The Cost of the Missing Bit: Communication Complexity with Help

László Babai [*]

Thomas P. Hayes [†]

Peter G. Kimmel [‡]

November 6, 2000

(FINAL VERSION)

Dedicated to the Memory of Paul Erdős

Abstract

We generalize the multiparty communication model of Chandra, Furst, and Lipton (1983) to functions with b-bit output ($b = 1$ in the CFL model). We allow the players to receive up to $b-1$ bits of information from an all-powerful benevolent Helper who can see all the input. Extending results of Babai, Nisan, and Szegedy (1992) to this model, we construct families of explicit functions for which $\Omega(n/c^k)$ bits of communication are required to find the "missing bit," where n is the length of each player's input and k is the number of players. As a consequence we settle the problem of separating the one-way vs. multiround communication complexities (in the CFL sense) for $k \le (1-\epsilon)\log n$ players, extending a result of Nisan and Wigderson (1991) who demonstrated this separation for $k = 3$ players. As a by-product we obtain $\Omega(n/c^k)$ lower bounds for the multiparty complexity (in the CFL sense) of new families of explicit boolean functions (not derivable from BNS).

[*] Department of Computer Science, University of Chicago, 1100 East 58th Street, Chicago, IL 60637. e-mail: `laci@cs.uchicago.edu`. Supported in part by NSA grant MSPR-96G-184

[†] Department of Mathematics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637. e-mail: `hayest@math.uchicago.edu`. Supported in part by an NSF Graduate Fellowship.

[‡] Department of Computer Science, Northeastern Illinois University, 5500 N. St. Louis Avenue, Chicago, IL 60625-4699. e-mail: `pgkimmel@neiu.edu`

The proofs exploit the interplay between two concepts of multicolor discrepancy; discrete Fourier analysis is the basic tool. We also include an unpublished lower bound by A. Wigderson regarding the one-way complexity of the 3-party pointer jumping function.

1 Introduction

1.1 Brief summary

Communication complexity is an abstract model of computation which focuses on the cost of inter-processor communication. It has been linked to a number of models of computation; it appears to be at the heart of complexity questions, especially in highly parallel models (shallow circuits) [13, 14, 25, 6, 20], space-bounded computation (including the related models of decision trees and branching programs) [8, 5, 11], and hardness results required for the construction of pseudorandom generators for shallow circuits [17] and for space-bounded computation [5]. Nisan applied communication complexity to threshold circuits with no depth restriction [19].

Lower bounds in communication complexity pose difficult questions; the *multiparty model*, introduced by Chandra, Furst, and Lipton [8], is notoriously hard. The strongest known lower bounds in the CFL [8] multiparty model were given by Babai, Nisan, and Szegedy [5], where two families of explicit functions were shown to require $\Omega(n/c^k)$ communication. Here n is the length of the input, k is the number of players and $2 \leq c \leq 4$ is a constant. We note that the value of c for one of these functions was subsequently improved by Chung and Tetali [9] from 4 to 2.

The multiparty communication model we consider extends the CFL model in two ways: the output to be computed has b bits ($b = 1$ in the CFL model); and the players receive "help:" a message of at most $b - 1$ bits from an all-powerful benevolent Helper who can see all the input. The help, of course, makes lower bound proofs more difficult.

Extending results of [5], we find strong ($\Omega(n/c^k)$) lower bounds on the communication cost of the "missing bit" for families of explicit functions with long outputs. Here n is the length of the input each player "misses," k is the number of players, and the output is up to n bits long.

Our main motivation was an application of these lower bounds to the original CFL model (boolean output): we extend a result of Nisan and Wigderson [20] on the separation of one-way vs. multiround communication complexity from three players to any constant number of players (in fact up to $k \leq (1 - \epsilon) \log n$ players).

As a by-product, we obtain $\Omega(n/c^k)$ lower bounds for the k-party communication complexity of interesting new families of boolean functions, including the "trace of matrix product mod 2" function. The importance of such results lies in the wide variety of applications of multiparty lower bounds that has been based on the $\Omega(n/c^k)$ lower bounds of [5].

The BNS [5] lower bounds were derived from upper bounds on the discrepancy over cylinder intersections of the two-coloring corresponding to the boolean function in question. We need to extend this technique to multicolor discrepancy. We consider two separate discrepancy concepts (strong and weak) both of which reduce to the ordinary discrepancy in the boolean case. Strong discrepancy is the straightforward combinatorial extension; the concept of *weak discrepancy* is more subtle and requires *characters of finite abelian groups* (the Discrete Fourier Transform). Our main technique is the interplay between these two concepts of multicolor discrepancy. A concept closely related to our "weak discrepancy" has previously been introduced in the context of multiparty communication complexity by Grolmusz [12].

1.2 Multiparty Communication with Help

Following Chandra, Furst, and Lipton [8], in all models to be considered we adopt the following basic assumptions and notation.

Players $A_1, \ldots, A_k$ wish to collaboratively evaluate the function $f : X_1 \times \cdots \times X_k \to B$ on input $\vec{x} = (x_1, \ldots, x_k)$. Note that in [8], the range B is $\{0,1\}$; here we shall allow arbitrary finite sets B (of variable size.) The function f is known to each player, and Player A_i sees all pieces of the input *except* x_i. We say that A_i *misses* $x_i \in X_i$. The players communicate by broadcasting messages to all other players. Let $P(\vec{x})$ be the string of messages broadcast by the players on input $\vec{x}$. P is said to compute f correctly if $f(\vec{x})$ is fully determined by $P(\vec{x})$ (i. e. $f(\vec{x})$ is a function of $P(\vec{x})$.) The communication complexity of f, denoted by $C(f)$, is the number of bits to be communicated under the best communication protocol, on the worst input: $C(f) = \min_P \max_{\vec{x}} |P(\vec{x})|$.

Our players will be aided by a "Helper" who can see the entire input and broadcasts a message $H(\vec{x})$ to the players before they begin to communicate. By sending $b := \log|B|$ bits, the Helper could announce the function value; therefore, we restrict the Helper to sending $r \leq b - 1$ bits.

In this model, a protocol P is a collection of protocols in the previous sense (without help): a protocol P^j for each possible help message j. The output of the protocol on input $\vec{x}$ is $(j, P^j(\vec{x}))$, where $j = H(\vec{x})$ is the help message for the given input. As before, we require the output to fully determine the value of the target function.

Definition 1.1 Let the space of inputs be $X = X_1 \times \cdots \times X_k$. A *protocol with help* consists of two functions: the *Help function* $H : X \to R$, where R is a finite set, and a function $P : R \to \mathcal{P}$, where $\mathcal{P}$ denotes the set of k-party protocols over X (without help).

Let f be a function with domain X. If there is a function g such that for every input $\vec{x}$, $f(\vec{x}) = g(H(\vec{x}), P(H(\vec{x}))(\vec{x}))$, then we say that (H, P) computes f. In this case we call (H, P) a *communication protocol for f, with help from the set R.*

We denote the set of such protocols (H, P) by $\mathcal{P}^{\text{help}}(R)$.

Definition 1.2 The *cost* of a communication protocol with help, (H, P) is the combined length of the longest message pair (j, s), where $j = H(\vec{x})$ is the help message, and $s = P(H(\vec{x}))(\vec{x})$ is the communication string sent by the players.

The *communication complexity of f with help,* denoted by $C^{\text{help}}(r, f)$, is the minimum cost of a protocol with r bits of help for f.

In most but not all cases we assume $r = b - 1$ and suppress r in the notation: $C^{\text{help}}(f) := C^{\text{help}}(b - 1, f)$. (A notable exception is described in Section 9.3.)

Remark 1.3 Another reasonable definition for the cost of a protocol with help would be the maximum length of a Help string, $\log |R|$, plus the maximum cost of a protocol P^j. In fact, this definition was used in a preliminary version of this paper. However, all our results hold using the stronger Definition 1.2. We are grateful to one of the referees for suggesting this modification.

1.3 Main Results

Our main result is the following $\Omega(n)$ lower bound on the complexity of multiparty communication with help for a family of explicit functions.

Theorem 1.4 *There exists an NC-computable class of functions $\{f_{n,b,k} : n \geq b \geq 1, k \geq 2\}$ where $f_{n,b,k}$ is a function of k n-bit arguments with b-bit output such that $C^{\text{help}}(f_{n,b,k}) \geq \Omega(n/c^k)$, for some constant $c > 1$.*

Remark 1.5 In the statement above, NC denotes, as usual, the class of functions computable by a uniform family of polynomial-size boolean circuits of polylogarithmic depth (this is a subclass of P, the class of polynomial-time computable functions). In fact, the function $\text{GIP}_{q,s,k}$ which will be used to establish the theorem (see Definition 4.11) is computable with bounded fan-in boolean circuits of depth $O(\log n \log k)$. In particular, it is in NC^1 for constant k. The same remark applies to Theorem 1.7 below. As we shall see, these lower bounds hold in the randomized and distributional (random input) models as well. We do not know any AC^0 family of functions satisfying a similar lower bound.

As an application, we generalize a result of Nisan and Wigderson [20].

Definition 1.6 A special case of the standard communication model is *one-way communication*, in which each player may speak only once, and they proceed in a specified order.

Nisan and Wigderson [20] demonstrate an exponential gap between the power of three-party one-way protocols for boolean functions depending on which player speaks first. We extend their result to $3 \le k \le (1-\epsilon)\log n$ players.

Theorem 1.7 *There exists an NC-computable class of boolean functions* $\{g_{n,k} : n \ge 1, k \ge 2\}$ *where* $g_{n,k}$ *is a function of* k n*-bit arguments such that the one-way communication complexity of* $g_{n,k}$ *when Player* k *speaks first is* $\Omega(\sqrt{n}/c^k)$*, for some constant* $c > 1$*, and the one-way communication complexity of* $g_{n,k}$ *when any other player speaks first is* $O(\log n)$.

This result resolves a problem which was open even for $k = 4$, and complements another separation result (one-way vs. oblivious, [4, 21]).

As a by-product of this work, we obtain a new family of explicit *boolean* functions whose k-party communication complexity satisfies the BNS [5] lower bounds $\Omega(n/c^k)$.

Definition 1.8 The *trace of matrix product mod* 2 function, $\text{TMP}_{2,d,k}$, takes k square ($d \times d$) $(0,1)$-matrices as inputs, and returns the trace of their product modulo 2.

Theorem 1.9 $\text{TMP}_{2,\sqrt{n},k}$ *has* k*-party communication complexity* $\Omega(n/k^2 2^k)$. *(*n *is the number of bits each player misses.)*

1.4 Organization of the Paper

Section 2 introduces two notions of *discrepancy* for multivalued functions, and derives a basic inequality between them (Lemma 2.9). Theorem 8.2 exhibits an inequality relating the two notions of discrepancy in the opposite direction.

In Section 3 we link these discrepancies, considered over the family of *cylinder intersections*, to *communication complexity with help*. The key lemma to our complexity lower bounds is Lemma 3.3.

Theorem 1.4 and Theorem 1.9 are proved in Section 4, which investigates the discrepancy of *multilinear functions* over a finite field, extending the results of [5] on the "generalized inner product" (GIP) function. The key result of this section is a rather general upper bound on the discrepancy of multilinear functions over cylinder intersections (Theorem 4.6).

In Section 5, we obtain similar discrepancy bounds for the "multiplicative coset of sum" (MCS) function, extending the results of [5] on the "quadratic character of sum" function. Section 7 shows how to compute a subclass of the MCS functions in polynomial time.

Following Nisan and Wigderson, in Section 6 we apply our results to one-way communication complexity, proving Theorem 1.7. A more technical statement appears as Corollary 6.3.

In Section 9 we reproduce three $\Omega(n)$ lower bounds for *two-party communication with help* (yielding in each case $\Omega(\sqrt{n})$ lower bounds on three-player one-way complexity for functions with $O(\log n)$ communication complexity if the order of players is changed).

The subject of Section 9.1 is the MCS function. In Section 9.2 we reproduce a result of Nisan and Wigderson [20] on *hash-functions* in our framework. That result provided the initial motivation to this paper.

Section 9.3 gives an exposition of an unpublished result of Avi Wigderson: a lower bound for the *pointer-jumping* function to which the discrepancy approach is not applicable.

A preliminary version of this paper appeared in STOC '98 [3].

2 Multicolor Discrepancy

Discrepancy is usually discussed in the context of two-colorings. We need a discrepancy concept for any function $f : X \to B$ (coloring with $|B|$ colors).

Let f be a function, $f : X \to B$, and let $S \subseteq X$. If $B = \{\pm 1\}$, then the common meaning of the discrepancy of f on S is the sum $|\sum_{x \in S} f(x)|$. Our concept of *strong* discrepancy generalizes this notion to arbitrary sets B. Another, weaker generalization of the 2-color discrepancy suggests itself in the case when B is a finite abelian group.

After defining both types of multicolor discrepancy, we prove that they are closely related. We will make good use of this relationship in Section 4.

Definition 2.1 For a function $f : X \to B$, (the "B-coloring of X"), a subset $S \subseteq X$ and $y \in B$, we define the *excess* $\gamma(y)$ of color y in S under coloring f by

$$\gamma(y) := \left(|f^{-1}(y) \cap S| - |S|/|B|\right) / |X|. \tag{1}$$

We define the *strong discrepancy of f in S* by

$$\Gamma(f, S) := \max_{y \in B} |\gamma(y)| \tag{2}$$

Definition 2.2 *(strong discrepancy for set systems)* For a set system $\mathcal{F}$ over the universe X, we define the *strong discrepancy of f over $\mathcal{F}$* by

$$\Gamma(f, \mathcal{F}) := \max_{S \in \mathcal{F}} \Gamma(f, S).$$

For the definitions of weak discrepancy, we require elementary facts about characters of finite abelian groups (see, for example, [1]).

Definition 2.3 Let G be a finite abelian group with m elements, with operation $+$ and identity element 0. A *character* of G is a homomorphism from G to the multiplicative group of complex roots of unity. The characters of G form a finite group under elementwise multiplication, which is denoted by $\widehat{G}$. The identity element of $\widehat{G}$ is the *principal character*, χ_0, defined by $\chi_0(g) = 1$ for all $g \in G$.

The following Proposition lists some basic facts about characters we will use.

Proposition 2.4 *Let G be a finite abelian group.*

1. $|\widehat{G}| = |G|$. *(Indeed, $\widehat{G} \cong G$, but we will not use this.)*
2. *For any* $0 \neq g \in G$, $\displaystyle\sum_{\chi \in \widehat{G}} \chi(g) = 0.$
3. *For any* $\chi_0 \neq \chi \in \widehat{G}$, $\displaystyle\sum_{g \in G} \chi(g) = 0.$

We define the concept of "weak discrepancy" using characters, for colorings $f : X \to G$, where G is a finite abelian group.

Definition 2.5 *(weak discrepancy)* For a function $f : X \to G$, a subset $S \subseteq X$ and a character $\chi \in \widehat{G}$, we define the *weak χ-discrepancy of f in S* by

$$\Gamma_\chi^{\text{weak}}(f, S) := \frac{1}{|X|} \left| \sum_{x \in S} \chi(f(x)) \right|. \tag{3}$$

We define the *weak discrepancy of f in S* by

$$\Gamma^{\text{weak}}(f,S) := \frac{1}{m-1} \sum_{\chi\in\widehat{G},\chi\neq\chi_0} \Gamma_\chi^{\text{weak}}(f,S). \tag{4}$$

Remark 2.6 The definition of $\Gamma_\chi^{\text{weak}}(f,S)$ generalizes a multicolor discrepancy concept introduced by V. Grolmusz [12]. In our language, Grolmusz considers the case when G is cyclic (which would suffice for our applications as well) and makes a specific choice of χ: $\chi(i) = \omega^i$, where ω is the m-th root of unity minimizing $|1+\omega|$. Rather than specifying a particular character, we consider the average over all non-principal characters, Γ^{weak}. This allows us to prove the crucial inequality (5) linking strong and weak discrepancies for arbitrary functions.

Definition 2.7 *(weak discrepancy for set systems)* For a set system $\mathcal{F}$ over universe X, we define the *weak discrepancy of f over $\mathcal{F}$* by

$$\Gamma^{\text{weak}}(f,\mathcal{F}) \ := \ \max_{S\in\mathcal{F}} \Gamma^{\text{weak}}(f,S).$$

Remark 2.8 It is easy to see that the inequality $\Gamma_\chi^{\text{weak}}(f,S) \leq m\ \Gamma(f,S)$ holds (see Proposition 8.1). From this, it is immediate that $\Gamma^{\text{weak}}(f,S) \leq m\ \Gamma(f,S)$, justifying the terms "weak" and "strong" discrepancy. Theorem 8.2 will improve this bound. Here, however, we are interested in a somewhat surprising inequality in the opposite direction.

Lemma 2.9 *Let G be a finite abelian group with m elements, $f : X \to G$ a function, and $S \subseteq X$. Then*

$$\Gamma(f,S) \ \leq \ (1-1/m)\,\Gamma^{\text{weak}}(f,S). \tag{5}$$

Remark 2.10 This inequality provides the tool for turning upper bounds on weak discrepancy into the upper bounds on strong discrepancy we need for our communication complexity lower bounds (see Lemma 3.3 below).

First we derive an expression for the weak discrepancy in terms of the $\gamma(y)$'s.

Proposition 2.11 *For any $S \subseteq X$ and any $\chi \in \widehat{G}, \chi \neq \chi_0$, we have*

$$\Gamma_\chi^{\text{weak}}(f,S) = \left|\sum_{y\in G} \chi(y)\gamma(y)\right|. \tag{6}$$

Proof: Consider equation (3). Reindex the summation by the values $y = f(x) \in G$:

$$\begin{aligned}\Gamma_\chi^{\text{weak}}(f,S) &= \frac{1}{|X|}\left|\sum_{y\in G}\chi(y)|f^{-1}(y)\cap S|\right| \\ &= \left|\sum_{y\in G}\chi(y)\left(\gamma(y)+\frac{|S|}{m|X|}\right)\right|.\end{aligned}$$

Since $\sum_{y\in G}\chi(y) = 0$, the $|S|/(m|X|)$ terms cancel. ■

Notation 2.12 The function $\delta : G \to \{0,1\}$ is defined by

$$\delta(g) = \begin{cases} 1 & \text{if } g = 0 \\ 0 & \text{otherwise.}\end{cases}$$

Proof of Lemma 2.9: Fix $z \in G$. Then $\gamma(z) = \sum_{y\in G}\gamma(y)\delta(y-z)$. From Proposition 2.4, it follows that, for every $g \in G$, $\delta(g) = \frac{1}{m}\sum_{\chi\in\widehat{G}}\chi(g)$. (This is the *Fourier expansion* of δ.) Hence

$$\begin{aligned}|\gamma(z)| &= \left|\frac{1}{m}\sum_{y\in G}\sum_{\chi\in\widehat{G}}\gamma(y)\chi(y-z)\right| \\ &= \left|\frac{1}{m}\sum_{y\in G}\sum_{\chi\neq\chi_0}\gamma(y)\chi(y-z)\right|,\end{aligned}$$

because $\sum_{y\in G}\gamma(y) = 0$. Now, by the multiplicativity of χ, this is equal to

$$\begin{aligned} &= \left|\frac{1}{m}\sum_{\chi\neq\chi_0}\chi(-z)\sum_{y}\gamma(y)\chi(y)\right| \\ &\leq \frac{1}{m}\sum_{\chi\neq\chi_0}\left|\sum_{y}\gamma(y)\chi(y)\right|,\end{aligned}$$

by the triangle inequality, observing that $|\chi(-z)| = 1$. By Proposition 2.11, the above expression equals:

$$= \frac{1}{m}\sum_{\chi\neq\chi_0}\Gamma_\chi^{\text{weak}}(f,S)$$

$$= \quad \frac{m-1}{m}\,\Gamma^{\text{weak}}(f,S).$$

(The last step is the definition of Γ^{weak}.) Since $\Gamma(f,S)$ is defined (Definition 2.1) as the maximum of $|\gamma(y)|$, the proof is complete. ■

3 Communication with Help vs. Multicolor Discrepancy over Cylinder Intersections

Definition 3.1 Let $X = X_1 \times \cdots \times X_k$. A subset $S \subseteq X$ is called a *cylinder in the* i^{th} *dimension* if membership in S does not depend on the i^{th} coordinate. A subset of X is called a *cylinder intersection* if it is an intersection of cylinders. We shall use $\mathcal{C}$ to denote the set of all cylinder intersections in X (with respect to the given Cartesian decomposition $X = X_1 \times \ldots \times X_k$).

As in the boolean case (cf. [5, Lemma 2.2]), upper bounds on multicolor discrepancy over cylinder intersections yield lower bounds on communication complexity in the "communication with help" model. This remains true even in the distributional sense:

Definition 3.2 The *ϵ-distributional communication complexity* of f with help, denoted by $C^{\text{help},\epsilon}(f)$, is the minimum cost of a protocol with help that computes f correctly on at least a $(1-\epsilon)$-fraction of X.

Lemma 3.3 *Let $f : X \to B$ be a function. For $0 \leq \epsilon < 1/2$,*

$$C^{\text{help},\epsilon}(f) \;\geq\; \log\left(\frac{1/2-\epsilon}{\Gamma(f,\mathcal{C})}\right)$$

Proof: Let $\gamma := C^{\text{help},\epsilon}(f)$. Let g be a function that agrees with f on at least a $(1-\epsilon)$-fraction of X, such that g is computed by a protocol (H,P) of cost γ. Let $E := \{x \in X | f(x) = g(x)\}$, so that by our choice of g, $|E| \geq (1-\epsilon)|X|$. For each help message j, and communication string s, let $X(j,s)$ denote those inputs in X which cause help message j and communication string s to be sent. In other words, $X(j,s) = H^{-1}(j) \cap (P^j)^{-1}(s)$. Let $E(j,s) = E \cap X(j,s)$.

Choose a help message j and a communication string s that maximize

$$|E(j,s)| - \frac{(P^j)^{-1}(s)}{|B|}.$$

If we sum this quantity over all pairs (j,s), the left term contributes $|E|$, since the sets $E(j,s)$ form a partition of E, while the right term contributes $-|X||R|/|B|$, since for each choice of j, the sets $(P^j)^{-1}(s)$ partition X. Since there are at most 2^γ pairs (j,s), the maximum of the quantity under consideration is at least $(|E|-|X||R|/|B|)2^{-\gamma}$.

Since the pair (j,s) determines g, g is constant on $X(j,s)$, and so f is constant on $E(j,s)$. As observed in [5], the set $(P^j)^{-1}(s)$ must be a cylinder intersection. Hence,

$$\begin{aligned}\Gamma(f,\mathcal{C}) &\geq \frac{1}{|X|}\left(|E(j,s)|-\frac{(P^j)^{-1}(s)}{|B|}\right)\\ &\geq (|E|/|X|-|R|/|B|)2^{-\gamma}\\ &\geq (1-\epsilon-1/2)2^{-\gamma},\end{aligned}$$

where the last step uses the fact that $|R|/|B| \leq 1/2$, part of the definition of communication with help. Taking logarithms,

$$C^{\text{help},\epsilon}(f)=\gamma\geq\log\left(\frac{1/2-\epsilon}{\Gamma(f,\mathcal{C})}\right),$$

as claimed. ■

Remark 3.4 It would be reasonable to relax the restriction that the Helper can send at most $b-1$ bits of information, and merely require that $|R|\leq|B|-1$. In this case, the proof of Lemma 3.3 implies that

$$C^{\text{help},\epsilon}(f)\geq\log\left(\frac{1-(|R|/|B|)-\epsilon}{\Gamma(f,\mathcal{C})}\right),$$

Remark 3.5 In [5], an upper bound on the maximum volume of a homogeneous cylinder intersection suffices for the lower bound on deterministic communication complexity. While a discrepancy bound implies a bound on this volume, the full power of the discrepancy bounds is only needed in [5] for the distributional result. In contrast, for the “communication with help” model, we need bounds on the strong discrepancy even for deterministic complexity.

4 The Discrepancy of Multilinear Functions

In this section, we extend the techniques of [5] to prove upper bounds on the weak discrepancy for a large class of functions which includes their “generalized inner product” function, and the “trace of matrix product” function, both of which we will define later.

Definition 4.1 Let V be a vector space over a field $\mathbb{F}$. A function $f : V^k \to \mathbb{F}$ is *k-linear* if whenever we fix $k-1$ of the k input vectors, the resulting function $g : V \to \mathbb{F}$ is a vector space homomorphism.

Our main tool is the following lemma, which was inspired by Lemma 2.3 of [5]. This result relates the weak discrepancy of a multilinear function over an arbitrary cylinder intersection to its weak discrepancy over the entire input space.

Lemma 4.2 *Let $X_1, X_2, \ldots, X_k, G$ be finite abelian groups. Let $X = X_1 \times \cdots \times X_k$. Let $f : X \to G$ be a homomorphism in each coordinate, let χ be a character of G, and for each $i \in \{1, \ldots, k\}$ let $\phi_i : X \to G$ be the characteristic function of a cylinder in the i^{th} dimension. Then*

$$\left| \mathop{\mathrm{E}}_{x \in X} \chi(f(x))\phi_1(x) \cdots \phi_k(x) \right| \leq \left(\mathop{\mathrm{E}}_{x \in X} \chi(f(x)) \right)^{2^{1-k}}, \tag{7}$$

where the expectation is taken with respect to the uniform distribution over X.

Remark 4.3 A direct imitation of the BNS [5] proof does not seem to work for this result. Instead, we present Lemma 4.4, a rather technical generalization of Lemma 4.2 to which an extension of the BNS argument is applicable.

Lemma 4.4 *Let $X_1, X_2, \ldots, X_k, G$ be finite abelian groups and Ω a probability space. Let $X = X_1 \times \cdots \times X_k$. For each $\lambda \in \Omega$, let $f^\lambda : X \to G$ be a homomorphism in each coordinate, let χ^λ be a character of G, and for each $i \in \{1, \ldots, k\}$ let $\phi_i^\lambda : X \to G$ be the characteristic function of a cylinder in the i^{th} dimension. Then*

$$\left| \mathop{\mathrm{E}}_{\lambda \in \Omega, x \in X} \chi^\lambda(f^\lambda(x))\phi_1^\lambda(x) \cdots \phi_k^\lambda(x) \right| \leq \left(\mathop{\mathrm{E}}_{\lambda \in \Omega, x \in X} \chi^\lambda(f^\lambda(x)) \right)^{2^{1-k}}, \tag{8}$$

where the expectation is taken with respect to the uniform distribution over X^k and the given probability measure over Ω.

Note 4.5 Throughout the proof, we will omit the superscript λ, which should be assumed everywhere, whether λ is bound or free. The space Ω of events λ is necessary for the inductive step.

Proof: First let us see that the expression

$$\mathop{\mathrm{E}}_{\lambda\in\Omega, x\in X} \chi(f(x))$$

is always real-valued and in the interval $[0,1]$. Fix some $\lambda \in \Omega$ and $u \in X_2 \times \cdots \times X_k$. By hypothesis, the function $f^u : X_1 \to G$ defined by $f^u(x) = f(x,u)$ is a homomorphism. Hence the composition $\chi \circ f^u$ is a character of X_1. Therefore $\mathop{\mathrm{E}}_{x\in X_1} \chi(f(x,u))$ is either 1 or 0 according to whether $\chi \circ f^u$ is or is not the principal character, which establishes the claim. This shows that the desired inequality is well-defined.

We proceed by induction on k. For $k=1$, the result is trivial.

Let $k>1$. For readability, we will omit the arguments to the ϕ_i, which are the same as the arguments to $\chi(f())$. Since ϕ_k does not depend on x_k, and $|\phi_k| \le 1$, we observe that

$$\begin{aligned} &\left| \mathop{\mathrm{E}}_{\lambda\in\Omega} \mathop{\mathrm{E}}_{x\in X} \chi(f(x))\phi_1 \ldots \phi_k \right| \\ \le \quad & \mathop{\mathrm{E}}_{\lambda\in\Omega} \mathop{\mathrm{E}}_{\widetilde{x}\in X_1\times\cdots\times X_{k-1}} \left| \mathop{\mathrm{E}}_{x_k\in X_k} \chi(f(\widetilde{x}, x_k))\phi_1 \ldots \phi_{k-1} \right| . \end{aligned}$$

Henceforth we will denote $X_1 \times \cdots \times X_{k-1}$ by $\widetilde{X}$.

Now we make use of the Cauchy-Schwarz inequality in the form $|E(z)|^2 \le E(|z|^2)$.

$$\begin{aligned} &\mathop{\mathrm{E}}_{\lambda\in\Omega} \mathop{\mathrm{E}}_{\widetilde{x}\in\widetilde{X}} \left| \mathop{\mathrm{E}}_{x_k\in X_k} \chi(f(\widetilde{x}, x_k))\phi_1 \ldots \phi_{k-1} \right| \\ \le \quad & \left(\mathop{\mathrm{E}}_{\lambda\in\Omega} \mathop{\mathrm{E}}_{\widetilde{x}\in\widetilde{X}} \left| \mathop{\mathrm{E}}_{x_k\in X_k} \chi(f(\widetilde{x}, x_k))\phi_1 \ldots \phi_{k-1} \right|^2 \right)^{1/2} . \end{aligned}$$

Expressing the squared term using complex conjugation, we obtain

$$\left(\mathop{\mathrm{E}}_{\lambda\in\Omega} \mathop{\mathrm{E}}_{\widetilde{x}\in\widetilde{X}} \left(\left(\mathop{\mathrm{E}}_{u\in X_k} \chi(f(\widetilde{x}, u))\phi_1^u \ldots \phi_{k-1}^u \right) \overline{\left(\mathop{\mathrm{E}}_{v\in X_k} \chi(f(\widetilde{x}, v))\phi_1^v \ldots \phi_{k-1}^v \right)} \right) \right)^{1/2} ,$$

where ϕ_i^u stands for $\phi_i(\widetilde{x}, u)$, and $\phi_i^v = \phi_i(\widetilde{x}, v)$. Since $\overline{\chi(f(\widetilde{x}, v))} = \chi(f(\widetilde{x}, -v))$, and the ϕ_i^v are $\{0,1\}$-valued, this expression may be rewritten as

$$\left(\mathop{\mathrm{E}}_{\lambda\in\Omega} \mathop{\mathrm{E}}_{\widetilde{x}\in\widetilde{X}} \mathop{\mathrm{E}}_{u,v\in X_k} \chi(f(\widetilde{x}, u-v))\phi_1^u\phi_1^v \ldots \phi_{k-1}^u\phi_{k-1}^v \right)^{1/2} ,$$

Next we switch the order of summation in the last expression. Define $\widetilde{\Omega}$ to be the space of events (λ, u, v), where u and v are distributed uniformly in X_k, and λ, u, v are independent. For each $\widetilde{\lambda} = (\lambda, u, v) \in \widetilde{\Omega}$ and $\widetilde{x} \in \widetilde{X}$, define $f^{\widetilde{\lambda}}(\widetilde{x}) := f^{\lambda}(\widetilde{x}, u - v)$ and $\phi_i^{\widetilde{\lambda}}(\widetilde{x}) := \phi_i^{\lambda}(\widetilde{x}, u)\phi_i^{\lambda}(\widetilde{x}, v)$. Notice that, for fixed $\widetilde{\lambda}$, $f^{\widetilde{\lambda}}$ is a $(k-1)$-linear function in $\widetilde{x}$, and that $\phi_i^{\widetilde{\lambda}}$ is a $\{0,1\}$-valued function which does not depend on the i^{th} coordinate of $\widetilde{x}$. This allows us to apply the inductive hypothesis, using $\widetilde{\Omega}$ for our new probability space. After re-reindexing, and noting that the difference $u-v$ is uniformly distributed in X_k, we obtain the desired result. Thus

$$
\begin{aligned}
& \left(\mathop{\mathrm{E}}_{\lambda \in \Omega} \mathop{\mathrm{E}}_{\widetilde{x} \in \widetilde{X}} \mathop{\mathrm{E}}_{u,v \in X_k} \chi(f(\widetilde{x}, u - v)) \phi_1^u \phi_1^v \ldots \phi_{k-1}^u \phi_{k-1}^v \right)^{1/2} \\
= & \left(\mathop{\mathrm{E}}_{\widetilde{\lambda} \in \widetilde{\Omega}} \mathop{\mathrm{E}}_{\widetilde{x} \in \widetilde{X}} \chi^{\lambda}(f^{\widetilde{\lambda}}(\widetilde{x})) \phi_1^{\widetilde{\lambda}} \ldots \phi_{k-1}^{\widetilde{\lambda}} \right)^{1/2} \\
\leq & \left(\mathop{\mathrm{E}}_{\widetilde{\lambda} \in \widetilde{\Omega}} \mathop{\mathrm{E}}_{\widetilde{x} \in \widetilde{X}} \chi^{\lambda}(f^{\widetilde{\lambda}}(\widetilde{x})) \right)^{2^{1-k}} \\
= & \left(\mathop{\mathrm{E}}_{\lambda \in \Omega} \mathop{\mathrm{E}}_{u,v \in X_k} \mathop{\mathrm{E}}_{\widetilde{x} \in \widetilde{X}} \chi(f(\widetilde{x}, u - v)) \right)^{2^{1-k}} \\
= & \left(\mathop{\mathrm{E}}_{\lambda \in \Omega} \mathop{\mathrm{E}}_{x \in X} \chi(f(x)) \right)^{2^{1-k}} . \qquad \blacksquare
\end{aligned}
$$

Theorem 4.6 *Let q be a prime power and $X_1, \ldots, X_k$ finite dimensional vector spaces over $\mathbb{F}_q$, the field of q elements. Let $X = X_1 \times \ldots \times X_k$, and let $f : X \to \mathbb{F}_q$ be a k-linear function. Then*

$$
\Gamma^{\text{weak}}(f, \mathcal{C}) \ \leq \ \left(\Pr_{u \in \widetilde{X}} [\forall x \in X_1 \quad f(x, u) = 0] \right)^{2^{1-k}}, \tag{9}
$$

where $\widetilde{X} := X_2 \times \cdots \times X_k$. (As before, $\mathcal{C}$ denotes the set of cylinder intersections over X.)

Proof: The left hand side of inequality (7) is exactly the normalized weak discrepancy $\Gamma_{\chi}^{\text{weak}}(f, S)$, where S is a cylinder intersection with characteristic function $\phi_1 \cdots \phi_k$. Lemma 4.2 establishes the inequality

$$
\Gamma_{\chi}^{\text{weak}}(f, \mathcal{C}) \ \leq \ \left(\mathop{\mathrm{E}}_{x \in X} \chi(f(x)) \right)^{2^{1-k}} . \tag{10}
$$

Suppose that $\chi \neq \chi_0$; in this case we claim that the right hand sides of Equations (9) and (10) are equal. To see this, we observe that

$$\mathop{\mathrm{E}}_{x \in X} \chi(f(x)) = \mathop{\mathrm{E}}_{u \in \widetilde{X}} \mathop{\mathrm{E}}_{x \in X_1} \chi(f(x,u)).$$

The inner expectation is 0 except when $f(\cdot, u)$ is identically zero, in which case it is 1. Hence the right hand side is simply $\Pr_{u \in \widetilde{X}}[\forall x \in X_1 \quad f(x,u) = 0]$. Averaging over all non-principal characters χ yields the desired inequality. ■

4.1 Application: Lower bounds for TMP and GIP

Theorem 4.6 shows that, to get a lower bound on communication complexity with help for a multilinear function, it suffices to give an upper bound on the probability that a related *vector-valued* function has value zero. The "trace of matrix product" (TMP) and "generalized inner product" (GIP) functions, which we will define shortly, have the "typical" property that the associated vector-valued functions are "reasonably uniform." In the case of TMP, the associated function is the matrix product function. In the case of GIP, the associated function is the k-fold componentwise product function. Because these vector-valued functions have large ranges, "reasonable uniformity" is enough to prove good upper bounds on the probability that the function outputs the zero vector.

Definition 4.7 Let q be a prime power, and let d be a positive integer. Let M be the space of $d \times d$ matrices over $\mathbb{F}_q$. The function $\mathrm{TMP}_{q,d,k} : M^k \to \mathbb{F}_q$ is defined by

$$\mathrm{TMP}_{q,d,k}(A_1, A_2, \ldots, A_k) := \mathrm{Tr}(A_1 \cdot A_2 \cdots A_k)$$

and is called the *trace of matrix product* function. (The *trace* of a matrix $(a_{i,j})$ is $\sum a_{i,i}$.)

Corollary 4.8 *$C^{\mathrm{help}}(\mathrm{TMP}_{q,d,k}) \geq \Omega(n/k^2 2^k)$, where $n = d^2 \log q$ is the number of bits each player misses.*

Remark 4.9 *Specializing to the case $q = 2$ we obtain Theorem 1.9.*

Proof: We first introduce a notation for the matrix product function.

Notation 4.10 Let q be a prime power, and let d be a positive integer. Let $M = M(d,q)$ be the space of $d \times d$ matrices over $\mathbb{F}_q$. The function $\mathrm{MP}_{q,d,k} : M^k \to M$ is defined by

$$\mathrm{MP}_{q,d,k}(A_1, A_2, \ldots, A_k) := A_1 \cdot A_2 \cdots A_k$$

and is called the *matrix product* function.

The result will follow from the following three claims, each of which is easy to see.

Consistent with our previous notation, let $X_1, \ldots, X_k$ each denote a copy of $M = M(d, q)$. Let $\widetilde{X}$ denote the space $X_2 \times \ldots \times X_k$ of inputs visible to Player 1. As before, all probabilities will be taken with respect to the uniform distribution over the space indicated under the "Pr" symbol.

Claim. Let $u \in \widetilde{X}$.

$$((\forall x \in X_1) \quad \text{TMP}_{q,d,k}(x, u) = 0) \Leftrightarrow (\text{MP}_{q,d,k-1}(u) = \mathbf{0}).$$

Claim.

$$\Pr_{u \in \widetilde{X}}[\text{MP}(u) = \mathbf{0}] \leq (k-1) \cdot \Pr_{x \in M}[\text{rank}\, x \leq d - d/(k-1)].$$

Claim. For every integer $r \geq 0$.

$$\Pr_{x \in M}[\text{rank}\, x \leq r] \leq \binom{d}{r} q^{-(d-r)^2}.$$

Combining Theorem 4.6 and these three claims, we have:

$$\Gamma^{\text{weak}}(\text{TMP}_{q,d,k}, \mathcal{C}) \ \leq \ \left((k-1)\binom{d}{\lceil d/(k-1) \rceil} q^{-\lceil d/(k-1) \rceil^2}\right)^{2^{1-k}} \qquad .$$

Applying Lemmas 2.9 and 3.3 and simplifying, we obtain

$$C^{\text{help}}(\text{TMP}_{q,d,k}) \geq 2^{(1-k)} \left(\lceil d/(k-1) \rceil^2 \log q - \log\left((k-1)\binom{d}{\lceil d/(k-1) \rceil}\right)\right) - 1.$$

Substituting $n = \lceil d^2 \log q \rceil$, and observing that the negative terms on the right hand side are negligible when $n > 2^k$, we conclude that

$$C^{\text{help}}(\text{TMP}_{q,d,k}) \geq \Omega\left(\frac{n}{k^2 2^k}\right). \qquad \blacksquare$$

Next we define the "generalized inner product" function over $\mathbb{F}_q$, and prove a lower bound on its communication complexity with help.

Definition 4.11 For a prime power q, positive integers s and k, we define the *generalized inner product*

$$\text{GIP}_{q,s,k} : (\mathbb{F}_q^s)^k \to \mathbb{F}_q$$

for $x_1, \ldots, x_k \in \mathbb{F}_q^s$ by

$$\text{GIP}_{q,s,k}(x_1, \ldots, x_k) = \sum_{i=1}^{s} x_{1,i} \cdot x_{2,i} \cdots x_{k,i}.$$

Corollary 4.12 *$C^{\text{help}}(\text{GIP}_{q,s,k}) \geq \Omega(n/4^k)$, where $n = \lceil s \log q \rceil$ is the number of bits each player misses. Moreover, if $q \geq k^{1+\epsilon}$ for some fixed $\epsilon > 0$, then*

$$C^{\text{help}}(\text{GIP}_{q,s,k}) \geq \Omega(n/2^k),$$

Proof: First we define the componentwise product function.

Definition 4.13 For a prime power q, positive integers s and k, we define the *componentwise product*

$$\text{CWP}_{q,s,k} : (\mathbb{F}_q^s)^k \to \mathbb{F}_q^s$$

for $x_1, \ldots, x_k \in \mathbb{F}_q^s$ by

$$(\text{CWP}_{q,s,k}(x_1, \ldots, x_k))_i = x_{1,i} \cdot x_{2,i} \cdots x_{k,i}.$$

Consistent with our previous notation, let $X_1, \ldots, X_k$ each denote the vector space $\mathbb{F}_q^s$. Let $\widetilde{X}$ denote the space $X_2 \times \ldots \times X_k$ of inputs visible to Player 1. As always, all probabilities will be taken with respect to the uniform distribution over the space indicated below the $\Pr$ symbol.

The result will follow from the following two claims, both of which are easy to see.

Claim. Let $u \in \widetilde{X}$.

$$((\forall x \in X_1) \quad \text{GIP}_{q,s,k}(x, u) = 0) \Leftrightarrow (\text{CWP}_{q,s,k-1}(u) = \mathbf{0}).$$

Claim.

$$\Pr_{u \in \widetilde{X}}[\text{CWP}_{q,s,k-1}(u) = \mathbf{0}] = (1 - (1 - 1/q)^{k-1})^s.$$

Applying Theorem 4.6 and Lemmas 2.9 and 3.3, we obtain

$$C^{\text{help}}(\text{GIP}_{q,s,k}) \geq -2^{(1-k)} \left(s \log \left(1 - (1 - 1/q)^{k-1} \right) \right) - 1.$$

Applying the inequality $1 - x < e^{-x}$ for $x = (1 - 1/q)^{k-1}$, we obtain the lower bound

$$C^{\text{help}}(\text{GIP}_{q,s,k}) \geq \Omega\left(\frac{s}{c(q)^k} \right) = \Omega\left(\frac{n}{c(q)^k \log q} \right)$$

where $c(q) := 2q/(q-1)$ and $n = \lceil s \log q \rceil$. Note that when $q = 2$, $c(q)^k \log q = 4^k$, and we are done. When $2 < q$, we have $c(q) \leq 3$, so as long as $2 < q < 2^{(4/3)^k}$,

$c(q)^k \log q < 4^k$. Since $q > 2^{(4/3)^k}$ implies $q \geq k^{1+\epsilon}$, it will suffice to prove the second part of the corollary.

Assume $q \geq k^{1+\epsilon}$. In this case, we may apply the inequality

$$(1 - 1/q)^{k-1} \geq 1 - (k-1)/q$$

to obtain

$$\begin{aligned} C^{\text{help}}(\text{GIP}_{q,s,k}) &\geq 2^{1-k} s(\log q - \log(k-1)) - 1 \\ &\geq 2^{1-k} s \frac{\epsilon}{1+\epsilon} \log q - 1 \\ &= \Omega(n/2^k). \end{aligned}$$

This completes the proof of Corollary 4.12. ■

5 The Discrepancy of the Multiplicative Coset of Sum

In this section, we prove upper bounds on the multicolor discrepancy of another class of explicit functions over cylinder intersections. The analysis of these "multiplicative coset of sum" (MCS) functions depends on A. Weil's character sum estimates, through a lemma proved by Babai-Nisan-Szegedy [5]. The MCS functions are more difficult to compute than the GIP and TMP functions (cf. Section 7). Besides their mathematical appeal, the significance of the MCS functions may lie in their potential to work beyond $\log n$ players.

Notation 5.1 For q a prime power, let $\mathbb{F}_q$ denote the field of q elements, and let $\mathbb{F}_q^\times$ denote $\mathbb{F}_q \setminus \{0\}$, the multiplicative group of $\mathbb{F}_q$. For $x = (x_1, \ldots, x_k) \in \mathbb{F}_q^k$, let Σx denote $\sum_{i=1}^k x_i$.

Definition 5.2 Let q be a prime power, let u divide $q-1$, and let $s := (q-1)/u$. Then the mapping $y \mapsto y^s$ maps the cyclic group $\mathbb{F}_q^\times$ onto its (unique) subgroup G of index s and order u.

We define the *multiplicative coset of sum* function

$$\text{MCS}_{q,u,k} : (\mathbb{F}_q)^k \to G$$

for $x = (x_1, \ldots, x_k) \in \mathbb{F}_q^k$, by

$$\text{MCS}_{q,u,k}(x) = \begin{cases} * & \text{if } \Sigma x = 0 \\ (\Sigma x)^s & \text{otherwise,} \end{cases} \tag{11}$$

where the "$*$" stands for "undefined." Our discrepancy bound will hold for any mapping of the undefined entries to G.

Remark 5.3 Note that G is in one-to-one correspondence with the cosets by the (unique) subgroup H of order s of $\mathbb{F}_q^\times$. In fact, we really only care about which coset of H contains Σx. The discrepancy bound we give holds for any encoding of the cosets.

Lemma 5.4 *Let $\mathcal{C}$ be the set of cylinder intersections in $\mathbb{F}_q^k$. Then*

$$\Gamma(\mathrm{MCS}_{q,u,k}, \mathcal{C}) \leq 3q^{-2^{-k}} + 1/q.$$

Our chief tool in proving this lemma is a special case of Lemma 2.6 of [5], which involves multiplicative characters.

Definition 5.5 A *multiplicative character* χ of $\mathbb{F}_q$ is a homomorphism of $\mathbb{F}_q^\times$ to the group of complex roots of unity. The principal character χ_0 is the character that maps all of $\mathbb{F}_q^\times$ to the value 1. Characters are usually extended to the entire field by setting $\chi(0) = 0$.

Lemma 5.6 *(BNS [5], Lemma 2.6) Let q be a prime power, and let $\psi : \mathbb{F}_q \to \mathbb{C}$ be a non-principal multiplicative character. For any positive integer k and any cylinder intersection S in $\mathbb{F}_q^k$, we have*

$$\left|\sum_{x\in S} \psi(\Sigma x)\right| \;\leq\; 3 \cdot q^{k-2^{-k}}. \qquad \blacksquare$$

Remark 5.7 Note that the proof of Lemma 5.6 is based on a character sum estimate by A. Weil (cf. [23]).

Proof of Lemma 5.4: Let χ be a non-principal character of G. To simplify notation, set $f = \mathrm{MCS}_{q,u,k}$. We shall estimate the weak χ-discrepancy of f over any cylinder intersection $S \subseteq \mathbb{F}_q^k$. Let ψ be the multiplicative character of $\mathbb{F}_q$ defined by $\psi(y) = \chi(y^s)$ for $y \in \mathbb{F}_q^\times$ and $\psi(0) = 0$. Since the map $y \mapsto y^s$ maps $\mathbb{F}_q^\times$ onto G, it follows that ψ is not the principal character.

Let $V = \{x \in \mathbb{F}_q^k : \Sigma x = 0\}$. Note that $|V| = q^{k-1}$.

We have

$$\Gamma_\chi^{\text{weak}}(f) = q^{-k}\left|\sum_{x\in S}\chi(f(x))\right| \leq q^{-k}\left(|V| + \left|\sum_{x\in S}\chi((\Sigma x)^s)\right|\right),$$

because the two expressions differ only for $x \in V$, in which case $\chi((\Sigma x)^s) = 0$ and $|\chi(f(x))| = 1$.

The right hand side can be estimated using Lemma 5.6:

$$q^{-k}\left|\sum_{x\in S}\chi((\Sigma x)^s)\right| = q^{-k}\left|\sum_{x\in S}\psi(\Sigma x)\right| \le 3\cdot q^{-2^{-k}}.$$

Combining the last two inequalities we obtain

$$\Gamma_\chi^{\text{weak}}(f) \le 3\cdot q^{-2^{-k}} + 1/q.$$

Since this inequality holds for every $\chi \neq \chi_0$ and every cylinder intersection S, it follows by Lemma 2.9 that the right hand side bounds the strong discrepancy as well, completing the proof of Lemma 5.4. ■

The lower bound for the communication complexity with help of MCS follows directly from Lemmas 3.3 and 5.4:

Theorem 5.8 $C^{\text{help},\epsilon}(\text{MCS}_{q,u,k}) \ \ge (\log q)/2^k + \log(1/2-\epsilon) - 2.$ ■

6 Bounds on One-way Communication via Help

In this section, we apply the concept of "communication with help" to derive a bound on one-way complexity of a special class of functions, when a particular player speaks first. The basic trick is inspired by Nisan and Wigderson, [20], where the 3-party case is considered. We extend their result to $k \le (1-\epsilon)\log n$ players.

In a one-way communication protocol, the players each speak exactly once, in a prespecified order. For our result, the exact order in which the players speak will not matter; we will only need that the first player to speak is only allowed to speak once. With this in mind, we define ${}^iC(f)$ to be the communication complexity of f when Player i speaks first, and only once. ${}^iC_1(f)$ is the one-way communication complexity of f when Player i speaks first, and the other players speak once in some order.

Now we define the class of functions for which we will derive one-way communication bounds. Nisan and Wigderson [20] used the following construction to obtain a k-party boolean function from a $(k-1)$-party multibit-output function.

Constuction 6.1 *Let $B = \{0,1\}^b$, and let $f : X_2 \times \ldots \times X_k \to B$ be any function. We define $\widetilde{f} : \{1,\ldots,b\} \times X_2 \times \ldots \times X_k \to \{0,1\}$ by $\widetilde{f}(i,x_2,\ldots,x_k) = f(x_2,\ldots,x_k)_i$. In other words, Player 1's input specifies a single bit of the output of f.*

Lemma 6.2 *Let $f, \widetilde{f}$ be as described above. Then*

$$^1C(\widetilde{f}) \;\geq\; \min\left\{b,\; C^{\text{help}}(f)/b\right\}.$$

Proof: Suppose we are given a communication protocol that computes f, that begins with Player 1 sending at most $b-1$ bits, but never speaking again. We use this protocol to construct a $(k-1)$-party protocol with help to compute the function f itself. In this protocol, on input $(x_2,\ldots,x_k)$, the Helper sends the same message Player 1 would send in the protocol for $\widetilde{f}$. Players 2 through k now compute $\widetilde{f}(i,x_2,\ldots,x_k)$ for each possible value of i, using the given protocol. After this, every bit of $f(x_1,\ldots,x_k)$ has been found, and at most $b \cdot {}^1C(\widetilde{f})$ bits of communication have been used. ■

Corollary 6.3 *Let $f = \text{GIP}_{2^{\sqrt{n}},\sqrt{n},k-1}$, and let $\widetilde{f}$ be defined as above. Then ${}^1C(\widetilde{f}) = \Omega(\sqrt{n}/2^k)$, while for $2 \leq i \leq k$, ${}^iC_1(\widetilde{f}) \leq \log n + 1$. Also, $\widetilde{f}$ can be computed by bounded fan-in boolean circuits of depth $O(\log n \log k)$.*

Proof: The lower bound on ${}^1C(\widetilde{f})$ follows directly from Corollary 4.12 and Lemma 6.2. The upper bound on ${}^iC_1(\widetilde{f})$ for $2 \leq i \leq k$ is achieved by the protocol in which Player i first sends all the bits of x_1, then Player 1 sends the output.

To obtain a small-depth circuit, observe that two elements of $\mathbb{F}_{2^{\sqrt{n}}}$ can be multiplied in depth $O(\log n)$. Therefore a single component of the componentwise product can be computed in depth $O(\log n \log k)$. Adding the terms takes only an additional $O(\log n)$ depth. ■

Theorem 1.7 is an immediate consequence of Corollary 6.3. ■

A similar result holds for the MCS function:

Corollary 6.4 *Let $f = \text{MCS}_{q,u,k}$ and let $\widetilde{f}$ be defined as above. For all n, one can choose u and q such that $\widetilde{f}$ is polynomial time computable, $\log q = \Theta(n)$ and ${}^1C(\widetilde{f}) \geq \Omega(\sqrt{n}/2^{k/2})$, while for $2 \leq i \leq k$, ${}^iC_1(\widetilde{f}) \leq O(\log n)$ for any q and u.*

Proof: We choose u to be a prime power, and q to be an integer power of u, such that $n = \lceil \log q \rceil \approx 2^k(\log u)^2$. The lower bound on ${}^iC_1(\widetilde{f})$ follows from Theorem 5.8 and Lemma 6.2.

To establish the claim that $\widetilde{f}$ is polynomial-time computable, we first need to address the issue of how the input and output of MCS should be represented by binary strings. We defer the discussion of this to Section 7. Our choice of parameters makes the group G the multiplicative subgroup of a subfield of $\mathbb{F}_q$; it will follow by Theorem 7.3 that $\widetilde{f}$ is polynomial-time computable. ■

7 MCS: How Explicit?

The function $\mathrm{MCS}_{q,u,k}$ given in Definition 5.2 is clearly computable in polynomial time ($(k-1)$ additions and $O(\log q/u)$ multiplications in $\mathbb{F}_q$), assuming $\mathbb{F}_q$ is given explicitly as a $\log q$-bit part of the input.

What is not reasonable about this view of the function, however, is, that the output is encoded by $\log q$ bits while its information content is only $\log u$ bits. Optimal encoding of the output (hashing) makes no difference in the context of communication complexity where the players have unlimited computational power. However, it does make a difference when talking about Turing-complexity such as polynomial time.

In this section we show how to accomplish the optimal encoding in polynomial time for the MCS function under certain restrictions on the parameters.

The special case we consider is when G is the multiplicative group of a subfield $\mathbb{F}_{u+1}$ of $\mathbb{F}_q$ and the output is optimally encoded as an element of the subfield.

We shall say that $\mathbb{F}_q$ is *explicitly represented* if we are given an irreducible polynomial $f \in \mathbb{F}_p[x]$ of degree w over $\mathbb{F}_p$ where p is the characteristic and $q = p^w$. In this case, we treat $\mathbb{F}_q$ as the field $\mathbb{F}_p[x]/(f)$. In this case the element ϑ of $\mathbb{F}_q$ corresponding to the polynomial x is a generator of $\mathbb{F}_q$, i. e., $\mathbb{F}_q = \mathbb{F}_p[\vartheta]$ and f is the minimal polynomial of ϑ over $\mathbb{F}_p$. The powers $1, \vartheta, \vartheta^2, \ldots, \vartheta^{w-1}$ form a basis of $\mathbb{F}_q$ over $\mathbb{F}_p$.

We make the polynomial f part of the input, thereby increasing the length of the input by at most a factor of 2.

An optimal encoding of the elements of a subfield $K \subset \mathbb{F}_q$ is an encoding by binary strings of length $\lceil \log |K| \rceil$.

Lemma 7.1 *Let $\mathbb{F}_q$ be explicitly represented, where $q = p^w$. Let $r \mid w$ and let K be the (unique) subfield of $\mathbb{F}_q$ of order p^r. Then an optimal encoding of the elements of K can be computed in polynomial time* $((\log q)^{O(1)})$.

Proof: Let $v := |K| = p^r$. The *trace* function, $\mathrm{Tr} : \mathbb{F}_q \to K$ is defined by $\mathrm{Tr}(g) = g + g^v + g^{v^2} + \ldots + g^{v^{t-1}}$, where $t = w/r$. The trace function is K-linear and it is easily seen to be onto (cf. Ireland and Rosen [15, p. 145]). Consequently, the elements $\mathrm{Tr}(1), \mathrm{Tr}(\vartheta), \mathrm{Tr}(\vartheta^2), \ldots, \mathrm{Tr}(\vartheta^{w-1})$ span K as a linear space over $\mathbb{F}_p$. Take the first r of these that are linearly independent over $\mathbb{F}_p$ to be the "canonical" basis for K over $\mathbb{F}_p$, and represent every element of K by its coordinates with respect to this basis. This can be converted to a binary string of length $\lceil w \log q \rceil = \lceil \log |K| \rceil$. ■

We now define the "explicit MCS function" (EMCS):

Definition 7.2 Let q be a prime power, let k, r, t be positive integers, let $w = rt$, let $q = p^w$, let $u = p^r - 1$, and let $s = (q-1)/u$. We define the function $\mathrm{EMCS}_{q,u,k} : \{0,1\}^{(k+1)\lceil w \log p \rceil} \to \{0,1\}^{\lceil r \log p \rceil}$ as follows. The first $\lceil w \log p \rceil$ bits define a degree w polynomial $f \in \mathbb{F}_p[x]$. If this polynomial is reducible, output 1. Otherwise, we represent $\mathbb{F}_q$ as $\mathbb{F}_q = \mathbb{F}_p[\mathbf{x}]/(f)$, and interpret the next $k\lceil w \log p \rceil$ bits as $x_1, \ldots, x_k \in \mathbb{F}_q$, in the usual way. Compute $\mathrm{MCS}_{q,u,k}(x_1, \ldots, x_k)$ and output its canonical representation as a string of length $\lceil r \log p \rceil$ in accordance with Lemma 7.1.

Theorem 7.3 *Let q, u, k be as in Definition 7.2. Then* $\mathrm{EMCS}_{q,u,k}$ *is polynomial-time computable.*

Proof: Berlekamp's algorithm [7] decides whether a polynomial of degree d over $\mathbb{F}_p$ is irreducible, in time $O(d^2 \log^2 d \log\log d \log p)$ (cf. Rabin [22]).

The other computations require only a polynomial number of arithmetic operations over $\mathbb{F}_p$, since we can use repeated squaring when evaluating the trace function. ■

This completes the justification of Corollary 6.4.

To make our output a bit more appealing, we might wish to represent the subfield by an irreducible polynomial. This, too, can be done in polynomial time, as shown in [2].

8 Appendix A: Strong versus Weak Discrepancies

We now prove the inequality that justifies the terms "strong" and "weak" for the two kinds of discrepancies (see Remark 2.8).

We use the notation introduced in Definition 2.3.

Proposition 8.1 *For any $S \subseteq X$ and any $\chi \in \widehat{G}, \chi \neq \chi_0$, we have $\Gamma_\chi^{\text{weak}}(f, S) \leq m \cdot \Gamma(f, S)$, where $m = |G|$.*

Proof: Follows from Definition 2.1 and Proposition 2.11. ■

From Proposition 8.1, it follows by averaging that $m \cdot \Gamma(f, S)$ is also an upper bound on $\Gamma^{\text{weak}}(f, S)$. However, we can improve this upper bound by a $\sqrt{m-1}$ factor.

Theorem 8.2 $\Gamma^{\text{weak}}(f, S) \;\leq\; (m/\sqrt{m-1}) \cdot \Gamma(f, S)$.

Proof: By Proposition 2.11, we have

$$\Gamma^{\text{weak}}(f,S) \;=\; \frac{1}{(m-1)} \sum_{\chi\neq\chi_0} \left| \sum_{y\in G} \chi(y)\gamma(y) \right|,$$

where the outer sum is taken over all nonprincipal characters χ of G, and the excess, $\gamma(y)$ is defined in Definition 2.1. By the Cauchy-Schwarz inequality we obtain

$$(\Gamma^{\text{weak}}(f,S))^2 \leq \frac{1}{(m-1)} \sum_{\chi\neq\chi_0} \left| \sum_{y\in G} \chi(y)\gamma(y) \right|^2 .$$

We rewrite this inequality, and proceed to simplify it.

$$\begin{aligned}
(m-1)(\Gamma^{\text{weak}}(f,S))^2 &\leq \sum_{\chi\neq\chi_0} \left| \sum_{y\in G} \chi(y)\gamma(y) \right|^2 \\
= \;& \sum_{\chi\neq\chi_0} \sum_{y,z\in G} \chi(y-z)\gamma(y)\gamma(z) \\
= \;& \sum_{\chi\neq\chi_0} \left(\sum_{\substack{y,z\in G\\ y\neq z}} \chi(y-z)\gamma(y)\gamma(z) + \sum_{y\in G} \gamma(y)^2 \right).
\end{aligned}$$

Reversing the order of summation and simplifying, we have

$$\begin{aligned}
= \;& \sum_{\substack{y,z\in G\\ y\neq z}} -\gamma(y)\gamma(z) + (m-1)\sum_{y\in G} \gamma(y)^2 \\
= \;& \frac{1}{2} \sum_{\substack{y,z\in G\\ y\neq z}} (\gamma(y)-\gamma(z))^2 \leq m^2\Gamma(f,S)^2,
\end{aligned}$$

where the last inequality follows by Proposition 2.11 and by choosing $\alpha = 2$ and $a_i = \gamma(y_i)/\Gamma(f,S)$ in Proposition 8.3 below. ■

Proposition 8.3 *Let $A = \{a_1,\ldots,a_m\}$ be a multiset of real numbers such that, for all i, $|a_i| \leq 1$. Then for $\alpha \geq 1$, we have*

$$h(A) := \sum_{1\leq i<j\leq m} |a_i - a_j|^\alpha \;\leq\; \lfloor m^2/4 \rfloor \cdot 2^\alpha.$$

Proof: For such a multiset A, let $N^+(A)$ denote the multiplicity of $+1$ in A, and let $N^-(A)$ denote the multiplicity of -1 in A. The claimed upper bound for $h(A)$ is achieved if $\{N^-(A), N^+(A)\} = \{\lfloor m/2 \rfloor, \lceil m/2 \rceil\}$. Otherwise, suppose without loss of generality that $N^-(A) < N^+(A)$. Let $b = \min\{a \in A \mid a > -1\}$. Then replacing b by -1 yields a new multiset A' of size m, such that $h(A') > h(A)$. This shows that $h(A)$ is maximal if and only if $\{N^-(A), N^+(A)\} = \{\lfloor m/2 \rfloor, \lceil m/2 \rceil\}$, proving the Proposition. ■

9 Appendix B: Two Players with Help

For the case of two players, cylinder intersections correspond to cartesian products ("rectangles") which are much easier to handle. This fact allows us to derive a strong upper bound on the discrepancy of MCS with two players without requiring mathematics as deep as Weil's character sum estimates. We need only some basic facts about the Fourier transform over finite abelian groups (cf. [1]).

We also reproduce the three-player boolean function one-way lower bound of Nisan and Wigderson using the two-player model with help. Finally, with the author's permission, we include a lower bound for the one-way complexity of the three-player pointer jumping function due to A. Wigderson.

9.1 MCS for Two Players with Help

Let $f = \text{MCS}_{q,u,2}$ where q is a prime power and $u|q-1$ (cf. Section 5). Let $s = (q-1)/u$. Recall the definition of this function:

$$f(x,y) := \begin{cases} * & \text{if } x+y=0 \\ (x+y)^s & \text{otherwise,} \end{cases}$$

where the "$*$" stands for "undefined." As before, the discrepancy bound will hold for any mapping of the undefined entries to the range $G := \{z^s \mid z \in \mathbb{F}_q^\times\}$.

To obtain an upper bound on the discrepancy of f we will use the following result about Fourier transforms over finite abelian groups (see [1, Thms. 4.1 and 6.8]).

Theorem 9.1 *Let q be a prime power, $s \mid q-1$, and $S_1, S_2 \subseteq \mathbb{F}_q$. For each $z \in \mathbb{F}_q^\times$, let N_z be the number of solutions $(x,y) \in S_1 \times S_2$ to the equation*

$$(x+y)^s = z. \tag{12}$$

Then, for each $z \in \mathbb{F}_q^\times$,

$$|N_z - |S_1||S_2|s/q| < \sqrt{|S_1||S_2|q}. \qquad ■ \tag{13}$$

Let $\mathcal{C}$ be the set of cylinder intersections (rectangles) in $\mathbb{F}_q \times \mathbb{F}_q$.

Lemma 9.2 $\Gamma(f, \mathcal{C}) < 2q^{-1/2}$.

Proof: Let $V = \{(x, y) \in \mathbb{F}_q \times \mathbb{F}_q : x + y = 0\}$. Let S be a rectangle in $\mathbb{F}_q \times \mathbb{F}_q$, i. e., $S = S_1 \times S_2$, where $S_1, S_2 \subseteq \mathbb{F}_q$. Let $z \in G$.

From Definition 2.1 and Theorem 9.1, we have

$$\begin{aligned}
\gamma(z) = \left| \frac{|f^{-1}(z) \cap S| - |S|/u}{|X|} \right| &\leq \frac{1}{q^2} \left(\left| N_z - \frac{|S|}{u} \right| + |V| \right) \\
&\leq \frac{1}{q^2} \left(\left| N_z - \frac{|S|\,(q-1)}{uq} \right| + \frac{|S|}{uq} + q \right) \\
&< \frac{1}{q^2} \left(\sqrt{|S|q} + \frac{|S|}{uq} + q \right) \\
&\leq q^{-1/2} + (uq)^{-1} + q^{-1} \\
&< 2q^{-1/2}.
\end{aligned}$$

Thus, by definition, $\Gamma(f, \mathcal{C}) \leq 2q^{-1/2}$. ■

Plugging this into Lemma 3.3 and choosing q to be an n-bit prime yields the following theorem:

Theorem 9.3 $C^{\text{help}}(f) \geq n/2 - 2$. ■

Applying Lemma 6.2 to f, we find

Corollary 9.4 ${}^1C(\widetilde{f}) \geq \sqrt{n/2} - 1$, *while when* $i = 2$ *or* $i = 3$, ${}^iC_1(\widetilde{f}) \leq \log n + 1$.

This result is analogous to the three-player one-way bound of Nisan and Wigderson [20] for their "bits of hash value" function (see Section 9.2).

9.2 Universal Hash Functions and the Nisan–Wigderson One-way Bound

Nisan and Wigderson [20] prove an $\Omega(\sqrt{n})$ lower bound for the 3-party one-way complexity of the "bits of hash value" (BHV) function derived from universal families of hash functions. Our limited understanding of their proof served as the initial motivation for this paper. In this section we present a version of their proof in the formal framework of "communication with help." We replace their somewhat cryptic handling of certain conditional probabilities with the analysis via multicolor discrepancy. Nisan and Wigderson invoke the "Leftover Hash Lemma" by Y. Mansour, N. Nisan, and P. Tiwari [18] to estimate a conditional probability; we rephrase this idea in terms of a multicolor discrepancy bound.

Definition 9.5 Let $X = \{0,1\}^{2n}, Y = Z = \{0,1\}^n$. We may think of X as a 2-universal family of hash functions, $x : Y \to Z$ (see Carter and Wegman [10]).

Let $1 \leq \ell \leq n$. Let

$$f_\ell : X \times Y \to \{0,1\}^\ell$$

be defined by $f_\ell(x,y) =$ the first ℓ bits of $x(y)$. We shall also use the notation $\mathrm{HV}_{n,\ell} = f_\ell$ (for ℓ-truncated n-bit hash value).

Recall that Construction 6.1 creates a function $\widetilde{f_\ell} : \{1,\dots,\ell\} \times X \times Y \to \{0,1\}$ defined by $\widetilde{f_\ell}(i,x,y) := f_\ell(x,y)_i = (x(y))_i$. We shall also use the notation $\mathrm{BHV}_{n,\ell} = \widetilde{f_\ell}$ ("bits of hash value"). The main result explained in this section is the following 3-player one-way lower bound.

Theorem 9.6 (Nisan, Wigderson [20]) ${}^1C(\mathrm{BHV}_{n,\sqrt{n/2}}) \geq \sqrt{n/2}$.

Following Nisan and Wigderson [20] we employ the "Leftover Hash Lemma" [18]. In our framework, this lemma implies an upper bound on the discrepancy of f_ℓ over the set of cylinder intersections (rectangles) in $X \times Y$.

Lemma 9.7 ("Leftover Hash Lemma," Mansour, Nisan, Tiwari [18]) *Let* $Y_0 \subseteq Y, Z_0 \subseteq Z, X_0 \subseteq X$, *and* $p = |Z_0|/|Z|$. *Then*

$$|\Pr[x(y) \in Z_0 \mid x \in X_0,\, y \in Y_0] \ - p| \ \leq \sqrt{p|X|/(|X_0||Y_0|)}.$$

Lemma 9.8 $\Gamma(f_\ell, \mathcal{C}) \leq 1/2^{(n+\ell)/2}$.

Proof: Let S be a cylinder intersection in $X \times Y$. For two dimensions, observe that cylinder intersections correspond to cartesian products of subsets, so there exist $X_S \subseteq X$ and $Y_S \subseteq Y$ such that $S = X_S \times Y_S$.

In the notation of Definition 2.1, let $\alpha \in \{0,1\}^\ell$ maximize $|\gamma(\alpha)|$ for the function f_ℓ and the set S. Then

$$\Gamma(f_\ell, S) = |\gamma(\alpha)| = \frac{1}{|X||Y|}\left|\left|f_\ell^{-1}(\alpha) \cap S\right| - \frac{|S|}{m}\right|.$$

Let $Z_0 := \{z \in \{0,1\}^n \ \mid$ the first ℓ bits of z are $\alpha\}$ so $p = 2^{-\ell}$. With this notation,

$$\begin{aligned}\Gamma(f_\ell, S) &= \frac{1}{|X||Y|} \cdot \left||f_n^{-1}(Z_0) \cap S| - \frac{|S|}{m}\right| \\ &= \frac{|S|}{|X||Y|}\left(\Pr_{(x,y)\in S}[x(y) \in Z_0] - p\right)\end{aligned}$$

(Here $f_n(x,y) = x(y)$, untruncated.) Since $S = X_S \times Y_S$, we infer from Lemma 9.7 that

$$\begin{aligned}\Gamma(f_\ell, S) &\le \sqrt{\frac{p|X|}{|S|}} \cdot \frac{|S|}{|X||Y|} = \sqrt{\frac{p|S|}{|X||Y|^2}} \\ &\le \sqrt{p/|Y|} = 2^{-(\ell+n)/2}. \qquad \blacksquare\end{aligned}$$

This bound and Lemma 3.3 imply the $\Omega(n)$ lower bound for the 2-party communication complexity with help of the HV function:

Theorem 9.9 $C^{\text{help}}(\text{HV}_{n,\ell}) \ge (n+\ell)/2 - 1.$ ■

Now an application of Lemma 6.2 to $\widetilde{f_\ell} = \text{BHV}_{n,\ell}$ with $\ell = \lceil\sqrt{n/2}\rceil$, together with Theorem 9.9, yield Theorem 9.6. ■

9.3 One-way separation for pointer jumping

In this section we present an unpublished result of A. Wigderson, analyzing the one-way complexity of the 3-party "pointer jumping" function via two-party communication with help.

Definition 9.10 For a positive integer n, we set $[n] := \{1, \ldots, n\}$. Let $m_0, m_1, \ldots, m_k$ be positive integers. The *k-party composition function* takes as input a k-tuple $(f_1, \ldots, f_k)$ of functions, where $f_1 : [m_0] \to [m_1], f_2 : [m_1] \to [m_2], \ldots, f_k : [m_{k-1}] \to [m_k]$, and returns their composition, $f_k \circ f_{k-1} \circ \ldots \circ f_1$. In the special case when $m_0 = 1, m_k = 2$, this function is boolean, and is called the *k-party pointer jumping function*.

We consider the 3-party pointer jumping function, where $n = m_2 = m_1^2$. In this case, the inputs f_1, f_2, f_3 can be represented by strings of $\log n/2, \sqrt{n}\log n/2, n$ bits, respectively (for convenience, we assume all quantities are integral).

Theorem 9.11 (Wigderson) *For the 3-party pointer jumping function, where $n = m_2 = m_1^2$, the one-way communication complexity is $\Omega(\sqrt{n})$ when the players speak in the order $1, 2, 3$, but is $O(\log n)$ for any other order.*

Proof: We follow Wigderson's outline [24]. For the k-party pointer jumping function, if the communication order is not $1, 2, \ldots, k$, then there exist $i < j$ such that Player j speaks before Player i. Let Player j's message encode the value $f_{j-1} \circ \ldots \circ f_1(1)$. This takes at most $\log m_{j-1}$ bits. Since Player i is given $f_j, \ldots, f_k$, she can compute the output without further communication. This shows that the one-way communication complexity for pointer jumping is $O(\log n)$ except possibly for the ascending communication order, where n is the bit-length of the longest input.

Now consider the 3-party pointer jumping function with $m_1 = \sqrt{n}$ and $m_2 = n$. As in Section 6, we relate the one-way communication complexity of the 3-party pointer jumping function to a restricted version of communication complexity with help for a 2-party composition function. Specifically, we consider the communication game where Player 2 sees input f_3, Player 3 sees input f_2, and Player 1 sees both functions. (There is no longer an input f_1.) The goal is to compute the composition $f_3 \circ f_2$. Communication occurs as in the two-way "communication with help" model, with Player 1 acting as Helper, except that after the help message, communication is one-way, with Player 2 speaking before Player 3. Additionally, we will restrict Player 1's help message to have length less than $\sqrt{n}/4$.

Suppose there were a one-way communication protocol of cost less than $\sqrt{n}/6$ for the 3-party pointer jumping function (computing $f_3 \circ f_2 \circ f_1$). From this, we construct a protocol of cost less than $n/6$ for the 2-party composition function (computing $f_3 \circ f_2$): Player 1's message is used as the help message. For each $1 \leq i \leq m_1$, Player 2 sends the message he would have sent if $f_1(1)$ were i. The total cost is less than $n/6$, and Player 3 can compute $f_3 \circ f_2(i)$ for every i, and so can output the answer.

We now develop the machinery that will allow us to deduce lower bounds for the two-party problem described above.

Definition 9.12 Let Y be a set and let n be a positive integer. For any n-tuple $y \in Y^n$ and non-negative integer r, we define the *Hamming ball of radius r centered at y*, $B_r(y)$, to be the set of all points $z \in Y$ whose Hamming distance from y is less than or equal to r. We say that S is a *near-Hamming ball* centered at y if there exists an r such that $B_r(y) \subseteq S \subseteq B_{r+1}(y)$.

Notation 9.13 For the rest of this Appendix, we adopt the following notational conventions: m_1, m_2, m_3 are arbitrary positive integers, f is an element of $[m_2]^{[m_1]}$, g, g', g_0 are elements of $[m_3]^{[m_2]}$, S, S_0, T are subsets of $[m_3]^{[m_2]}$, b is a function $b : [m_2]^{[m_1]} \to [m_3]^{[m_1]}$. Moreover, when f and g are selected at random, f is always assumed to be uniformly distributed over $[m_2]^{[m_1]}$ and g uniformly distributed over S.

Definition 9.14 We say that (b, S) form a *standard pair* if there exists a function g_0 such that for every f, $b(f) = g_0 \circ f$, and such that S is a near-Hamming ball centered at g_0.

Definition 9.15 We set

$$p(b,S,f) := \Pr_g[b(f) = g \circ f],$$

and

$$p(b,S) := \Pr_{f,g}[b(f) = g \circ f].$$

The following combinatorial lemma is our main technical tool.

Lemma 9.16 *Let k be a positive integer. Then $\max\{p(b,S) \mid |S| = k\}$ is attained by the standard pairs (b,S) such that $|S| = k$.*

Proof:

Claim 1: Fix b and suppose there exists g_0 such that for all f, $b(f) = g_0 \circ f$. Then $p(b,S)$ is maximized, subject to the constraint $|S| = k$, when S is any near-Hamming ball centered at g_0.

To see this, fix g and consider $\Pr_f[b(f) = g \circ f]$. This probability is $(1 - d(g,g_0)/m_2)^{m_1}$, where $d(g,g_0)$ denotes the Hamming distance between g and g_0. Since this function is decreasing, the result is clear.

Claim 2: For all pairs (b,S), for all g_0, there exists S_0 such that $|S_0| = |S|$ and for all f, $p(b,S,f) \leq p(b_0,S_0,f)$, where b_0 denotes the function $b_0(f) = g_0 \circ f$.

The proof is by a shifting argument. The following construction incrementally modifies the pair (b,S) in m_2 rounds, to produce the output (b_0,S_0). At the end of each iteration, for every f, $p(b,S,f)$ has increased or remained the same, and $|S|$ also remains the same. From this, the claim follows. Here is the construction:

```
for $j := 1$ to $m_2$
    $T := \emptyset$
    for $g \in S$
        Let $g'$ be defined by $g'(i) = \begin{cases} g_0(j) & \text{if } i = j \\ g(i) & \text{otherwise.} \end{cases}$
        if $g' \notin S$ then add $g'$ to $T$
                else add $g$ to $T$
    endfor
    $S := T$
    Redefine $b$ so that, for all $f$, $\begin{cases} b(f)(i) = g_0(j) \text{ when } f_2(i) = j \\ b(f)(i) \text{ is unchanged otherwise.} \end{cases}$
endfor
```

It is clear that the output (b_0, S_0) produced by this algorithm satisfies $|S_0| = |S|$ and for every f, $b_0(f) = g_0 \circ f$. To see that, for any f, $p(b, S, f)$ does not decrease at step j, we observe that after round j, the number of functions $g \in S$ such that $g(j) = g_0(j)$ is at least as great as the number of instances of any previous value of $g(j)$; in particular, for any f, i such that $f(i) = j$, it is at least as great as the number of instances that $g(j) = b(f)(i)$. This proves Claim 2.

Now, suppose we are given a pair (b, S) maximizing $p(b, S)$. By Claim 2, we can replace this by a pair (b_0, S_0), where there exists g_0 such that b is defined by $b(f) = g_0 \circ f$, without decreasing $p(b, S, f)$ for any f. In particular, $p(b, S)$ must still be maximal. Now, by Claim 1, we know that S_0 must be a near-Hamming ball, and so (b_0, S_0) is a regular pair. ■

Remark 9.17 One can show that, under the conditions of Lemma 9.16, the *only* pairs (b, S) attaining the maximum value are the standard pairs.

Corollary 9.18 *For all* b, S *such that* $|S| = k$*, we have*

$$p(b, S) \leq e^{-m_1/4} + 2^{5m_2/6}(m_3 - 1)^{m_2/4}/k.$$

Proof: By Lemma 9.16, we may assume that b is defined by $b(f) = g_0 \circ f$ for some g_0. Suppose $g \in S$ has Hamming distance i from g_0. Then

$$\Pr_f[g \circ f = g_0 \circ f] = (1 - i/m_2)^{m_1} \leq e^{-im_1/m_2}.$$

It follows that

$$p(b, S) \leq \frac{1}{k} \sum_{i=0}^{m_2} |S_i| e^{-im_1/m_2},$$

where S_i denotes the set of elements of S at Hamming distance i from g_0. Those terms for which $i \geq m_2/4$ contribute at most $ke^{-m_1/4}$ total to the sum. On the other hand, there are exactly $\binom{m_2}{i}(m_3 - 1)^i$ functions g at Hamming distance i from g_0. It follows, since the value of the binary entropy function $H(1/4) = 0.8113... < 5/6$, that the number of functions g with Hamming distance at most $m_2/4$ from g_0 is less than $2^{5m_2/6}(m_3 - 1)^{m_2/4}$. Adding these estimates completes the proof. ■

We now complete the proof of Theorem 9.11. Recall that, on input (f_2, f_3), Player 1 sends help message $H(f_2, f_3)$. Let $M(h)(f_3)$ denote the message Player 2 would send on input f_3 under help message h.

For any input pair (f_2, f_3), let $h = H(f_2, f_3)$ denote the message sent by the Helper, Player 1, and let $s = M(h)(f_3)$ denote the message sent by Player 2. For such a pair of messages (h, s), we define sets

$$\begin{aligned} X_{h,s} &:= \{(f_2, f_3) \mid h = H(f_2, f_3), s = M(h)(f_3)\} \\ S_{h,s} &:= M(h)^{-1}(s) = \{f_3 \mid s = M(h)(f_3)\}. \end{aligned}$$

In the one-way communication protocol with help, after the message pair (h, s) has been sent, Player 3 computes the output $f_3 \circ f_2$. Hence the communication protocol defines, for each message pair (h, s), a function $b_{h,s} : [m_2]^{[m_1]} \to [m_3]^{[m_1]}$, defined by $b_{h,s}(f_2) :=$ Player 3's output given input f_2, when message h is sent by the Helper, and message s is sent by Player 2. (We may assume that the protocol specifies an output for Player 3, even given communication strings which could not be sent for an actual input pair.) Observe that

$$X_{h,s} \subseteq \{(f_2, f_3) \mid b_{h,s}(f_2) = f_3 \circ f_2, f_3 \in S_{h,s}\},$$

and thus

$$|X_{h,s}| \leq \Pr[b_{h,s}(f_2) = f_3 \circ f_2] \cdot m_2^{m_1} \cdot |S_{h,s}|,$$

where $f_2 \in [m_2]^{[m_1]}$ and $f_3 \in S_{h,s}$ are chosen uniformly. It follows, by Corollary 9.18, with $m_1 = \sqrt{n}, m_2 = n, m_3 = 2, S = S_{h,s}, b = b_{h,s}$, that

$$|X_{h,s}| \leq (e^{-\sqrt{n}/4}|S_{h,s}| + 2^{5n/6})n^{\sqrt{n}}$$

The sets $X_{h,s}$ partition the input space $[n]^{[\sqrt{n}]} \times [2]^{[n]}$. Hence,

$$\begin{aligned} 2^n n^{\sqrt{n}} &= \sum_{h,s} |X_{h,s}| \\ &\leq \sum_{h,s} (e^{-\sqrt{n}/4}|S_{h,s}| + 2^{5n/6})n^{\sqrt{n}}. \end{aligned}$$

Now, since for fixed h, the sets $S_{h,s}$ partition $[2]^{[n]}$, we have

$$2^n \leq \sum_h e^{-\sqrt{n}/4} 2^n + \sum_{h,s} 2^{5n/6}.$$

It follows that either the number of help messages h is at least $2^{\sqrt{n}(\log e/4)-1}$ or the number of message pairs (h, s) is at least $2^{n/6-1}$. As discussed earlier, this proves that the original 3-party protocol had cost at least $\sqrt{n}/6$. This proves Theorem 9.11. ■

Acknowledgment.

We wish to thank Avi Wigderson for allowing us to include his bound for the pointer-jumping function in this paper (Section 9.3).

References

[1] L. Babai. *The Fourier Transform and Equations Over Finite Abelian Groups*, Lecture Notes, manuscript, version 1.2, December 1989.

[2] L. Babai, T. Hayes. *Finding a generator for a subfield of a finite field.* Manuscript, 2000. .

[3] L. Babai, T. Hayes, and P. Kimmel. *The Cost of the Missing Bit: Communication Complexity with Help.* (Preliminary version) 30th ACM STOC (1998) 673–682.

[4] L. Babai, P. Kimmel, and S. V. Lokam. *Simultaneous Messages vs. Communication.* 12th Symp. Theor. Asp. Comp. Sci., 1995, Springer Lecture Notes in Comp. Sci. **900** (1995) 361–372. Updated version: L. Babai, A. Gál, P. Kimmel, and S. V. Lokam. *Simultaneous Messages vs. Communication.* SIAM J. Comp., submitted.

[5] L. Babai, N. Nisan, and M. Szegedy. *Multiparty Protocols, Pseudorandom Generators for Logspace and Time-Space Trade-offs.* J. Comp. Sys. Sci. **45**, 1992, 204–232. (Preliminary version: 20th ACM STOC (1988) 539–550.)

[6] R. Beigel and J. Tarui. *On ACC.* 32nd IEEE FOCS (1991) 783–792.

[7] E.R. Berlekamp. *Factoring polynomials over large finite fields.* Math. Comput. **24** (1970) 713–735.

[8] A.K. Chandra, M.L. Furst, and R.J. Lipton. *Multiparty Protocols.* 15th ACM STOC (1983) 94–99.

[9] F.R.K. Chung and P. Tetali. *Communication complexity and quasi randomness.* SIAM J. Discrete Math. **6** (1993) 110–123.

[10] L. Carter and M. Wegman, *Universal Hash Functions*. J. Comp. Sys. Sci. **18** (1979) 143–154.

[11] H. D. Gröger, G. Turán. *On linear decision trees computing boolean functions.* 18th ICALP, Springer Lect. Notes in Comp. Sci. **510** (1991) 707–719.

[12] V. Grolmusz. *Separating the Communication Complexities of MOD m and MOD p Circuits.* J. Comp. Sys. Sci. **51** (1995) 307–313. (Preliminary version: 33rd IEEE FOCS (1992) 278–287.)

[13] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, G. Turán. *Threshold circuits of bounded depth.* 28th IEEE FOCS (1987) 99–110.

[14] J. Håstad and M. Goldmann. *On the Power of Small-Depth Threshold Circuits.* Computational Complexity **1** (1991) 113–129.

[15] K. Ireland and M. Rosen. *A Classical Introduction to Modern Number Theory.* Springer-Verlag, 1982.

[16] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, 1997.

[17] M. Luby, B. Veličković, and A. Wigderson. *Deterministic Approximate Counting of Depth-2 Circuits.* Proc. Second Israel Symp. on Theory and Computing Systems, IEEE Comp. Soc. Press, 1993, 18–24.

[18] Y. Mansour, N. Nisan, and P. Tiwari. *The Computational Complexity of Universal Hashing.* Theor. Comp. Sci. **107** (1993) 121–133.

[19] N. Nisan. *The communication complexity of threshold gates.* In: Combinatorics, Paul Erdős is Eighty (D. Miklós, T. Szőnyi, V. T. Sós, eds.), Vol. 1. Bolyai Society Mathematical Studies 1, Budapest 1993 (distributed by the A. M. S.), 301–315.

[20] N. Nisan, A. Wigderson. *Rounds in Communication Complexity Revisited.* SIAM J. Comp. **22** (1993) 211–219. (Preliminary version: 23rd ACM STOC (1991) 419–429.)

[21] P. Pudlák, V. Rödl, and J. Sgall. *Boolean Circuits, Tensor Ranks and Communication Complexity.* SIAM J. Comp. **26** (1997) 605–633.

[22] M. Rabin, *Probabilistic Algorithms in Finite Fields.* SIAM J. Comp. **9** (1980) 273–280.

[23] W. M. Schmidt: *Equations over Finite Fields: An Elementary Approach.* Lect. Notes in Math. **536**, Springer 1976.

[24] A. Wigderson. Personal communication. April 20, 1997.

[25] A. C-C. Yao. *On ACC and Threshold Circuits.* 31st IEEE FOCS (1990) 619–627.