Interleaved group products*

W. T. Gowers† Emanuele Viola‡

December 18, 2017

Abstract

Let G be the special linear group $\mathrm{SL}(2, q)$. We show that if $(a_1, \ldots, a_t)$ and $(b_1, \ldots, b_t)$ are sampled uniformly from large subsets A and B of G^t then their interleaved product $a_1 b_1 a_2 b_2 \cdots a_t b_t$ is nearly uniform over G. This extends a result of the first author [Gow08], which corresponds to the independent case where A and B are product sets. We obtain a number of other results. For example, 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. Extensions to other groups are also discussed.

We obtain closely related results in communication complexity, which is the setting where some of these questions were first asked by Miles and Viola [MV13]. For example, suppose 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 G. 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 $g \in G$, and their goal is to determine which of the two the product is equal to. We show that for all fixed k and all sufficiently large t the communication is $\Omega(t \log |G|)$, which is tight. Even for $k = 2$ the previous best lower bound was $\Omega(t)$. As an application, we establish the security of the leakage-resilient circuits studied by Miles and Viola [MV13] in the "only computation leaks" model.

*Preliminary versions of this paper have appeared as [GV15, GV16].

†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 in part while visiting the Simons institute for the theory of computing. Email: viola@ccs.neu.edu.

1 Introduction and our results

Let G be a finite group. All our results are asymptotic in the size of the group, so G should be considered large. Suppose we have m probability distributions X_i over G, each of high entropy. For this discussion, we can think of each X_i as being uniform over a constant fraction of G. We will first consider the case where the X_i are independent, and later we will throw in dependencies.

If we sample x_i from X_i and output the product $x_1 \dots x_m$, the resulting distribution $D = \prod_{i \le m} X_i$ is the convolution of the distributions X_i. Our aim is to show that D closely approximates the uniform distribution on G. More precisely, we ask for the approximation to be uniform: we would like to show that

$$|\mathbb{P}[D = g] - 1/|G|| \le \epsilon/|G|,$$

for every $g \in G$. Such a bound guarantees in particular that D is supported over the entire group. It also immediately implies that D is ϵ-close to uniform in statistical distance, that is, in the ℓ_1 norm.

The above goal has many applications in group theory, see for example [Gow08, BNP08] and the citations therein. As discussed later, it is also closely related to problems in communication complexity and cryptography.

As a warm-up, consider the case $m = 2$. Here we have only two distributions X and Y, and it is easy to see that XY does not mix, no matter which group is considered. Indeed, let X be uniform over an arbitrary subset S of G of density $1/2$, and let Y be (uniform over) the set of the same density consisting of all the elements in G except the inverses of the elements in S, i.e., $Y := G \setminus S^{-1}$. It is easy to see that XY never equals 1_G. (In some groups we get a good ℓ_2 approximation, so when we say that XY does not mix we mean in the uniform sense mentioned above.)

Now consider the case $m = 3$, so we have three distributions X, Y, and Z. Here the answer depends on the group G. It is easy to see that if G has a large non-trivial subgroup H then $D := XYZ$ does not mix. Indeed, we can just let each distribution be uniform over H. It is also easy to see that $X + Y + Z$ does not mix over the abelian group Z_p. For example, if $X = Y = Z$ are uniform over $\{0, 1, \dots, p/4\}$ then $X + Y + Z$ is never equal to $p - 1$.

However, for other groups it is possible to establish a good bound. This was shown by Gowers [Gow08]. The following version of the result was given by Babai, Nikolov, and Pyber. The inequality can be stated in terms of the 2-norm of the probability distributions. In this paper we shall use two different normalizations of the 2-norm, so to avoid confusion we shall use different notation for the two.

Notation 1.1. *Let G be a finite group and let $v : G \to \mathbb{R}$. We denote by $\|v\|_2$ the ℓ_2 norm $\sqrt{\sum_x v(x)^2}$ of v and by $\|v\|_{L_2}$ the L_2 norm $\sqrt{\mathbb{E}_x v(x)^2}$ of v. Here $\mathbb{E}$ denotes the average over G.*

The inequality of Babai, Nikolov, and Pyber is the following.

Theorem 1.2 ([BNP08])**.** *Let G be a finite group and let X and Y be two independent distributions over G with product D. Let U be the uniform distribution over G. Then*

$$\|D - U\|_2 \leq \|X\|_2 \|Y\|_2 \sqrt{|G|/d},$$

where d is the minimum dimension of a non-trivial representation of G.

It is also essentially present in [Gow08] as Lemma 3.2: there the inequality is stated (in an equivalent form) in the case where one of the two distributions is uniform over a subset of G, but the proof yields the more general result with hardly any modification. (Babai, Nikolov and Pyber give a slightly different argument, however, and it was subsequently observed by several people that there is a short and natural proof using non-abelian Fourier analysis.)

From Theorem 1.2 and the Cauchy-Schwarz inequality one can immediately deduce the following inequality for three distributions, where a uniform bound is obtained.

Theorem 1.3 ([BNP08])**.** *Let G be a group, and let g be an element of G. Let X, Y, and Z be three independent distributions over G. Then*

$$|\mathbb{P}[XYZ = g] - 1/|G|| \leq \|X\|_2 \|Y\|_2 \|Z\|_2 \sqrt{|G|/d},$$

where d is the minimum dimension of a non-trivial representation of G.

When each distribution is uniform over a constant fraction of G, the right-hand side becomes

$$O(d^{-1/2})/|G|.$$

Note that the parameter ϵ in our goal above then becomes $O(d^{-1/2})$. We mention that for any non-abelian simple group we have $d \geq \sqrt{\log |G|}/2$, whereas for G the special linear group $\mathrm{SL}(2, q)$ we have $d \geq |G|^{1/3}$, cf. [Gow08]. In particular, for $G = \mathrm{SL}(2, q)$ we have that XYZ is ϵ-close to uniform over the group, where $\epsilon = 1/|G|^{-\Omega(1)}$. Later we shall also give an alternative proof of this last bound.

Dependent distributions. In this paper we consider the seemingly more difficult case where there may be dependencies across the X_i. To set the scene, consider three distributions A, Y, and A', where A and A' may be dependent, but Y is independent of (A, A'). Must the distribution AYA' mix? It is not hard to see that the answer is no. Indeed, let Y be uniformly distributed over an arbitrary set S of density $1/2$. We may now define (A, A') to be uniformly distributed over all pairs (x, y) such that $1_G \notin xSy$. Then the marginal distributions A and A' are both uniform over the whole of G, but AYA' is never equal to 1_G.

One of our main results is that mixing does, however, occur for distributions of the form $ABA'B'$, where A and A' are dependent, and B and B' are also dependent, but (A, A') and (B, B') are independent. Moreover, if we look at interleaved products of longer tuples, then the bound scales in the desired way with the length t of the tuples.

Definition 1.4. *The* interleaved product $a \bullet b$ *of two tuples* $(a_1, a_2, \ldots, a_t)$ *and* $(b_1, b_2, \ldots, b_t)$ *is defined as*

$$a \bullet b := a_1 b_1 a_2 b_2 \cdots a_t b_t.$$

Theorem 1.5. *Let* $G = SL(2, q)$. *Let* $A, B \subseteq G^t$ *have densities* α *and* β *respectively. Let* $g \in G$. *If* a *and* b *are selected uniformly from* A *and* B *we have*

$$|\mathbb{P}[a \bullet b = g] - 1/|G|| \leq (\alpha\beta)^{-1}|G|^{-\Omega(t)}/|G|.$$

In particular, the distribution $a \bullet b$ is at most $(\alpha\beta)^{-1}|G|^{-\Omega(t)}$ away from uniform in statistical distance. Here $\Omega(t)$ denotes a function that is bounded below by ct for some $c > 0$.

For the case of $t = 2$ we obtain a result that applies to arbitrary distributions and is sharper: the factor $1/\alpha\beta$ is improved to $\sqrt{1/\alpha\beta}$.

Theorem 1.6. *Let* G *be the group* $SL(2, q)$. *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* $g \in G$,

$$|\mathbb{P}_{a,b}[a \bullet b = g] - 1/|G|| \leq \gamma|G| \cdot \|u\|_2 \|v\|_2,$$

where γ *can be taken to be of the form* $|G|^{-\Omega(1)}$.

To get a feel for what this bound is saying, note that if u and v are uniform over subsets of G^2 of densities α and β, respectively, then $\|u\|_2 = (\alpha|G|^2)^{-1/2}$ and $\|v\|_2 = (\beta|G|^2)^{-1/2}$, and so the upper bound is $(\alpha\beta)^{-1/2}\gamma/|G|$. Thus, in general we get a good uniform bound provided that $\alpha\beta$ is significantly greater than γ^2, so for the γ above we can take α and β as small as $|G|^{-\Omega(1)}$.

From Theorem 1.6 we obtain a number of other results which we now describe. 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 close to uniform over the entire space G^m.

Theorem 1.7. *Let* $G = SL(2, q)$. *For every* $m \geq 2$ *there exists* r *such that the following is true. Let* $\mu_1, \ldots, \mu_r$ *be pairwise uniform distributions on* G^m. *Let* μ *be the distribution obtained by taking the pointwise product of random samples from 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 each μ_i to be uniform on the set of points of the form $(x, x, \ldots, x)$.

As mentioned earlier, our results also imply a special case of Theorem 1.3. Recall that the latter bounds the distance between XYZ and uniform. Our Theorem 1.5 with $t = 2$ immediately implies a similar result but with four distributions, i.e., a bound on the distance

of $WXYZ$ from uniform. To obtain a result about three distributions like Theorem 1.3 we make a simple and general observation that mixing in four steps implies mixing in three, see §6. Thus we recover, up to polynomial factors, the bound in Theorem 1.3 for the special case of $G = \mathrm{SL}(2, q)$. Unlike the original proofs, ours avoids representation theory.

A more significant benefit of our proof is that it applies to other settings. For example, in the next theorem we can prove the XYZ result even if one of the distribution is dense only within a conjugacy class, as opposed to the whole group.

Theorem 1.8. *Let $G = SL(2, q)$. For all but $O(1)$ conjugacy classes S of G the following holds. Let A be a subset of S of density $|A|/|S| = \alpha$. Let B and C be subsets of G of densities $|B|/|G| = \beta$ and $|C|/|G| = \gamma$. Pick a, b, and c uniformly and independently from A, B, and C respectively. Let $g \in G$. Then $|\mathbb{P}[abc = g] - 1/|G|| \leq (\alpha\beta\gamma)^{-1}|G|^{-\Omega(1)}/|G|$.*

We show that theorems 1.5, 1.6, and 1.7 above, and Theorem 1.11 in §1.1 hold for any group G that satisfies a certain condition about conjugacy classes. We then prove that that condition is satisfied for $\mathrm{SL}(2, q)$. In order to state the condition, we need some notation.

Notation 1.9. *Let G be a group. For $x \in G$ we write $\mathbf{C}(x)$ for the uniform distribution on the conjugacy class of x, i.e., the distribution $u^{-1}xu$ for uniform $u \in G$. Different occurrences of $\mathbf{C}$ correspond to independent choices of u.*

The theorem below states that the condition is satisfied (so the condition is the conclusion of the theorem).

Theorem 1.10. *Let $G = SL(2, q)$ and let $a \in G$. Then*

$$\mathbb{E}_{b,b' \in G}\mathbb{P}[\mathbf{C}(ab^{-1})\mathbf{C}(b) = \mathbf{C}(ab'^{-1})\mathbf{C}(b')] \leq (1 + \gamma)/|G|$$

with $\gamma = |G|^{-\Omega(1)}$.

Actually for our results we only need this theorem when a is uniform over G. We note that the left-hand side of the inequality above is the collision probability of the distribution $\mathbf{C}(ab^{-1})\mathbf{C}(b)$ for uniform $b \in G$.

We conjectured [GV15, GV16] that Theorem 1.10 holds for all groups of Lie type of bounded rank, and that a weaker version of Theorem 1.10 holds for all non-abelian simple groups. After our work, Shalev confirmed this [Sha16]. As a consequence, theorems 1.5, 1.6, 1.7, and 1.11 hold as stated for groups of Lie type of bounded rank, and weaker versions of theorems 1.5, 1.6, 1.7, and 1.11 hold for all non-abelian simple groups. This improves on a result in [GV15] which also applied to all non-abelian simple groups. For a concrete example consider the alternating group. For this group [Sha16] proves a bound of the same form as that of Theorem 1.10 but with $\gamma = 1/\log^{\Omega(1)}|G|$. This yields Theorem 1.6 for the alternating group with this weaker γ. Miles and Viola [MV13] show that this result is best possible up to the constant in the $\Omega(1)$, and so the same applies to the result in [Sha16] about the alternating group.

Our results above are closely related to results in *communication complexity*, which is the setting in which some of them were originally asked [MV13]. We discuss this perspective next.

1.1 Communication complexity

Computing the product $\prod_{i \le t} g_i$ of a given tuple $(g_1, \ldots, g_t)$ of elements from a group G is a fundamental task. This is for two reasons. First, depending on the group, this task is complete for various complexity classes [KMR66, CM87, Bar89, BC92, IL95, Mil14]. For example, Barrington's famous result [Bar89] shows that it is complete for NC^1 whenever the group is non-solvable, a result which disproved previous conjectures. Moreover, the reduction in this result is very efficient: a projection. The second reason is that such group products can be randomly self-reduced [Bab87, Kil88], again in a very efficient way. The combination of completeness and self-reducibility makes group products extremely versatile, see e.g. [FKN94, AIK06, GGH$^+$08, MV13].

Still, some basic open questions remain regarding the complexity of iterated group products. Here we study a communication complexity question raised by Miles and Viola in [MV13], which was the starting point of our work. This question is interesting already in Yao's basic 2-party communication model [Yao79]. However we will be able to answer it even in the multiparty number-on-forehead model [CFL83]. So we now describe the latter. For background, see the book [KN97]. There are k parties $A_1, \ldots, A_k$ who 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). They 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. This is reflected in the sequence of papers [GV15, GV16] leading to the present one.

In this paper we consider the following problem, posed in [MV13]. Each x_i is a sequence $(a_{i1}, \ldots, a_{it})$ of t 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 certain other groups a communication lower bound of $t/2^{O(k)}$ follows via [Bar89] from the lower bound in [CG88, BNS92] for generalized inner product; cf. [MV13]. However, this bound does not improve with the size of the group. In particular it is far from the (trivial) upper bound of $O(t \log |G|)$, and it

gives nothing when $t = O(1)$. We stress that no better results were known even for the case of $k = 2$ parties.

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.

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 1.11. *There is a constant b such that the following holds. Let $G = SL(2, q)$. 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 $(a_{i,j})_{i \leq k, j \leq t}$ such that $\prod_{j=1}^{t} a_{1j} \dots a_{kj} = g$. For any two $g, h \in G$ we have:*

(1) For any k, if $t \geq b^{2^k}$ then $|p_g - p_h| \leq 2^c \cdot |G|^{-t/b^{2^k}}$.

(2) For $k = 2$, if $t \geq 2$ then $|p_g - p_h| \leq 2^c \cdot |G|^{-t/b}$.

We note that the problem for $t = 1$ can be solved with $O(1)$ communication using the public-coin protocol for equality, cf. [KN97]. The same technique solves with $O(1)$ communication the variant of the $t = 2$ case where one element, say a_1, is fixed to the identity. For $k = 2$ parties we show that the case $t = 2$ is hard; for $k > 2$ it remains open to determine what is the smallest t for which the problem is hard.

We conjecture that the doubly-exponential b^{2^k} terms in Theorem 1.11 can be replaced with the singly exponential b^k. This would match the state-of-the-art lower bounds [BNS92]. In fact, we make a much bolder conjecture. Let us first review the context. A central open problem in number-on-forehead communication complexity is to prove lower bounds when the number of players is more than logarithmic in the input length, cf. [KN97]. Moreover, there is a shortage of candidate hard functions, thanks to the many clever protocols that have been obtained [Gro94, BGKL03, PRS97, Amb96, AL00, ACFN12, CS14], which in some cases show that previous candidates are easy. One candidate by Raz [Raz00] that still stands is computing one entry of the multiplication of k $n \times n$ matrices over GF(2). He proves [BNS92]-like bounds for it, and further believes that this remains hard even for k much larger than $\log n$. Our setting is different, for example because we multiply more than k matrices and the matrices can be smaller.

We conjecture that over any non-abelian simple group, the interleaved product remains hard even when the number of parties is more than logarithmic in the input length. We note that this conjecture is interesting even for a group of constant size and for deterministic protocols that compute the whole product (as opposed to protocols that distinguish with some advantage tuples that multiply to g from those that multiply to h). We are unaware of upper bounds for this problem.

For context, we mention that the works [BGKL03, PRS97, Amb96, AL00] consider the so-called generalized addressing function. Here, the first $k - 1$ parties receive an element g_i from a group G, and the last party receives a map f from G to $\{0, 1\}$. The goal is to output

$f(g_1 g_2 \cdots g_{k-1})$. For any $k \geq 2$, this task can be solved with communication $\log |G| + 1$. Note that this is logarithmic in the input length to the function which is $|G| + (k-1)\log |G|$. By contrast, for interleaved products we prove and conjecture bounds that are linear in the input length. The generalized addressing function is more interesting in restricted communication models, which is the focus of those papers.

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 1.11 specialized to $k = 8$ parties holds.

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

Corollary 1.12. *The leakage-resilient construction in [MV13] is secure in the "only computation leaks" model.*

Proof. Combine Theorem 1.11 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.

Organization. This paper is organized as follows. In §2 we exhibit a series of reductions, valid in all groups, that reduce proving Theorem 1.6 to Theorem 1.10. In §3 we use Theorem 1.6 to prove Theorem 1.7. Then in §4 we use Theorem 1.7 to prove the communication-complexity lower bounds in Theorem 1.11. We also give some simple equivalences between mixing and communication complexity, yielding the mixing Theorem 1.5. Finally, Theorem

1.10 is proved in §5. In §6 we give an alternative proof of Theorem 1.3 for SL(2,q), and also prove Theorem 1.8.

2 Reducing Theorem 1.6 to Theorem 1.10

In this section we prove that Theorem 1.6 follows from Theorem 1.10. We need a definition and a couple of lemmas. Here it is convenient to work with L_2-norms instead of ℓ_2-norms, and other norms we discuss will also be defined using averages rather than sums.

Definition 2.1 (Box norm)**.** *Let $f : X_1 \times X_2 \times \cdots \times X_k \to \mathbb{R}$ be a function. Define the* box norm $\|f\|_\square$ *of f as*

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

In this section we only use the box norm with $k = 2$ but later we will need larger k as well.

The first lemma is standard and says that a function with small box norm has small correlation with functions of the form $(x, y) \mapsto u(x)v(y)$.

Lemma 2.2. *Let X and Y be finite sets, let $u : X \to \mathbb{R}$, let $v : Y \to \mathbb{R}$ and let $f : X\times Y \to \mathbb{R}$. Then*

$$|\mathbb{E}_{x,y} f(x, y)u(x)v(y)| \le \|f\|_\square \|u\|_{L_2}\|v\|_{L_2}.$$

Proof. The proof uses two applications of the Cauchy-Schwarz inequality. We have

$$\begin{aligned}
(\mathbb{E}_{x,y} f(x, y)u(x)v(y))^4 &= ((\mathbb{E}_x u(x)\mathbb{E}_y f(x, y)v(y))^2)^2\\
&\le ((\mathbb{E}_x u(x)^2)(\mathbb{E}_x(\mathbb{E}_y f(x, y)v(y))^2)^2\\
&= \|u\|_{L_2}^4 \,(\mathbb{E}_{y,y'} v(y)v(y')\mathbb{E}_x f(x, y)f(x, y'))^2\\
&\le \|u\|_{L_2}^4(\mathbb{E}_{y'y'} v(y)^2v(y')^2)\left(\mathbb{E}_{y,y'}(\mathbb{E}_x f(x, y)f(x, y'))^2\right)\\
&= \|u\|_{L_2}^4\|v\|_{L_2}^4\|f\|_\square^4.
\end{aligned}$$

The result follows on taking fourth roots. □

If we do not have a bound for $\|f\|_\square$, the next lemma shows that we can still bound the correlation by bounding $\|ff^T\|_\square$.

Lemma 2.3. *Let X and Y be finite sets, let $u : X \to \mathbb{R}$, let $v : Y \to \mathbb{R}$ and let $f : X\times Y \to \mathbb{R}$. Let $g : X \times X \to \mathbb{R}$ be defined by $g(x, x') = \mathbb{E}_y f(x, y)f(x', y)$. Then*

$$|\mathbb{E}_{x,y} f(x, y)u(x)v(y)| \le \|g\|_\square^{1/2}\|u\|_{L_2}\|v\|_{L_2}.$$

Proof. We have

$$\begin{aligned}(\mathbb{E}_{x,y}f(x,y)u(x)v(y))^2 &= (\mathbb{E}_y v(y)\mathbb{E}_x f(x,y)u(x))^2\\ &\le (\mathbb{E}_y v(y)^2)\mathbb{E}_y\left(\mathbb{E}_x f(x,y)u(x)\right)^2\\ &= \|v\|_{L_2}^2\,\mathbb{E}_{x,x'}(\mathbb{E}_y f(x,y)f(x',y))u(x)u(x').\\ &= \|v\|_{L_2}^2\,\mathbb{E}_{x,x'}g(x,x')u(x)u(x').\end{aligned}$$

But by Lemma 2.2 this is at most $\|v\|_{L_2}^2\|g\|_\square\|u\|_{L_2}^2$, which proves the lemma. □

We also use the following lemma about how the box norm is affected by adding a constant function.

Lemma 2.4. *Let X and Y be finite sets and let $F : X\times Y \to \mathbb{R}$. Suppose that $\mathbb{E}_y F(x,y) = \delta$ for every x and $\mathbb{E}_x F(x,y) = \delta$ for every y. For each $x \in X$ and $y \in Y$ let $f(x,y) = F(x,y) - \delta$. Then $\|f\|_\square^4 = \|F\|_\square^4 - \delta^4$.*

Proof. We have $F(x,y) = f(x,y) + \delta$ for every x and y. If we make this substitution into the expression

$$\mathbb{E}_{x,x',y,y'}F(x,y)F(x,y')F(x',y)F(x',y'),$$

then we obtain 16 terms, of which two are

$$\mathbb{E}_{x,x',y,y'}f(x,y)f(x,y')f(x',y)f(x',y')$$

and δ^4. All remaining terms involve at least one variable that occurs exactly once. Since $\mathbb{E}_y f(x,y) = 0$ for every x and $\mathbb{E}_x f(x,y) = 0$ for every y, all such terms are zero. The result follows. □

Now we are ready for the proof.

Proof of Theorem 1.6 assuming Theorem 1.10. It suffices to prove the theorem in the case where g is the identity element. Let us pick a and b uniformly, and note that what we want to bound equals

$$|G|^4|E_{a,b}(\Gamma(a,b) - 1/|G|)u(a)v(b)|,$$

where $\Gamma(a,b)$ is the indicator function of $a\bullet b = 1$. Letting

$$f(a,b) := \Gamma(a,b) - 1/|G|,$$

and

$$g(x,x') := \mathbb{E}_y f(x,y)f(x',y),$$

Lemma 2.3 gives an upper bound of

$$|G|^4\|g\|_\square^{1/2}\|u\|_{L_2}\|v\|_{L_2} = |G|^2\|g\|_\square^{1/2}\|u\|_2\|v\|_2.$$

Now let us define

$$\Delta(x,x') := \mathbb{E}_y\Gamma(x,y)\Gamma(x',y).$$

Note that for each x,

$$\mathbb{E}_{x'}\Delta(x,x') = \mathbb{E}_{x',y}\Gamma(x,y)\Gamma(x',y) = \mathbb{E}_y\Gamma(x,y)\mathbb{E}_{x'}\Gamma(x',y) = 1/|G|^2.$$

By symmetry, $\mathbb{E}_x\Delta(x,x') = 1/|G|^2$ for every x' as well. Moreover, g differs from Δ by a constant: $g(x,x') = \Delta(x,x') - 1/|G|^2$ for every x, x' because

$$\begin{aligned} g(x,x') = \mathbb{E}_y f(x,y)f(x',y) &= \mathbb{E}_y(\Gamma(x,y) - 1/|G|)(\Gamma(x',y) - 1/|G|) \\ &= \mathbb{E}_y\Gamma(x,y)\Gamma(x',y) - 1/|G|^2 = \Delta(x,x') - 1/|G|^2. \end{aligned}$$

Hence we can apply Lemma 2.4 with F replaced by Δ, δ replaced by $1/|G|^2$, and f replaced by g to obtain

$$\|g\|_\square^{1/2} \leq (\|\Delta\|_\square^4 - 1/|G|^8)^{1/8}.$$

Thus it remains to show

$$\|\Delta\|_\square^4 \leq (1+\gamma)/|G|^8.$$

Note that

$$\|\Delta\|_\square^4 = \mathbb{E}_{x,x'}(\mathbb{E}_z\Delta(x,z)\Delta(x',z))^2 = \mathbb{E}_{x,x'}(\mathbb{E}_z\Delta(x,z)\Delta(z,x'))^2,$$

where the first equality follows by the definition of the box norm, and the second by the fact that Δ is symmetric.

Now we fix x and x' and consider $\mathbb{E}_z\Delta(x,z)\Delta(z,x') = \mathbb{E}_{z,y,y'}\Gamma(x,y)\Gamma(z,y)\Gamma(z,y')\Gamma(x',y')$. This is the probability, for a randomly chosen z, y, y' that

$$x_1y_1x_2y_2 = z_1y_1z_2y_2 = z_1y_1'z_2y_2' = x_1'y_1'x_2'y_2' = e,$$

which is $|G|^{-2}$ times the probability that $x_1y_1x_2 = z_1y_1z_2$ and $z_1y_1'z_2 = x_1'y_1'x_2'$.

These last two equations can be rewritten as

$$\begin{aligned} y_1^{-1}z_1^{-1}x_1y_1x_2 &= z_2 \\ y_1'^{-1}x_1'^{-1}z_1y_1'z_2 &= x_2'. \end{aligned}$$

By plugging the first equation in the second, and right-multiplying by x_2^{-1}, we obtain that our probability is $1/|G|$ times the probability that

$$y_1'^{-1}x_1'^{-1}z_1y_1'y_1^{-1}z_1^{-1}x_1y_1 = x_2'x_2^{-1}.$$

We rewrite this as

$$\mathbf{C}(x_1'^{-1}z_1)\mathbf{C}(z_1^{-1}x_1) = x_2'x_2^{-1}.$$

So we have shown that for every x and x':

$$\mathbb{E}_z\Delta(x,z)\Delta(x',z) = \frac{1}{|G|^3}\mathbb{P}_z[\mathbf{C}(x_1'^{-1}z_1)\mathbf{C}(z_1^{-1}x_1) = x_2'x_2^{-1}].$$

And consequently

$$\|\Delta\|_\square^4 = |G|^{-6}\mathbb{E}_{x,x'}(\mathbb{P}_z[\mathbf{C}(x_1'^{-1}z_1)\mathbf{C}(z_1^{-1}x_1) = x_2'x_2^{-1}])^2.$$

Thus it remains to show that the expectation in the right-hand side is at most $(1+\gamma)/|G|^2$.

By introducing variables $c = x_2' x_2^{-1}$, $b = z_1^{-1} x_1$, and $a = x_1'^{-1} x_1$ we can rewrite the expectation as

$$E_{a,c}(\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c])^2.$$

Multiplying by $|G|$, it remains to show that

$$\sum_c E_a(\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c])^2 \leq (1+\gamma)/|G|.$$

This follows from Theorem 1.10, which concludes the proof. □

3 Boosting pairwise independence

In this section we prove Theorem 1.7 from the introduction. Actually, we prove a slightly stronger statement, Corollary 3.4 below, which will be used later. 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_∞, and ℓ_1 norms are respectively $\|f\|_\infty = \max_x |f(x)|$, and $\|f\|_1 = \sum_x |f(x)|$. Next we define the measure of closeness to uniform that we will work with.

Definition 3.1. *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.

The main technical result shows how to go from pairwise independence to three-wise independence. We write $*$ for convolution, and note 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 3.2. *There is an integer $d \geq 2$ such that the following holds. Let $\mu_1, \ldots, \mu_d$ 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 3.3. *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 3.2.

Corollary 3.4. *There is an integer $d \ge 2$ 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 3.4 from Theorem 3.2. 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 the corollary by induction on m. For $m = 3$ this is Theorem 3.2. 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 $\mathbb{P}[\nu_1 * \cdots * \nu_d = x]$, which equals

$$\mathbb{P}[\nu_1^1 * \cdots * \nu_d^1 = x^1 | \nu_1^0 * \dots \nu_d^0 = x^0] \cdot \mathbb{P}[\nu_1^0 * \cdots * \nu_d^0 = x^0]. \tag{1}$$

By induction, each ν_i^0 is $(1/n, m-3)$-good, so Lemma 3.3 gives us that $\nu_1^0 * \cdots * \nu_d^0$ is $(1/n^2, m-3)$-good. Thus, the second term in expression (1) lies 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 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 3.2 the convolution of the α_i is $(1/n^2, 3)$-good.

Putting together these bounds for the two factors in the expression (1) we get that the product 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 1.7 stated in the introduction.

The rest of this section is devoted to the proof of Theorem 3.2. 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 return to the ℓ_∞ norm thanks to the following simple fact.

Fact 3.5. *For any distributions μ and ν it holds that $\|\mu * \nu\|_\infty \leq \|\mu\|_2 \|\nu\|_2$.*

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

We now state and prove the flattening lemma.

Lemma 3.6. *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 $\nu(ay_1,y_2,w)$. 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 $\sum_b \beta(b,y)\gamma(b,z)$.

Our 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 1.6 can be applied. To apply it, we proceed to bound the norms of u and v.

Note that by our hypotheses on μ and ν we have for each w that

$$\sum_x u(x,w) = \sum_a \delta(a,w) \sum_x \alpha(a,x) = \sum_a \delta(a,w) \sum_x \mu(x_1 a, x_2, x) \leq n^{-2} \sum_a \delta(a,w) \leq n^{-4},$$

with three similar inequalities for summing over the other coordinate and for v. Hence we also have that $\sum_{x,w} u(x,w) \leq n^{-3}$. We also have for each x, w that

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

with a similar argument giving the same bound for $\|v\|_\infty$. Combining these two facts we get that $\sum_{x,w} u(x,w)^2 \leq \|\mu\|_\infty / n^5$. And we have a similar bound for v.

We apply Theorem 1.6 to the probability distributions $u/\|u\|_1$ and $v/\|v\|_1$, and then we multiply by $\|u\|_1 \|v\|_1$ to obtain that

$$\sum_{xy=zw} u(x,w)v(y,z) = n^{-1}\|u\|_1\|v\|_1 + n \cdot \gamma \|u\|_2 \|v\|_2 \leq n^{-1} \cdot n^{-3} \cdot n^{-3} + \gamma \sqrt{\|\mu\|_\infty \|\nu\|_\infty} / n^4.$$

This implies that

$$\|\mu * \nu\|_2^2 \leq n^{-3} + \gamma \sqrt{\|\mu\|_\infty \|\nu\|_\infty},$$

which proves the result. □

Corollary 3.7. *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 3.6. Suppose further that $\|\mu_i\|_\infty \leq \alpha$ for every $i \in \{1,2,3,4\}$.*

Then

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

Proof. This follows on applying Lemma 3.6 to μ_1 and μ_2 and to μ_3 and μ_4 and then applying Fact 3.5 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 3.8. *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 3.2. If we divide each μ_i by $(1 + 1/\sqrt{n})$, then we obtain functions ν_i that satisfy the conditions of Lemma 3.6 and such that $\|\nu_i\|_\infty$ is at most $1/n^2$. Applying Corollary 3.7 a constant number of times, using Lemma 3.8 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

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

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, where note the inequality is Fact 3.5:

$$\begin{aligned}\|\alpha * \beta - U\|_\infty^2 = \|(\alpha - U) * (\beta - U)\|_\infty^2 \\ \le \|\alpha - U\|_2^2 \|\beta - U\|_2^2 = (\|\alpha\|_2^2 - 1/n^3)(\|\beta\|_2^2 - 1/n^3).\end{aligned}$$

In our case we have $\|\mu\|_2^2 \le \|\mu\|_\infty \le n^{-3}(1 + n^{-\Omega(1)})$. So we convolve one more time and apply the above fact to obtain a distribution μ' such that $\|\mu' - U\|_\infty \le n^{-\Omega(1)}/n^3$. Hence, μ' is $(n^{-\Omega(1)}, 3)$-good. Now if we convolve another constant number of times and apply Lemma 3.3 we obtain a distribution which is $(n^{-2}, 3)$-good, as desired. □

4 Communication complexity lower bound

In this section we prove Theorem 1.11. We shall focus first on the case where t is at least a large enough constant. The only case that is not covered by this is the case of $t = k = 2$. We then establish some simple equivalences that give that case, and also Theorem 1.5.

A key idea is to obtain this theorem 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). The next theorem summarizes this result.

Theorem 4.1. *There is a constant b such that the following holds. Let k and t be integers, $G = SL(2,q)$, $m = 2^k$, and $t \ge b^m$. Let $x_1^0, x_1^1, \dots, x_k^0, x_k^1$ 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.

To apply Corollary 3.4 we need to show that our distributions can be written as the product of many pairwise-independent distributions. This is done by the following lemma.

Lemma 4.2. *Let μ be the distribution over G^m in Theorem 4.1, and let also t be as in Theorem 4.1. 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 $x_i^{\epsilon_i} = (a_{i1}^{\epsilon_i}, \ldots, a_{ik}^{\epsilon_i})$. Then for each $\epsilon \in \{0,1\}^k$ we have

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

Now for $1 \leq j \leq t$ let s_j be the m-tuple $(a_{1j}^{\epsilon_1} \ldots a_{kj}^{\epsilon_k})_{\epsilon \in \{0,1\}^k}$. 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 $a_{1j}^{\epsilon_1} \ldots a_{kj}^{\epsilon_k}$, we have that

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

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

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

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 $\epsilon_i \neq \eta_i$. Conditioning on the values of $u_j^{\epsilon_j}$ and $u_j^{\eta_j}$ for every $j \neq i$, we find that we are looking at two products of the form $a u_i^{\epsilon_i} b$ and $c u_i^{\eta_i} d$. For this to equal (g,h), we need $u_i^{\epsilon_i} = a^{-1} g b^{-1}$ and $u_i^{\eta_i} = c^{-1} h d^{-1}$. 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 4.1 follows easily.

Proof of Theorem 4.1. Let m be 2^k. Let d be the constant in Corollary 3.4. Write the distribution μ as the product of t independent distributions μ_i, each of which is pairwise uniform, using Lemma 4.2. Group the μ_i in consecutive blocks of length d^m. The convolution in each block is $(1/n, m)$-good by Corollary 3.4. By repeated applications of Lemma 3.3 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. □

Finally, the proof that Theorem 4.1 implies Theorem 1.11 is a technically simple application of the "box norm," given next.

Proof that Theorem 4.1 implies Theorem 1.11. 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

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

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

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

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

To conclude it remains to notice that Theorem 4.1 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.

□

4.1 The remaining claims

We now establish some simple equivalences that give Theorem 1.11.(2), and also Theorem 1.5. Specifically we shall show that Theorem 1.11.(2) and the mixing bound for flat distributions, Theorem 1.5, are both equivalent to the following version of the mixing bound. We identify sets with their characteristic functions.

Theorem 4.3. *Let $G = SL(2, q)$. Let $A, B \subseteq G^t$ have densities α and β respectively. Let $g \in G$. We have*

$$|\mathbb{E}_{a \bullet b = g} A(a) B(b) - \alpha\beta| \leq |G|^{-\Omega(t)},$$

where the expectation is over a and b such that $a \bullet b = g$.

Claim 4.4. *Theorems 1.11, 1.5, and 4.3 are equivalent.*

Given this claim we obtain Theorem 1.11.(2) from Theorem 1.6, and we obtain Theorem 1.5 from Theorem 1.11.(1).

Proof of Claim 4.4. The equivalence between the two versions of the mixing bound, Theorems 1.5 and 4.3, follows by Bayes' identity

$$\mathbb{P}[a \bullet b = g | a \in A, b \in B] = \frac{\mathbb{P}[a \in A, b \in B | a \bullet b = g]}{|G|\alpha\beta}.$$

We now show that Theorem 4.3 implies the communication bound, Theorem 1.11. By an averaging argument we can assume that the protocol P in Theorem 1.11 is deterministic. Now write

$$P(a, b) = \sum_{i \leq C} R_i(a, b)$$

where $C = 2^c$, the R_i are disjoint rectangles in $(G^t)^2$, i.e., $R_i = S_i \times T_i$ for some $S_i, T_i \subseteq G^t$, cf. [KN97], and we also write R_i for the characteristic function with output in $\{0,1\}$. For any g and h in G we then have, using the triangle inequality:

$$|p_g - p_h| = \left| \sum_{i \leq C} \left(\mathbb{E}_{a \bullet b = g} R_i(a,b) - |R_i|/|G|^{2t} + |R_i|/|G|^{2t} - \mathbb{E}_{a \bullet b = h} R_i(a,b) \right) \right| \leq 2^C |G|^{-\Omega(t)}.$$

To see the reverse direction, that Theorem 1.11 implies Theorem 4.3, suppose that we are given sets A and B. Consider the constant-communication protocol $P(a,b) = A(a)B(b)$, and note that $p_g = \mathbb{E}_{a \bullet b = g} A(b)B(b)$ and that $\mathbb{E}_h p_h = \alpha\beta$. So we have

$$|\mathbb{E}_{a \bullet b = g} A(a)B(b) - \alpha\beta| = |p_g - \mathbb{E}_h p_h| \leq \mathbb{E}_h |p_g - p_h| = O(|G|^{-\Omega(t)}).$$

□

5 SL(2,q): Proof of Theorem 1.10

In this section we prove Theorem 1.10. We start with a lemma from the literature.

Lemma 5.1 (Structure of $\mathrm{SL}(2,q)$.)**.** *The group $SL(2,q)$ has the following properties.*

1. It has size $q^3 - q$.

2. It has $q + O(1)$ conjugacy classes.

3. All but $O(1)$ conjugacy classes have size either $q(q+1)$ or $q(q-1)$.

4. Every conjugacy class has size $\Omega(q^2)$, except for the trivial classes $\{1_G\}$ and $\{-1_G\}$ which have size 1.

Property 1 is easy to verify. For more precise versions of Properties 2 and 3 see e.g. theorems 38.1 and 38.2 in [Dor71]. Next we state another lemma and then prove Theorem 1.10 from the lemmas.

Lemma 5.2. *Let $G = SL(2,q)$. Let $D = (D_1, D_2)$ be a distribution over G^2 such that D_1 is uniform and D_2 is uniform. With probability $1 - O(1/q)$ over a pair (g,h) sampled from D the following holds.*

(1) For any conjugacy class S of G, the probability that $g\mathbf{C}(h) \in S$ is $O(1/q)$.

(2) The distribution of $\mathbf{C}(g)\mathbf{C}(h)$ is $q^{-\Omega(1)}$-close to uniform over G.

Proof of Theorem 1.10 assuming Lemma 5.2. First, note that the bound claimed in the theorem is equivalent to

$$\sum_c (\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c] - 1/|G|)^2 \leq \gamma^{\Omega(1)}/|G|.$$

We proceed by case analysis. The $c = 1$ summand is, by Cauchy-Schwarz,

$$(\mathbb{P}_b[ab^{-1}\mathbf{C}(b) = 1] - 1/|G|)^2 \leq \mathbb{E}_b(\mathbb{P}[ab^{-1}\mathbf{C}(b) = 1] - 1/|G|)^2.$$

If $b \notin \{1,-1\}$ then the conjugacy class of b has size $\Omega(q^2)$, by Lemma 5.1. Hence, $\mathbb{P}[ab^{-1}\mathbf{C}(b) = 1] = O(1/q^2)$ and so $(\mathbb{P}[ab^{-1}\mathbf{C}(b) = 1] - 1/|G|)^2 \leq O(1/q^4)$. If instead $b \in \{1,-1\}$ then the conjugacy class of b is $\{b\}$, so the probability is 1 if $a = 1$, which happens with probability $1/|G|$, and 0 otherwise. Thus, the $c = 1$ summand is at most $O(1/q^4) + O(1/|G|^2) \leq \gamma^{\Omega(1)}/|G|$.

A similar argument gives the same bound for the $c = -1$ summand.

It remains to show that

$$\sum_{c \in G \setminus \{-1,1\}} (\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c] - 1/|G|)^2 \leq \gamma^{\Omega(1)}/|G|.$$

We bound above the left-hand side of this by

$$\left(\max_{c \in G \setminus \{-1,1\}} (\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c] - 1/|G|)\right)\left(\sum_{c \in G \setminus \{-1,1\}} (\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c] - 1/|G|)\right).$$

First we show that the maximum is $O(1/|G|)$. Except with probability $O(1/q)$ over the choice of b, Lemma 5.2 guarantees that $\mathbf{C}(ab^{-1})\mathbf{C}(b)$ is in the conjugacy class of c with probability $O(1/q)$. Because by Lemma 5.1 this class has size $\Omega(q^2)$, the probability that $\mathbf{C}(ab^{-1})\mathbf{C}(b) = \mathbf{C}(\mathbf{C}(ab^{-1})\mathbf{C}(b))$ equals c is $O(1/q)O(1/q^2) = O(1/|G|)$. In the event that we cannot apply the lemma, the probability is still at most $O(1/q^2)$. Hence overall the maximum is at most $O(1/|G|)$.

Thus, it remains to show that

$$\sum_{c \in G \setminus \{-1,1\}} |\mathbb{P}_b[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c] - 1/|G|| \leq \gamma^{\Omega(1)}.$$

The left-hand side of this is at most

$$\mathbb{E}_b \sum_c |\mathbb{P}[\mathbf{C}(ab^{-1})\mathbf{C}(b) = c] - 1/|G||.$$

Except with probability $O(1/q)$ over b, we have by Lemma 5.2 that the distribution $\mathbf{C}(ab^{-1})\mathbf{C}(b)$ is $q^{-\Omega(1)}$-close to uniform over G, in which case the sum is at most $q^{-\Omega(1)}$. Also, for every b the sum is at most 2. So overall we obtain an upper bound of $q^{-\Omega(1)} + O(1/q)$, as desired. □

To complete the proof of Theorem 1.10 we must prove Lemma 5.2. A key observation, which is also central to many other papers concerning conjugacy classes in SL(2,q), is that there is an approximate one-to-one correspondence between conjugacy classes and the traces of the matrices in the conjugacy class. In one direction this is trivial, since the trace is a conjugacy invariant. The other direction can be expressed in a form suitable for our purposes as the following claim.

Definition 5.3. *For a group element g we denote by* $\mathrm{Tr}(g)$ *the* trace *of the matrix corresponding to g, and by* $\mathrm{Class}(g)$ *the conjugacy class of g.*

Claim 5.4. *Let $G = SL(2,q)$, let D be a distribution over G and let U be the uniform distribution over G. Suppose that $\mathrm{Tr}(D)$ is ϵ-close to uniform over $\mathbb{F}_q$ in statistical distance. Then $\mathrm{Class}(D)$ and $\mathrm{Class}(U)$ are ϵ'-close, where $\epsilon' = O(\epsilon) + O(1/q)$.*

Proof. Let H be the set of conjugacy classes of G which have size in $\{q(q+1), q(q-1)\}$ and whose trace is not equal to that of any other conjugacy class. We have that $|H| \geq q - O(1)$, because all q field elements can arise as the trace of a conjugacy class, the number of conjugacy classes is $q + O(1)$ by Lemma 5.1, and all but $O(1)$ conjugacy classes have size in $\{q(q+1), q(q-1)\}$ again by Lemma 5.1.

Next we claim that the distribution of $\mathrm{Class}(U)$ is $O(1/q)$-close to the uniform distribution V over H. Indeed, $\mathbb{P}[\mathrm{Class}(U) = c] = (q^2 + e_c)/|G|$ where $|e_c| = q$ for any $c \in H$. And for any $c \notin H$ we have $\mathbb{P}[\mathrm{Class}(U) = c] = O(q^2/|G|) = O(1/q)$. Hence the statistical distance between $\mathrm{Class}(U)$ and V is at most

$$\sum_{c \notin H} O(1/q) + \sum_{c \in H} \left| \frac{q^2 + e_c}{|G|} - \frac{1}{|H|} \right| \leq O(1/q) + \sum_{c \in H} \left| \frac{q^2}{|G|} - \frac{1}{|H|} \right| \leq O(1/q).$$

Finally, we claim that $\mathrm{Class}(D)$ is ϵ'-close to V. The probability that $\mathrm{Class}(D) = c$ for a c in H is equal to the probability that $\mathrm{Tr}(D) = c$. Let B be the set of $q - |H| = O(1)$ values for the trace map that do not arise from classes in H. The statistical distance between $\mathrm{Class}(U)$ and V is at most

$$\mathbb{P}[\mathrm{Tr}(D) \in B] + \sum_{c \in H} |\mathbb{P}[\mathrm{Tr}(D) = c] - 1/|H||$$
$$\leq \epsilon + \sum_{c \in H} |\mathbb{P}[\mathrm{Tr}(D) = c] - 1/q| + \sum_{c \in H} |1/q - 1/|H|| \leq O(\epsilon) + O(1/q).$$

The result follows by summing the distance between $\mathrm{Class}(U)$ and V and between $\mathrm{Class}(D)$ and V. □

This correspondence allows us to derive Lemma 5.2 from the following lemma about the trace map.

Lemma 5.5. *Let $G = SL(2,q)$. Let v and w be two elements of $\mathbb{F}_q$. Suppose that either (i) q is even, or (ii) q is odd and $(v^2, w^2) \neq (-4, -4)$ and $(v, w) \neq (0, 0)$. Let D be the distribution of $Tr\left(\begin{pmatrix} 0 & 1 \\ 1 & w \end{pmatrix} \mathbf{C} \begin{pmatrix} v & 1 \\ 1 & 0 \end{pmatrix}\right)$. Then*

(1) D takes any value $x \in \mathbb{F}_q$ with probability $O(1/q)$, and
(2) D is $1/q^{\Omega(1)}$ close to uniform in statistical distance.

Proof of Lemma 5.2. Note that for every h and g the distribution of the trace of $hugu^{-1}$ for uniform u is the same as the distribution of the trace of $h'ug'u^{-1}$ for uniform u, for any h' that is conjugate to h and for any g' that is conjugate to g. This is true because if $g = xg'x^{-1}$ and $h = yh'y^{-1}$ then by the cyclic-shift property of the trace function we have

$$\mathrm{Tr}(yh'y^{-1}uxg'x^{-1}u^{-1}) = \mathrm{Tr}(h'y^{-1}uxg'x^{-1}u^{-1}y),$$

and the latter has the same distribution of the trace of $h'ug'u^{-1}$ for uniform u. Because of this fact, Lemma 5.5 applies to $\mathrm{Tr}(g\mathbf{C}(h))$ for any g except those in $O(1)$ conjugacy classes and similarly for any h except those in $O(1)$ conjugacy classes. Those conjugacy classes make up at most an $O(1/q)$ fraction of the group. Hence, with probability $1 - O(1/q)$ over (g, h) sampled from D, we can apply Lemma 5.5.

Property (1) in Lemma 5.5 immediately gives Property (1) in Lemma 5.2.

To verify Property (2), note that the distribution of $\mathbf{C}(g)\mathbf{C}(h)$ is the same as that of $\mathbf{C}(\mathbf{C}(g)\mathbf{C}(h))$. By Item (2) in Lemma 5.5, $\mathrm{Tr}(\mathbf{C}(g)\mathbf{C}(h))$ is $q^{-\Omega(1)}$-close to uniform. By Claim 5.4, $\mathrm{Class}(\mathbf{C}(g)\mathbf{C}(h))$ is $q^{-\Omega(1)}$-close to the distribution of the conjugacy class of a uniform element from G. Hence $\mathbf{C}(\mathbf{C}(g)\mathbf{C}(h))$ is $q^{-\Omega(1)}$-close to uniform. □

It remains to prove Lemma 5.5. This proof is somewhat technical and appears in the next subsection.

5.1 Proof of Lemma 5.5

In this subsection, if we use a letter such as a to refer to an element of G, we shall refer to its entries as $a_1, \ldots, a_4$. That is, we shall take a to be the matrix $\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$.

We begin by working out an expression for the trace that concerns us.

Claim 5.6. *Let a, u and g be 2×2 matrices in $SL(2, q)$. Then*

$$\begin{aligned}\mathrm{Tr}(augu^{-1}) = (a_1u_1 + a_2u_3)(g_1u_4 - g_2u_3) + (a_1u_2 + a_2u_4)(g_3u_4 - g_4u_3) \\ + (a_3u_1 + a_4u_3)(-g_1u_2 + g_2u_1) + (a_3u_2 + a_4u_4)(-g_3u_2 + g_4u_1).\end{aligned}$$

Proof. Note that $\begin{pmatrix} u_1 & u_2 \\ u_3 & u_4 \end{pmatrix}^{-1} = \begin{pmatrix} u_4 & -u_2 \\ -u_3 & u_1 \end{pmatrix}$. Now

$$au = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} u_1 & u_2 \\ u_3 & u_4 \end{pmatrix} = \begin{pmatrix} a_1u_1 + a_2u_3 & a_1u_2 + a_2u_4 \\ a_3u_1 + a_4u_3 & a_3u_2 + a_4u_4 \end{pmatrix}$$

and

$$gu^{-1} = \begin{pmatrix} g_1 & g_2 \\ g_3 & g_4 \end{pmatrix} \begin{pmatrix} u_4 & -u_2 \\ -u_3 & u_1 \end{pmatrix} = \begin{pmatrix} g_1u_4 - g_2u_3 & -g_1u_2 + g_2u_1 \\ g_3u_4 - g_4u_3 & -g_3u_2 + g_4u_1 \end{pmatrix}.$$

The result follows. □

Our proof of Lemma 5.5 uses the following well-known theorem from arithmetic geometry, due to Lang and Weil [LW54]. It can also be found as Theorem 5A, page 210, of [Sch04].

Theorem 5.7. *For every positive integer d there is a constant c_d such that the following holds: if $f(x_1, \ldots, x_n)$ is any absolutely irreducible polynomial over F_q of total degree d, with N zeros in F_q^n, then*

$$|N - q^{n-1}| \leq c_d q^{n-3/2}.$$

After these preliminaries we now present the proof of Lemma 5.5. First we remark that the calculation below for the trace in the case $v = w = 0$ shows that the condition $(v, w) \neq (0, 0)$ is necessary over odd characteristic.

From Claim 5.6 we obtain the following expression for the trace.

$$\begin{aligned} f'' := u_3(vu_4 - u_3) + u_4u_4 + (u_1 + wu_3)(-vu_2 + u_1) + (u_2 + wu_4)(-u_2) \\ = vu_3u_4 - u_3^2 + u_4^2 - vu_1u_2 + u_1^2 - vwu_2u_3 + wu_1u_3 - u_2^2 - wu_2u_4. \end{aligned}$$

We shall show that for all but $O(1)$ choices for s, the number of solutions to the system $f'' = -s$ and $u_1u_4 - u_2u_3 = 1$ has distance e_s from q^2 where $|e_s| \leq q^{2-\Omega(1)}$. And for the other $O(1)$ choices of s the number of solutions is $O(q^2)$. This will show Property (2), i.e., that the trace has statistical distance $1/q^{\Omega(1)}$ from uniform. Indeed, using that $|G| = q^3 - q$, the contribution to this distance from each of the $q - O(1)$ good values of s is $|(q^2 + e_s)/(q^3 - q) - 1/q| = |(1 + e_s)/(q^3 - q)| \leq 1/q^{1+\Omega(1)}$ because $|e_s| \leq q^{2-\Omega(1)}$. These add up to a contribution of $1/q^{\Omega(1)}$, while for each of the others the contribution is at most $O(1/q)$. Property (1) will then follow.

First, we consider the case when q is even and $v = w = 0$. In this case the trace becomes

$$(u_1 - u_2 - u_3 + u_4)^2.$$

Now note that the map $\begin{pmatrix} u_1 & u_2 \\ u_3 & u_4 \end{pmatrix} \to \begin{pmatrix} u_1 & u_2 \\ u_3 + u_1 & u_4 + u_2 \end{pmatrix}$ is a permutation on G. If we apply it, the expression of the trace simplifies to $(-u_3 + u_4)^2$ which is close to uniform, because squaring in characteristic 2 is a permutation, and $u_4 - u_3$ is approximately uniform.

As a next step we count the solutions with $u_1 = 0$. In this case the trace plus s is

$$vu_3u_4 - u_3^2 + u_4^2 - vwu_2u_3 - u_2^2 - wu_2u_4 + s.$$

The equation $u_1u_4 - u_2u_3 = 1$ gives us that $u_3 = -1/u_2$. For any choice of u_2, the above becomes a univariate polynomial in u_4 which is non-zero because of the u_4^2 term. Hence the total number of solutions with $u_1 = 0$ is $O(q)$. This amount does not affect the result, so from now on we count the solutions with $u_1 \neq 0$.

We can now eliminate $u_4 = (1 + u_2u_3)/u_1$ in f'. Renaming u_1, u_2, and u_3 as x, y, z, respectively, we get the expression

$$f' := vz(1 + yz)/x - z^2 + (1 + yz)^2/x^2 - vxy + x^2 - vwyz + wxz - y^2 - wy(1 + yz)/x.$$

First we note an upper bound of $O(q^2)$ on the number of solutions to $f' = s$, for any s. Indeed, after we pick x and y we are left with a quadratic polynomial in z which is not zero because of the z^2 term. Hence, this polynomial has at most two solutions.

Next we show the stronger bound for all but $O(1)$ values of s. Letting $f(x, y, z) := x^2(f' + s)$ and expanding and rearranging, we get the expression

$$\begin{aligned} f := x^4 - x^2y^2 - x^2z^2 + y^2z^2 + 2yz + 1 \\ + v(-x^3y + xz + xyz^2) + w(-xy - xy^2z + x^3z) - vwx^2yz + sx^2. \end{aligned}$$

We shall show that if f is not absolutely irreducible, then s takes one of $O(1)$ values. So if s is not one of those values, then we can apply Theorem 5.7. This will give the desired bound of $q^2 + e_s$ on the number of roots with $x, y, z \in F$. We actually just wanted to count the roots with $x \neq 0$. However, if $x = 0$ then f simplifies to $(1 + yz)^2$ which has $q - 1$ roots. So the bound is correct even if we insist that $x \neq 0$.

The function f is a polynomial of degree 4 in three variables. Suppose that it can be factorized as $f = PQ$. Note first that both P and Q must have a constant term because f has it. Also, neither P nor Q can have a power of y as a term, because f does not have it (but such a term would arise in the product between the highest-power such term in P and in Q, one of which could be the constant term). Similarly, neither can have a power of z as a term.

If $f = PQ$, then the sum of the degrees of P and Q is at most 4. If P has degree 3 then Q has degree 1. By the above, Q would be of the form $ax + b$. However in this case there would be no way to produce the term y^2z^2.

So both P and Q have degree at most 2, and we can write

$$\begin{aligned} P &= axy + byz + cxz + dx^2 + ex + f, \\ Q &= a'xy + b'yz + c'xz + d'x^2 + e'x + f'. \end{aligned}$$

Equating coefficients gives the systems of equations

$$\begin{aligned} xy^2z &\to ab' + a'b = -w \\ x^2yz &\to ac' + a'c + bd' + b'd = -vw \\ x^3y &\to ad' + a'd = -v \\ x^2y &\to ae' + a'e = 0 \\ xy &\to af' + a'f = -w \\ xyz^2 &\to bc' + b'c = v \\ xyz &\to be' + b'e = 0 \\ yz &\to bf' + b'f = 2 \\ x^3z &\to cd' + c'd = w \\ x^2z &\to ce' + c'e = 0 \\ xz &\to cf' + c'f = v \\ x^3 &\to de' + d'e = 0 \\ x^2 &\to df' + f'd + ee' = s \\ x &\to ef' + e'f = 0 \end{aligned}$$

and

$$\begin{aligned} x^2y^2 &\to aa' = -1 \\ y^2z^2 &\to bb' = 1 \\ x^2z^2 &\to cc' = -1 \\ x^4 &\to dd' = 1 \\ 1 &\to ff' = 1. \end{aligned}$$

Multiplying by bf the yz equation and using that $bb' = ff' = 1$, we find that

$$b^2ff' + bb'f^2 = b^2 + f^2 = 2bf.$$

Therefore, $(b-f)^2 = 0$ and so $b = f$. Since $bb' = ff' = 1$, we also get that $b' = f'$.

Now we claim that $e' = 0$. Assume for a contradiction that $e' \neq 0$. Multiplying by appropriate variables, the equations with right-hand side equal to zero become:

$$\begin{aligned} x^2y &\to a^2e' - e = 0 \\ xyz &\to b^2e' + e = 0 \\ x^2z &\to c^2e' - e = 0 \\ x^3 &\to d^2e' + e = 0. \end{aligned}$$

Summing the first two gives us that $(a^2 + b^2)e' = 0$, which implies that $a^2 + b^2 = 0$ because $e' \neq 0$. Repeating this argument we obtain that

$$a^2 + b^2 = a^2 + d^2 = b^2 + c^2 = c^2 + d^2 = 0.$$

Now multiplying the xy^2z equation by ab we get that $a^2 - b^2 = 2a^2 = -wab$. Dividing by $ab \neq 0$ we obtain that $2a/b = -w$. Because $a^2/b^2 = -1$, squaring we obtain that $w^2 = -4$. Similarly, multiplying the x^3y equation by ad we get that $a^2 - d^2 = 2a^2 = vad$ and we get that $v^2 = -4$ as well. For odd q, this contradicts our assumption that $(v^2, w^2) \neq (-4, -4)$. For even q we have $4 = 0$ and so $v = w = 0$ which we were also excluding. Therefore $e' = 0$. (From the equation for xyz we get that $e = 0$ as well, but we will not use this.)

We can now simplify some of the equations as follows:

$$\begin{aligned} x^2yz &\to ac' + a'c + s = -vw \\ x^2 &\to db' + d'b = s. \end{aligned}$$

Next, we handle the case of even q where exactly one of v or w is 0. If $w = 0$, then multiplying the xy^2z equation by ab we find that $a^2 - b^2 = 0$. So $a = b$ and the x^3y equation has the same left-hand side as the x^2 equation, which implies that $s = v$. Similarly, if $v = 0$, then the x^3y equation gives us that $a = d$. Now the xy^2z and the x^2 equation have the same left-hand side, giving us that $s = w$.

Now we continue the analysis for any q. Multiplying equations by appropriate quantities we get:

$$\begin{aligned} xy^2z &\to a^2 - b^2 = -wab \\ x^3y &\to a^2 - d^2 = -vad \\ xyz^2 &\to -b^2 + c^2 = vbc \\ x^3z &\to c^2 - d^2 = wcd. \end{aligned}$$

The first minus the second gives $-b^2 + d^2 = a(vd - wb)$; the third minus the fourth gives $-b^2 + d^2 = c(vb - wd)$. And so

$$a(vd - wb) = c(vb - wd).$$

Now assume that $vd - wb \neq 0$. Then by dividing by it and by $c \neq 0$ we get

$$\frac{a}{c} = \frac{vb - wd}{vd - wb}.$$

So we have that

$$\begin{aligned} \frac{a}{c} + \frac{c}{a} = \frac{(vb - wd)^2 + (vd - wb)^2}{(vd - wb)(vb - wd)} &= \frac{(b^2 + d^2)(v^2 + w^2) - 4vwbd}{-vw(b^2 + d^2) + (w^2 + v^2)bd} \\ &= \frac{(b^2 + d^2)(v^2 + w^2 - 4vw/s)}{(b^2 + d^2)(-vw + (w^2 + v^2)/s)} = \frac{s(v^2 + w^2) - 4vw}{-svw + w^2 + v^2}. \end{aligned}$$

Here we used the x^2 equation multiplied by bd, which is $bds = b^2 + d^2$, and then divided by s. So we are assuming that $s \neq 0$.

Now if we plug this expression into the x^2yz equation, which, using the fact that $aa' = cc' = -1$, can be transformed into the equation $-a/c - c/a + s = -vw$, we obtain that

$$\frac{s(v^2 + w^2) - 4vw}{-svw + w^2 + v^2} + s = -vw.$$

This expression can be satisfied by only a constant number of s. Indeed, taking the right-hand side to the left and multiplying by the denominator we obtain the equation

$$2s(v^2 + w^2) - 4vw - s^2vw - sv^2w^2 + vw(w^2 + v^2) = 0.$$

Now, if q is odd and if exactly one of v and w is 0 then all the terms vanish except the first one, yielding that $s = 0$. Together with our assumptions and previous analysis, we can assume at this point that $vw \neq 0$. In this case we obtain a quadratic polynomial in s which is not zero because of the $-s^2vw$ term. This polynomial has at most two roots.

The case we left out is when $vd - wb = 0$. In that case $d = bw/v$. From the x^2 equation and the fact that $bb' = dd' = 1$ we get that

$$v/w + w/v = s.$$

Altogether, we have shown that if the polynomial is not irreducible then s takes one of at most six possible values. These values are $0, v, w, v/w + w/v$, and the at most two roots of the quadratic polynomial above. Although it does not affect the result, we recall that these values of s correspond to values of $-s$ for the traces.

6 Miscellaneous results

In this section we first give an alternative proof of Theorem 1.3 for SL(2,q). Then we prove Theorem 1.8.

An immediate consequence of Theorem 4.3 with $t = 2$ is that the group $\mathrm{SL}(2, q)$ has the property that the product of any four dense sets is almost uniformly distributed. More precisely, we have the following result.

Theorem 6.1. *Let G be the group $SL(2, q)$, and let $A, B, C, D \subseteq G$ be subsets of density α, β, γ and δ, respectively. Then for every $g \in G$,*

$$|\mathbb{E}_{abcd=g} A(a)B(b)C(c)D(d) - \alpha\beta\gamma\delta| = O(|G|^{-c})$$

and

$$|\mathbb{P}[abcd = g | a \in A, b \in B, c \in C, d \in D] - |G|^{-1}| = (\alpha\beta\gamma\delta)^{-1} O(|G|^{-(1+c)}).$$

It turns out that from this result for four sets follows the same result for three sets.

Corollary 6.2. *Let G be the group $SL(2, q)$, and let $A, B, C \subset G$ be subsets of density α, β and γ, respectively. Then for every $g \in G$,*

$$|\mathbb{E}_{abc=g} A(a)B(b)C(c) - \alpha\beta\gamma| = O(|G|^{-c})$$

and

$$|\mathbb{P}[abc = g | a \in A, b \in B, c \in C] - |G|^{-1}| = (\alpha\beta\gamma)^{-1} O(|G|^{-(1+c)}.$$

Proof. For each a, let $f(a) = A(a) - \alpha$. Then

$$\begin{aligned}\mathbb{E}_{abc=g} A(a)B(b)C(c) &= \alpha \mathbb{E}_{abc=g} B(b)C(c) + \mathbb{E}_{abc=g} f(a)B(b)C(c) \\ &= \alpha\beta\gamma + \mathbb{E}_{abc=g} f(a)B(b)C(c).\end{aligned}$$

But

$$\begin{aligned}(\mathbb{E}_{abc=g} f(a)B(b)C(c))^2 &\leq (\mathbb{E}_c C(c)^2)(\mathbb{E}_c(\mathbb{E}_{ab=gc^{-1}} f(a)B(b))^2) \\ &= \gamma \mathbb{E}_c \mathbb{E}_{ab=a'b'=gc^{-1}} f(a)B(b)f(a')B(b') \\ &= \gamma \mathbb{E}_{abb'^{-1}a^{-1}=e}(A(a) - \alpha)B(b)B(b')(A(a') - \alpha).\end{aligned}$$

There are four terms that make up the expectation. Each term that involves at least one α is equal to $\pm\alpha^2\beta^2$, with two minus signs and one plus sign. The remaining term is $\alpha^2\beta^2 + O(|G|^{-c})$, by Theorem 6.1. The first statement follows. Once again, the second statement is equivalent to it by a simple application of Bayes's theorem, together with the observation that $\mathbb{E}_{abc=g} A(a)B(b)C(c) = \mathbb{P}[a \in A, b \in B, c \in C \mid abc = g]$. □

Proof of Theorem 1.8. It suffices to prove the theorem in the case where g is the identity e. Let a', b', and c' be selected independently and uniformly from S, G, and G, respectively.

By Bayes' rule we can write $\mathbb{P}[abc = e] = \mathbb{P}[a'b'c' = e | a' \in A, b' \in B, c' \in C] = \mathbb{P}[a' \in A, b' \in B, c' \in C | a'b'c' = e] \cdot \frac{1}{\alpha\beta\gamma|G|}$.

Thus our goal is to show

$$|\mathbb{P}[a' \in A, b' \in B, c' \in C | a'b'c' = e] - \alpha\beta\gamma| \leq |G|^{-\Omega(1)}.$$

Rewrite the difference as

$$|\mathbb{E}_{a \in S, b \in G} A(a)B(b)D(a,b)|$$

where $D(a,b) = C^{-1}(ab) - \gamma$. Thinking of D as a function defined on $S \times G$, we can use Lemma 2.2 to bound the fourth power of this expression by

$$\|A\|_{L_2}^4 \|B\|_{L_2}^4 \|D\|_{\square}^4 \leq \|D\|_{\square}^4 = \mathbb{E}_{b,b'} \mathbb{E}_{a,a' \in S} D(a,b)D(a,b')D(a',b)D(a',b').$$

If we change variables by premultiplying b by a^{-1} and b' by a'^{-1}, then we can rewrite the right-hand side as

$$\mathbb{E}_{b,b'} \mathbb{E}_{a,a' \in S} D(1,b)D(a, a'^{-1}b')D(a', a^{-1}b)D(1,b').$$

We claim that the quadruple $(b, b', aa'^{-1}b', a'a^{-1}b)$ is ϵ-close in statistical distance to $(v, w, x, wx^{-1}v)$, for $\epsilon \leq 1/|G|^{\Omega(1)}$, where v, w, and x are uniform in G. It suffices to show that the first three coordinates are jointly ϵ-close to uniform. But the distance is at most that of aa'^{-1} from uniform. It therefore follows by Lemmas 5.1 and 5.2.(2) that except for $O(1)$ conjugacy classes S, the first three coordinates are indeed ϵ-close to uniform.

Hence the value of the expression is at most ϵ plus

$$\mathbb{E}_{v,w,x}(C^{-1}(v) - \gamma)(C^{-1}(w) - \gamma)(C^{-1}(x) - \gamma)(C^{-1}(wx^{-1}v) - \gamma).$$

If we expand the product, in any term with at most three copies of C^{-1} we can replace those copies by γ. Hence, by direct calculation or say the binomial theorem, we can rewrite it as

$$\mathbb{E}_{v,w,x} C^{-1}(v)C^{-1}(w)C^{-1}(x)C^{-1}(wx^{-1}v) - \gamma^4.$$

By suitably redefining the copies of C we can put this in the form of Theorem 6.1 and thereby obtain a bound of $|G|^{-\Omega(1)}$. □

Acknowledgments. We thank the anonymous referees for their useful feedback. We also thank Laci Pyber for pointing out Theorem 2.5 in [Sha08], which we used in [GV15]. Emanuele Viola is very grateful to Eric Miles for extensive discussions during the early stages of this project. He also thanks Laci Babai for an email exchange, and Swastik Kopparty for pointing out the book [Sch04].

References

[ACFN12] Anil Ada, Arkadev Chattopadhyay, Omar Fawzi, and Phuong Nguyen. The NOF multiparty communication complexity of composed functions. In *Coll. on Automata, Languages and Programming (ICALP)*, pages 13–24, 2012.

[AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC^0. *SIAM J. on Computing*, 36(4):845–888, 2006.

[AL00] Andris Ambainis and Satyanarayana V. Lokam. Imroved upper bounds on the simultaneous messages complexity of the generalized addressing function. In *Latin American Symposium on Theoretical Informatics (LATIN)*, pages 207–216, 2000.

[Amb96] Andris Ambainis. Upper bounds on multiparty communication complexity of shifts. In *Symp. on Theoretical Aspects of Computer Science (STACS)*, pages 631–642, 1996.

[Bab87] László Babai. Random oracles separate PSPACE from the polynomial-time hierarchy. *Information Processing Letters*, 26(1):51–53, 1987.

[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.

[BC92] Michael Ben-Or and Richard Cleve. Computing algebraic formulas using a constant number of registers. *SIAM J. on Computing*, 21(1):54–58, 1992.

[BGKL03] László Babai, Anna Gál, Peter G. Kimmel, and Satyanarayana V. Lokam. Communication complexity of simultaneous messages. *SIAM J. on Computing*, 33(1):137–166, 2003.

[BNP08] László Babai, Nikolay Nikolov, and László Pyber. Product growth and mixing in finite groups. In *ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 248–257, 2008.

[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.

[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.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. on Computing*, 17(2):230–261, 1988.

[CM87] Stephen A. Cook and Pierre McKenzie. Problems complete for deterministic logarithmic space. *J. Algorithms*, 8(3):385–394, 1987.

[CS14] Arkadev Chattopadhyay and Michael E. Saks. The power of super-logarithmic number of players. In *Workshop on Randomization and Computation (RANDOM)*, pages 596–603, 2014.

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

[Dor71] Larry Dornhoff. *Group representation theory. Part A: Ordinary representation theory.* Marcel Dekker, Inc., New York, 1971. Pure and Applied Mathematics, 7.

[FKN94] Uriel Feige, Joe Kilian, and Moni Naor. A minimal model for secure computation (extended abstract). In *ACM Symp. on the Theory of Computing (STOC)*, pages 554–563, 1994.

[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.

[GGH+08] Shafi Goldwasser, Dan Gutfreund, Alexander Healy, Tali Kaufman, and Guy Rothblum. A (de)constructive approach to program checking. In *40th ACM Symposium on Theory of Computing (STOC)*, pages 143–152, 2008.

[Gow08] W. T. Gowers. Quasirandom groups. *Combinatorics, Probability & Computing*, 17(3):363–387, 2008.

[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.

[Gro94] Vince Grolmusz. The BNS lower bound for multi-party protocols in nearly optimal. *Inf. Comput.*, 112(1):51–54, 1994.

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

[GV16] W. T. Gowers and Emanuele Viola. The multiparty communication complexity of interleaved group products. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2016.

[IL95] Neil Immerman and Susan Landau. The complexity of iterated multiplication. *Inf. Comput.*, 116(1):103–116, 1995.

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

[Kil88] Joe Kilian. Founding cryptography on oblivious transfer. In *ACM Symp. on the Theory of Computing (STOC)*, pages 20–31, 1988.

[KMR66] Kenneth Krohn, W. D. Maurer, and John Rhodes. Realizing complex Boolean functions with simple groups. *Information and Control*, 9:190–195, 1966.

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

[LW54] Serge Lang and André Weil. Number of points of varieties in finite fields. *American Journal of Mathematics*, 76:819–827, 1954.

[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.

[PRS97] Pavel Pudlák, Vojtěch Rödl, and Jiří Sgall. Boolean circuits, tensor ranks, and communication complexity. *SIAM J. on Computing*, 26(3):605–633, 1997.

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

[Sch04] Wolfgang Schmidt. *Equations Over Finite Fields: An Elementary Approach*. Kendrick Press, 2004.

[Sha08] Aner Shalev. Mixing and generation in simple groups. *J. Algebra*, 319(7):3075–3086, 2008.

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

[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.