The multiparty communication complexity of interleaved group products

W. T. Gowers* Emanuele Viola†

August 10, 2016

Abstract

Party A_i of k parties $A_1, \ldots, A_k$ receives on its forehead a t-tuple $(a_{i1}, \ldots, a_{it})$ of elements from the group $G = \mathrm{SL}(2, q)$. The parties are promised that the interleaved product $a_{11} \ldots a_{k1} a_{12} \ldots a_{k2} \ldots a_{1t} \ldots a_{kt}$ is equal either to the identity e or to some other fixed element $g \in G$. Their goal is to determine which of e and g the interleaved product is equal to, using the least amount of communication.

We show that for all fixed k and all sufficiently large t the communication is $\Omega(t \log |G|)$, which is tight. As an application, we establish the security of the leakage-resilient circuits studied by Miles and Viola (STOC 2013) in the "only computation leaks" model. Our main technical contribution is of independent interest. We show that if X is a probability distribution on G^m such that any two coordinates are uniform in G^2, then a pointwise product of s independent copies of X is nearly uniform in G^m, where s depends on m only.

*Royal Society 2010 Anniversary Research Professor.

†Supported by NSF grant CCF-1319206. Work done in part while a visiting scholar at Harvard University, with support from Salil Vadhan's Simons Investigator grant, and while at the Simons Institute for the Theory of Computing. Email: viola@ccs.neu.edu.

1 Introduction and our results

In the multiparty number-on-the-forehead model of communication complexity [CFL83] (cf. [Yao79, KN97]), k parties $A_1, \ldots, A_k$ wish to compute the value $f(x_1, \ldots, x_k)$ of some function of k variables, where each x_i belongs to some set X_i. The party A_i knows the values of all the x_j apart from x_i (one can think of x_i as being written on A_i's forehead) and the aim is for the parties to determine $f(x_1, \ldots, x_k)$. To do this, they are allowed to write bits on a blackboard according to some protocol: the *communication complexity* of f is the smallest number of bits they will need to write in the worst case.

The overlap of information makes proving lower bounds in this model useful and challenging. Useful, because such bounds find a striking variety of applications; see for example the book [KN97] for some of the earlier ones. This paper adds to the list an application to cryptography. Challenging, because obtaining a lower bound even for $k = 3$ parties typically requires different techniques from those that may work for $k = 2$ parties. For example, consider the disjointness function. After the two-party lower bounds [KS92, Raz92] it took more than fifteen years to obtain multiparty lower bounds [LS09, CA08], and required new techniques, cf. [She08]. In fact, despite substantial effort, tight bounds are not known even for $k = 3$ parties: see [She14] for the state of the art. The lower bounds in this paper are proved via a new technique: *boosting pairwise uniformity*.

In this paper we consider the following problem, posed in [MV13]. Each x_i is a sequence $(a_{i1}, \ldots, a_{it})$ of group elements, and we define their *interleaved product* to be

$$x_1 \bullet x_2 \bullet \cdots \bullet x_k = a_{11} \ldots a_{k1} a_{12} \ldots a_{k2} \ldots a_{1t} \ldots a_{kt},$$

which we shall sometimes write as $\prod_{j=1}^{t} a_{1j} \ldots a_{kj}$. In other words, the entire input is a $k \times t$ matrix of elements from G, party i knows all the elements except those in row i, and the interleaved product is the product in column order. The parties are told that $x_1 \bullet \cdots \bullet x_k$ is equal either to the identity e or to a specified group element g, and their job is to determine which.

If the group is abelian the problem can be solved with communication $O(1)$ by just two players, using the public-coin protocol for equality. Over any non-solvable group, a communication lower bound of $t/2^{O(k)}$ follows via [Bar89] from the lower bound in [BNS92] for generalized inner products; see [MV13] for details. However, this bound is far from the (trivial) upper bound of $O(t \log |G|)$, and it gives nothing when $t = O(1)$. Motivated by a cryptographic application which is reviewed below, the paper [MV13] asks whether a lower bound that grows with the size of the group, ideally $\Omega(t \log |G|)$, can be established over some group G for $k = 8$ parties.

In the case of $k = 2$ parties, such $\Omega(t \log |G|)$ bounds have been recently established by the authors [GV15] over the group $G = \mathrm{SL}(2, q)$ of 2×2 matrices with determinant 1 over the field with q elements.

Here we show that if $t \geq b^{2^k}$ where b is a certain constant, then the communication is at least $(t/b^{2^k}) \log |G|$, even for randomized protocols that are merely required to offer a small advantage over random guessing. In particular, for all fixed k and all sufficiently large t we obtain an $\Omega(t \log |G|)$ lower bound, which is tight.

Theorem 1.1. *There is constant b such that the following holds. Let $G = SL(2, q)$ for a prime power q. Let $P : G^{k\times t} \to \{0, 1\}$ be a c-bit k-party number-on-forehead communication protocol. For $g \in G$ denote by p_g the probability that P outputs 1 over a uniform input $(a_{i,j})_{i\le k, j\le t}$ such that $\prod_{j=1}^{t} a_{1j} \dots a_{kj} = g$.*

If $t \ge b^{2^k}$ then for any two $g, h \in G$ we have $|p_g - p_h| \le 2^c \cdot |G|^{-t/b^{2^k}}$.

To prove this theorem we establish a result about mixing of tuples of group elements which is of independent interest. Call a distribution over G^m *pairwise uniform* if any two coordinates are uniform in G^2. We show that the product of a sufficiently large number of pairwise uniform distributions over G^m is approximately uniform over the entire space G^m.

Theorem 1.2. *Let $G = SL(2, q)$. For every $m \ge 2$ there exists r such that the following is true. Let $\mu_1, \dots, \mu_r$ be pairwise uniform distributions on G^m. Let μ be the component-wise product of the μ_i. Then μ is $1/|G|$ close in statistical distance to the uniform distribution over G^m.*

As we shall see later, the parameter $1/|G|$ is not too important in the sense that it can be made smaller by making r larger. Note that the assumption that the distributions are pairwise uniform cannot be relaxed to the assumption that each coordinate is uniform. A simple counterexample is to take $m = 2$ and every distribution to be uniform on the set of points of the form (x, x).

Application to leakage-resilient cryptography. We now informally describe the application to cryptography we alluded to before – for formal definitions and extra discussion we refer the reader to [MV13]. Also motivated by successful attacks on cryptographic hardware, an exciting line of work known as *leakage-resilient cryptography* considers models in which the adversary obtains more information from cryptographic algorithms than just their input/output behavior. Starting with [ISW03], a general goal in this area is to compile any circuit into a new "shielded" circuit that is secure even if an adversary can obtain partial information about the values carried by the internal wires of the circuit during the computation on inputs of their choosing. This partial information can be modeled in two ways.

One way is the "only computation leaks" model [MR04]. Here the compiled circuit is partitioned (by the compiler) into topologically ordered sets of wires, i.e. in such a way that the value of each wire depends only on wires in its set or in sets preceding it. Goldwasser and Rothblum [GR12] give a construction that is secure against any leakage function that operates separately on each set, as long as the function has bounded output length.

Another way is the "computationally bounded model," where the leakage function has access to all wires simultaneously but it is computationally restricted [FRR+10].

The paper [MV13] gives a new construction of leakage-resilient circuits based on group products. This construction enjoys strong security properties in the "computationally bounded model" [MV13, Mil14]. Moreover, the construction was shown to be secure even in the "only computation leaks" model assuming that a lower bound such as that in Theorem 1.1 specialized to $k = 8$ parties holds.

In this work we obtain such bounds and thus we also obtain the following corollary.

Corollary 1.3. *The leakage-resilient construction in [MV13] is secure in the "only computation leaks" model.*

Proof. Combine Theorem 1.1 with Theorem 1.7 in [MV13]. □

This corollary completes the program of proving that the construction in [MV13] is secure in both the "only computation leaks" and the "computationally bounded" models. It seems to be the first construction to achieve this.

1.1 Overview of techniques

In this section we give an overview of our techniques. A key idea is to obtain our main Theorem 1.1 by showing that a certain collection of group products are jointly close to uniform. The group products to consider arise naturally from an application of the "box norm" (a.k.a. the multiparty norm). We state this result as Theorem 1.5 below, after a definition of the measure of closeness to uniform that we will work with.

Definition 1.4. *A distribution D on G^m is* (ϵ, k)-good *if for any $1 \leq i_1 < \cdots < i_k \leq m$ and any $g_1, \ldots, g_k \in G$, the probability, when x is sampled randomly from D, that $x_{i_j} = g_j$ for $j = 1, \ldots, k$ is between $(1-\epsilon)n^{-k}$ and $(1+\epsilon)n^{-k}$.*

To relate this definition to that of statistical distance, note that if a distribution D on G^m is (ϵ, k)-good then the projection of D to any k coordinates is ϵ-close to uniform in statistical distance.

Theorem 1.5. *There is constant b such that the following holds. Let k and t be integers, $G = SL(2, q)$, $m = 2^k$, and $t \geq b^m$.*

Let $x_1^0, x_1^1, \ldots, x_k^0, x_k^1$ be chosen independently and uniformly from G^t and consider the distribution μ on G^m whose coordinate $\epsilon \in \{0,1\}^k$ is the interleaved product

$$x_1^{\epsilon_1} \bullet x_2^{\epsilon_2} \bullet \cdots \bullet x_k^{\epsilon_k}.$$

Then μ is $(1/|G|^{t/b^m}, m)$-good.

The proof that Theorem 1.5 implies Theorem 1.1 is a technically simple application of the "box norm."

To prove Theorem 1.5 we begin by writing μ as a pointwise product of t independent distributions μ_i over G^m. This pointwise product can also be regarded as convolution $*$ in G^m, or simply as sampling from each distribution and then outputting the product. An important observation now is that each μ_i is pairwise uniform: the marginal distribution on any two coordinates will be uniform over G^2. This is the only property of the distributions that we shall use: We show (cf. Theorem 1.2) that the product of a sufficiently large number of *pairwise uniform* distributions over G^m is approximately uniformly distributed over the entire space G^m. To prove this latter fact, we focus on the case $m = 3$; larger values of m are handled using induction. We establish a "flattening lemma" that show that multiplying

four pairwise uniform distributions on G^3 increases their min-entropy (equivalently, reduces their infinity norm). The proof of this relies on a result from our previous work [GV15], which is all that we use about the group. We note however that the arguments in [GV15] and in this paper are rather different. The technique of boosting pairwise uniformity is not present in [GV15]. Another difference is reflected in the length t of the tuples to which the bounds apply. In [GV15] $t = 2$ suffices, but here the lower bound only kicks in when t is sufficiently large.

The authors' previous work [GV15] has already sparked interest among other group theorists: Shalev proved [Sha16] some of the conjectures in [GV15] obtaining also results related to a conjecture by Thompson in group theory. We hope that the present paper will stimulate further interaction with other group theorists. Some specific open problems are (1) to understand the minimum length t of the tuples for which a lower bound holds, (2) to improve the dependency on k, and (3) to extend the results to other groups. Regarding the latter point, we note that in the case of $k = 2$ parties and tuples of length $t = 2$ a lower bound holds for all finite (non-abelian) simple groups [GV15]. For sharper bounds see [Sha16]. But it is still not clear how to prove a similar extension for other settings of t or k.

Organization. In §2 we show that Theorem 1.5 implies the communication lower bound Theorem 1.1. The next sections are then devoted to the proof of Theorem 1.5 and Theorem 1.2. Specifically, §3 reduces our task to that of going from pairwise to three-wise uniformity. And then we establish the latter in §4, where our flattening lemma is proved.

2 Proof that Theorem 1.5 implies Theorem 1.1

Consider the function $d : G^{tk} \to \{0, 1, -1\}$ that maps $x = (x_1, \ldots, x_k)$ to 1 if $x_1 \bullet \cdots \bullet x_k = e$, to -1 if $x_1 \bullet \cdots \bullet x_k = g$, and to 0 otherwise. Then we have

$$|p_g - p_h| = 0.5 \cdot n \cdot |E_x(-1)^{P(x)} d(x)|.$$

Following previous work [BNS92, CT93, Raz00, VW08], we bound the latter expectation using the *box norm* $\|d\|_\square$ of d, defined as

$$\|d\|_\square^{2^k} = \mathbb{E}_{x_1^0, x_1^1, x_2^0, x_2^1, \ldots, x_k^0, x_k^1} \prod_{\epsilon \in \{0,1\}^k} d(x_1^{\epsilon_1}, \ldots, x_k^{\epsilon_k}).$$

Specifically, by e.g. Corollary 3.11 in [VW08], we have

$$0.5 \cdot n \cdot |E_x(-1)^{P(x)} d(x)| \leq 0.5 \cdot n \cdot 2^c \cdot \|d\|_\square.$$

To conclude it remains to notice that Theorem 1.5 allows us to bound $\|d\|_\square$. First, note that the product in $\|d\|_\square^{2^k}$ is equal to zero unless each of the 2^k interleaved products $x_1^{\epsilon_1} \bullet \cdots \bullet x_k^{\epsilon_k}$ is equal to 1 or g, in which case it is 1 if the number of products equal to g is even and -1 if it is odd. If the 2^k products were uniform and independent, then the expectation of the product would be zero. If instead they are $(\alpha, 2^k)$-good then $\|d\|_\square^{2^k} \leq 2^k \alpha / n^{2^k}$, and so $\|d\|_\square \leq 2\alpha^{1/2^k}/n$. Plugging in $\alpha = 1/|G|^{t/b^m}$ for $m = 2^k$ completes the proof.

3 Going from 2-wise uniformity to 3-wise suffices

In this section we prove Theorem 1.5 assuming the next theorem which will be proved in the next section. Recall that we write $*$ for convolution, and that the convolution of two distributions μ and ν is the same as the distribution obtained by sampling independently from μ and ν and outputting the product.

Theorem 3.1. *There is a constant d such that the following holds. Let $\mu_1, \dots, \mu_d$ be $(1/\sqrt{n}, 2)$-good probability distributions on G^3. Then $\mu_1 * \cdots * \mu_d$ is $(1/n^2, 3)$-good.*

The choice of polynomials $1/\sqrt{n}$ and $1/n^2$ will be convenient in a later proof by induction, but is not too important. Indeed, any bound of this type can be quickly improved if one makes the products slightly longer, as shown by the following lemma which we will use several times.

Lemma 3.2. *Let μ and ν be (ϵ, k)-good probability distributions on G^m. Then $\mu * \nu$ is (ϵ^2, k)-good.*

Proof. The convolution of the projection of the distributions on to any k coordinates is the same as the projection of the convolution, so it is enough to consider the case $m = k$. Let H be any finite group (we shall be interested in the case $H = G^k$), let U be the uniform distribution on H, and let μ and ν be distributions on H such that $\|\mu - U\|_\infty$ and $\|\nu - U\|_\infty$ are both at most ϵ/n. Let $\alpha = \mu - U$ and $\beta = \nu - U$. Then for every x we have

$$\mu * \nu(x) = \sum_{yz=x} \left(\frac{1}{n} + \alpha(y)\right)\left(\frac{1}{n} + \beta(z)\right) = \frac{1}{n} + \sum_{yz=x} \alpha(y)\beta(z).$$

where the second inequality follows from the fact that α and β are functions that sum to zero.

But $|\sum_{yz=x} \alpha(y)\beta(z)| \le n(\epsilon/n)^2 = \epsilon^2/n$, from which it follows that $\|\mu * \nu - U\|_\infty \le \epsilon^2/n$. Applying this when $H = G^k$ we obtain the result. □

Using the above two results and induction we obtain the following simple corollary of Theorem 3.1.

Corollary 3.3. *There is a constant d such that the following holds. Let $m \ge 3$, and let $\mu_1, \dots, \mu_{d^m}$ be $(1/n, 2)$-good probability distributions on G^m, where $|G| = n$. Then $\mu_1 * \cdots * \mu_{d^m}$ is $(1/n, m)$-good.*

Proof of Corollary 3.3 from Theorem 3.1. In the proof we use that (ϵ, k)-good implies $(\epsilon, k-1)$-good, as can be seen by summing on one coordinate. We prove this by induction on m. For $m = 3$ this is Theorem 3.1. Now we assume the corollary for $m - 1$ and prove it for m. Let ν_i, $i = 1, \dots, d$, be the product of d^{m-1} consecutive μ_i. For an element y of G^m we write y^0 for the first $m - 3$ coordinates, and y^1 for the other three. Pick any $x \in G^m$. We need to bound

$$\Pr[*_{i \le d} \nu_i = x] = \Pr[*_{i \le d} \nu_i^1 = x^1 | *_{i \le d} \nu_i^0 = x^0] \cdot \Pr[*_{i \le d} \nu_i^0 = x^0]. \quad (1)$$

By induction, each ν_i^0 is $(1/n, m-3)$-good, and so by Lemma 3.2 $*_{i\le d}\nu_i^0$ is $(1/n^2, m-3)$-good. Thus, the rightmost term in Equation (1) is between $(1-1/n^2)/n^{m-3}$ and $(1+1/n^2)/n^{m-3}$.

Now we bound the conditional probability in Equation (1). Let α_i be the distribution v_i^1 on G^3 conditioned on any fixing of v_i^0. We claim that α_i is $(1/\sqrt{n}, 2)$-good. Indeed, by the assumption the probability p that two coordinates of α_i equal any fixed pair satisfies

$$\frac{(1-1/n)/n^{m-1}}{(1+1/n)/n^{m-3}} \le p \le \frac{(1+1/n)/n^{m-1}}{(1-1/n)/n^{m-3}}$$

which implies

$$\frac{1-1/\sqrt{n}}{n^2} \le p \le \frac{1+1/\sqrt{n}}{n^2}$$

for large enough n.

Hence, by Theorem 3.1 the convolution of the α_i is $(1/n^2, 3)$-good.

Putting together these bounds for the two factors in the right-hand side of Equation (1) we get that the probability in Equation (1) lies between $(1-1/n^2)^2/n^m$ and $(1+1/n^2)^2/n^m$, from which the result follows. □

We remark that this corollary implies Theorem 1.2 stated in the introduction.

The last piece that we need to prove Theorem 1.5 is the following lemma.

Lemma 3.4. *Let μ be the distribution over G^m in Theorem 1.5, and let also t be as in Theorem 1.5. Then μ is the component-wise product of t independent distributions, each of which is pairwise uniform.*

Proof. Recall that $m = 2^k$. Let us write $x_i^{\epsilon_i} = (a_{i1}^{\epsilon_i}, \dots, a_{ik}^{\epsilon_i})$. Then for each $\epsilon \in \{0,1\}^k$ we have

$$x_1^{\epsilon_1} \bullet x_2^{\epsilon_2} \bullet \cdots \bullet x_k^{\epsilon_k} = a_{11}^{\epsilon_1} \dots a_{k1}^{\epsilon_k} a_{12}^{\epsilon_1} \dots a_{k2}^{\epsilon_k} \dots a_{1t}^{\epsilon_1} \dots a_{kt}^{\epsilon_k}.$$

Now for $1 \le j \le t$ let s_j be the m-tuple $(a_{1j}^{\epsilon_1} \dots a_{kj}^{\epsilon_k})_{\epsilon\in\{0,1\}^k}$. Then the m-tuple $(x_1^{\epsilon_1} \bullet x_2^{\epsilon_2} \bullet \cdots \bullet x_k^{\epsilon_k})_{\epsilon\in\{0,1\}^k}$ is the pointwise product of the s_j. That is, writing $s_j(\epsilon)$ for $a_{1j}^{\epsilon_1} \dots a_{kj}^{\epsilon_k}$, we have that

$$x_1^{\epsilon_1} \bullet x_2^{\epsilon_2} \bullet \cdots \bullet x_k^{\epsilon_k} = s_1(\epsilon)s_2(\epsilon)\dots s_t(\epsilon)$$

for every $\epsilon \in \{0,1\}^k$.

Note that the m-tuples s_j are independent and distributed as follows. We choose elements $u_1^0, u_1^1, u_2^0, u_2^1, \dots, u_k^0, u_k^1$ uniformly and independently at random from G and we form an m-tuple s by setting $s(\epsilon) = u_1^{\epsilon_1} u_2^{\epsilon_2} \dots u_k^{\epsilon_k}$.

We note that s is pairwise uniform. That is, if you take any pair of distinct elements ϵ, η in $\{0,1\}^k$, then the pair $(s(\epsilon), s(\eta))$ is uniformly distributed in G^2. To see this, choose some i such that $\epsilon_i \ne \eta_i$. Conditioning on the values of $u_j^{\epsilon_j}$ and $u_j^{\eta_j}$ for every $j \ne i$, we find that we are looking at two products of the form $au_i^{\epsilon_i}b$ and $cu_i^{\eta_i}d$. For this to equal (g,h), we need $u_i^{\epsilon_i} = a^{-1}gb^{-1}$ and $u_i^{\eta_i} = c^{-1}hd^{-1}$. Since $u_i^{\epsilon_i}$ and $u_i^{\eta_i}$ are independent and uniformly distributed, this happens with probability $1/|G|^2$. □

We note that the distribution s in the proof is far from being uniformly distributed, since there are only n^{2k} possible 2^k-tuples of this form.

Now Theorem 1.5 follows easily.

Proof of Theorem 1.5. Let m be 2^k. Let d be the constant in Corollary 3.3. Write the distribution μ as the product of t independent distributions μ_i, each of which is pairwise uniform, using Lemma 3.4. Group the μ_i in consecutive blocks of length d^m. The convolution in each block is $(1/n, m)$-good by Corollary 3.3. By repeated applications of Lemma 3.2 we obtain that the final distribution is $(1/n^{t/b^m}, m)$-good. The change of the constant from d to b is to handle the case in which t/d^m is not a power of two. □

4 From 2-wise to 3-wise uniformity

In this section we prove Theorem 3.1. First we fix some notation. Throughout the section G is the group $\mathrm{SL}(2, q)$ and $|G| = n$. For a real-valued function f, its ℓ_∞, ℓ_1, and ℓ_2 norms are respectively $\|f\|_\infty = \max_x |f(x)|$, $\|f\|_1 = \sum_x |f(x)|$, and $\|f\|_2 = (\sum_x f(x)^2)^{1/2}$. To prove the theorem we show that if we convolve pairwise-uniform distributions over G^3, then we reduce their ℓ_∞ norm. To get a sense of the parameters, note that the assumption of pairwise uniformity implies an upper bound of $1/n^2$ on this norm, and that the minimum possible value is $1/n^3$. So we are aiming to use convolutions to get down from $1/n^2$ to about $1/n^3$. Actually, it is more convenient to work with the ℓ_2 norm, but by convolving again we can turn to the ℓ_∞ norm thanks to the following simple fact.

Fact 4.1. *For any distributions μ and ν it holds that $\|\mu * \nu\|_\infty \le \|\mu\|_2 \|\nu\|_2$.*

Proof. For any x, $\mu * \nu(x) = \sum_y \mu(y)\nu(y^{-1}x) \le \sqrt{\sum_y \mu(y)^2}\sqrt{\sum_y \nu(y)^2}$, using the Cauchy-Schwarz inequality. □

We rely on the following result from our previous work.

Theorem 4.2 ([GV15])**.** *Let $G = SL(2, q)$. Let u and v be two independent distributions over G^2. Let a be sampled according to u and b according to v. Then, for every $g \in G$:*

$$|\Pr_{a,b}[a \bullet b = g] - 1/n| \le n^{-\Omega(1)} \cdot n \cdot \|u\|_2 \cdot \|v\|_2.$$

To get a sense of the parameters, note that if u and v are uniform over an α and β fraction of G^2 respectively, then $\|u\|_2 = (\alpha n^2)^{-1/2}$ and $\|v\|_2 = (\beta n^2)^{-1/2}$, and so the upper bound is $(\alpha\beta)^{-1/2} n^{-\Omega(1)}/n$. Thinking of α and β as constants, and summing over all elements of G, we can then deduce that the distribution of $a \bullet b$ is $1/n^{\Omega(1)}$ close to uniform over G in statistical distance. However, jumping ahead we will apply this theorem not to distributions but to non-negative functions.

We now state and prove the flattening lemma.

Lemma 4.3. *Let μ and ν be two non-negative functions defined on G^3 and suppose that however you fix two coordinates of one of the functions and sum over the third, the total is at most n^{-2}. Then $\|\mu * \nu\|_2^2 \leq n^{-3} + n^{-\Omega(1)}\sqrt{\|\mu\|_\infty \|\nu\|_\infty}$.*

Proof. Expanding out the definition of $\|\mu * \nu\|_2^2$ we obtain

$$\sum_{x_1y_1=z_1w_1} \sum_{x_2y_2=z_2w_2} \sum_{x_3y_3=z_3w_3} \mu(x_1,x_2,x_3)\nu(y_1,y_2,y_3)\mu(z_1,z_2,z_3)\nu(w_1,w_2,w_3).$$

We can rewrite this as

$$\sum_{a,b,x_1,x_2,y_1,y_2} \sum_{x_3y_3=z_3w_3} \mu(x_1a,x_2,x_3)\nu(y_1,by_2,y_3)\mu(x_1,x_2b,z_3)\nu(ay_1,y_2,w_3).$$

By averaging, it follows that there exist x_1, x_2, y_1, y_2 such that

$$\|\mu * \nu\|_2^2 \leq n^4 \sum_{a,b} \sum_{x_3y_3=z_3w_3} \mu(x_1a,x_2,x_3)\nu(y_1,by_2,y_3)\mu(x_1,x_2b,z_3)\nu(ay_1,y_2,w_3). \qquad (2)$$

Define $\alpha(a,x)$ to be $\mu(x_1a,x_2,x)$, $\beta(b,y)$ to be $\nu(y_1,by_2,y)$, $\gamma(b,z)$ to be $\mu(x_1,x_2b,z)$ and $\delta(a,w)$ to be $\nu(ay_1,y_2,w_3)$. Then we can rewrite this inequality as

$$\|\mu * \nu\|_2^2 \leq n^4 \sum_{a,b} \sum_{xy=zw} \alpha(a,x)\beta(b,y)\gamma(b,z)\delta(a,w).$$

Now let us set $u(x,w)$ to be $\sum_a \alpha(a,x)\delta(a,w)$ and $v(y,z)$ to be $\sum_b \beta(b,y)\gamma(b,z)$. Note that by our hypotheses on μ and ν,

$$\sum_x u(x,w) = \sum_a \delta(a,w) \sum_x \alpha(a,x) = \sum_a \delta(a,w) \sum_x \mu(x_1a,x_2,x) \leq n^{-2} \sum_a \delta(a,w) \leq n^{-4}, \qquad (3)$$

with three similar inequalities for summing over the other coordinate and for v. We also have for each x, w that

$$u(x,w) \leq \|\alpha\|_\infty \sum_a \delta(a,w) = n^{-\Omega(1)}\|\mu\|_\infty n^{-2},$$

with a similar argument giving the same bound for $\|v\|_\infty$.

Our earlier bound (2) can be rewritten as

$$\|\mu * \nu\|_2^2 \leq n^4 \sum_{xy=zw} u(x,w)v(y,z).$$

On the right-hand side there is an interleaved product of the kind to which Theorem 4.2 can be applied. Since by (3) we have $\sum_x u(x,w) \leq n^{-4}$ for each w, we also have that

$\sum_{x,w} u(x,w) \le n^{-3}$. From this and our estimate for $\|u\|_\infty$ it follows that $\sum_{x,w} u(x,w)^2 \le \|\mu\|_\infty/n^5$. And we have a similar bound for v. Therefore, Theorem 4.2 implies that

$$\sum_{xy=zw} u(x,w)v(y,z) = n^{-1}\|u\|_1\|v\|_1 + n \cdot n^{-\Omega(1)}\|u\|_2\|v\|_2 \le n^{-\Omega(1)}\sqrt{\|\mu\|_\infty\|\nu\|_\infty}/n^4.$$

Note that here we are applying the theorem to the probability distributions $u/\|u\|_1$ and $v/\|v\|_1$, and then we multiply by $\|u\|_1\|v\|_1$. This implies that

$$\|\mu * \nu\|_2^2 \le n^{-3} + n^{-\Omega(1)}\sqrt{\|\mu\|_\infty\|\nu\|_\infty},$$

which proves the result. □

Corollary 4.4. *Let $\mu_1, \mu_2, \mu_3, \mu_4$ be non-negative functions defined on G^3 and suppose that they all satisfy the condition that μ and ν satisfy in Lemma 4.3. Further suppose that $\|\mu_i\|_\infty \le \alpha$ for every $i \in \{1,2,3,4\}$.*

Then

$$\|\mu_1 * \mu_2 * \mu_3 * \mu_4\|_\infty \le n^{-3} + n^{-\Omega(1)}\alpha.$$

Proof. This follows by applying Lemma 4.3 to μ_1 and μ_2 and to μ_3 and μ_4 and then applying Fact 4.1 to $\mu_1 * \mu_2$ and $\mu_3 * \mu_4$. □

The next lemma shows that convolution preserves one of the main properties we used.

Lemma 4.5. *Let μ and ν be non-negative functions defined on G^3 and suppose that whenever you fix two coordinates of μ or ν and sum over the other, you get at most n^{-2}. Then the same is true of $\mu * \nu$.*

Proof. For each $(z_1, z_2, z_3) \in G$ we have

$$\mu * \nu(z_1, z_2, z_3) = \sum_{x_1}\sum_{x_2}\sum_{x_3} \mu(x_1, x_2, x_3)\nu(x_1^{-1}z_1, x_2^{-1}z_2, x_3^{-1}z_3).$$

If we fix z_1 and z_2 and sum over z_3 we obtain

$$\sum_{x_1}\sum_{x_2}\sum_{x_3,z_3} \mu(x_1, x_2, x_3)\nu(x_1^{-1}z_1, x_2^{-1}z_2, x_3^{-1}z_3).$$

But for each x_1, x_2, we have

$$\sum_{x_3,z_3} \mu(x_1, x_2, x_3)\nu(x_1^{-1}z_1, x_2^{-1}z_2, x_3^{-1}z_3) = \sum_x \mu(x_1, x_2, x)\sum_y \nu(x_1^{-1}z_1, x_2^{-1}z_2, y) \le n^{-4}.$$

The result follows on summing over x_1 and x_2. □

Proof of Theorem 3.1. If we divide each μ_i by $(1 + 1/\sqrt{n})$, then we obtain functions ν_i that satisfy the conditions of Lemma 4.3 and such that $\|\nu_i\|_\infty$ is at most $1/n^2$. Applying Corollary 4.4 a constant number of times, using Lemma 4.5 to argue that the assumptions are satisfied throughout, we deduce that a convolution of a constant number ℓ of such functions has infinity norm at most $n^{-3}(1 + n^{-\Omega(1)})$. If we now multiply one such convolution by $(1 + 1/\sqrt{n})^\ell$ we obtain a probability distribution μ with

$$\|\mu\|_\infty \leq n^{-3}(1 + n^{-\Omega(1)})(1 + 1/\sqrt{n})^\ell \leq n^{-3}(1 + n^{-\Omega(1)})$$

for large enough n.

This is close to our goal of bounding $\|\mu - U\|_\infty$. To achieve the goal, we use the following fact about any two probability distributions α and β over G^3:

$$\|\alpha * \beta - U\|_\infty^2 = \|(\alpha - U) * (\beta - U)\|_\infty^2 \leq \|\alpha - U\|_2^2 \|\beta - U\|_2^2 = (\|\alpha\|_2^2 - 1/n^3)(\|\beta\|_2^2 - 1/n^3).$$

In our case we have $\|\mu\|_2^2 \leq \|\mu\|_\infty \leq n^{-3}(1 + n^{-\Omega(1)})$. So we convolve one more time and apply the above fact to obtain a distribution μ' such that $\|\mu' - U\|_\infty \leq n^{-\Omega(1)}/n^3$. Hence, μ' is $(n^{-\Omega(1)}, 3)$-good. Now if we convolve another constant number of times and apply Lemma 3.2 we obtain a distribution which is $(n^{-2}, 3)$-good, as desired. □

References

[Bar89] David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *J. of Computer and System Sciences*, 38(1):150–164, 1989.

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. of Computer and System Sciences*, 45(2):204–232, 1992.

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity*, Technical Report TR08-002, 2008.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *15th ACM Symp. on the Theory of Computing (STOC)*, pages 94–99, 1983.

[CT93] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM Journal on Discrete Mathematics*, 6(1):110–123, 1993.

[FRR+10] Sebastian Faust, Tal Rabin, Leonid Reyzin, Eran Tromer, and Vinod Vaikuntanathan. Protecting circuits from leakage: the computationally-bounded and noisy cases. In *Int. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 135–156, 2010.

[GR12] Shafi Goldwasser and Guy N. Rothblum. How to compute in the presence of leakage. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2012.

[GV15] W. T. Gowers and Emanuele Viola. The communication complexity of interleaved group products. In *ACM Symp. on the Theory of Computing (STOC)*, 2015.

[ISW03] Yuval Ishai, Amit Sahai, and David Wagner. Private circuits: Securing hardware against probing attacks. In *Int. Cryptology Conf. (CRYPTO)*, pages 463–481, 2003.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[KS92] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[LS09] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[Mil14] Eric Miles. Iterated group products and leakage resilience against NC^1. In *ACM Innovations in Theoretical Computer Science conf. (ITCS)*, 2014.

[MR04] Silvio Micali and Leonid Reyzin. Physically observable cryptography. In *Theory of Cryptography Conf. (TCC)*, pages 278–296, 2004.

[MV13] Eric Miles and Emanuele Viola. Shielding circuits with groups. In *ACM Symp. on the Theory of Computing (STOC)*, 2013.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[Raz00] Ran Raz. The BNS-Chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

[Sha16] Aner Shalev. Mixing, communication complexity and conjectures of Gowers and Viola. *Combinatorics, Probability and Computing*, pages 1–13, 6 2016. arXiv:1601.00795.

[She08] Alexander A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[She14] Alexander A. Sherstov. Communication complexity theory: Thirty-five years of set disjointness. In *Symp. on Math. Foundations of Computer Science (MFCS)*, pages 24–43, 2014.

[VW08] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4:137–168, 2008.

[Yao79] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing. In *11th ACM Symp. on the Theory of Computing (STOC)*, pages 209–213, 1979.