique distribution with these properties. For a distribution  $\mathcal{D} \in \Delta([q_1] \times \dots \times [q_k])$  and  $\mathbf{v} \in [q_1] \times \dots \times [q_k]$ , we define  $\mathcal{D}(\mathbf{v}) = \Pr_{\mathbf{c} \sim \mathcal{D}}[\mathbf{c} = \mathbf{v}]$ . Note that it suffices to prove that if  $\mathcal{D}' \in \Delta([q_1] \times \dots \times [q_k])$  is supported on a chain and  $\boldsymbol{\mu}(\mathcal{D}') = \boldsymbol{\mu}$ , then  $\mathcal{D}'(q_1, \dots, q_k) = \tau$ . Clearly  $\mathcal{D}'(q_1, \dots, q_k) \leq \tau$ . Let  $\mathbf{u}$  be lexicographically the largest vector smaller than  $(q_1, \dots, q_k)$  in the



support of  $\mathcal{D}'$ . Let  $r$  be an index where  $\mathbf{u}_r < q_r$ . Since  $\mathcal{D}'$  is supported on a chain,  $\mathcal{D}'(\mathbf{v}) = 0$  for  $\mathbf{v} \in [q_1] \times \cdots \times [q_k]$  such that  $\mathbf{v}_r = q_r$  and  $\mathbf{v} \neq (q_1, \dots, q_k)$ . Hence  $\mu_{r, q_r} = \mathcal{D}'(q_1, \dots, q_k)$ . Since  $\tau = \min_{i \in [k]} \mu_{i, q_i}$ , we have  $\tau \leq \mu_{r, q_r} = \mathcal{D}'(q_1, \dots, q_k)$ .  $\square$

For  $\mathbf{u}, \mathbf{v} \in \mathcal{Q}$ , let  $\mathbf{u}' = \min\{\mathbf{u}, \mathbf{v}\} \triangleq (\min\{u_1, v_1\}, \dots, \min\{u_k, v_k\})$  and let  $\mathbf{v}' = \max\{\mathbf{u}, \mathbf{v}\} \triangleq (\max\{u_1, v_1\}, \dots, \max\{u_k, v_k\})$ . We say  $\mathbf{u}$  and  $\mathbf{v}$  are incomparable if  $\mathbf{u} \not\leq \mathbf{v}$  and  $\mathbf{v} \not\leq \mathbf{u}$ . Note that if  $\mathbf{u}$  and  $\mathbf{v}$  are incomparable then  $\{\mathbf{u}, \mathbf{v}\}$  and  $\{\mathbf{u}', \mathbf{v}'\}$  are disjoint<sup>10</sup>.

**Definition 7.3** (Polarization (update) operator). *Given a distribution  $\mathcal{D} \in \Delta(\mathcal{Q})$  and incomparable elements  $\mathbf{u}, \mathbf{v} \in \mathcal{Q}$ , we define the  $(\mathbf{u}, \mathbf{v})$ -polarization of  $\mathcal{D}$ , denoted  $\mathcal{D}_{\mathbf{u}, \mathbf{v}}$ , to be the distribution as given below. Let  $\varepsilon = \min\{\mathcal{D}(\mathbf{u}), \mathcal{D}(\mathbf{v})\}$ .*

$$\mathcal{D}_{\mathbf{u}, \mathbf{v}}(\mathbf{b}) = \begin{cases} \mathcal{D}(\mathbf{b}) - \varepsilon & , \mathbf{b} \in \{\mathbf{u}, \mathbf{v}\} \\ \mathcal{D}(\mathbf{b}) + \varepsilon & , \mathbf{b} \in \{\mathbf{u}', \mathbf{v}'\} \\ \mathcal{D}(\mathbf{b}) & , \text{otherwise.} \end{cases}$$

We refer to  $\varepsilon(\mathcal{D}, \mathbf{u}, \mathbf{v}) = \min\{\mathcal{D}(\mathbf{u}), \mathcal{D}(\mathbf{v})\}$  as the polarization amount.

It can be verified that the polarization operator preserves the marginals, i.e.,  $\boldsymbol{\mu}(\mathcal{D}) = \boldsymbol{\mu}(\mathcal{D}_{\mathbf{u}, \mathbf{v}})$ . Note also that this operator is non-trivial, i.e.,  $\mathcal{D}_{\mathbf{u}, \mathbf{v}} = \mathcal{D}$ , if  $\{\mathbf{u}, \mathbf{v}\} \not\subseteq \text{supp}(\mathcal{D})$ .

**Theorem 7.4** (Indistinguishability of the polarization step). *Let  $n, k, q \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  where  $k, q, \alpha$  are constants with respect to  $n$  and  $\alpha n$  is an integer less than  $n/k$ . For a distribution  $\mathcal{D} \in \Delta([q]^k)$ , incomparable vectors  $\mathbf{u}, \mathbf{v} \in [q]^k$ , and  $\delta > 0$ , there exists  $\tau > 0$  such that every protocol for  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD achieving advantage  $\delta$  requires  $\tau\sqrt{n}$  bits of communication.*

We defer the proof of this theorem to [Section 8.2](#) and focus instead on the number of steps

## 7.1 Finite upper bound on the number of polarization steps

In this section we prove that there is a finite upper bound on the number of polarization steps needed to move from a distribution  $\mathcal{D} \in \Delta(\mathcal{Q})$  to the canonical distribution with marginal  $\boldsymbol{\mu}(\mathcal{D})$ , i.e.,  $\mathcal{D}_{\boldsymbol{\mu}(\mathcal{D})}$ . Together with the indistinguishability result from [Theorem 7.4](#) this allows us to complete the proof of [Theorem 5.4](#) by going from  $\mathcal{D}_Y$  to  $\mathcal{D}_{\boldsymbol{\mu}(\mathcal{D}_Y)} = \mathcal{D}_{\boldsymbol{\mu}(\mathcal{D}_N)}$  and then to  $\mathcal{D}_N$  by using the triangle inequality for indistinguishability.

In this section we extend our considerations to functions  $A : \mathcal{Q} \rightarrow \mathbb{R}^{\geq 0}$ . Let  $\mathcal{F}(\mathcal{Q}) = \{A : \mathcal{Q} \rightarrow \mathbb{R}^{\geq 0}\}$ . For  $A \in \mathcal{F}(\mathcal{Q})$  and  $i \in [k]$ , let  $\mu_0(A) = \sum_{\mathbf{a} \in \mathcal{Q}} A(\mathbf{a})$ . Note  $\Delta(\mathcal{Q}) \subseteq \mathcal{F}(\mathcal{Q})$  and  $A \in \Delta(\mathcal{Q})$  if and only if  $A \in \mathcal{F}(\mathcal{Q})$  and  $\mu_0(A) = \sum_{\mathbf{a} \in \mathcal{Q}} A(\mathbf{a}) = 1$ . We extend the definition of marginals, support, canonical distribution, and polarization operators to  $\mathcal{F}(\mathcal{Q})$ . In particular we let  $\boldsymbol{\mu}(A) = (\mu_0, (\mu_{i, \sigma})_{i \in [k], \sigma \in [q_i]})$  where  $\mu_{i, \sigma} = \sum_{\mathbf{a} \in \mathcal{Q} : \mathbf{a}_i = \sigma} A(\mathbf{a})$ . We also define canonical function and polarization operators so as to preserve  $\boldsymbol{\mu}(A)$ . So given arbitrary  $A$ , let  $\mathcal{D} = \frac{1}{\mu_0(A)} \cdot A$ . Note  $\mathcal{D} \in \Delta(\mathcal{Q})$ . For  $\boldsymbol{\mu} = (\mu_0, (\mu_{i, \sigma})_{i \in [k], \sigma \in [q_i]})$  where  $\forall i, \sum_{\sigma \in [q_i]} \mu_{i, \sigma} = \mu_0$ , we define  $A_{\boldsymbol{\mu}} = \mu_0 \cdot \mathcal{D}_{\boldsymbol{\mu}'}$  where  $\boldsymbol{\mu}' = (\mu_{i, \sigma} / \mu_0)_{i \in [k], \sigma \in [q_i]}$  to be the canonical function associated with  $\boldsymbol{\mu}$ .

**Definition 7.5** (Polarization length). *For distribution  $A \in \mathcal{F}(\mathcal{Q})$ , where  $\mathcal{Q} = [q_1] \times \cdots \times [q_k]$ , let  $N(A)$  be the smallest  $t$  such that there exists a sequence  $\mathbf{A} = A_0, A_1, \dots, A_t$  such that  $A_0 = A$ ,*

<sup>10</sup>To see this, suppose  $\mathbf{u} = \mathbf{u}'$ , then we have  $u_j = \min\{u_j, v_j\}$  for all  $j \in [k]$  and hence  $\mathbf{u} \leq \mathbf{v}$ , which is a contradiction. The same analysis works for the other cases.

$A_t = A_{\mu(A)}$  is canonical and for every  $i \in [t]$  it holds that there exists incomparable  $\mathbf{u}_i, \mathbf{v}_i \in \text{supp}(A_{i-1})$  such that  $A_i = (A_{i-1})_{\mathbf{u}_i, \mathbf{v}_i}$ . If no such finite sequence exists then let  $N(A)$  be infinite. Let  $N(k, q_1, \dots, q_k) = \sup_{A \in \mathcal{F}(\mathcal{Q})} \{N(A)\}$ , and  $\tilde{N}(Q) = \max_{k, q_1, \dots, q_k \mid \sum_i q_i = Q} N(k, q_1, \dots, q_k)$ . Again, if  $N(A) = \infty$  for some  $A$  or if no finite upper bound exists,  $\tilde{N}(Q)$  is defined to be  $\infty$ .

Note that if  $\mathcal{D} \in \Delta(\mathcal{Q})$ , so is every element in the sequence, so the polarization length bound below applies also to distributions. Our main lemma in this subsection is the following:

**Lemma 7.6** (A finite upper bound on  $\tilde{N}(Q)$ ).  *$\tilde{N}(Q)$  is finite for every finite  $Q$ . Specifically  $\tilde{N}(Q) \leq (Q^2 + 3)\tilde{N}(Q - 1)$ . Consequently for every  $k, q_1, \dots, q_k$ ,  $N(q_1, \dots, q_k)$  is finite as well.*

We prove [Lemma 7.6](#) constructively in the following four steps.

**Step 1: The algorithm POLARIZE.** Let us start with some notations. For  $A \in \mathcal{F}([q_1] \times \dots \times [q_k])$  we let  $A|_{x_\ell=q_\ell}$  denote the function  $A$  restricted to the domain  $[q_1] \times \dots \times [q_{\ell-1}] \times \{q_\ell\} \times [q_{\ell+1}] \times \dots \times [q_k]$ . Note that  $A|_{x_\ell=q_\ell}$  is effectively a  $(k-1)$ -dimensional function. We also define  $A|_{x_\ell < q_\ell}$  as the restriction of  $A$  to the domain  $[q_1] \times \dots \times [q_{\ell-1}] \times [q_\ell - 1] \times [q_{\ell+1}] \times \dots \times [q_k]$ .

---

**Algorithm 2** POLARIZE( $\cdot$ )

---

**Input:**  $A \in \mathcal{F}([q_1] \times \dots \times [q_k])$ .

- 1: **if**  $k=1$  OR  $\nexists i : q_i \geq 2$  **then**
  - 2:     **Output:**  $A$ .
  - 3: **WLOG**, let  $q_k \geq 2$ .
  - 4:  $t \leftarrow 0$ ;  $Q^- \leftarrow \sum_{i=1}^k (q_i - 1) - 1$ ;  $Q^+ \leftarrow \sum_{i=1}^{k-1} (q_i - 1)$
  - 5:  $(A_0)|_{x_k < q_k} \leftarrow \text{POLARIZE}(A|_{x_k < q_k})$ ;  $(A_0)|_{x_k = q_k} \leftarrow \text{POLARIZE}(A|_{x_k = q_k})$
  - 6: Let  $(1)^k = \mathbf{a}_t(0) < \dots < \mathbf{a}_t(Q^-) = (q_1, \dots, q_{k-1}, q_k - 1)$  be a chain supporting  $(A_t)|_{x_k < q_k}$ .
  - 7: Let  $((1)^{k-1}, q_k) = \mathbf{b}_t(0) < \dots < \mathbf{b}_t(Q^+) = (q_1, \dots, q_k)$  be a chain supporting  $(A_t)|_{x_k = q_k}$ .
  - 8: **while**  $\exists (i, j)$  with  $j < Q^+$  s.t.  $\max\{\mathbf{a}_t(i), \mathbf{b}_t(j)\} = (q_1, \dots, q_k)$  and  $A_t(\mathbf{a}_t(i)), A_t(\mathbf{b}_t(j)) > 0$  **do**
  - 9:     Let  $(i_t, j_t)$  be the lexicographically smallest such pair  $(i, j)$ .
  - 10:      $B_t \leftarrow (A_t)_{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)}$ .
  - 11:      $(A_{t+1})|_{x_k < q_k} \leftarrow \text{POLARIZE}(B_t|_{x_k < q_k})$ ;  $(A_{t+1})|_{x_k = q_k} \leftarrow (B_t)|_{x_k = q_k}$ .
  - 12:      $t \leftarrow t + 1$ .
  - 13:     Let  $(1)^k = \mathbf{a}_t(0) < \dots < \mathbf{a}_t(Q^-) = (q_1, \dots, q_k - 1)$  be a chain supporting  $(A_t)|_{x_k < q_k}$ .
  - 14:     Let  $((1)^{k-1}, q_k) = \mathbf{b}_t(0) < \dots < \mathbf{b}_t(Q^+) = (q_1, \dots, q_k)$  be a chain supporting  $(A_t)|_{x_k = q_k}$ .
  - 15: Let  $\ell \in [k]$  be such that for every  $\mathbf{a} \in [q_1] \times \dots \times [q_k] \setminus \{(q_1, \dots, q_k)\}$  we have  $A_t(\mathbf{a}) > 0 \Rightarrow a_\ell < q_\ell$ .
  - 16:  $(A_{t+1})|_{x_\ell < q_\ell} \leftarrow \text{POLARIZE}(A_t|_{x_\ell < q_\ell})$ ;  $(A_{t+1})|_{x_\ell = q_\ell} \leftarrow (A_t)|_{x_\ell = q_\ell}$ .
  - 17: **Output:**  $A_{t+1}$ .
- 

The goal of the rest of the proof is to show that [Algorithm 2](#) terminates after a finite number of steps and outputs  $A_{\mu(A)}$ .

**Step 2: Correctness assuming POLARIZE terminates.**

**Claim 7.7** (Correctness condition of POLARIZE). *For every  $A \in \mathcal{F}([q_1] \times \dots \times [q_k])$ , if POLARIZE terminates, then  $\text{POLARIZE}(A) = A_{\mu(A)}$ . In particular,  $\text{POLARIZE}(A)$  has the same marginals as  $A$  and is supported on a chain.*

*Proof.* First, by the definition of the polarization operator (Definition 7.3), the marginals of  $A_t$  are the same for every  $t$ . So in the rest of the proof, we focus on inductively showing that if POLARIZE terminates, then POLARIZE( $A$ ) is supported on a chain.

The base case where  $k = 1$  is trivially supported on a chain as desired.

When  $k > 1$ , note that when the algorithm enters the Clean-up stage, if we let  $m$  and  $n$  denote the largest indices such that  $A_t(\mathbf{a}_t(m)), A_t(\mathbf{b}_t(n)) > 0$  and  $A_t(\mathbf{b}_t(n)) \neq (q_1, \dots, q_k)$ , then the condition that  $\max\{\mathbf{a}_t(m), \mathbf{b}_t(n)\} \neq (q_1, \dots, q_k)$  implies that there is a coordinate  $\ell$  such that  $\mathbf{a}_t(m)_\ell < q_\ell$  and  $\mathbf{b}_t(n)_\ell < q_\ell$ . Since every  $\mathbf{c}$  such that  $A_t(\mathbf{c}) > 0$  and  $c_k < q_k$  satisfies  $\mathbf{c} \leq \mathbf{a}_t(m)$ , we have  $A_t(\mathbf{c}) > 0$  implies  $c_\ell < q_\ell$ . Similarly for every  $\mathbf{c} \neq (q_1, \dots, q_k)$  such that  $c_k = q_k$ , we have  $A_t(\mathbf{c}) > 0$  implies  $c_\ell < q_\ell$ . We conclude that  $A_t$  is supported on  $\{(q_1, \dots, q_k)\} \cup \{\mathbf{c} \mid c_\ell < q_\ell\}$ . Thus, by the induction hypothesis, after polarizing  $(A_t)|_{x_\ell < q_\ell}$  and leaving  $(A_t)|_{x_\ell = q_\ell}$  unchanged, we get that the resulting function  $A_{t+1}$  is supported on a chain as desired and complete the induction. We conclude that if POLARIZE terminates, we have  $\text{POLARIZE}(A) = A_{\mu(A)}$ .  $\square$

**Step 3: Invariant in POLARIZE.** Now, in the rest of the proof of Lemma 7.6, the goal is to show that for every input  $A$ , the number of iterations of the while loop in Algorithm 2 is finite. The key claim (Claim 7.11) here asserts that the sequence of pairs  $(i_t, j_t)$  is monotonically increasing in lexicographic order. Once we establish this claim, it follows that there are at most  $Q^- \cdot Q^+$  iterations of the while loop and so  $\tilde{N}(Q) \leq (Q^2 + 3)\tilde{N}(Q - 1)$ , proving Lemma 7.6. Before proving Claim 7.11, we establish the following properties that remain invariant after every iteration of the while loop.

**Claim 7.8.** *For every  $t \geq 0$ , we have  $(A_t)|_{x_k = q_k}$  and  $(A_t)|_{x_k < q_k}$  are both supported on chains.*

*Proof.* For  $(A_t)|_{x_k < q_k}$ , the claim follows from the correctness of the recursive call to POLARIZE. For  $(A_t)|_{x_k = q_k}$ , we claim by induction on  $t$  that the supporting chain  $\mathbf{b}_t(0) < \dots < \mathbf{b}_t(Q^+)$  never changes (with  $t$ ). To see this, note that  $\mathbf{b}_t(k-1) = (q_1, \dots, q_k)$  is the only point in the support of  $(A_t)|_{x_k = q_k}$  that increases in value, and this is already in the supporting chain. Thus  $\mathbf{b}_t(0) < \dots < \mathbf{b}_t(Q^+)$  continues to be a supporting chain for  $(A_{t+1})|_{x_k = q_k}$ .  $\square$

For  $\mathbf{c} \in [q_1] \times \dots \times [q_k]$ , we say that a function  $A : [q_1] \times \dots \times [q_k] \rightarrow \mathbb{R}^{\geq 0}$  is  $\mathbf{c}$ -respecting if for every  $\mathbf{c}'$  such that  $A(\mathbf{c}') > 0$ , we have  $\mathbf{c}' \geq \mathbf{c}$  or  $\mathbf{c}' \leq \mathbf{c}$ . We say that  $A$  is  $\mathbf{c}$ -downward-respecting if  $A$  is  $\mathbf{c}$ -respecting and the points in the support of  $A$  above  $\mathbf{c}$  form a partial chain, specifically, if  $\mathbf{u}, \mathbf{v} > \mathbf{c}$  have  $A(\mathbf{u}), A(\mathbf{v}) > 0$ , then either  $\mathbf{u} \geq \mathbf{v}$  or  $\mathbf{v} \geq \mathbf{u}$ .

Note that if  $A$  is supported on a chain then  $A$  is  $\mathbf{c}$ -respecting for every point  $\mathbf{c}$  in the chain. Conversely, if  $A$  is supported on a chain and  $A$  is  $\mathbf{c}$ -respecting, then  $A$  is supported on a chain that includes  $\mathbf{c}$ .

**Claim 7.9.** *Let  $A$  be a  $\mathbf{c}$ -respecting function and let  $\tilde{A}$  be obtained from  $A$  by a finite sequence of polarization updates, as in Definition 7.3. Then  $\tilde{A}$  is also  $\mathbf{c}$ -respecting. Furthermore if  $A$  is  $\mathbf{c}$ -downward-respecting and  $\mathbf{w} > \mathbf{c}$  then  $\tilde{A}$  is also  $\mathbf{c}$ -downward-respecting and  $A(\mathbf{w}) = \tilde{A}(\mathbf{w})$ .*

*Proof.* Note that it suffices to prove the claim for a single update by a polarization operator since the rest follows by induction. So let  $\tilde{A} = A_{\mathbf{u}, \mathbf{v}}$  for incomparable  $\mathbf{u}, \mathbf{v} \in \text{supp}(A)$ . Since  $A$  is  $\mathbf{c}$ -respecting, and  $\mathbf{u}, \mathbf{v}$  are incomparable, either  $\mathbf{u} \leq \mathbf{c}, \mathbf{v} \leq \mathbf{c}$  or  $\mathbf{u} \geq \mathbf{c}, \mathbf{v} \geq \mathbf{c}$ . Suppose the former is true, then  $\max\{\mathbf{u}, \mathbf{v}\} \leq \mathbf{c}$  and  $\min\{\mathbf{u}, \mathbf{v}\} \leq \mathbf{c}$ , and hence,  $\tilde{A}$  is  $\mathbf{c}$ -respecting. Similarly, in the case when  $\mathbf{u} \geq \mathbf{c}, \mathbf{v} \geq \mathbf{c}$ , we can show that  $\tilde{A}$  is  $\mathbf{c}$ -respecting. The furthermore part follows by

noticing that for  $\mathbf{u}$  and  $\mathbf{v}$  to be incomparable if  $A$  is  $\mathbf{c}$ -downward-respecting and  $A(\mathbf{u}), A(\mathbf{v}) > 0$ , then  $\mathbf{u}, \mathbf{v} \leq \mathbf{c}$ , and so the update changes  $A$  only at points below  $\mathbf{c}$ .  $\square$

The following claim asserts that in every iteration of the while loop, by the lexicographically minimal choice of  $(i_t, j_t)$ , there exists a coordinate  $h \in [k-1]$  such that every vector  $\mathbf{c} < \mathbf{a}_t(i_t)$  in the support of  $A_t, B_t$ , or  $A_{t+1}$  has  $c_h < q_h$ , and every vector  $\mathbf{c} \neq (q_1, \dots, q_k)$  in the support of  $(A_t)|_{x_k=q_k}$  has  $c_h < q_h$ .

**Claim 7.10.** *For every  $t \geq 0$ ,  $\exists h \in [k-1]$  such that  $\forall \mathbf{c} \in [q_1] \times \dots \times [q_k]$ , if  $\mathbf{c} \in \text{supp}(A_t) \cup \text{supp}(B_t) \cup \text{supp}(A_{t+1})$ , then the following hold:*

- If  $\mathbf{c} < \mathbf{a}_t(i_t)$ , then  $c_h < q_h$ .
- If  $c_k = q_k$  and  $\mathbf{c} \neq (q_1, \dots, q_k)$ , then  $c_h < q_h$ .

*Proof.* Since  $(i_t, j_t)$  is lexicographically the smallest incomparable pair in the support of  $A_t$ , for  $i < i_t, j < Q^+$ , and  $A_t(\mathbf{a}(i)), A_t(\mathbf{b}(j)) > 0$ , we have  $\max\{\mathbf{a}(i), \mathbf{b}(j)\} \neq (q_1, \dots, q_k)$ . Let  $m$  be the largest index smaller than  $i_t$  such that  $A_t(\mathbf{a}_t(m)) > 0$ . Similarly, let  $n < Q^+$  be the largest index such that  $A_t(\mathbf{b}_t(n)) > 0$ . Then the fact that  $\max\{\mathbf{a}_t(m), \mathbf{b}_t(n)\} \neq (q_1, \dots, q_k)$  implies that there exists  $h \in [k-1]$  such that  $\mathbf{a}_t(m)_h < q_h$  and  $\mathbf{b}_t(n)_h < q_h$ . Now, using the fact (from Claim 7.8) that  $(A_t)|_{x_k < q_k}$  is supported on a chain, we conclude that for every  $\mathbf{c} < \mathbf{a}_t(i_t)$ ,  $A_t(\mathbf{c}) > 0$  implies that  $\mathbf{c} \leq \mathbf{a}_t(m)$  and hence,  $c_h < q_h$ . Similarly, for every vector  $\mathbf{c} \neq (q_1, \dots, q_k)$  in the support of  $(A_t)|_{x_k=q_k}$ , by the maximality of  $n$ , we have  $c_h < q_h$ .

We now assert that the same holds for  $B_t$ . First, recall that  $\text{supp}(B_t) \subset \text{supp}(A_t) \cup \{(q_1, \dots, q_k), \min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\}\}$  since  $B_t = (A_t)_{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)}$ . Next, note that the only point (other than  $(q_1, \dots, q_k)$ ) where  $B_t$  is larger than  $A_t$  is  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\}$ . It suffices to show that  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\}_h < q_h$ . We have  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\} \leq \mathbf{b}_t(j_t) \leq \mathbf{b}_t(n)$  and hence  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\}_h < q_h$ .

Finally, we assert that same holds also for  $A_{t+1}$ . Since  $A_{t+1}|_{x_k=q_k} = B_t|_{x_k=q_k}$ , the second item in the claim follows trivially. To prove the first item, let us consider  $\mathbf{a}' \in [q_1] \times \dots \times [q_k]$  defined as follows:  $\mathbf{a}'_h = q_h - 1$  and  $\mathbf{a}'_r = \mathbf{a}_t(i_t)_r$  for  $r \neq h$ . Note that  $B_t|_{x_k < q_k}$  is  $\mathbf{a}_t(i_t)$ -respecting since potentially the only new point in its support (compared to  $A_t|_{x_k < q_k}$ ) is  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\} \leq \mathbf{a}_t(i_t)$ . From the previous paragraph we also have that if  $B_t(\mathbf{c}) > 0$  and  $\mathbf{c} < \mathbf{a}_t(i_t)$ , then  $c_h < q_h$  and hence,  $\mathbf{c} \leq \mathbf{a}'$ . On the other hand, if  $B_t(\mathbf{c}) > 0$  and  $\mathbf{c} \geq \mathbf{a}_t(i_t)$ , then  $\mathbf{c} \geq \mathbf{a}'$ . Therefore,  $B_t|_{x_k < q_k}$  is  $\mathbf{a}'$ -respecting. By applying Claim 7.9, we conclude that  $(A_{t+1})|_{x_k < q_k}$  is also  $\mathbf{a}'$ -respecting. It follows that if  $\mathbf{c} < \mathbf{a}_t(i_t)$  and  $A_{t+1}(\mathbf{c}) > 0$ , then  $\mathbf{c} \leq \mathbf{a}'$  and so  $c_h < q_h$ .  $\square$

**Step 4: Proof of Lemma 7.6.** The following claim establishes that the while loop in the POLARIZE algorithm terminates after a finite number of iterations.

**Claim 7.11.** *For every  $t \geq 0$ ,  $(i_t, j_t) < (i_{t+1}, j_{t+1})$  in lexicographic ordering.*

*Proof.* Consider the chain  $\mathbf{a}_{t+1}(0) < \dots < \mathbf{a}_{t+1}(Q^-)$  supporting  $A_{t+1}|_{x_k < q_k}$ . Note that for  $i \geq i_t$ ,  $A_{t+1}|_{x_k < q_k}$  is  $\mathbf{a}_t(i)$ -respecting (since  $A_t|_{x_k < q_k}$  and  $B_t|_{x_k < q_k}$  were also so). In particular,  $A_t|_{x_k < q_k}$  is  $\mathbf{a}_t(i)$ -respecting because it is supported on a chain containing  $\mathbf{a}_t(i)$ . Next  $B_t|_{x_k < q_k}$  is  $\mathbf{a}_t(i)$ -respecting since potentially the only new point in its support is  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\} \leq \mathbf{a}_t(i)$ . Finally,  $A_{t+1}|_{x_k < q_k}$  is also  $\mathbf{a}_t(i)$ -respecting using Claim 7.9. Thus we can build a chain containing  $\mathbf{a}_t(i)$  that supports  $A_{t+1}|_{x_k < q_k}$ . It follows that we can use  $\mathbf{a}_{t+1}(i) = \mathbf{a}_t(i)$  for  $i \geq i_t$ . Now consider  $i < i_t$ . We

must have  $\mathbf{a}_{t+1}(i) < \mathbf{a}_{t+1}(i_t) = \mathbf{a}_t(i_t)$ . By [Claim 7.10](#), there exists  $h \in [k-1]$  such that for  $i < i_t$ ,  $\mathbf{a}_{t+1}(i)_h < q_h$ .

We now turn to analyzing  $(i_{t+1}, j_{t+1})$ . Note that by definition,  $A_{t+1}(\mathbf{a}_{t+1}(i_{t+1})) > 0$  and  $A_{t+1}(\mathbf{b}_{t+1}(b_{t+1})) > 0$ . First, let us show that  $i_t \leq i_{t+1}$ . On the contrary, let us assume that  $i_{t+1} < i_t$ . It follows from the above paragraph that  $\mathbf{a}_{t+1}(i_{t+1})_h < q_h$ . Also, for every  $\mathbf{b}_{t+1}(j)$  with  $j < Q^+$  and  $A_{t+1}(\mathbf{b}_{t+1}(j)) > 0$ , we have  $\mathbf{b}_{t+1}(j)_h < q_h$ . Therefore,  $\max\{\mathbf{a}(i_{t+1}), \mathbf{b}(j_{t+1})\} \neq (q_1, \dots, q_k)$  (in particular  $\max\{\mathbf{a}(i_{t+1}), \mathbf{b}(j_{t+1})\}_h < q_h$ ), which is a contradiction.

Next, we show that if  $i_{t+1} = i_t$ , then  $j_{t+1} \geq j_t$ . By the minimality of  $(i_t, j_t)$  in the  $t$ -th round, for  $j < j_t$  such that  $A_t(b_t(j)) > 0$ , we have  $\max\{a_t(i_t), b_t(j)\} \neq (q_1, \dots, q_k)$ . Since  $i_{t+1} = i_t$ ,  $a_{t+1}(i_{t+1}) = a_{t+1}(i_t) = a_t(i_t)$ . We already noted in the proof of [Claim 7.8](#) that  $\mathbf{b}_t(0) < \dots < \mathbf{b}_t(Q^+)$  is also a supporting chain for  $(A_{t+1})|_{x_k=q_k}$ . The only point where the function  $A_{t+1}|_{x_k=q_k}$  has greater value than  $A_t|_{x_k=q_k}$  is  $(q_1, \dots, q_k)$ . Therefore, for  $j < j_t$  such that  $A_{t+1}(b_{t+1}(j)) > 0$ , we have  $\max\{a_{t+1}(i_{t+1}), b_{t+1}(j)\} \neq (q_1, \dots, q_k)$  and hence,  $j_{t+1} \geq j_t$ .

So far, we have established that  $(i_{t+1}, j_{t+1}) \geq (i_t, j_t)$  in lexicographic ordering. Finally, we will show that  $(i_{t+1}, j_{t+1}) \neq (i_t, j_t)$  by proving that at least one of  $A_{t+1}(\mathbf{a}_{t+1}(i_t))$  and  $A_{t+1}(\mathbf{b}_{t+1}(j_t))$  is zero. The polarization update ensures that at least one of  $B_t(\mathbf{a}_t(i_t))$  and  $B_t(\mathbf{b}_t(j_t))$  is zero. If  $B_t(\mathbf{b}_t(j_t)) = 0$ , then by definition, we have  $A_{t+1}(\mathbf{b}_{t+1}(j_t)) = A_{t+1}(\mathbf{b}_t(j_t)) = 0$ . Finally to handle the case  $B_t(\mathbf{a}_t(i_t)) = 0$ , let us again define  $\mathbf{a}'$  as:  $\mathbf{a}'_h = q_h - 1$  and  $\mathbf{a}'_r = \mathbf{a}_t(i_t)_r$  for  $r \neq h$ , where  $h$  is as given by [Claim 7.10](#). We assert that  $B_t|_{x_k < q_k}$  is  $\mathbf{a}'$ -downward-respecting. As shown in the proof of [Claim 7.10](#), we have  $B_t|_{x_k < q_k}$  is  $\mathbf{a}'$ -respecting. The support of  $B_t|_{x_k < q_k}$  is contained in  $\{\mathbf{a}_t(0), \dots, \mathbf{a}_t(Q^-)\} \cup \{\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\}\}$  and  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\} < \mathbf{a}_t(i_t)$ , and by [Claim 7.10](#),  $\min\{\mathbf{a}_t(i_t), \mathbf{b}_t(j_t)\} \leq \mathbf{a}'$ . It follows that  $B_t|_{x_k < q_k}$  is  $\mathbf{a}'$ -downward-respecting. Finally, by the furthermore part of [Claim 7.9](#) applied to  $B_t|_{x_k < q_k}$  and  $\mathbf{w} = \mathbf{a}_t(i_t)$ , we get that  $A_{t+1}(\mathbf{a}_{t+1}(i_t)) = A_{t+1}(\mathbf{a}_t(i_t)) = B_t(\mathbf{a}_t(i_t)) = 0$ . It follows that  $(i_{t+1}, j_{t+1}) \neq (i_t, j_t)$ .  $\square$

*Proof of [Lemma 7.6](#).* By [Claim 7.7](#), we know that if [Algorithm 2](#) terminates, we have  $\text{POLARIZE}(A) = A_{\mu(A)}$ . Hence, the maximum number of polarization updates used in  $\text{POLARIZE}$  (on input from  $\mathcal{F}([q_1] \times \dots \times [q_k])$ ) serves as an upper bound for  $\tilde{N}(Q)$ , for  $Q = \sum_{i=1}^k q_k$ . By [Claim 7.11](#), we know that there are at most  $Q^2$  iterations of the while loop and so  $\tilde{N}(Q) \leq (Q^2 + 3)\tilde{N}(Q - 1)$  as desired.  $\square$

## 7.2 Reduction from single function to a family of functions

In this subsection, we prove the following lemma that reduces an SD problem for a single function to an SD problem for a family of functions.

**Lemma 7.12.** *Suppose there exists  $\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N$ ,  $\delta > 0$  with  $\mu(\mathcal{D}_Y) = \mu(\mathcal{D}_N)$  and a  $c = c(n)$ -communication protocol achieving advantage  $\delta$  solving  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD on instances of length  $n$  for every  $n \geq n_0$ . Then there exist  $\mathcal{D}_1, \mathcal{D}_2 \in \Delta([q]^k)$  with  $\mu(\mathcal{D}_1) = \mu(\mathcal{D}_2)$ ,  $\delta' > 0$ ,  $n'_0$ , and a  $c$ -communication protocol achieving advantage  $\delta'$  solving  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD on instances of length  $n \geq n'_0$  using  $O(s)$  bits of communication.*

We prove the lemma by a hybrid argument, where we slowly change the distribution  $\mathcal{D}_Y$  to  $\mathcal{D}_N$  by considering one function from  $\mathcal{F}$  at a time. The crux of the lemma is in showing that two adjacent steps in this sequence are at least as hard as some single-function SD problem, which follows from the following lemma.

**Lemma 7.13.** Let  $n, k, q \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  where  $k, q, \alpha$  are constants with respect to  $n$  and  $\alpha n$  is an integer less than  $n/k$ . Let  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ . For every  $\varepsilon, \delta \in (0, 1)$ , there exist  $n' = \Omega(n)$  and constants  $\alpha', \delta' \in (0, 1)$  such that the following holds. For every distributions  $\mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2 \in \Delta(\mathcal{F} \times [q]^k)$  such that  $\mathcal{D}_Y = (1 - \varepsilon)\mathcal{D}_0 + \varepsilon\mathcal{D}_1$  and  $\mathcal{D}_N = (1 - \varepsilon)\mathcal{D}_0 + \varepsilon\mathcal{D}_2$  and for every  $c \in \mathbb{N}$ , suppose there exists a protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD with parameters  $n$  and  $\alpha$  using  $c$  bits of communication with advantage  $\delta$ , then there exists a protocol for  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD with parameters  $n'$  and  $\alpha'$  using  $c$  bits of communication with advantage  $\delta'$ .

The proof idea of [Lemma 7.13](#) is very similar to that of [Theorem 7.4](#). We defer the proof to [Section 8.2](#) and turn to showing how [Lemma 7.12](#) follows.

*Proof of [Lemma 7.12](#).* Let  $\mathbf{ALG}(\mathbf{x}^*; M, \mathbf{z})$  be the  $c$ -bit protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD achieving advantage  $\delta$  guaranteed to exist by the theorem statement. Let  $\mathcal{F} = \{f_1, \dots, f_\ell\}$ . Since  $\mu(\mathcal{D}_Y) = \mu(\mathcal{D}_N)$  for each  $i \in [m]$ , we have that  $\Pr[f = f_i : (f, \mathbf{b}) \sim \mathcal{D}_Y] = \Pr[f = f_i : (f, \mathbf{b}) \sim \mathcal{D}_N]$ . Let us denote this probability by  $w^{(i)}$ ,  $w^{(i)} = \Pr[f = f_i : (f, \mathbf{b}) \sim \mathcal{D}_Y] = \Pr[f = f_i : (f, \mathbf{b}) \sim \mathcal{D}_N]$  for each  $i \in [\ell]$ . For each  $i \in [\ell]$ , let  $\mathcal{D}_Y^{(i)}$  be the distribution of a random variable  $\mathbf{b} \in [q]^k$  that is sampled from  $(f, \mathbf{b}) \sim \mathcal{D}_Y$  conditioned on  $f = f_i$ . Similarly, for each  $i \in [\ell]$ , let  $\mathcal{D}_N^{(i)}$  be the distribution of  $\mathbf{b} \in [q]^k$  from  $(f, \mathbf{b}) \sim \mathcal{D}_N$  conditioned on  $f = f_i$ . This way we have that  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  are the mixture distributions:  $\mathcal{D}_Y = \sum_{i \in [\ell]} w^{(i)} \cdot \mathcal{D}_Y^{(i)}$  and  $\mathcal{D}_N = \sum_{i \in [\ell]} w^{(i)} \cdot \mathcal{D}_N^{(i)}$ .

For every  $i \in \{0, \dots, \ell\}$ , we define a distribution  $\mathcal{D}^{(i)}$  as the following mixture distribution:

$$\mathcal{D}^{(i)} = \sum_{j \in \{1, \dots, i\}} w^{(j)} \cdot \mathcal{D}_N^{(j)} + \sum_{j \in \{i+1, \dots, \ell\}} w^{(j)} \cdot \mathcal{D}_Y^{(j)}.$$

Let  $p_i = \Pr[\mathbf{ALG}(\mathbf{x}^*; M, \mathbf{z}) = \mathbf{YES} : (f, \mathbf{b}) \sim \mathcal{D}^{(i)}]$  for every  $i \in \{0, \dots, \ell\}$ . Observe that  $p_0 = \Pr[\mathbf{ALG}(\mathbf{x}^*; M, \mathbf{z}) = \mathbf{YES} : (f, \mathbf{b}) \sim \mathcal{D}_Y]$  and  $p_\ell = \Pr[\mathbf{ALG}(\mathbf{x}^*; M, \mathbf{z}) = \mathbf{YES} : (f, \mathbf{b}) \sim \mathcal{D}_N]$ . Since the advantage of  $\mathbf{ALG}$  in distinguishing  $\mathcal{D}_Y$  and  $\mathcal{D}_N$  is at least  $\delta$ , we have that

$$\delta = |p_0 - p_\ell| = \left| \sum_{i \in \{0, \dots, \ell-1\}} (p_i - p_{i+1}) \right| \leq \sum_{i \in \{0, \dots, \ell-1\}} |p_i - p_{i+1}|.$$

Let  $\delta' = \delta/\ell$ . We have that at least one term of this sum is  $|p_i - p_{i+1}| \geq \delta'$ . From this we conclude that for some  $i \in \{0, \dots, \ell-1\}$ ,  $\mathbf{ALG}$  achieves advantage at least  $\delta'$  for  $(\mathcal{F}, \mathcal{D}^{(i)}, \mathcal{D}^{(i+1)})$ -SD.

It remains to show that if one can distinguish  $\mathcal{D}^{(i)}$  and  $\mathcal{D}^{(i+1)}$  that differ only for  $(f, \mathbf{b})$  with  $f = f_{i+1}$ , then one can also distinguish  $\mathcal{D}_1 = \mathcal{D}_Y^{(i+1)}$  and  $\mathcal{D}_2 = \mathcal{D}_N^{(i+1)}$ . Since  $\mu(\mathcal{D}_1) = \mu(\mathcal{D}_2)$ , this will finish the proof. We show that  $\mathcal{D}_1$  and  $\mathcal{D}_2$  are distinguishable using [Lemma 7.13](#).

Let us define  $\varepsilon = w_{i+1}$ ,  $\mathcal{D} = \frac{1}{1-\varepsilon} \left( \sum_{j \in \{1, \dots, i\}} w^{(j)} \cdot \mathcal{D}_N^{(j)} + \sum_{j \in \{i+2, \dots, \ell\}} w^{(j)} \cdot \mathcal{D}_Y^{(j)} \right)$ . Now observe that  $\mathcal{D}^{(i)} = (1 - \varepsilon)\mathcal{D} + \varepsilon\mathcal{D}_1$  and  $\mathcal{D}^{(i+1)} = (1 - \varepsilon)\mathcal{D} + \varepsilon\mathcal{D}_2$ . Now by [Lemma 7.13](#), a protocol that distinguishes  $\mathcal{D}^{(i)}$  and  $\mathcal{D}^{(i+1)}$  implies a protocol for  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD with advantage  $\delta'' > 0$  and communication complexity  $O(s)$ .  $\square$

### 7.3 Putting it together

We now have the ingredients in place to prove [Theorem 5.4](#) which we recall below for convenience.



**Theorem 5.4** (Communication lower bound for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD). *For every  $k, q$ , every finite set  $\mathcal{F}$ , every pair of distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$  with  $\boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$  there exists  $\alpha_0 > 0$  such that for every  $0 < \alpha \leq \alpha_0$  and  $\delta > 0$  there exists  $\tau > 0$  such that the following holds: Every protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD achieving advantage  $\delta$  on instances of length  $n$  requires  $\tau\sqrt{n}$  bits of communication.*

*Proof of Theorem 5.4.* Fix  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  and distributions  $\mathcal{D}_Y, \mathcal{D}_N \in \Delta(\mathcal{F} \times [q]^k)$  with  $\boldsymbol{\mu} = \boldsymbol{\mu}(\mathcal{D}_Y) = \boldsymbol{\mu}(\mathcal{D}_N)$ . Lemma 7.12, applied to  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ , gives us  $n_0, \delta'$ , and distributions  $\mathcal{D}'_Y, \mathcal{D}'_N \in \Delta([q]^k)$  with  $\boldsymbol{\mu}' = \boldsymbol{\mu}(\mathcal{D}'_Y) = \boldsymbol{\mu}(\mathcal{D}'_N)$  such that any  $c$ -communication protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD with advantage  $\delta$  implies a  $c$ -communication protocol for  $(\mathcal{D}'_Y, \mathcal{D}'_N)$ -SD with advantage  $\delta'$  for all  $n \geq n_0$ . Now we'll focus on proving a lower bounds for the problem  $(\mathcal{D}'_Y, \mathcal{D}'_N)$ -SD.

Lemma 7.6, applied to  $\mathcal{D}'_Y$ , gives us  $\mathcal{D}_0 = \mathcal{D}'_Y, \mathcal{D}_1, \dots, \mathcal{D}_t = \mathcal{D}_{\boldsymbol{\mu}'}$  such that  $\mathcal{D}_{i+1} = (\mathcal{D}_i)_{\mathbf{u}(i), \mathbf{v}(i)}$ , i.e.,  $\mathcal{D}_i$  is an update of  $\mathcal{D}_i$ , with  $t \leq \tilde{N}(Q) < \infty$ , for  $Q = \sum_{i=1}^k q_k$ . Similarly Lemma 7.6, applied to  $\mathcal{D}'_N$ , gives us  $\mathcal{D}'_0 = \mathcal{D}'_N, \mathcal{D}'_1, \dots, \mathcal{D}'_{t'} = \mathcal{D}_{\boldsymbol{\mu}'}$  such that  $\mathcal{D}'_{i+1} = (\mathcal{D}'_i)_{\mathbf{u}'(i), \mathbf{v}'(i)}$  with  $t' \leq \tilde{N}(Q) < \infty$ .

Applying Theorem 7.4 with  $\delta'' = \delta'/(2\tilde{N}(Q))$  to the pairs  $\mathcal{D}_i$  and  $\mathcal{D}_{i+1}$ , we get that there exists  $\tau_i$  such that every protocol for  $(\mathcal{D}_i, \mathcal{D}_{i+1})$ -SD requires  $\tau_i\sqrt{n}$  bits of communication to achieve advantage  $\delta''$ . Similarly applying Theorem 7.4 again with  $\delta'' = \delta'/(2\tilde{N}(Q))$  to the pairs  $\mathcal{D}'_i$  and  $\mathcal{D}'_{i+1}$ , we get that there exists  $\tau'_i$  such that every protocol for  $(\mathcal{D}'_i, \mathcal{D}'_{i+1})$ -SD requires  $\tau'_i\sqrt{n}$  bits of communication to achieve advantage  $\delta''$ .

Letting  $\tau' = \min\{\min_{i \in [t]}\{\tau_i\}, \min_{i \in [t']}\{\tau'_i\}\}$ , we get, using the triangle inequality for indistinguishability, that every protocol  $\Pi'$  for  $(\mathcal{D}'_Y, \mathcal{D}'_N)$ -SD achieving advantage  $(t + t')\delta'' \leq \delta'$  requires  $\tau'\sqrt{n}$  bits of communication. Finally, by Lemma 7.12, every protocol  $\Pi$  for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD achieving advantage  $\delta$  requires  $\tau'\sqrt{n}$  bits of communication.  $\square$

## 8 Indistinguishability of the Polarization Step

Recall that in Definition 7.3 we define a polarization operator that polarizes a distribution  $\mathcal{D} \in \Delta([q]^k)$  to  $\mathcal{D}_{\mathbf{u}, \mathbf{v}} \in \Delta([q]^k)$  for every incomparable pair  $(\mathbf{u}, \mathbf{v})$ . In this section, we show that  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD requires  $\Omega(\sqrt{n})$  communication.

**Theorem 7.4** (Indistinguishability of the polarization step). *Let  $n, k, q \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  where  $k, q, \alpha$  are constants with respect to  $n$  and  $\alpha n$  is an integer less than  $n/k$ . For a distribution  $\mathcal{D} \in \Delta([q]^k)$ , incomparable vectors  $\mathbf{u}, \mathbf{v} \in [q]^k$ , and  $\delta > 0$ , there exists  $\tau > 0$  such that every protocol for  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD achieving advantage  $\delta$  requires  $\tau\sqrt{n}$  bits of communication.*

Let  $\mathbf{u} \vee \mathbf{v}, \mathbf{u} \wedge \mathbf{v} \in [q]^k$  be given by  $u_i \vee v_i = \max\{u_i, v_i\}$  and  $u_i \wedge v_i = \min\{u_i, v_i\}$ . Let  $\mathcal{A}_Y = \text{Unif}(\{\mathbf{u}, \mathbf{v}\})$  and  $\mathcal{A}_N = \text{Unif}(\{\mathbf{u} \vee \mathbf{v}, \mathbf{u} \wedge \mathbf{v}\})$ . We prove Theorem 7.4 in two steps. First, we use the Boolean hardness in Theorem 6.4 to show in Lemma 8.1 that the hardness holds for the special case  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD. Next, we use a similar reduction as in the proof for Lemma 7.7 in [CGSV21] to reduce  $(\mathcal{A}_Y, \mathcal{A}_N)$ -advice-SD to  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD for arbitrary distribution  $\mathcal{D} \in \Delta([q]^k)$ .

### 8.1 Reduce a Boolean SD problem to a non-Boolean SD problem

In this subsection, we consider a special case of  $\mathbf{u}, \mathbf{v} \in [q]^k$  where  $u_i \neq v_i$  for every  $i \in [k]$ . The following key lemma of this subsection establishes the hardness of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD via a reduction from a Boolean SD problem to a non-Boolean version.



**Lemma 8.1.** Let  $n, k, q \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  where  $k, q, \alpha$  are constants with respect to  $n$  and  $\alpha n$  is an integer less than  $n/k$ . For  $\mathbf{u}, \mathbf{v} \in [q]^k$  satisfying  $u_i \neq v_i$  for all  $i \in [k]$  and  $\delta > 0$ , there exists  $\tau > 0$  such that every protocol for  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD achieving advantage  $\delta$  requires  $\tau\sqrt{n}$  bits of communication.

We prove [Lemma 8.1](#) by a reduction. For such  $\mathbf{u}, \mathbf{v}$ , let  $\bar{\mathbf{u}}, \bar{\mathbf{v}} \in \{0, 1\}^k$  be the Boolean version given by  $(\bar{u}_i, \bar{v}_i) = (0, 1)$  if  $u_i < v_i$  and  $(\bar{u}_i, \bar{v}_i) = (1, 0)$  if  $u_i > v_i$ . Let  $\bar{\mathcal{A}}_Y = \text{Unif}(\{\bar{\mathbf{u}}, \bar{\mathbf{v}}\})$  and  $\bar{\mathcal{A}}_N = \text{Unif}(\{\bar{\mathbf{u}} \vee \bar{\mathbf{v}}, \bar{\mathbf{u}} \wedge \bar{\mathbf{v}}\})$ . Note that both  $\bar{\mathcal{A}}_Y$  and  $\bar{\mathcal{A}}_N$  are distributions on Boolean domain with uniform marginals. Thus, [Theorem 6.4](#) shows that any protocol for  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD requires  $\Omega(\sqrt{n})$  bits of communication. In the rest of this subsection, we reduce  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD to  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD.

For every  $\bar{n}, k, \bar{\alpha}, q, \delta$ , let  $n = 2q\bar{n}$  and  $\alpha = q^{k-1}2^{-(k+2)}\bar{\alpha}$ . Let  $\bar{I} = (\bar{\mathbf{x}}, \bar{\Gamma}, \bar{\mathbf{b}}, \bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$  denote an instance of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD of length  $\bar{n}$  with parameter  $\bar{\alpha}$ . We show below how Alice and Bob can use their inputs and shared randomness to generate an instance  $I = (\mathbf{x}, \Gamma, \mathbf{b}, M, \mathbf{z}, \mathbf{a})$  of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -advice-SD of length  $n$  with parameter  $\alpha$  “locally” and “nearly” according to the correct distributions. Namely, we show that with high probability if  $\bar{I}$  is a Yes (resp. No) instance of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD, then  $I$  will be a Yes (resp. No) instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD.

**Step 1: Specify the shared randomness.** The common randomness between Alice and Bob is an instances  $I_R = (\mathbf{x}_R, \Gamma_R, \mathbf{b}_R, M_R, \mathbf{z}_R, \mathbf{a}_R)$  drawn according to the Yes<sup>11</sup> distribution of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -advice-SD of length  $n$  with parameter  $\alpha$ . For  $j \in [\alpha n]$ , let  $V_j$  denote the set of variables in the  $j$ -th constraint, i.e.,  $V_j = \{\ell \in [n] \mid \Gamma_R(\ell) \in \{k(j-1)+1, \dots, k(j-1)+k\}\}$ . For  $i \in [k]$ , let  $T_i$  be the set of variables that are in the  $i$ -th partition and take on values in  $\{u_i, v_i\}$ , i.e.,  $T_i = \{j \in [n] \mid a_j = i \ \& \ (\mathbf{x}_R)_i \in \{u_i, v_i\}\}$ . Let  $U \subseteq [\alpha n]$  be the set of constraints that work on variables in  $T_i$ , i.e.,  $U = \{j \in [\alpha n] \mid V_j \subseteq \cup_i T_i\}$ . See [Fig. 2](#) for an example.

$V_1$	2	8	$(\mathbf{u}_1, \mathbf{v}_1) = (1, 3); (\mathbf{u}_2, \mathbf{v}_2) = (3, 2)$																																	
$V_2$	7	6																																		
$V_3$	1	4																																		
$V_4$	3	5																																		
	12	11																																		
	10	9																																		
	<table style="border-collapse: collapse; margin-left: auto; margin-right: auto;"> <thead> <tr> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_1</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_2</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_3</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_4</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_5</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_6</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_7</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_8</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)_9</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)__{10}</math></th> <th style="border-right: 1px solid black; padding: 2px 5px;"><math>(\mathbf{x}_R)__{11}</math></th> <th style="padding: 2px 5px;"><math>(\mathbf{x}_R)__{12}</math></th> </tr> </thead> <tbody> <tr> <td style="border-right: 1px solid black; text-align: center;">1</td> <td style="border-right: 1px solid black; text-align: center;">3</td> <td style="border-right: 1px solid black; text-align: center;">2</td> <td style="border-right: 1px solid black; text-align: center;">3</td> <td style="border-right: 1px solid black; text-align: center;">1</td> <td style="border-right: 1px solid black; text-align: center;">2</td> <td style="border-right: 1px solid black; text-align: center;">3</td> <td style="border-right: 1px solid black; text-align: center;">1</td> <td style="border-right: 1px solid black; text-align: center;">1</td> <td style="border-right: 1px solid black; text-align: center;">3</td> <td style="border-right: 1px solid black; text-align: center;">2</td> <td style="text-align: center;">2</td> </tr> </tbody> </table>												$(\mathbf{x}_R)_1$	$(\mathbf{x}_R)_2$	$(\mathbf{x}_R)_3$	$(\mathbf{x}_R)_4$	$(\mathbf{x}_R)_5$	$(\mathbf{x}_R)_6$	$(\mathbf{x}_R)_7$	$(\mathbf{x}_R)_8$	$(\mathbf{x}_R)_9$	$(\mathbf{x}_R)__{10}$	$(\mathbf{x}_R)__{11}$	$(\mathbf{x}_R)__{12}$	1	3	2	3	1	2	3	1	1	3	2	2
$(\mathbf{x}_R)_1$	$(\mathbf{x}_R)_2$	$(\mathbf{x}_R)_3$	$(\mathbf{x}_R)_4$	$(\mathbf{x}_R)_5$	$(\mathbf{x}_R)_6$	$(\mathbf{x}_R)_7$	$(\mathbf{x}_R)_8$	$(\mathbf{x}_R)_9$	$(\mathbf{x}_R)__{10}$	$(\mathbf{x}_R)__{11}$	$(\mathbf{x}_R)__{12}$																									
1	3	2	3	1	2	3	1	1	3	2	2																									
	$T_1 = \{1, 2, 7, 10\}$ $X_1 = \{1, 7\}$ $W_1 = \{2, 10\}$ $S_1 = \{1, 7, 10\}$ $T_2 = \{4, 5, 6, 11\}$ $X_2 = \{4, 6\}$ $W_2 = \{5, 11\}$ $S_2 = \{4, 6, 11\}$ $U = \{2, 3\}$																																			

Figure 2: An example of shared randomness used in [Lemma 8.1](#). Here  $n = 12$ ,  $k = 2$ ,  $q = 3$ , and  $\alpha = 1/3$ . The value of  $\mathbf{x}_R \in [q]^n$  is listed in a table. Consider  $(u_1, v_1) = (1, 3)$  and  $(u_2, v_2) = (3, 2)$ . The variables in sets  $T_1, T_2$  are marked grey. The variables correspond to the set  $U$  are circled with red lines and the variables correspond to sets  $S_1, S_2$  are circled with yellow dashed lines.

If  $|U| \geq \bar{\alpha}\bar{n}$  we say an error of type (1) has occurred. For  $i \in [k]$ , let  $X_i \subseteq T_i$  be the set of variables that operate on constraints in  $U$ , i.e.,  $X_i = T_i \cap (\cup_{j \in U} V_j)$ . Let  $W_i \subseteq T_i$  be a set of variables that do not participate in any constraint, i.e.,  $W_i = T_i \setminus (\cup_{j \in [\alpha n]} V_j)$ . Finally let  $S_i$  be any

<sup>11</sup>The reduction also works if we used No distribution. However, the mapping between Yes and No instances would get flipped. Namely, if  $\bar{I}$  is a Yes (resp. No) instance of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD, then  $I$  will be a No (resp. Yes) instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD.

set satisfying  $|S_i| = \bar{n}/k$  with  $X_i \subseteq S_i \subseteq X_i \cup W_i$  if such a set exists. If no such set exists we say an error of type (2) has occurred.

**Step 2: Specify the reduction.** If there is an error, we simply set  $I = I_R$ . If no errors have occurred, our reduction will embed  $\bar{I}$  into  $I_R$  by replacing the constraints in  $U$  and the variables in  $\cup_i S_i$  as described next. Note that we have to specify variables  $(\mathbf{x}, \Gamma, \mathbf{b}, M, \mathbf{z})$ . In particular, we want the private inputs can be computed locally. We verify the local property of the reduction in [Claim 8.2](#) and prove the correctness of the reduction in [Claim 8.3](#).

- $\mathbf{x}$ : Let  $\rho : [\bar{n}] \rightarrow \cup_i S_i$  be a bijection satisfying  $\bar{a}_j = i \Rightarrow \rho(j) \in S_i$ . We now define  $\mathbf{x} \in [q]^n$  as follows:

$$x_j = \begin{cases} (\mathbf{x}_R)_j & \text{if } j \notin \cup_{i \in [k]} S_i \\ u_i & \text{if } j \in S_i \text{ for some } i \in [k] \text{ and } u_i < v_i \text{ and } \bar{x}_j = 0 \\ u_i & \text{if } j \in S_i \text{ for some } i \in [k] \text{ and } u_i > v_i \text{ and } \bar{x}_j = 1 \\ v_i & \text{if } j \in S_i \text{ otherwise} \end{cases}$$

- $\Gamma$  and  $M$ : Let  $V = \{V(1), \dots, V(\bar{n})\}$  with  $V(j) < V(j+1)$  be such that  $V = \{j \in [n] \mid \Gamma_R(j) \in \cup_{i \in [k]} S_i\}$ . For  $j \in [n]$  we let

$$\Gamma(j) = \begin{cases} \Gamma_R(j) & \text{if } j \notin V \\ \rho(\bar{\Gamma}(\bar{j})) & \text{if } j = V(\bar{j}) \end{cases}$$

It may be verified that  $\Gamma$  is a permutation and furthermore the constraints in  $\Gamma$  corresponding to  $j \in U$  are derived from constraints of  $\bar{I}$ .  $M$  is then defined as the partial permutation matrix capturing  $\Gamma^{-1}(j)$  for  $j \in [k\alpha n]$ .

- $\mathbf{b}$ : Since  $\mathbf{b}$  is a hidden variable and won't be given to Alice and Bob, we postpone the specification of  $\mathbf{b}$  to the proof of [Claim 8.3](#).
- $\mathbf{z}$ : Let  $\mathbf{z}(j) = \bar{\mathbf{z}}(V(j))$  if  $j \in U$  and  $\mathbf{z}(j) = \mathbf{z}_R(j)$  otherwise.

**Step 3: Correctness of the reduction assuming no error occurs.**

**Claim 8.2** (The reduction can be computed locally). *Let  $\bar{I} = (\bar{\mathbf{x}}, \bar{\Gamma}, \bar{\mathbf{b}}, \bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$  be an instance of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD and  $I_R = (\mathbf{x}_R, \Gamma_R, \mathbf{b}_R, M_R, \mathbf{z}_R, \mathbf{a}_R)$  be the shared randomness of Alice and Bob. The above reduction satisfies the following local properties:*

- Alice can compute  $\mathbf{x}$  using  $I_R$  and  $(\bar{\mathbf{x}}, \bar{\mathbf{a}})$ .
- Bob can compute  $(M, \mathbf{z})$  using  $I_R$  and  $(\bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$ .

*Proof.*

- Note that from the construction, it suffices to have  $\{S_i\}$ ,  $\bar{\mathbf{a}}$ ,  $\bar{\mathbf{x}}$  to compute  $\mathbf{x}$ . Since  $\{S_i\}$  can be obtained from  $I_R$ , we conclude that Alice can compute  $\mathbf{x}$  using  $I_R$  and  $(\bar{\mathbf{x}}, \bar{\mathbf{a}})$ .
- Note that from the construction, it suffices to have  $\{S_i\}$ ,  $\Gamma_R$ ,  $\bar{\Gamma}(j)$  where  $j \in [k\alpha n]$  to compute  $M$ . Since  $\bar{\Gamma}(j)$  is encoded in  $\bar{M}$  for every  $j \leq k\alpha n$ , and the other information can be obtained from  $I_R$ , we know that  $M$  can be computed from  $I_R$  and  $\bar{M}$ . Finally, since  $\mathbf{z} = \mathbf{z}'$ ,  $\mathbf{z}$  can also be computed from  $I_R$ . We conclude that Bob can compute  $(M, \mathbf{z})$  using  $I_R$  and  $(\bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$ .

□

**Claim 8.3** (The distribution of  $I$ ). *Let  $\bar{I} = (\bar{\mathbf{x}}, \bar{\Gamma}, \bar{\mathbf{b}}, \bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$  be an instance drawn from either the Yes or No distribution of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD and  $I_R = (\mathbf{x}_R, \Gamma_R, \mathbf{b}_R, M_R, \mathbf{z}_R, \mathbf{a}_R)$  be a instance drawn from the Yes distribution of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -advice-SD. Let  $I = (\mathbf{x}, \Gamma, \mathbf{b}, M, \mathbf{z}, \mathbf{a})$  be the result of applying the above reduction on  $\bar{I}$  and  $I_R$ . Then the following hold.*

- $\mathbf{x} \sim \text{Unif}([q]^n)$ .
- $M$  is a uniformly random partial permutation matrix as required in the item 3 of [Definition 6.3](#).
- Suppose there is no error happening in the reduction.
  - If  $\bar{I}$  is a Yes instance, then  $\Pr[\mathbf{z}(j) = 1] = \Pr_{\mathbf{b}(j) \sim \mathcal{A}}[(M\mathbf{x})(j) = \mathbf{b}(j)]$  for every  $j \in [\alpha n]$ .
  - If  $\bar{I}$  is a No instance, then  $\Pr[\mathbf{z}(j) = 1] = \Pr_{\mathbf{b}(j) \sim \mathcal{A}'}[(M\mathbf{x})(j) = \mathbf{b}(j)]$  for every  $j \in [\alpha n]$ .

Namely, if  $\bar{I}$  is a Yes (resp. No) instance of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD, then  $I$  is a Yes (resp. No) instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD.

*Proof.*

- To prove  $\mathbf{x} \sim \text{Unif}([q]^n)$ , observe that  $\mathbf{x}$  is obtained from  $\mathbf{x}_R$  by flipping some of the  $u_i$  to  $v_i$  (and vice versa). In particular, (i)  $\mathbf{x}_R \sim \text{Unif}([q]^n)$  and (ii) the flipping is decided by  $\bar{\mathbf{x}}$ , which is uniformly sampled from  $\{0, 1\}^{\bar{n}}$  and is independent to  $\mathbf{x}_R$ . Note that for a fixed  $\mathbf{x}_R$ ,  $S_i$ , and  $j \in S_i$ , the probability of  $x_j$  being set to  $u_i$  is the same as being set to  $v_i$ . As a result, by symmetry of  $u_i$  and  $v_i$ , we conclude that  $\mathbf{x} \sim \text{Unif}([q]^n)$ .
- By the symmetry of the  $n$  variables,  $M$  is a uniformly random partial permutation matrix as required in the item 3 of [Definition 6.3](#).
- Suppose there is no error happening in the reduction. We consider the following two cases:
  - (i)  $j \in [\alpha n] \setminus U$  and (ii)  $j \in U$ .
    - (i) For each  $j \in [\alpha n] \setminus U$ , by the construction we have  $\mathbf{z}(j) = \mathbf{z}_R(j)$  and hence when fixing  $\mathbf{x}_R, M_R$ , we have  $\Pr[\mathbf{z}(j) = 1] = \Pr[\mathbf{z}_R(j) = 1] = \Pr_{\mathbf{b}_R(j) \sim \mathcal{A}}[(M_R \mathbf{x}_R)(j) = \mathbf{b}_R(j)]$ . We set  $\mathbf{b}(j) = \mathbf{b}_R(j)$  and note that  $\mathbf{b}(j) \sim \mathcal{A}_Y$  (resp.  $\mathbf{b}(j) \sim \mathcal{A}_N$ ) if  $\bar{\mathbf{b}}(j) \sim \bar{\mathcal{A}}_Y$  (resp.  $\bar{\mathbf{b}}(j) \sim \bar{\mathcal{A}}_N$ ) for every  $j \in U$ . Finally, since  $j \notin U$ , there exists  $i \in [k]$  such that  $(M_R \mathbf{x}_R(j))_i = (M \mathbf{x}(j))_i \notin \{u_i, v_i\}$  and hence  $\Pr_{\mathbf{b}_R(j) \sim \mathcal{A}}[(M_R \mathbf{x}_R)(j) = \mathbf{b}_R(j)] = \Pr[(M \mathbf{x})(j) = \mathbf{b}(j)] = 0$ . So we have  $\Pr[\mathbf{z}(j) = 1] = \Pr_{\mathbf{b}(j) \sim \mathcal{A}_Y}[(M \mathbf{x})(j) = \mathbf{b}(j)]$  (resp.  $\Pr[\mathbf{z}(j) = 1] = \Pr_{\mathbf{b}(j) \sim \mathcal{A}_N}[(M \mathbf{x})(j) = \mathbf{b}(j)]$ ) if  $\bar{I}$  is a Yes (resp. No) instance as desired.
    - (ii) For each  $j \in U$ , by construction we have  $\mathbf{z}(j) = \bar{\mathbf{z}}(V(j))$ . We set

$$\mathbf{b}(j)_i = \begin{cases} u_i & \text{if } u_i < v_i \text{ and } \bar{\mathbf{b}}(V(j))_i = 0 \\ u_i & \text{if } u_i > v_i \text{ and } \bar{\mathbf{b}}(V(j))_i = 1 \\ v_i & \text{otherwise.} \end{cases}$$

First, observe that  $\mathbf{z}(j) = 1$  iff  $(M \mathbf{x})(j) = \mathbf{b}(j)$ . To see this, note that

$$\begin{aligned} \mathbf{z}(j) = 1 &\Leftrightarrow \bar{\mathbf{z}}(V(j)) = 1 \\ &\Leftrightarrow (\bar{M} \bar{\mathbf{x}})(V(j)) = \bar{\mathbf{b}}(V(j)). \end{aligned}$$

For each  $i \in [k]$ , if  $u_i < v_i$  and  $\bar{\mathbf{b}}(V(j))_i = (\bar{M}\bar{\mathbf{x}})(V(j))_i = 0$ , we have  $\mathbf{b}(j)_i = (M\mathbf{x})(j)_i = u_i$ . Similarly, for all the other situations we have  $\mathbf{b}(j)_i = (M\mathbf{x})(j)_i$  and hence the equation becomes

$$\Leftrightarrow (M\mathbf{x})(j) = \mathbf{b}(j)$$

as desired.

Next, observe that if  $\bar{I}$  is a Yes (resp. No) instance, then  $\mathbf{b}(j) \sim \mathcal{A}_Y$  (resp.  $\mathbf{b}(j) \sim \mathcal{A}_N$ ). We analyze the two cases as follows.

- If  $\bar{I}$  is a Yes instance, we have  $\bar{\mathbf{b}}(V(j)) \sim \bar{\mathcal{A}}_Y = \text{Unif}(\{\bar{\mathbf{u}}, \bar{\mathbf{v}}\})$ . Recall that  $(\bar{u}_i, \bar{v}_i) = (0, 1)$  if  $u_i < v_i$  and  $(\bar{u}_i, \bar{v}_i) = (1, 0)$  otherwise. Now observe that, by the above choice of  $\mathbf{b}(j)$ , we have  $\bar{\mathbf{b}}(V(j)) = \bar{\mathbf{u}}$  iff  $\mathbf{b}(j) = \mathbf{u}$  (resp.  $\bar{\mathbf{b}}(V(j)) = \bar{\mathbf{v}}$  iff  $\mathbf{b}(j) = \mathbf{v}$ ). Thus, we have  $\mathbf{b}(j) \sim \mathcal{A}_Y$  as desired.
- If  $\bar{I}$  is a No instance, we have  $\bar{\mathbf{b}}(V(j)) \sim \bar{\mathcal{A}}_N = \text{Unif}(\{\bar{\mathbf{u}} \vee \bar{\mathbf{v}}, \bar{\mathbf{u}} \wedge \bar{\mathbf{v}}\})$ . Recall that for each  $i \in [k]$ ,  $u_i \vee v_i = \max\{u_i, v_i\}$  and  $u_i \wedge v_i = \min\{u_i, v_i\}$ . Now observe that, by the above choice of  $\mathbf{b}(j)$ , we have  $\bar{\mathbf{b}}(V(j)) = \bar{\mathbf{u}} \vee \bar{\mathbf{v}}$  iff  $\mathbf{b}(j) = \mathbf{u} \vee \mathbf{v}$  (resp.  $\bar{\mathbf{b}}(V(j)) = \bar{\mathbf{u}} \wedge \bar{\mathbf{v}}$  iff  $\mathbf{b}(j) = \mathbf{u} \wedge \mathbf{v}$ ). Thus, we have  $\mathbf{b}(j) \sim \mathcal{A}_N$  as desired.

To sum up, for each  $j \in U$ , we have  $\Pr[\mathbf{z}(j) = 1] = \Pr_{\mathbf{b}(j) \sim \mathcal{A}_Y}[(M\mathbf{x})(j) = \mathbf{b}(j)]$  (resp.  $\Pr[\mathbf{z}(j) = 1] = \Pr_{\mathbf{b}(j) \sim \mathcal{A}_N}[(M\mathbf{x})(j) = \mathbf{b}(j)]$ ) if  $\bar{I}$  is a Yes (resp. No) instance as desired.  $\square$

#### Step 4: An error occurs with low probability.

**Claim 8.4.** *When  $n$  is sufficiently large, the probability of an error happening in the reduction is at most  $2^{-\Omega((2/q)^{2k}\alpha n)}$ .*

*Proof.* Recall that for given  $\bar{n}, k, \bar{\alpha}, q, \delta$ , we let  $n = 2q\bar{n}$  and  $\alpha = q^{k-1}2^{-(k+2)}\bar{\alpha}$ .

Note that  $U$  is a sum of  $\alpha n$  i.i.d.  $\text{Bern}((2/q)^k)$ . So by concentration inequality, we have  $\Pr[|U| > 2(2/q)^k\alpha n] < 2^{-\Omega((2/q)^{2k}\alpha n)}$ . By the choice of parameters, we have  $2(2/q)^k\alpha n \leq \bar{\alpha}\bar{n}$ . Thus, type (1) error happens with probability at most  $2^{-\Omega((2/q)^{2k}\alpha n)}$ .

Note that by the choice of parameters, we have  $|X_i| = |U| \leq \bar{n}/k$  and hence type (2) error happens only when  $|U| + |W_i| < \bar{n}/k$  for some  $i \in [k]$ . For each  $i \in [k]$ , note that  $|W_i|$  is a sum of  $n/k - \alpha n$  i.i.d.  $\text{Bern}(2/q)$ . So by concentration inequality, we have  $\Pr[|W_i| < (n/k - \alpha n)/q] < 2^{-\Omega((1/q)^2(n/k - \alpha n))}$ . By the choice of parameters, we have  $(n/k - \alpha n)/q \geq \bar{n}/k$ . Thus, type (2) error happens with probability at most  $2^{-\Omega((1/q)^2(n/k - \alpha n))} \leq 2^{-\Omega((2/q)^{2k}\alpha n)}$ .  $\square$

#### Step 5: Proof of Lemma 8.1.

*Proof of Lemma 8.1.* For every  $\bar{n}, k, \bar{\alpha}, q, \delta$ , we let  $n = 2q\bar{n}$  and  $\alpha = q^{k-1}2^{-(k+2)}\bar{\alpha}$ . Suppose there is a protocol for  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD using  $C(n)$  bits of communication and achieving advantage  $\delta$ . We show that how to get a protocol  $\bar{\Pi}$  for  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD with parameters  $(\bar{n}, \bar{\alpha})$  using  $C(n)$  bits of communication and achieving advantage  $\delta/2$ .

Let  $\bar{I} = (\bar{\mathbf{x}}, \bar{\Gamma}, \bar{\mathbf{b}}, \bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$  be an instance drawn from either the Yes or No distribution of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD where  $(\bar{\mathbf{x}}, \bar{\mathbf{a}})$  is Alice's private input and  $(\bar{M}, \bar{\mathbf{z}}, \bar{\mathbf{a}})$  is Bob's private input. The protocol  $\bar{\Pi}$  works as follows. Alice and Bob first use their private input and the shared randomness

to compute  $\mathbf{x}$  and  $(M, \mathbf{z})$  respectively. This can be done due to [Claim 8.2](#). Next, Alice and Bob simply invoke the protocol  $\bar{\Pi}$  on the new instance  $\mathbf{x}$  and  $(M, \mathbf{z})$  and output the result accordingly.

It is immediate to see that  $\bar{\Pi}$  only uses  $C(n)$  bits of communication. To show that  $\bar{\Pi}$  has advantage at least  $\delta/2$ , we first show that the joint distribution of  $(\mathbf{x}, M, \mathbf{z})$  is the same as that from an instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD if there is no error in the reduction. By [Claim 8.3](#),  $\mathbf{x} \sim \text{Unif}([q]^n)$  and  $M$  follows the distribution as required in the item 3 of [Definition 6.3](#). When there is no error in the reduction and  $\bar{I}$  is sampled from the Yes (resp. No) distribution of  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD, [Claim 8.3](#) implies that  $\mathbf{z}$  follows the conditional distribution (conditioned on  $\mathbf{x}$  and  $M$ ) of a Yes (resp. No) instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD as required in the item 5 of [Definition 6.3](#). Next, [Claim 8.4](#) shows that the probability of an error happening in the reduction is at most  $\delta/2$ . Finally, by triangle inequality, we conclude that  $\bar{\Pi}$  has advantage at least  $\delta/2$  in solving  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD.

To conclude, by [Theorem 6.4](#), any protocol for  $(\bar{\mathcal{A}}_Y, \bar{\mathcal{A}}_N)$ -advice-SD with advantage  $\delta/2$  requires  $\bar{\tau}\sqrt{n}$  bits of communication. Thus, we have  $C(n) \geq \bar{\tau}\sqrt{n} \geq \tau\sqrt{n}$  for some constant  $\tau > 0$ .  $\square$

## 8.2 Indistinguishability of shifting distributions

In this subsection, we prove the following lemma which was used in [Section 7.2](#) for reducing a single-function SD to a multi-function SD, and will be used in [Section 8.3](#) for reductions between various SD problems.

**Lemma 7.13.** *Let  $n, k, q \in \mathbb{N}$ ,  $\alpha \in (0, 1)$  where  $k, q, \alpha$  are constants with respect to  $n$  and  $\alpha n$  is an integer less than  $n/k$ . Let  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$ . For every  $\varepsilon, \delta \in (0, 1]$ , there exist  $n' = \Omega(n)$  and constants  $\alpha', \delta' \in (0, 1)$  such that the following holds. For every distributions  $\mathcal{D}_Y, \mathcal{D}_N, \mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2 \in \Delta(\mathcal{F} \times [q]^k)$  such that  $\mathcal{D}_Y = (1 - \varepsilon)\mathcal{D}_0 + \varepsilon\mathcal{D}_1$  and  $\mathcal{D}_N = (1 - \varepsilon)\mathcal{D}_0 + \varepsilon\mathcal{D}_2$  and for every  $c \in \mathbb{N}$ , suppose there exists a protocol for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD with parameters  $n$  and  $\alpha$  using  $c$  bits of communication with advantage  $\delta$ , then there exists a protocol for  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD with parameters  $n'$  and  $\alpha'$  using  $c$  bits of communication with advantage  $\delta'$ .*

*Proof.* Given the parameters  $n, \alpha$  and  $\varepsilon \in (0, 1)$ , define  $n' = \varepsilon n$  and  $\alpha' = 2\alpha$ .

Let  $(\mathbf{x}', M', \mathbf{b}', \mathbf{z}')$  be an instance of the  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD problem where  $\mathbf{x}' \in [q]^{n'}$ ,  $M' \in \{0, 1\}^{k\alpha'n' \times n'}$ ,  $\mathbf{b}' \in [q]^{k\alpha'n'}$ ,  $\mathbf{z}' \in \{0, 1\}^{\alpha'n'}$ . Let  $R'$  be the shared randomness defined later. We specify the map  $(\mathbf{x}', M', \mathbf{b}', \mathbf{z}', R') \mapsto (\mathbf{x}, M, \mathbf{b}, \mathbf{z})$  where  $\mathbf{x} \in [q]^n$ ,  $M \in \{0, 1\}^{k\alpha n \times n}$ ,  $\mathbf{b} \in [q]^{k\alpha n}$ ,  $\mathbf{z} \in \{0, 1\}^{\alpha n}$ .

### A reduction from $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD to $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD

Let  $\mathbf{y} \sim \text{Unif}([q]^{n-n'})$ ,  $\mathbf{w} \sim \text{Bern}(2\varepsilon)^{\alpha n}$ . Let  $\Gamma \in \{0, 1\}^{n \times n}$  be a uniform permutation matrix. Let  $\mathbf{c} = (\mathbf{c}(1), \dots, \mathbf{c}((n - n')/k))$  where  $\mathbf{c}(i) \sim \mathcal{D}$  are chosen independently.

- Let  $R' = (\mathbf{y}, \mathbf{w}, \Gamma, \mathbf{c})$  be the shared randomness.

Let  $\#_w(i) = |\{j \in [i] \mid w_j = 1\}|$  denote the number 1's among the first  $i$  coordinates of  $\mathbf{w}$ . If  $\#_w(\alpha n) \geq \alpha'n'$  or if  $\alpha n - \#_w(\alpha n) \geq (n - n')/k$  we declare an error, Note  $\mathbb{E}[\#_w(n)] = \alpha'n'/2$  so the probability of error is negligible (specifically it is  $\exp(-n)$ ).

Given  $(\mathbf{x}', M', \mathbf{b}', \mathbf{z}', R')$ , we now define  $(\mathbf{x}, M, \mathbf{b}, \mathbf{z})$  as follows.

- Let  $\mathbf{x} = \Gamma(\mathbf{x}', \mathbf{y})$  so  $\mathbf{x}$  is a random permutation of the concatenation of  $\mathbf{x}'$  and  $\mathbf{y}$ .
- Let  $M' = (M'_1, \dots, M'_{\alpha'n'})$  where  $M'_i \in \{0, 1\}^{k \times n'}$ . We extend  $M'_i$  to  $N_i \in \{0, 1\}^{k \times n}$

by adding all-zero columns to the right. For  $i \in \{1, \dots, (n - n')/k\}$ , let  $P_i \in \{0, 1\}^{k \times n}$  be given by  $(P_i)_{j\ell} = 1$  if and only if  $\ell = n' + (i - 1)k + j$ . Next we define a matrix  $\tilde{M} \in \{0, 1\}^{k\alpha n \times n} = (\tilde{M}_1, \dots, \tilde{M}_{\alpha n})$  where  $\tilde{M}_i \in \{0, 1\}^{k \times n}$  is defined as follows: If  $w_i = 1$  then we let  $\tilde{M}_i = N_{\#_w(i)}$  else we let  $\tilde{M}_i = P_{i - \#_w(i)}$ . Finally we let  $M = \tilde{M} \cdot \Gamma^{-1}$ .

- Let  $\mathbf{b} = (\mathbf{b}(1), \dots, \mathbf{b}(\alpha n))$  where  $\mathbf{b}(i) = \mathbf{b}'(\#_w(i))$  if  $w_i = 1$ , otherwise  $\mathbf{b}(i) = \mathbf{c}(i - \#_w(i))$ .
- Let  $z_i = 1$  if and only if  $M_i \mathbf{x} = \mathbf{b}(i)$  for every  $i \in [\alpha n]$ .

Now, we verify that the reduction satisfies the following success conditions.

### Success conditions for the reduction

- (1) **The reduction is locally well-defined.** Namely, there exist random strings  $R'$  so that (i) Alice can get  $\mathbf{x}$  through a map  $(\mathbf{x}', R') \mapsto \mathbf{x}$  while Bob can get  $(M, \mathbf{z})$  through a map  $(M', \mathbf{z}', R') \mapsto (M, \mathbf{z})$ .
- (2) **The reduction is sound and complete.** Namely, (i)  $z_i = 1$  if and only if  $M_i \mathbf{x} = \mathbf{b}(i)$  for all  $i \in [\alpha n]$ . (ii) If  $\mathbf{b}' \sim \mathcal{D}_1^{\alpha' n'}$ , then  $\mathbf{b} \sim \mathcal{D}_Y^{\alpha n}$ . Similarly if  $\mathbf{b}' \sim \mathcal{D}_2^{\alpha' n'}$ , then  $\mathbf{b} \sim \mathcal{D}_N^{\alpha n}$ . (iii)  $\mathbf{x} \sim \text{Unif}([q]^n)$  and  $M$  is a uniformly random matrix conditioned on having exactly one “1” per row and at most one “1” per column.

**Claim 8.5.** *If  $\#_w(\alpha n) \leq \alpha' n'$  and  $\alpha n - \#_w(\alpha n) \leq (n - n')/k$ , then the second map in the reduction is locally well-defined, sound, and complete. In particular, the error event happens with probability at most  $\exp(-\Omega(n))$  over the randomness of  $R'$ .*

*Proof.* To see the reduction is locally well-defined, first note that Alice can compute  $\mathbf{x} = \Gamma(\mathbf{x}', \mathbf{y})$  from  $\mathbf{x}'$  and the shared randomness  $R'$  locally. As for Bob, note that the maximum index needed for  $N$  and  $\mathbf{b}'$  (resp.  $P$  and  $\mathbf{c}$ ) is at most  $\#_w(\alpha n)$  (resp.  $\alpha n - \#_w(i)$ ). Namely, if  $\#_w(\alpha n) \leq \alpha' n'$  and  $\alpha n - \#_w(\alpha n) \leq (n - n')/k$ , then  $M$  and  $\mathbf{b}$  are well-defined. Note that this happens with probability at least  $1 - 2^{-\Omega(n)}$ . Also, one can verify from the construction that  $M$  and  $\mathbf{b}$  can be locally computed by  $M'$ ,  $\mathbf{b}'$ , and the shared randomness  $R'$ .

To see the reduction is sound and complete, (i)  $z_i = 1$  if and only if  $M_i \mathbf{x} = \mathbf{b}(i)$  for every  $i \in [\alpha n]$  directly follows from the construction. As for (ii), if  $\mathbf{b}' \sim \mathcal{D}_1^{\alpha' n'}$ . Now, for each  $i \in [\alpha n]$ ,  $\mathbf{b}(i) = \mathbf{b}'(\#_w(i))$  with probability  $\varepsilon$  and  $\mathbf{b}(i) = \mathbf{c}(i - \#_w(i))$  with probability  $1 - \varepsilon$ . As  $\mathbf{b}'(i') \sim \mathcal{D}_1$  for every  $i' \in [\alpha' n']$  and  $\mathbf{c}(i') \sim \mathcal{D}_0$  for every  $i' \in [(n - n')/k]$ , we have  $\mathbf{b}(i) \sim (1 - \varepsilon)\mathcal{D}_0 + \varepsilon\mathcal{D}_1 = \mathcal{D}_Y$  as desired. Similarly, one can show that if  $\mathbf{b}' \sim \mathcal{D}_2^{\alpha' n'}$ , then for every  $i' \in [\alpha' n']$  we have  $\mathbf{b}(i') \sim \mathcal{D}_N$ . Finally, we have  $\mathbf{x} \sim \text{Unif}([q]^n)$  and  $M$  is a uniformly random matrix with exactly one “1” per row and at most one “1” per column (due to the application of a random permutation  $\Gamma$ ) by construction.

This completes the proof of the success conditions (1)-(2) for the reduction.  $\square$

To wrap up the proof of [Lemma 7.13](#), suppose there is a protocol  $\Pi$  for  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD with parameters  $n$  and  $\alpha$  using  $c$  bits of communication with advantage  $\delta$ . We describe a protocol  $\Pi'$  for  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD with parameters  $n'$  and  $\alpha'$  using  $c$  bits of communication with advantage at least



$\delta - 2^{-\Omega(n)}$ .

Let  $(\mathbf{x}', M', \mathbf{b}', \mathbf{z}')$  be an instance of the  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD problem where  $\mathbf{x}' \in [q]^{n'}$ ,  $M' \in \{0, 1\}^{k\alpha'n' \times n'}$ ,  $\mathbf{b}' \in [q]^{k\alpha'n'}$ ,  $\mathbf{z}' \in \{0, 1\}^{\alpha'n'}$ . Let  $R'$  be the shared randomness defined above. In the new protocol  $\Pi'$ , Alice and Bob computes their private inputs  $\mathbf{x}$  and  $(M, \mathbf{z})$  respectively. By [Claim 8.5](#), the computation can be done locally with their original private inputs and the shared randomness. Also, with probability at least  $1 - 2^{-\Omega(n)}$ , the Yes (resp. No) instance of  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$ -SD is mapped to the Yes (resp. No) instance of  $(\mathcal{F}, \mathcal{D}_Y, \mathcal{D}_N)$ -SD. Namely, by directly applying  $\Pi$  on the new inputs, Alice and Bob can achieve  $\delta - 2^{-\Omega(n)}$  advantage on  $(\mathcal{F}, \mathcal{D}_1, \mathcal{D}_2)$  using the same amount of communication as desired.  $\square$

### 8.3 Proof of Theorem 7.4

Let  $\mathbf{u}, \mathbf{v}$  be incomparable, let  $S = \{i \in [k] \mid u_i \neq v_i\}$ , and let  $k'' = |S|$ .

**Step 1: Specify the auxiliary distributions.**

- Let  $\mathcal{A}_Y = \text{Unif}(\{\mathbf{u}|_S, \mathbf{v}|_S\})$  and  $\mathcal{A}_N = \text{Unif}(\{\mathbf{u}|_S \vee (\mathbf{v}|_S), \mathbf{u}|_S \wedge (\mathbf{v}|_S)\})$ . By [Lemma 8.1](#),  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD requires  $\tau\sqrt{n}$  space.
- Let  $\mathcal{D}_1 = \text{Unif}(\{\mathbf{u}, \mathbf{v}\})$  and  $\mathcal{D}_2 = \text{Unif}(\{\mathbf{u} \vee \mathbf{v}, \mathbf{u} \wedge \mathbf{v}\})$ .
- Finally, there exists  $\mathcal{D}_0$  such that we have  $\mathcal{D} = (1 - 2\varepsilon)\mathcal{D}_0 + 2\varepsilon\mathcal{D}_1$  and  $\mathcal{D}_{\mathbf{u}, \mathbf{v}} = (1 - 2\varepsilon)\mathcal{D}_0 + 2\varepsilon\mathcal{D}_2$ .

In the following, we are going to describe reduction from  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD with parameters  $(n'', \alpha'', k'')$  to  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD with parameters  $(n', \alpha', k)$ . And by [Lemma 7.13](#), there exists a reduction from  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD with parameters  $(n', \alpha', k)$  to  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD with parameters  $(n, \alpha, k)$ .

**Step 2: Overview of the reduction from  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD to  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD.** Let  $\Pi$  be a protocol for  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD with parameter  $\alpha \leq 1/(200k)$  using  $C(n)$  communication bits to achieve advantage  $\delta$  on instances of length  $n$ . We let  $n'' = (k''\varepsilon/k)n$ ,  $\alpha'' = (2k/k'')\alpha$  and design a protocol  $\Pi''$  for  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD with parameter  $\alpha''$  achieving advantage at least  $\delta/2$  on instances of length  $n''$  using  $C''(n'') = C(n)$  communication. Thus, by [Lemma 8.1](#), there exists a constant  $\tau'' > 0$  such that  $C(n) = C''(n'') \geq \tau''\sqrt{n''} = \tau''\sqrt{(k''\varepsilon/k)}\sqrt{n}$  as desired.

To construct such reduction, we first reduce the above instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD to an instance of  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD with parameters  $n' = kn''/k''$  and  $\alpha' = \alpha''n''/n'$ . Next, we invoke [Lemma 7.13](#) to get a protocol  $\Pi'$  (from  $\Pi$ ) which achieves  $\delta - 2^{-\Omega(n)}$  advantage on  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD using  $C(n)$  communication.

Without loss of generality, we assume  $\Pi'$  is deterministic and our new protocol  $\Pi''$  for  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD uses shared randomness between Alice and Bob. The protocol  $\Pi''$  consists a map:  $(\mathbf{x}'', M'', \mathbf{b}'', \mathbf{z}'', R'') \mapsto (\mathbf{x}', M', \mathbf{b}', \mathbf{z}')$ .

Before describing the map, let us first state the desired conditions.

#### Success conditions for the reduction

- (1) **The reduction is locally well-defined.** Namely, there exist a random string  $R''$  so that (i) Alice can get  $\mathbf{x}'$  through the maps  $(\mathbf{x}'', R'') \mapsto \mathbf{x}'$  while Bob can get  $(M', \mathbf{z}')$  through the map  $(M'', \mathbf{z}'', R'') \mapsto (M', \mathbf{z}')$ .



- (2) **The reduction is sound and complete.** Namely, (i)  $z'_i = 1$  if and only if  $M'_i \mathbf{x}' = \mathbf{b}'(i)$  for all  $i \in [\alpha' n']$ . (ii) If  $\mathbf{b}'' \sim \mathcal{A}_Y^{\alpha'' n''}$  then  $\mathbf{b}' \sim \mathcal{D}_1^{\alpha' n'}$ . Similarly if  $\mathbf{b}'' \sim \mathcal{A}_N^{\alpha'' n''}$  then  $\mathbf{b}' \sim \mathcal{D}_2^{\alpha' n'}$ . (iii)  $\mathbf{x}' \sim \text{Unif}([q]^{n'})$  and  $M'$  is a uniformly random matrix conditioned on having exactly one “1” per row and at most one “1” per column.

**Step 3: Specify and analyze the reduction from  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD to  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD.** We now specify the first map mentioned above and prove that it satisfies conditions (1)-(2).

#### A reduction from $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD to $(\mathcal{D}_1, \mathcal{D}_2)$ -SD

- Let  $R'' \sim \text{Unif}([q]^{n' - n''})$  be the shared randomness.

Given  $(\mathbf{x}'', M'', \mathbf{b}'', \mathbf{z}'', R'')$  we define  $(\mathbf{x}', M', \mathbf{b}', \mathbf{z}')$  as follows. To get  $M', \mathbf{z}'$  and  $\mathbf{b}'$  we need some more notations. First, note that  $\alpha'' n'' = \alpha' n'$  due to the choice of parameters.

- Let  $\mathbf{x}' = (\mathbf{x}'', R'')$ .
- $M'$  can be viewed as the stacking of matrices  $M''_1, \dots, M''_{\alpha'' n''} \in \{0, 1\}^{k'' \times n''}$ . We first extend  $M''_i$  by adding all-zero columns at the end to get  $N''_i \in \{0, 1\}^{k'' \times n'}$ . We then stack  $N''_i$  on top of  $P''_i \in \{0, 1\}^{(k - k'') \times n'}$  to get  $M'_i$ , where  $(P''_i)_{j\ell} = 1$  if and only if  $\ell = n'' + (i - 1)k + j$ . We let  $M'$  be the stacking of  $M'_1, \dots, M'_{\alpha' n'}$ .
- Let  $\mathbf{b}'' = (\mathbf{b}''(1), \dots, \mathbf{b}''(\alpha'' n''))$ . Let  $\tilde{\mathbf{u}} = (u_{k''+1}, \dots, u_k)$  denote the common parts of  $\mathbf{u}$  and  $\mathbf{v}$ . We let  $\mathbf{b}'(i) = (\mathbf{b}''(i), \tilde{\mathbf{u}})$  and  $\mathbf{b}' = (\mathbf{b}'(1), \dots, \mathbf{b}'(\alpha' n'))$ .
- Let  $z'_i = 1$  if and only if  $M'_i \mathbf{x}' = \mathbf{b}'(i)$  for all  $i \in [\alpha' n']$  as required.

**Claim 8.6.** *The above reduction is locally well-defined, sound, and complete.*

*Proof.* To see the map is locally well-defined, note that Alice can compute  $\mathbf{x}' = (\mathbf{x}'', R'')$  locally. Similarly, Bob can compute  $M'$  locally by construction. As for  $\mathbf{z}'$ , note that for every  $i \in [\alpha' n']$ ,  $z'_i = 1$  if and only if  $z''_i = 1$  and  $P''_i \mathbf{x}' = \tilde{\mathbf{u}}$ . Since Bob has  $\mathbf{z}'$  and can locally compute  $P''_i \mathbf{x}'$  for every  $i$ , her can also compute  $\mathbf{z}'$  locally.

To see the map is sound and complete, (i)  $z'_i = 1$  if and only if  $M'_i \mathbf{x}' = \mathbf{b}'(i)$  follows from the construction. As for (ii), for each  $i \in [\alpha' n'] = [\alpha'' n'']$ , if  $\mathbf{b}''_i \sim \mathcal{A}_Y = \text{Unif}(\{\mathbf{u}|_S, \mathbf{v}|_S\})$ , then  $\mathbf{b}'_i \sim \text{Unif}(\{(\mathbf{u}|_S, \tilde{\mathbf{u}}), (\mathbf{v}|_S, \tilde{\mathbf{u}})\}) = \text{Unif}(\{\mathbf{u}, \mathbf{v}\}) = \mathcal{D}_1$  as desired. Similarly, one can show that if  $\mathbf{b}''_i \sim \mathcal{A}_N$ , then  $\mathbf{b}'_i \sim \mathcal{D}_1$ . Finally, we have  $\mathbf{x}' \sim \text{Unif}([q]^{n'})$  by construction and hence (iii) holds.

This completes the proof of conditions (1)-(2) for the reduction.  $\square$

#### Step 4: Proof of [Theorem 7.4](#).

*Proof of [Theorem 7.4](#).* Let us start with setting up the parameters. Given  $k \in (0, 1/(200k))$ ,  $\alpha, n, \mathcal{D}$ , and incomparable pair  $(\mathbf{u}, \mathbf{v}) \in \text{supp}(\mathcal{D})$  and polarization amount  $\varepsilon = \varepsilon(\mathcal{D}, \mathbf{u}, \mathbf{v})$ , let  $k'' = |\{i \in [k] \mid u_i \neq v_i\}|$ ,  $n'' = (k'' \varepsilon / k) n$ ,  $\alpha'' = (2k / k'') \alpha$ ,  $n' = kn'' / k''$ ,  $\alpha' = \alpha'' n' / n'$ , and  $\delta'' = \delta / 2$ .

Now, for the sake of contradiction, we assume that there exists a protocol  $\Pi = (\Pi_A, \Pi_B)$  for  $(\mathcal{D}, \mathcal{D}_{\mathbf{u}, \mathbf{v}})$ -SD with advantage  $\delta$  and at most  $\tau \sqrt{n}$  bits of communication.

First, by [Claim 8.6](#), if  $(\mathbf{x}'', M'', \mathbf{z}'')$  is a Yes (resp. No) instance of  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD, then the output of the reduction, i.e.,  $(\mathbf{x}', M', \mathbf{z}')$ , is a Yes (resp. No) instance of  $(\mathcal{D}_1, \mathcal{D}_2)$ -SD. Next, Alice and Bob run the protocol  $\Pi'$  from [Lemma 7.13](#) on  $(\mathbf{x}', M', \mathbf{z}')$ . By the correctness of the reduction as well as the protocol  $\Pi'$ , we know that Alice and Bob have advantage at least  $\delta - \exp(-\Omega(n)) \geq \delta/2 = \delta''$  in solving  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD with at most  $\tau\sqrt{n} = \tau\sqrt{(k/(k''\varepsilon))n''}$  bits of communication.

Finally, by [Lemma 8.1](#), we know that there exists a constant  $\tau_0 > 0$  such that any protocol for  $(\mathcal{A}_Y, \mathcal{A}_N)$ -SD with advantage  $\delta''$  requires at least  $\tau_0\sqrt{n''}$  bits of communication. This implies that  $\tau \geq \tau_0\sqrt{k''\varepsilon/k}$ . We conclude that any protocol for  $(\mathcal{D}, \mathcal{D}_{\mathbf{u},\mathbf{v}})$ -SD with advantage  $\delta$  requires at least  $\tau\sqrt{n}$  bits of communication.  $\square$

## 9 Dichotomy for exact Computation

In this section we prove [Theorem 2.22](#). For this, we will use tight bounds on the randomized communication complexity of the Disjointness (Disj) and Gap Hamming Distance (GHD) problems.

**Definition 9.1** (Disjointness (Disj)). *In the  $\text{Disj}_n$  problem, Alice and Bob receive binary strings  $x, y \in \{0, 1\}^n$  of Hamming weight  $\Delta(x) = \Delta(y) = n/4$ , respectively. If the Hamming distance  $\Delta(x, y) = n/2$  the players must output 1, if  $\Delta(x, y) < n/2$  they must output 0.*

**Definition 9.2** (Gap Hamming Distance (GHD)). *In the  $\text{GHD}_{n,t,g}$  problem, Alice and Bob receive binary strings  $x, y \in \{0, 1\}^n$ , respectively. If the Hamming distance  $\Delta(x, y) \geq t + g$  the players must output 1, if  $\Delta(x, y) \leq t - g$  they must output 0, otherwise they may output either 0 and 1.*

The following results give tight bounds on the randomized communication complexity of Disj and GHD.

**Theorem 9.3** ([\[KS92, Raz90\]](#)). *For all large enough  $n$ , any randomized protocol solving  $\text{Disj}_n$  with probability  $2/3$  must use  $\Omega(n)$  bits of communication.*

**Theorem 9.4** ([\[CR12, Vid12, She12\]](#)). *For every  $a \in (0, 1/2]$  and every  $g \geq 1$ , and all large enough  $n$  the following holds. If  $t \in [an, (1 - a)n]$ , then any randomized protocol solving  $\text{GHD}_{n,t,g}$  with probability  $2/3$  must use  $\Omega(\min\{n, n^2/g^2\})$  bits of communication.*

Equipped with these results, we are ready to prove [Theorem 2.22](#).

**Theorem 2.22.** *For every  $q, k \in \mathbb{N}$ , every family of functions  $\mathcal{F} \subseteq \{f : [q]^k \rightarrow \{0, 1\}\}$  the following hold:*

1. *If  $\mathcal{F}$  is constant satisfiable, then there exists a deterministic dynamic streaming algorithm that uses  $O(\log n)$  space and solves  $\text{Max-CSP}(\mathcal{F})$  exactly optimally.*
2. *If  $\mathcal{F}$  is not constant satisfiable, then the following hold in the insertion-only and dynamic settings:*
  - (a) *Every probabilistic algorithm solving  $\text{Max-CSP}(\mathcal{F})$  exactly requires  $\Omega(n)$  space.*
  - (b) *For every  $\varepsilon = \varepsilon(n) > 0$ ,  $(1, 1 - \varepsilon)$ - $\text{Max-CSP}(\mathcal{F})$  requires  $\Omega(\min\{n, \varepsilon^{-1}\})$ -space<sup>12</sup> on sufficiently large inputs.*

<sup>12</sup> The constant hidden in the  $\Omega$  depends on  $\mathcal{F}$ , but (obviously) not on  $\varepsilon$ .

- (c) For  $\rho_{\min}(\mathcal{F})$  defined in [Definition 2.11](#), for every  $\rho_{\min}(\mathcal{F}) < \gamma < 1$  and every  $\varepsilon = \varepsilon(n) > 0$ ,  $(\gamma, \gamma - \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) requires  $\Omega(\min\{n, \varepsilon^{-2}\})$ -space<sup>7</sup> on sufficiently large inputs.

While this theorem doesn't give tight bounds on the space complexity of Max-CSP( $\mathcal{F}$ ) in terms of  $n$ , the dependence on  $\varepsilon$  is tight. For every family of functions  $\mathcal{F}$ , if we sample  $O(n/\varepsilon^2)$  random constraints, then by the Chernoff bound we preserve the values of all assignments within a factor of  $1 \pm \varepsilon$ .

*Proof.* For the first item of the theorem, we note that the maximum number of simultaneously satisfiable constraints in a  $\sigma$ -satisfiable formula is the number of non-zero constraints  $f \in \mathcal{F} \setminus \{\mathbf{0}\}$  in it. This can be computed in space  $O(\log n)$ .

Now we turn to the proof of the second item of the theorem in the insertion-only setting (and this implies the same lower bound in the dynamic setting). To this end, first we prove that there exists an unsatisfiable instance  $I$  of Max-CSP( $\mathcal{F} \setminus \{\mathbf{0}\}$ ). Let  $I$  be an instance on  $kq$  variables that has every constraint from  $\mathcal{F} \setminus \{\mathbf{0}\}$  applied to every (unordered)  $k$ -tuple of distinct variables. Any assignment  $\nu \in [q]^{kq}$  has at least  $k$  equal coordinates. That is, there exists  $\sigma \in [q]$  such that  $\Sigma = \{i: \nu_i = \sigma\}$  has size  $|\Sigma| \geq k$ . Since  $\mathcal{F}$  is not  $\sigma$ -satisfiable, there exists a function  $f \in \mathcal{F} \setminus \{\mathbf{0}\}$  that  $f(\sigma^k) \neq 1$ . Thus, the corresponding constraint of  $I$  is not satisfied by  $\nu$ .

Now we pick a minimal unsatisfied formula  $J$  on  $kq$  variables with constraints from  $\mathcal{F} \setminus \{\mathbf{0}\}$ , that is a formula such that all proper subsets of the constraints of  $J$  can be simultaneously satisfied. Since  $J$  doesn't have zero-constraints,  $J$  must have at least two constraints. We partition  $J$  into two arbitrary non-empty subsets of constraints  $J = J_A \sqcup J_B$ . Note that by minimality of  $J$ ,  $J_A$  and  $J_B$  are both satisfiable.

Observe that item 2(a) of the theorem follows from 2(b) by setting  $\varepsilon = \Theta(1/n)$ . In order to prove the item 2(b), we reduce Disj $_m$  for  $m = |J|^{-1}\varepsilon^{-1}$  to Max-CSP( $\mathcal{F}$ ) on  $n$  variables. We can assume that  $\varepsilon \geq \frac{kq}{n|J|}$ , as for smaller  $\varepsilon$  the optimal lower bound of  $\Omega(n)$  is implied by this setting. We partition the  $n$  variables of Max-CSP( $\mathcal{F}$ ) into at least  $m$  groups of size  $kq$ . Let  $x, y \in \{0, 1\}^m$  be the inputs of Alice and Bob in the Disj $_m$  problem. If  $x_i = 1$ , then Alice applies the constraints  $J_A$  to the  $i$ th block of  $kq$  variables of the formula. Similarly, if  $y_i = 1$ , then Bob applies the constraints  $J_B$  to the  $i$ th block of  $kq$  variables. Let  $C_A$  and  $C_B$  be the sets of constraints produced by Alice and Bob, respectively, and let  $\Psi = C_A \cup C_B$ . Since  $\Delta(x) = \Delta(y) = m/4$ , the total number of constraints in the formula  $|\Psi| = |J|m/4$ . Note that  $\Psi$  is satisfiable if and only if Disj( $x, y$ ) = 1. Therefore, if  $x$  and  $y$  are disjoint, then  $\text{val}(C_A \cup C_B) = 1$ , otherwise

$$\text{val}(\Psi) \leq 1 - \frac{4}{|J|m} < 1 - \varepsilon.$$

Any streaming algorithm that receives constraints  $C_A$  and  $C_B$  and solves  $(1, 1 - \varepsilon)$ -Max-CSP( $\mathcal{F}$ ) with probability  $2/3$ , also solves the Disj $_m$  problem. Therefore, by [Theorem 9.3](#), such an algorithm must use space  $\Omega(m) = \Omega(1/\varepsilon)$ .

In order to prove item 2(c), we reduce the GHD $_{n,t,g}$  problem to Max-CSP( $\mathcal{F}$ ) on  $nkq + O(1) = O(n)$  variables, where  $t = n(1 - \gamma)$  and  $g \geq 1$  will be determined later. We will create two groups of constraints: the first group of constraints  $C_A \cup C_B$  will have value  $1 - O(\Delta(x, y)/n)$ , and the second group of constraints will have value close to  $\rho_{\min}$ . By taking a weighed combination of these two groups we will get a formula whose value is less than  $\gamma - \varepsilon$  for  $\Delta(x, y) \geq t + g$ , and value is at least  $\gamma$  for  $\Delta(x, y) \leq t$ .

Again, we start with a minimal unsatisfiable formula on  $kq$  variables. If  $|J| = 2d$  is even, then we arbitrarily partition  $J$  into two sets of  $d$  constraints  $J_A$  and  $J_B$ . If  $|J|$  is odd, then we add one constraint to  $|J|$  as follows. By minimality, there is an assignment that satisfies  $|J| - 1$  constraints of  $J$ , let  $c$  be one of these constraints. We add another copy of  $c$  to  $J$ , and partition  $J$  into two sets of  $d$  constraints  $J_A$  and  $J_B$ . Note that while  $J_A$  and  $J_B$  are satisfiable, only  $2d - 1$  constraints of  $J_A \cup J_B$  can be satisfied simultaneously.

Let  $x, y \in \{0, 1\}^n$  be the inputs of Alice and Bob in the  $\text{GHD}_{n,t,g}$  problem. If  $x_i = 1$ , then Alice applies the constraints  $J_A$  to the  $i$ th block of  $kq$  variables of the formula, otherwise Alice applies the constraint  $J_B$  to these variables. Similarly, if  $y_i = 1$  or  $y_i = 0$ , then Bob applies the constraints  $J_A$  or  $J_B$  to the  $i$ th block of  $kq$  variables. Let  $C_A$  and  $C_B$  be the sets of constraints produced by Alice and Bob, respectively. Observe that  $|C_A| = |C_B| = nd$ . The set of constraints added by Alice and Bob when processing their  $i$ th coordinates is satisfiable if and only if  $x_i = y_i$ . When  $x_i \neq y_i$ , then by the construction of  $J$ , exactly  $2d - 1$  constraints are satisfiable. Therefore,

$$\text{val}(C_A \cup C_B) = 1 - \frac{\Delta(x, y)}{2dn}.$$

Let  $\gamma' = (\gamma + \rho_{\min})/2 < \gamma$ . By the definition of  $\rho_{\min}(\mathcal{F})$ , there exists  $n_0$  and a formula  $\Phi'$  of  $\text{Max-CSP}(f)$  such that  $\text{val}(\Phi') = \gamma'$ . By taking several copies of  $\Phi'$  on the same  $n_0$  variables, we get an instance  $\Phi$  with  $D = |\Phi| = \frac{n(2d-1)(1-\gamma)}{\gamma-\gamma'} = \Theta(n)$  constraints and value  $\text{val}(\Phi) = \gamma'$ .

Now we output an instance  $\Psi$  of  $\text{Max-CSP}(\mathcal{F})$  on  $nkq + n_0$  variables that is simply a union of  $C_A \cup C_B$  and  $\Phi$  on disjoint sets of variables. By construction,

$$\text{val}(\Psi) = \frac{(2dn - \Delta(x, y)) + \gamma'D}{2dn + D}.$$

In the case when  $\Delta(x, y) \leq t = (1 - \gamma)n$ , we have

$$\text{val}(\Psi) \geq \frac{2dn - (1 - \gamma)n + \gamma'D}{2dn + D} = \gamma.$$

And for the case of  $\Delta(x, y) \geq t + g = (1 - \gamma)n + g$ , we have that

$$\text{val}(\Psi) \leq \frac{(2dn - (1 - \gamma)n - g) + \gamma'D}{2dn + D} = \gamma - \frac{g}{2dn + D} = \gamma - \varepsilon$$

for  $g = \varepsilon(2dn + D) = \Theta(n\varepsilon)$ .

Therefore, any streaming algorithm for  $(\gamma, \gamma - \varepsilon)$ - $\text{Max-CSP}(\mathcal{F})$  will imply a protocol for the  $\text{GHD}_{n,t,g}$  problem. By [Theorem 9.4](#), any such streaming algorithm must use at least  $\Omega(\min\{n, n^2/g^2\}) = \Omega(\min\{n, \varepsilon^{-2}\})$  bits of communication.  $\square$

## Acknowledgments

Thanks to Alex Andoni for pointers to  $\ell_{1,\infty}$  norm estimation algorithms.

## References

[AHLW16] Yuqing Ai, Wei Hu, Yi Li, and David P. Woodruff. New characterizations in turnstile streams with applications. In *CCC 2016*, 2016.

- [AKO11] Alexandr Andoni, Robert Krauthgamer, and Krzysztof Onak. Streaming algorithms via precision sampling. In *FOCS 2011*, pages 363–372. IEEE, 2011.
- [And20] Alexandr Andoni. Personal communication, 24 December 2020.
- [BPR06] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*. Springer, 2006.
- [BV04] Stephen P. Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [CGSV21] Chi-Ning Chou, Alexander Golovnev, Madhu Sudan, and Santhoshini Velusamy. Approximability of all Boolean CSPs in the dynamic streaming setting. *CoRR*, abs/2102.12351v3, 14th April 2021.
- [CGV20] Chi-Ning Chou, Alexander Golovnev, and Santhoshini Velusamy. Optimal streaming approximations for all Boolean Max-2CSPs and Max- $k$ SAT. In *FOCS 2020*, pages 330–341. IEEE, 2020.
- [CR12] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM J. Comput.*, 41(5):1299–1317, 2012.
- [GKK<sup>+</sup>09] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2009.
- [GT19] Venkatesan Guruswami and Runzhou Tao. Streaming hardness of unique games. In *APPROX 2019*, pages 5:1–5:12. LIPIcs, 2019.
- [GVV17] Venkatesan Guruswami, Ameya Velingker, and Santhoshini Velusamy. Streaming complexity of approximating Max 2CSP and Max Acyclic Subgraph. In *APPROX 2017*. LIPIcs, 2017.
- [Ind00] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. In *FOCS 2000*, pages 189–197. IEEE, 2000.
- [KK19] Michael Kapralov and Dmitry Krachun. An optimal space lower bound for approximating MAX-CUT. In *STOC 2019*, pages 277–288. ACM, 2019.
- [KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on Boolean functions. In *FOCS 1988*, pages 68–80. IEEE, 1988.
- [KKS15] Michael Kapralov, Sanjeev Khanna, and Madhu Sudan. Streaming lower bounds for approximating MAX-CUT. In *SODA 2015*, pages 1263–1282. SIAM, 2015.
- [KKSV17] Michael Kapralov, Sanjeev Khanna, Madhu Sudan, and Ameya Velingker.  $(1 + \omega(1))$ -approximation to MAX-CUT requires linear space. In *SODA 2017*, pages 1703–1722. SIAM, 2017.
- [KNW10] Daniel M. Kane, Jelani Nelson, and David P. Woodruff. On the exact space complexity of sketching and streaming small norms. In *SODA 2010*, pages 1161–1178. SIAM, 2010.

- [KS92] Bala Kalyanasundaram and Georg Schintger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.
- [LNW14] Yi Li, Huy L. Nguyen, and David P. Woodruff. Turnstile streaming algorithms might as well be linear sketches. In *STOC 2014*, pages 174–183, 2014.
- [O’D14] Ryan O’Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.
- [Raz90] Alexander A. Razborov. On the distributional complexity of disjointness. In *ICALP 1990*, pages 249–253. Springer, 1990.
- [She12] Alexander A. Sherstov. The communication complexity of gap hamming distance. *Theory Comput.*, 8(1):197–208, 2012.
- [SSV21] Noah Singer, Madhu Sudan, and Santhoshini Velusamy. Personal communication, January 2021.
- [Vid12] Thomas Vidick. A concentration inequality for the overlap of a vector on a large set, with application to the communication complexity of the gap-hamming-distance problem. *Chicago J. Theor. Comput. Sci.*, 18(1):1–12, 2012.