Electronic Colloquium on Computational Complexity, Report No. 44 (2015)

The communication complexity of interleaved group products

W. T. Gowers* Emanuele Viola†

April 2, 2015

Abstract

Alice receives a tuple $(a_1, \ldots, a_t)$ of t elements from the group $G = \mathrm{SL}(2, q)$. Bob similarly receives a tuple $(b_1, \ldots, b_t)$. They are promised that the interleaved product $\prod_{i \le t} a_i b_i$ equals to either g and h, for two fixed elements $g, h \in G$. Their task is to decide which is the case.

We show that for every $t \ge 2$ communication $\Omega(t \log |G|)$ is required, even for randomized protocols achieving only an advantage $\epsilon = |G|^{-\Omega(t)}$ over random guessing. This bound is tight, improves on the previous lower bound of $\Omega(t)$, and answers a question of Miles and Viola (STOC 2013). An extension of our result to 8-party number-on-forehead protocols would suffice for their intended application to leakage-resilient circuits.

Our communication bound is equivalent to the assertion 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 is nearly uniform over $G = \mathrm{SL}(2, q)$. This extends results by Gowers (Combinatorics, Probability & Computing, 2008) and by Babai, Nikolov, and Pyber (SODA 2008) corresponding to the independent case where A and B are product sets. We also obtain an alternative proof of their result that the product of three independent, high-entropy elements of G is nearly uniform. Unlike the previous proofs, ours does not rely on representation theory.

*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. Email: `viola@ccs.neu.edu`.

ISSN 1433-8092

1 Introduction and our results

Computing the iterated product $\prod_{i \leq t} g_i$ of a given tuple $(g_1, \ldots, g_t)$ of elements from a group G is a fundamental task. This is due to 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 [Yao79, KN97] question raised in [MV13]. First we give a definition.

Definition 1.1. *Let G be a group, let t be a positive integer, and let $a = (a_1, a_2, \ldots, a_t)$ and $b = (b_1, b_2, \ldots, b_t)$ be elements of G^t. The* interleaved product $a \bullet b$ *of a and b is the element $\prod_{i \leq t} a_i b_i = a_1 b_1 a_2 b_2 \ldots a_t b_t$ of G.*

We consider the following promise [ESY84] problem. Alice receives a tuple $a \in G^t$, and Bob similarly receives a tuple $b \in G^t$. They are guaranteed that the interleaved product $a \bullet b$ is equal to either g or h, where g and h are two fixed elements in G. They wish to decide which is the case.

A communication lower bound of $\Omega(t)$ over non-solvable groups follows by the lower bound for inner product [CG88], because, again, inner product can be reduced to iterated group product via [Bar89]. However, this bound is far from the (trivial) upper bound of $O(t \log |G|)$, and it gives nothing when $t = O(1)$.

The authors of [MV13] asked if a lower bound of $\omega(t)$, or ideally $\Omega(t \log |G|)$, can be established over any group. They arrived at this question through a study of leakage-resilient circuits. Specifically, they proposed a construction of such circuits based on group products, and showed that it resists leakage from various classes of circuits. (For recent progress, see [Mil14]). They also showed that the same construction remains secure in the "only computation leaks" model [MR04], if a lower bound of $\omega(t)$ holds for the extension of the above problem to 8-party number-on-forehead [CFL83] protocols, discussed below.

In this work we answer their question affirmatively in the 2-party case. We give a tight lower bound of $\Omega(t \log |G|)$ when $G = \mathrm{SL}(2, q)$ is the special linear group of 2×2 matrices with determinant 1 over the field $\mathbb{F}_q$. The lower bound holds even against public-coin protocols which achieve a small advantage over random guessing.

Theorem 1.2. *Let G be the group SL$(2, q)$. Let $P : G^t \times G^t \to \{0, 1\}$ be a (randomized, public-coin) c-bit communication protocol. For g in G denote by p_g the probability that $P(a, b)$ outputs 1 over uniform tuples a and b such that $a \bullet b = g$.*

For any $g, h \in G$, $|p_g - p_h| \leq 2^c |G|^{-\Omega(t)}$.

In this paper $\Omega(t)$ denotes a function bounded below by ct for an absolute constant c. In particular, c is independent of t and $|G|$. Similarly, $O(t)$ denotes a function bounded above by Ct for an absolute constant C.

We mention three variants of the problem in Theorem 1.2 that can be solved with $O(1)$ communication using the public-coin protocol for equality, cf. [KN97]. First, there is the case in which the group is abelian. Thus, our theorem provides a strong separation between interleaved group products over abelian groups and over $\mathrm{SL}(2,q)$. Second, there is the case $t = 1$. Third, there is a generalization of the second case, where $t = 2$ but one element, say a_1, is fixed to the identity. To see the latter, note that the problem reduces to checking whether $a_2 = b_1^{-1} g b_2^{-1}$. Thus, the case $t = 2$ appears to be the simplest case that is hard.

We conjecture that $\omega(t \log\log |G|)$ bounds hold for any non-abelian simple group. (The group $\mathrm{SL}(2,q)$ with odd q is not simple because it has a normal subgroup of size 2. This is not an obstacle for our result and we believe it never is, but for simplicity we state the conjecture for simple groups only.)

Conjecture 1.3. *Let G be a non-abelian simple group. Define p_g as in Theorem 1.2. For any $g, h \in G$, $|p_g - p_h| \le 2^c (\log |G|)^{-\Omega(t)}$.*

The bound in Conjecture 1.3 cannot be improved for the (non-abelian, simple) alternating group, see [MV13].

Multiparty protocols. We put forth several conjectures regarding extending our results to the number-on-forehead communication model [CFL83]. We denote by $G^{t\times k}$ a $t \times k$ matrix of elements in G. For $a \in G^{t\times k}$ we denote by $a_{i,j}$ the (i,j) entry. Consider the following problem on input $a \in G^{t\times k}$. There are k parties, where party i knows all the elements except those in column i. They are guaranteed that the *k-party t-tuple interleaved group product*

$$\prod_{i\le t} \prod_{j \le k} a_{i,j}$$

is equal to either g or h, where g and h are two fixed elements in G. They wish to decide which is the case.

Over any non-solvable group, a communication lower bound of $t/2^{O(k)}$ follows by reduction from the lower bound in [BNS92] for generalized inner product. We conjecture that improvements corresponding to those in Theorem 1.2 and Conjecture 1.3 hold in the multiparty setting:

Conjecture 1.4. *Let $P : G^{t\times k} \to \{0,1\}$ 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 $(a_{i,j})_{i\le t, j\le k}$ such that $\prod_{i\le t}\prod_{j\le k} a_{i,j} = g$. Then, for any two $g, h \in G$:*

1. *$|p_g - p_h| \le 2^c |G|^{-t/2^{O(k)}}$ if $G = SL(2,q)$, and*
2. *$|p_g - p_h| \le 2^c (\log |G|)^{-t/2^{O(k)}}$ if G is non-abelian and simple.*

Conjecture 1.4.1 with $k = 6$ would show that the aforementioned construction of leakage-resilient circuits in [MV13] tolerates a polynomial amount of leakage, as achieved by e.g. [GR12]. Conjecture 1.3 is the special case of Conjecture 1.4.2 with $k = 2$.

A central open problem in number-on-forehead communication complexity is to prove lower bounds when then 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 the top-left 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 entry 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 make the following conjecture. For concreteness we focus on the specific setting of parameters of polylogarithmic parties and communication.

Conjecture 1.5. *Let G be a non-abelian simple group, and let $c > 0$ be a constant. Then there is no protocol for the k-party t-tuple interleaved group product over G with $k = \log^c t$ parties and communication $\log^c t$.*

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).

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 $f(g_1 \cdot g_2 \cdot \ldots \cdot g_{k-1})$. 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.

The bound in our main result, Theorem 1.2, is equivalent to a bound on the mixing of interleaved distributions in groups, which is of independent interest. The next section discusses this perspective.

1.1 Mixing in groups

We consider the following general setup. G is a group which, as in the previous section, should be considered large. We have m distributions X_i over G, where each X_i has 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.

Our goal is to show that the distribution $D := \prod_{i \leq m} X_i$ (i.e., sample from each X_i and output the product, a.k.a. convolution) mixes, i.e., is nearly uniform over G. We will focus on an L_∞ bound. Specifically, we aim to show that D is equal to any fixed $g \in G$ with probability $1/|G|$ up to a multiplicative factor of $(1+\epsilon)$ for a small $|\epsilon|$:

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

Such a bound guarantees that D is supported over the entire group. By summing over all elements, we also infer that D is ϵ-close to uniform in statistical distance.

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

As a warm-up, consider the case $m = 2$. That is, we have two distributions X and Y. In this case, 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.

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 $X+Y+Z$ do not mix over the abelian group Z_p. For example, if $X = Y = Z$ are uniform over $\{0, 1, \ldots, p/4\}$ then $X+Y+Z$ is never equal to $p-1$.

However, for other groups it is possible to establish a good L_∞ bound. This was shown by Gowers [Gow08]. A sharper version of the result was proved by Babai, Nikolov, and Pyber, who established the following inequality.

Theorem 1.6 ([BNP08]). *Let G be a group, and let g be an element of 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.

In our example setting where 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 is equal to $O(d^{-1/2})$. We mention that for any non-abelian simple group we have $d \geq \sqrt{\log|G|}/2$, whereas for $G = \mathrm{SL}(2, q)$ 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 $\epsilon = 1/|G|^{-\Omega(1)}$. Jumping ahead, one of our contributions is to give an alternative proof of the latter bound which avoids representation theory.

Dependent distributions. We now consider the seemingly more difficult case where there may be dependencies across the X_i. As a warm-up, consider three distributions A, Y, and A', where A and A' may be dependent, but Y is independent from (A, A'). Does the distribution AYA' mix? It is not hard to see that the answer is negative. Indeed, let Y be uniform over an arbitrary set S of density $1/2$. Further let A be the uniform distribution over G. And define A' conditioned on the value of A as $A' := G \setminus S^{-1}A^{-1}$. It is easy to see that AYA' is never equal to 1_G. (This example corresponds to one mentioned after Theorem 1.2.)

Our main result 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, the bound scales in the desired way with the length t of the tuple.

Theorem 1.7. *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 $a \bullet b$ has distance at most $(\alpha\beta)^{-1}|G|^{-\Omega(t)}$ from uniform in statistical distance.

Mixing in three steps. As mentioned earlier, our results imply a special case of Theorem 1.6. Recall that the latter bounds the distance between XYZ and uniform. Our Theorem 1.7 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 1.6 we make a simple and general observation that mixing in four steps implies mixing in three, see §5.4. Thus we recover, up to polynomial factors, the bound in Theorem 1.6 for the special case of $G = \mathrm{SL}(2, q)$. This is of some interest because, unlike the original proofs, ours avoids representation theory.

1.2 Overview of techniques

In this section we give an overview of our techniques. First, it is easy to see that our communication bound, Theorem 1.2, and the mixing bound, Theorem 1.7, are both equivalent to the following version of the mixing bound, which is what we will work with. Here and elsewhere in the paper we identify sets with their characteristic functions.

Theorem 1.8. *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 1.9. *Theorems 1.2, 1.7, and 1.8 are equivalent.*

Proof. The equivalence between the two versions of the mixing bound, theorems 1.7 and 1.8, follows by Bayes' equality:

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

We now show that Theorem 1.8 implies the communication bound, Theorem 1.2. By an averaging argument we can assume that the protocol P in Theorem 1.2 is deterministic. Now write

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

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:

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

To see the reverse direction, that Theorem 1.2 implies Theorem 1.8, suppose that we are given sets A and B. Consider the constant-communication protocol $P(a,b) := A(a)B(b)$, and note that $p_g = E_{a \bullet b = g} A(b)B(b)$ and that $E_h p_h = \alpha\beta$. So we have

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

□

In the remainder of this section we explain the main ideas in the proof of Theorem 1.8. This theorem is proved in a somewhat different manner depending on whether $t = 2$ or t is large. Indeed, jumping ahead, for the case $t = 2$ we have two proofs, but we only know how to extend one of them to the case of larger t. We now focus on the case $t = 2$ and later we explain the case of larger t.

Our main technical lemma is the following result saying that the product of two typical conjugacy classes in $\mathrm{SL}(2,q)$ is nearly uniform over the group. Recall that the conjugacy class of an element g of a group G is the set of elements $u^{-1}gu$ for $u \in G$. We use the notation $C(g)$ to denote a uniform element from this set, i.e., $U^{-1}gU$ for a uniformly chosen U in G. Different occurrences of C correspond to different, independent U.

Lemma 1.10. *Let $G = SL(2,q)$. With probability $1 - 1/|G|^{\Omega(1)}$ over uniform a and b in G, the distribution $C(a)C(b)$ is $1/|G|^{\Omega(1)}$ close to uniform in statistical distance.*

This lemma relies on the specific choice of the group $G = \mathrm{SL}(2,q)$. But other than this, our proof applies to any group. So if a lemma like 1.10 can be established for other groups, Theorem 1.8 with $t = 2$ would follow for those groups too.

As it may not be apparent, we sketch why Lemma 1.10 is sufficient to prove Theorem 1.8, and then we explain how we prove Lemma 1.10.

Lemma 1.10 implies Theorem 1.8 with $t = 2$. Note that the quantity to bound in Theorem 1.8 can be written as

$$\mathbb{E}_b \mathbb{E}_{a : a \bullet b = 1} (A(a) - \alpha) B(b).$$

We can now use Cauchy-Schwarz and some simple manipulations to bound this from above by

$$\mathbb{E}_b \mathbb{E}^2_{a:a\bullet b=1} A(a) - \alpha^2,$$

up to polynomial factors. Since the inner expectation is squared, the whole expectation is equivalent to choosing b and then two values for a, both subject to $a \bullet b = 1$. Recalling $a \bullet b = a_1 b_1 a_2 b_2$, and averaging over b_2, we can rewrite the above expression as

$$\mathbb{E}_{a_1 b_1 a_2 = a'_1 b_1 a'_2} A(a_1, a_2) A(a'_1, a'_2) - \alpha^2.$$

Note that if a_1, a_2, a'_1, a'_2 were uniform this difference would be 0.

The fact that the same variable b_1 occurs on both sides of the equation $a_1 b_1 a_2 = a'_1 b_1 a'_2$ is what gives rise to conjugacy classes. Indeed, this equation can be rewritten as

$$a_2 = b_1^{-1}(a_1^{-1} a'_1) b_1 a'_2 = C(a_1^{-1} a'_1) a'_2.$$

Changing names of variables, we see that we are considering the following random walk on G^2. Pick a uniformly in G^2, and then move to ah, where h is the uniform distribution on pairs $(y, C(y))$. We need to show that the probability that a lands in A and ah also lands in A is close to α^2. As a final step, we make a general observation, again proved via Cauchy-Schwarz, that a result such as this follows from the result for a and ahh, cf. §5.1. The latter is given by Lemma 1.10, because hh is the distribution $(yz, C(y)C(z))$.

Proof of Lemma 1.10. There is an extensive literature on the growth of products of conjugacy classes in groups, see e.g. the book [AH85] and the papers citing it. However, existing results appear to be insufficient to obtain Lemma 1.10. The main reason for this is that papers in the literature mostly focus on the size of the support of $C(a)C(b)$, while we need to bound the distance from the uniform distribution. For the group $G = \mathrm{SL}(2, q)$ the size of the support is analyzed in a paper by Adan-Bante and Harris [ABH12] that was important for this work.

To prove Lemma 1.10 we begin with the observation that, for any a and b, the distribution of $C(a)C(b)$ is equal to the distribution of $C(C(a)C(b))$, i.e., we are allowed to take one extra conjugation at the end "for free." Now we critically rely on the structure of the conjugacy classes of the group $G = \mathrm{SL}(2, q)$. Essentially, G is a group of size q^3 with q classes of size q^2, cf. Lemma 3.1. In particular, except for a constant number of classes, every class in the group has roughly the same size. Coupled with the above observation, this means that it will be sufficient to show that $C(a)C(b)$ lands in each of the roughly q conjugacy classes with probability equal to $1/q$ up to a multiplicative factor $(1 + \epsilon)$ for $|\epsilon| \leq 1/q^{\Omega(1)}$.

To show the latter, we rely on the fact that there is an approximately 1-1 correspondence between the conjugacy class of an element g and its trace in $\mathbb{F}_q$. This key fact is also central to the aforementioned paper [ABH12] and to many other papers concerning conjugacy classes in $\mathrm{SL}(2, q)$. Thus, it suffices to show that for typical a and b, the trace of $C(a)C(b)$ is nearly uniform over $\mathbb{F}_q$. Because the traces of xy and yx are the same, the distribution of the trace of $C(a)C(b)$ is the same as the distribution of the trace of $aC(b)$. To show that $aC(b)$ is nearly uniform, we write this trace as a polynomial R whose variables are the four entries of U in the expression $aC(b) = aU^{-1}bU$. Our goal is to show that this polynomial R is equidistributed over the field $\mathbb{F}_q$, in the multiplicative sense above, cf. Lemma 4.1.

Whereas in some cases we can give an elementary proof of this equidistribution, in general we have to rely on classical results in algebraic geometry, specifically the Lang-Weil [LW54] multidimensional generalization of Weil's bound. This result shows that if a polynomial is absolutely irreducible, i.e., irreducible over any field extension, then it will take the value 0 with probability $1/q$ up to a multiplicative factor $(1+\epsilon)$ for $|\epsilon| \leq 1/q^{\Omega(1)}$. By showing that for all but $O(1)$ values $D \in \mathbb{F}_q$ the polynomial $R - D$ is absolutely irreducible, we conclude that the polynomial is close to uniformly distributed over $\mathbb{F}_q$. This concludes the proof sketch of Lemma 1.10.

Alternative proofs. As mentioned already, our proof avoids representation theory and as such departs from the approach in [Gow08, BNP08]. We have however an alternative proof of our main Theorem 1.8 for the case $t = 2$ which uses representation theory but avoids Lang-Weil. In a nutshell, this alternative proof starts along the way described above. However, one obtains a weaker equidistribution result, simply saying that $aC(b)$ lands in any fixed conjugacy class with probability $O(1/q)$ (but possibly misses a constant fraction of the classes). This weaker result, for which Schwartz-Zippel is sufficient, can then be boosted via the representation-theory inequality in Theorem 1.6 to obtain Lemma 1.10.

Longer tuples. We now briefly explain the proof of Theorem 1.8 for larger t. Applying Cauchy-Schwarz in a manner similar to that discussed in the above subsection "Lemma 1.10 implies Theorem 1.8," we reduce the problem to that of understanding a random walk over G that is obtained by alternating multiplication by a group element s_i and conjugation:

$$C(s_t \ldots C(s_2 C(s_1)) \ldots).$$

We prove that with probability $1-1/|G|^{\Omega(t)}$ over the choice of the s_i, the resulting distribution is $1/|G|^{\Omega(t)}$-close to uniform in statistical distance. This in turn follows by the next result, which shows that a constant number of steps in such a walk reduces the statistical distance to uniform of any distribution by a factor $1/|G|^{\Omega(1)}$:

Lemma 1.11. *Let D be a distribution over $G = SL(2,q)$. With probability $1 - 1/|G|^{\Omega(1)}$ over $s_1, s_2 \in G$, we have:*

$$|C(s_1 C(s_2 C(D))) - U|_1 \leq |D - U|_1/|G|^{\Omega(1)},$$

where U is the uniform distribution over G.

Organization. This paper is organized as follows. In §2 we exhibit a series of reductions, valid in all groups, that show that a statement similar to Lemma 1.11 – Lemma 2.5 – is sufficient to obtain the main mixing result, Theorem 1.8, in the case of large t. This lemma is then proved appealing to specific properties of the group in sections 3 and 3.3. The case $t = 2$ is then proved in §5. In §5.4 we explain how we recover a special case of Theorem 1.6.

2 Reductions of the theorem that are valid in all groups

In this section we shall give a sequence of reductions of what we need to prove until we end up with a simple statement about products of conjugacy classes in SL$(2, q)$.

2.1 Formulation in terms of quasirandom graphs

It turns out that what we are trying to prove is that a certain graph is quasirandom, in a sense that goes back to important papers of Thomason and Chung, Graham and Wilson in the 1980s [Tho87, CGW89]. Their main insight was that several properties of graphs, all of which say that a graph is "random-like" in some sense, are approximately equivalent. One of these properties is known as having low discrepancy. Given a graph Γ of density δ and two subsets A, B of $V(\Gamma)$ of densities α and β, we would expect the number of pairs (a, b) with $a \in A$ and $b \in B$ to be approximately $\alpha\beta\delta$ if Γ was random. The *discrepancy* of Γ is the maximum over all subsets A and B of $V(\Gamma)$, of the quantity

$$|\mathbb{E}_{x,y\in\Gamma}\Gamma(x, y)A(x)B(y) - \delta\alpha\beta|,$$

where α and β are the densities of A and B and we are using the letter Γ for the adjacency matrix of the graph as well as for the graph itself.

A very similar definition applies to bipartite graphs: the only difference is that this time if Γ has vertex sets X and Y, then we define the density of Γ to be $\mathbb{E}_{x\in X,y\in Y}\Gamma(x, y)$ and to define the discrepancy we take the maximum over all subsets $A \subset X$ and $B \subset Y$.

If we now let $G =$SL$(2, q)$ and $g \in G$ as above, and define a bipartite graph Γ with two copies X and Y of G^t as its vertex sets by joining $x \in X$ to $y \in Y$ if and only if $x \bullet y = g$, then the statement that

$$\mathbb{E}_{a\bullet b=g}A(a)B(b) - \alpha\beta$$

is small for any two sets $A, B \subset G^t$ of densities α and β is precisely the statement that the graph Γ has small discrepancy. Indeed, if we divide through by $|G|$, then the quantity we wish to bound becomes

$$\mathbb{E}_{a,b}\Gamma(a, b)A(a)B(b) - \alpha\beta|G|^{-1}.$$

Since the density of Γ is $|G|^{-1}$, this is of the right form, and our aim will be to prove that the discrepancy of Γ is at most $|G|^{-ct-1}$.

For convenience, we now state and prove the main (standard) fact about quasirandom graphs that we shall need. Given a bipartite graph Γ with finite vertex sets X and Y, define the *4-cycle density* of Γ to be the quantity

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

This is the probability that the quadruple (x, y, x', y') forms a (possibly degenerate) 4-cycle when x and x' are chosen independently and uniformly from X and y and y' are chosen independently and uniformly from Y. More generally, define the *4-cycle norm* $\|f\|_\square$ of a function $f : X \times Y \to \mathbb{R}$ by the formula

$$\|f\|_\square^4 = \mathbb{E}_{x,x',y,y'}f(x, y)f(x, y')f(x', y)f(x', y').$$

It can be proved that this does indeed define a norm, though we shall not need that fact here. In the next few results we define the L_2 norm using expectations: that is, if $f : X \to \mathbb{R}$, then $\|f\|_2 = (\mathbb{E}_x f(x)^2)^{1/2}$.

The following inequality tells us that a function with small 4-cycle norm has small correlation with functions of the form $(x, y) \mapsto u(x)v(y)$.

Lemma 2.1. *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*

$$|\mathbb{E}_{x,y} f(x, y)u(x)v(y)| \leq \|f\|_\Box \|u\|_2 \|v\|_2.$$

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

$$\begin{aligned}(\mathbb{E}_{x,y} f(x, y)u(x)v(y))^4 &= ((\mathbb{E}_x u(x)\mathbb{E}_y f(x, y)v(y))^2)^2 \\ &\leq ((\mathbb{E}_x u(x)^2)(\mathbb{E}_x(\mathbb{E}_y f(x, y)v(y))^2)^2 \\ &= \|u\|_2^4 \, (\mathbb{E}_{y,y'} v(y)v(y')\mathbb{E}_x f(x, y)f(x, y'))^2 \\ &\leq \|u\|_2^4 (\mathbb{E}_{y'y'} v(y)^2 v(y')^2)(\mathbb{E}_{y,y'}(\mathbb{E}_x f(x, y)f(x, y'))^2 \\ &= \|u\|_2^4 \|v\|_2^4 \|f\|_\Box^4.\end{aligned}$$

The result follows on taking fourth roots. □

Lemma 2.2. *Let Γ be a bipartite graph with finite vertex sets X and Y and density δ. Suppose that every vertex in X has degree $\delta|Y|$ and every vertex in Y has degree $\delta|X|$. For each $x \in X$ and $y \in Y$ let $f(x, y) = \Gamma(x, y) - \delta$. Then*

$$\|f\|_\Box^4 = \|\Gamma\|_\Box^4 - \delta^4.$$

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

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

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. □

Armed with these two lemmas, we can now show that if Γ is a bipartite graph of density δ that is regular in the sense of Lemma 2.2 (this assumption is not necessary, but it is convenient, and holds for our application), and if the 4-cycle density is not much larger than δ^4, then Γ has small discrepancy.

Corollary 2.3. *Let Γ be as in Lemma 2.2, let $c > 0$, and suppose that the 4-cycle density of Γ is at most $\delta^4(1 + c^4)$. Then for any two sets $A \subset X$ and $B \subset Y$ of densities α and β, respectively, we have the discrepancy estimate*

$$|\mathbb{E}_{x,y}\Gamma(x, y)A(x)B(y) - \delta\alpha\beta| \leq c\delta(\alpha\beta)^{1/2}.$$

Proof. Let $f(x,y)=\Gamma(x,y)-\delta$. Then by Lemma 2.2 we know that $\|f\|_\square \leq c^4\delta^4$. Therefore, by Lemma 2.1 we have

$$|\mathbb{E}_{x,y}f(x,y)A(x)B(y)| \leq c\delta(\alpha\beta)^{1/2},$$

since $\|A\|_2=\alpha^{1/2}$ and $\|B\|_2=\beta^{1/2}$. This is equivalent to the statement we wish to prove. $\square$

It will therefore be enough to prove that the 4-cycle density of Γ is at most $|G|^{-4}(1+|G|^{-ct})$ for some positive absolute constant c.

A sufficient condition for this to hold is that for all but a proportion $|G|^{-ct}$ of pairs of vertices a, a', the intersection of the neighbourhoods of a and a' has density within $|G|^{-ct}$ of $|G|^{-2}$, again for some absolute constant $c>0$. (It is simple to show that the condition is necessary as well, but we shall not need that fact.) To put this more analytically, we would like to show that if a and a' are chosen randomly from G^t, then

$$\mathbb{P}[|\mathbb{E}_b\Gamma(a,b)\Gamma(a',b)-|G|^2| \geq |G|^{-ct}] \leq |G|^{-ct},$$

where $\Gamma(a,b)=1$ if and only if $a\bullet b=g$ for some specified element g.

Our next task will be to understand this condition in a little more detail.

2.2 Reduction to the very rapid mixing of a certain random walk

A simple preliminary remark is that it is enough to prove the result when g is the identity e. Indeed, if we define $\phi: G^t\to G^t$ by $\phi:(b_1,\dots,b_t)\mapsto(b_1,\dots,b_{t-1},b_tg^{-1})$ and we let $B'=\phi(B)$, then B' has the same density as B, and

$$\mathbb{E}_{a\bullet b=g}A(a)B(b)=\mathbb{E}_{a\bullet\phi(b)=e}A(a)B(b)=\mathbb{E}_{a\bullet b=e}A(a)B(\phi^{-1}b)=\mathbb{E}_{a\bullet b=e}A(a)B'(b).$$

So from now on we shall restrict attention to the case $g=e$, which means that Γ is the graph where ab is an edge if and only if $a_1b_1a_2b_2\dots a_tb_t=e$.

As explained in the last section, we wish to show that for almost all pairs a, a' the proportion of b such that $a\bullet b=a'\bullet b=e$ is approximately $|G|^{-2}$. But this proportion is $|G|^{-1}$ times the proportion of $(b_1,\dots,b_{t-1})\in G^{t-1}$ such that

$$a_1b_1a_2b_2\dots a_{t-1}b_{t-1}a_t=a_1'b_1a_2'b_2\dots a_{t-1}'b_{t-1}a_t',$$

since for each such $(b_1,\dots,b_{t-1})$ there is a unique b_t such that if $b=(b_1,\dots,b_t)$, then $a\bullet b=a'\bullet b=e$.

Let us rearrange this equation as

$$a_t^{-1}b_{t-1}^{-1}a_{t-1}^{-1}\dots b_1^{-1}a_1^{-1}a_1'b_1a_2'b_2\dots a_{t-1}'b_{t-1}a_t'=e.$$

We are regarding a and a' as fixed, and $b_1,\dots,b_{t-1}$ as independent elements of G, chosen uniformly. So the left-hand side of the above equation is obtained as follows. We begin with the identity. Then we premultiply by a_1^{-1} and postmultiply by a_1', obtaining an element u_1. Next, we conjugate it by a random element of G to obtain an element v_1. Then we premultiply by a_2^{-1} and postmultiply by a_2' to obtain an element u_2, and pick a random conjugate v_2 of u_2, and so on.

Let us define c_i to be $a_i'a_i^{-1}$ for each i. Then for each i we know that $u_i = a_i^{-1}v_{i-1}a_i'$ is conjugate to $a_i'a_i^{-1}v_{i-1} = c_iv_{i-1}$. Therefore, a random conjugate of u_i has the same distribution as a random conjugate of c_iv_{i-1}.

It follows that we can consider a slightly simpler process instead. We have a fixed sequence $(c_1, \dots, c_{t-1})$ and two elements a_t and a_t'. We begin with the identity and alternately multiply by the next c_i and take a random conjugate. After the $(t-1)$st stage of this process, we premultiply by a_t^{-1} and postmultiply by a_t'. And we would like to prove that for almost all choices of $c_1, \dots, c_{t-1}, a_t, a_t'$, the probability that the resulting element is the identity is within $|G|^{-ct}$ of $|G|^{-1}$, where "almost all" means all but a proportion at most $|G|^{-ct}$.

Clearly, if we want to show that the final element has a probability very close to $|G|^{-1}$ of being the identity, it is sufficient to show that it is almost exactly uniformly distributed (in an L_∞ sense). And this will be the case if and only if it is the case before the final premultiplication by a_t^{-1} and postmultiplication by a_t'.

Given an element $c \in G$, define a linear map $T_c : \mathbb{R}^G \to \mathbb{R}_G$ that corresponds to the process of multiplying by c and taking a random conjugate. That is, writing $\sim$ for "is conjugate to",

$$T_c f(g) = \mathbb{E}_{ch\sim g} f(h).$$

If f is a probability distribution over G, then $T_c f$ is the probability distribution obtained by picking a random element according to the distribution f, multiplying it by c (it doesn't matter on which side), and taking a random conjugate. In particular, $T_g\delta_e$ is the uniform distribution on the conjugacy class of g.

It is therefore sufficient to prove the following result.

Theorem 2.4. *Let $c_1, \dots, c_{t-1}$ be chosen independently and uniformly from G. Then with probability at least $1 - |G|^{-ct}$, we have*

$$\|T_{c_{t-1}} \dots T_{c_2}T_{c_1}\delta_e - u\|_\infty \le |G|^{-ct},$$

where $c > 0$ is an absolute constant, δ_e is the point distribution at the identity, and u is the uniform distribution on G.

2.3 Reduction to the case $t = 3$

It is straightforward to show that Theorem 2.4 for sufficiently large t follows from a slightly stronger statement for $t = 3$. In this short section we prove this further reduction.

Lemma 2.5. *Suppose that there exists an absolute constant $c > 0$ such that if c_1, c_2, c_3 are chosen uniformly at random from G, then with probability at least $1 - |G|^{-c}$ we have the bound*

$$\|T_{c_3}T_{c_2}T_{c_1}\delta_g - u\|_1 \le |G|^{-c}$$

for every $g \in G$, where δ_g is the point distribution at g. Then Theorem 2.4 holds.

Proof. Let $f : G \to \mathbb{R}$ be a function with $\sum_x f(x) = 0$. Then

$$f = \sum_{g \in G} f(g)\delta_g = \sum_{g \in G} f(g)(\delta_g - u).$$

Let us call (c_1, c_2, c_3) *good* if the bound in the statement of the theorem holds, and let us call the operator $T_{c_3}T_{c_2}T_{c_1}$ good if and only if the triple (c_1, c_2, c_3) is good. (Whether or not a triple is good does not depend on c_1, but that will not concern us too much.)

Since $T_c u = u$ for every c, it follows that for every good triple (c_1, c_2, c_3) and every function f that sums to zero,

$$\|T_{c_3}T_{c_2}T_{c_1}f\|_1 \le \sum_{g \in G} |f(g)| \|T_{c_3}T_{c_2}T_{c_1}\delta_g - u\|_1 \le |G|^{-c}\|f\|_1.$$

That is, if we put the ℓ_1 norm on the space of functions that sum to zero, then the operator norm of $T_{c_3}T_{c_2}T_{c_1}$ is at most $|G|^{-c}$ for every good triple (c_1, c_2, c_3).

We also have that

$$\|T_c f\|_1 = \sum_g |\mathbb{E}_{ch \sim g} f(h)| \le \sum_g \mathbb{E}_{ch \sim g}|f(h)| = \sum_h |f(h)| \mathbb{E}_{g \sim ch} 1 = \|f\|_1$$

for every $c \in G$ and every function $f : G \to \mathbb{R}$.

Now let $t = 3m + 1$ be sufficiently large, and let $c_1, \dots, c_t$ be chosen uniformly and independently from G. We can write the operator $T_{c_{t-1}}T_{c_{t-2}} \dots T_{c_2}T_{c_1}$ as a composition of m operators $S_i = T_{c_{3i}}T_{c_{3i-1}}T_{c_{3i-2}}$, which each have a probability at least $1 - |G|^{-c}$ of being good, these events being independent.

The probability that at least half of the operators S_i are bad is at most $2^m|G|^{-cm}$, which is at most $|G|^{-c't}$ for some absolute constant $c' > 0$. If the number of bad S_i is less than $m/2$, then for every function f that sums to zero,

$$\|S_m S_{m-1} \dots S_1 f\|_1 \le |G|^{-c(t-1)/3}\|f\|_1,$$

which is again at most $|G|^{-c't}$ for some absolute constant $c' > 0$. Setting $f = \delta_e - u$ and using the fact that $S_i u = u$ for every i, we can deduce that

$$\|T_{c_{t-1}} \dots T_{c_2}T_{c_1}\delta_e - u\|_1 \le |G|^{-c't}.$$

Since the ℓ_∞ norm is at most the ℓ_1 norm, this implies the conclusion of Theorem 2.4. □

Lemma 2.5 is therefore sufficient to prove our main result, Theorem 1.8, when t is sufficiently large. This means that our remaining tasks are to prove Lemma 2.5 and to prove Theorem 1.8 when $t = 2$. (It is straightforward to prove that the bound does not get worse as t gets larger: we shall give the simple argument after proving the $t = 2$ case.)

3 Reductions that involve properties of the group

In order to reduce our problem (for large t) to Lemma 2.5, we have not needed to use the fact that we are working in the group $G = \mathrm{SL}(2,q)$. It is now, when we wish to prove the lemma, that particular properties of the group G become important. The main properties of $\mathrm{SL}(2,q)$ that we shall use are contained in the following two lemmas, the first of which is standard and the second of which is newer but related to work that has been done on products of conjugacy classes. In this section we shall explain why the given properties are sufficient, and then in the next section we shall prove the second lemma.

Lemma 3.1. *Let $G = SL(2,q)$. Then*

1. $|G| = q^3(1 + O(q^{-1}))$.
2. *G has $q(1 + O(q^{-1}))$ conjugacy classes.*
3. *If g is chosen uniformly at random from G, then with probability $1 - O(q^{-1})$ the conjugacy class of g has size $q^2(1 + O(q^{-1}))$.*

Lemma 3.1 is a special case of results on $\mathrm{SL}(2,q)$ that go back to Schur [Sch07]. For a more recent account see theorems 38.1 and 38.2 in [Dor71]. The case of even q is also discussed on pages 444-445 of [LS01].

It will be convenient to use the following notation.

Definition 3.2. *If g is an element of a group G, let $C(g)$ denote the conjugacy class of g. If we write (a.e. x) $P(x)$ or say that $P(x)$ holds for almost every x, then this will mean that if x is chosen uniformly at random, then the probability that $P(x)$ holds is $1 - O(q^{-1})$.*

Lemma 3.3. *Let $G = SL(2,q)$. Then*

$$(\text{a.e. } g)\ (\text{a.e. } h)\ \|T_h T_g \delta_e - u\|_1 = O(q^{-1/2}).$$

Claim 3.4. *Lemmas 3.1 and 3.3 imply the assumption of Lemma 2.5.*

Proof. Let us call a conjugacy class $C(g)$ *good* if it has size $q^2(1 + O(q^{-1}))$ and if

$$(\text{a.e. } h)\ \|T_h T_g \delta_e - u\|_1 = O(q^{-1/2}).$$

The first condition holds for almost every g, since this is Property 3, and the second condition also holds for almost every g, by Lemma 3.3. Thus, for almost every g, the conjugacy class $C(g)$ is good.

Now let $(c_1, c_2, c_3) \in G^3$ and let us find a sufficient condition for $\|T_{c_3} T_{c_2} T_{c_1} \delta_g - u\|_1$ to be small for every g. The probability distribution $T_{c_3} T_{c_2} T_{c_1} \delta_g$ can be defined as follows. We pick an element uniformly at random from the conjugacy class $C(c_1 g)$, multiply it on the left by c_2, take a random conjugate of that, multiply on the left by c_3, and take a random conjugate of *that*.

Suppose first that $C(c_1g)$ is a good conjugacy class. Then for almost every c_2 we have that $\|T_{c_2}T_{c_1g}\delta_e - u\|_1 = O(q^{-1/2})$. But $T_{c_1g}\delta_e = T_{c_1}\delta_g$. Also, $T_{c_3}u = u$ and T_{c_3} does not increase ℓ_1 norms. It follows that $\|T_{c_3}T_{c_2}T_{c_1}\delta_g - u\|_1 = O(q^{-1/2})$.

Now suppose that $C(c_1g)$ is a bad conjugacy class. Since the union B of all bad conjugacy classes has size $O(q^2)$, if we choose c_2 uniformly at random, the expected size of $c_2C(c_1g)\cap B$ is $O(q^{-1})|c_2C(c_1g)|$. Therefore, by Markov's inequality, the probability that $c_2C(c_1g)\cap B$ has size greater than $q^{-1/2}|c_2C(c_1g)|$ is $O(q^{-1/2})$.

If c_2 does not have this property, and h is a random element of $C(c_1g)$, then c_2h belongs to a good conjugacy class with probability $1-O(q^{-1/2})$. But then the proof in the first case gives us that with probability $1-O(q^{-1})$, $\|T_{c_3}T_{c_2}\delta_h - u\|_1 = O(q^{-1/2})$. If c_2h belongs to a bad conjugacy class, we still have that $\|T_{c_3}T_{c_2}h - u\|_1 \le 2$.

Therefore, $\|T_{c_3}T_{c_2}T_{c_1}\delta_g - u\|_1 = O(q^{-1/2})$ when $C(c_1g)$ is a bad conjugacy class too. □

4 Proof of Lemma 3.3

When it comes to proving Lemma 3.3, a key observation, which is also central to the argument of Adan-Bante and Harris [ABH12] (and for many other papers concerning conjugacy classes in $\mathrm{SL}(2,q)$), 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. For the other, note that a matrix of determinant 1 can have any of the q possible traces, and recall from Lemma 3.1 that the group has $q+O(1)$ conjugacy classes.

Thus, if we want to prove that some distribution over $\mathrm{SL}(2,q)$ is approximately close to uniform, it is enough to prove that the trace of a random matrix from that distribution is approximately uniformly distributed. In the case of Lemma 3.3 this means showing that for almost every g and almost every h, the trace of $hugu^{-1}$ is approximately uniformly distributed for uniform u. Moreover, 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. Hence, we can work with representatives of our choosing, as done in the next key lemma. Recall that we use C to stand for "the conjugacy class of".

Lemma 4.1. *Let $G = SL(2,q)$. Then the distribution of $Tr\left(\begin{pmatrix} 0 & 1 \\ 1 & w \end{pmatrix} C\begin{pmatrix} v & 1 \\ 1 & 0 \end{pmatrix}\right)$ is $1/q^{\Omega(1)}$ close to uniform in statistical distance if either (i) q is even, or (ii) q is odd and $(v^2,w^2) \neq (-4,-4)$ and $(v,w) \neq (0,0)$.*

In the proof that follows, if we use a letter such as a to refer to an element of G, we shall refer to its entries as $a_1,\dots,a_4$. More precisely, we shall take a to be the matrix $\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$.

We begin with working out a simple expression for the trace.

Claim 4.2. *Let a, u and g be 2×2 matrices in $SL(2,q)$. Then*

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

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 4.1 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 4.3. *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*

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

The rest of the section is devoted to the proof of Lemma 4.1. First we remark that the calculation below for the trace in the case $v = w = 0$ shows that the condition $(v, w) \ne (0, 0)$ is necessary over odd characteristic.

From Claim 4.2 we obtain the following expression for the trace.

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

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 $|e_s| \le q^{2-\Omega(1)}$. And for the other $O(1)$ choices of s the number of solutions is $O(q^2)$. This will show that the trace has statistical distance $1/q^{\Omega(1)}$ from uniform. Indeed, using that $|G| = q^3 - q$, the contribution to this distance of each of the aforementioned $q - O(1)$ values of s is $|(q^2 + e_s)/(q^3 - q) - 1/q| = |(1 + e_s)/(q^3 - q)| \le 1/q^{1+\Omega(1)}$ because $|e_s| \le q^{2-\Omega(1)}$. These add up to a contribution of $1/q^{\Omega(1)}$, while for each of the others the contribution is at most $O(1/q)$.

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 4.3. 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 $x \neq 0$.

Note that 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:

$$x^2yz \to ac' + a'c + s = -vw$$
$$x^2 \to db' + d'b = s.$$

Now 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:

$$xy^2z \to a^2 - b^2 = -wab$$
$$x^3y \to a^2 - d^2 = -vad$$
$$xyz^2 \to -b^2 + c^2 = vbc$$
$$x^3z \to c^2 - d^2 = wcd.$$

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

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

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 now assume 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.

5 The case $t = 2$

In this section we prove that if A, B are subsets of G^2 of density α and β, respectively, then

$$|\mathbb{E}_{a_1b_1a_2b_2=e}A(a)B(b) - \alpha\beta| \leq |G|^{-c}$$

for some absolute constant $c > 0$. Unfortunately, we cannot use Theorem 2.4 here, because that would require us to prove that with high probability $\|T_g\delta_e - u\|_\infty \leq |G|^{-c}$ when g is chosen randomly from G. Since $T_g\delta_e$ is the uniform distribution over the conjugacy class of g, this is clearly false.

5.1 Proving low discrepancy using 8-cycles

To get round this, we prove a variant of Lemma 2.1.

Lemma 5.1. *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*

$$|\mathbb{E}_{x,y}f(x, y)u(x)v(y)| \leq \|g\|_\square^{1/2}\|u\|_2\|v\|_2.$$

Proof. Note that up to normalization, g is just ff^T. So the lemma is saying that if we do not have a bound for $\|f\|_\square$ we can still get a discrepancy bound by bounding $\|ff^T\|_\square$.

To prove it, we begin more or less as we began the proof of Lemma 2.1.

$$\begin{aligned}(\mathbb{E}_{x,y}f(x,y)u(x)v(y))^2 &= (\mathbb{E}_y v(y)\mathbb{E}_x f(x,y)u(x))^2 \\ &\leq (\mathbb{E}_y v(y)^2)(\mathbb{E}_y(\mathbb{E}_x f(x,y)u(x))^2 \\ &= \|v\|_2^2\,\mathbb{E}_{x,x'}(\mathbb{E}_y f(x,y)f(x',y))u(x)u(x'). \\ &= \|v\|_2^2\,\mathbb{E}_{x,x'}g(x,x')u(x)u(x').\end{aligned}$$

But by Lemma 2.1 this is at most $\|v\|_2^2\|g\|_\square\|u\|_2^2$, which proves the lemma. □

We also need a slight generalization of Lemma 2.2.

Lemma 5.2. *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. The proof is identical to that of Lemma 2.2 except that Γ is replaced by F. □

Suppose now that Γ is a bipartite graph with finite vertex sets X and Y such that each vertex in X has degree $\delta|Y|$ and each vertex in Y has degree $\delta|X|$. Let $f = \Gamma - \delta$. Then for every $x, x' \in X$ we have

$$\mathbb{E}_y f(x,y)f(x',y) = \mathbb{E}_y(\Gamma(x,y) - \delta)(\Gamma(x',y) - \delta) = \mathbb{E}_y\Gamma(x,y)\Gamma(x',y) - \delta^2,$$

since $\mathbb{E}_y\Gamma(x,y) = \delta$ for every x and $\mathbb{E}_x\Gamma(x,y) = \delta$ for every y.

Let $\Delta(x,x') = \mathbb{E}_y\Gamma(x,y)\Gamma(x,y')$ for every $x, x' \in X$. We also have that for each x,

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

By symmetry, $\mathbb{E}_x\Delta(x,x') = \delta^2$ for every x' as well.

It follows from these observations that we can apply Lemma 5.2 with F replaced by Δ, δ replaced by δ^2, and f replaced by g (where g is as in Lemma 5.1), to deduce that

$$\|g\|_\square^4 = \|\Delta\|_\square^4 - \delta^8.$$

5.2 Application to interleaved products

Let Γ be the bipartite graph that has vertex sets equal to G^2, where (a_1, a_2) is joined to (b_1, b_2) if and only if $a_1b_1a_2b_2 = e$. Then the density of Γ is $|G|^{-1}$, so what we want to prove is equivalent to the estimate

$$|\mathbb{E}_{a,b}\Gamma(a,b)A(a)B(b) - \alpha\beta|G|^{-1}| \leq |G|^{-(1+c)},$$

which in turn is equivalent to the estimate

$$|\mathbb{E}_{a,b}f(a,b)A(a)B(b)| \leq |G|^{-(1+c)}.$$

For this it is sufficient, by Lemma 5.1 to prove that $\|g\|_\square \le |G|^{-(2+c)}$. Since $\|g\|_\square^4 = \|\Delta\|_\square^4 - \delta^8$ and in our case $\delta = |G|^{-1}$, it is enough to show that $\|\Delta\|_\square^4 \le |G|^{-8}(1+|G|^{-c})$. That is, $\|\Delta\|_\square$ must be within a factor $1+|G|^{-c}$ of the smallest possible value it can take. (Note that in this discussion $c > 0$ is an absolute constant that can change from line to line.)

Now

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

so our aim is to find an upper bound for the variance of $\mathbb{E}_z\Delta(x,z)\Delta(z,x')$. But

$$\mathbb{E}_z\Delta(x,z)\Delta(z,x') = \mathbb{E}_{z,y,y'}\Gamma(x,y)\Gamma(z,y)\Gamma(z,y')\Gamma(x',y'),$$

which 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 $y_1^{-1}z_1^{-1}x_1y_1 = z_2x_2^{-1}$ and $y_1'^{-1}x_1'^{-1}z_1y_1' = x_2'z_2^{-1}$. Let us introduce variables $u_1 = z_1^{-1}x_1$, $u_2 = z_2x_2^{-1}$, $v_1 = x_1'^{-1}z_1$ and $v_2 = x_2'z_2^{-1}$. Then the constraints on the u_i and v_i are that $v_1u_1 = x_1'^{-1}x_1$ and $v_2u_2 = x_2'x_2^{-1}$, and they are uniformly distributed subject to those constraints (as z varies with x and x' fixed). Therefore, the probability that $y_1^{-1}z_1^{-1}x_1y_1 = z_2x_2^{-1}$ and $y_1'^{-1}x_1'^{-1}z_1y_1' = x_2'z_2^{-1}$ is equal to

$$\mathbb{P}[y_1^{-1}u_1y_1 = u_2 \wedge y_1'^{-1}v_1y_1' = v_2 \,|\, v_1u_1 = x_1'^{-1}x_1 \wedge v_2u_2 = x_2'x_2^{-1}].$$

The two events in the above conditional probability both have probability $|G|^{-2}$, so by Bayes's theorem we can rewrite this conditional probability as

$$\mathbb{P}[v_1u_1 = x_1'^{-1}x_1 \wedge v_2u_2 = x_2'x_2^{-1} \,|\, y_1^{-1}u_1y_1 = u_2 \wedge y_1'^{-1}v_1y_1' = v_2].$$

In words, we pick u_1 and v_1 uniformly and then let u_2 and v_2 be random conjugates of u_1 and v_1, respectively. We then want to estimate the probability that $v_1u_1 = x_1'^{-1}x_1$ and $v_2u_2 = x_2'x_2^{-1}$.

For the variance mentioned earlier to be small, we need this probability to be close to $|G|^{-2}$ for almost all x and x'. In other words, the information that u_2 is a conjugate of u_1 and v_2 is a conjugate of v_1 should make almost no difference to the distribution of the pair (v_1u_1, v_2u_2).

Our question can therefore be rephrased as follows. We are given some $x \in G$. We then randomly choose u and v such that $uv = x$, and we randomly choose conjugates gug^{-1} and hvh^{-1} of u and v. What can we say about the distribution of $gug^{-1}hvh^{-1}$?

The work we have done in previous sections quickly answers this question. First, observe that $gug^{-1}hvh^{-1} = gu(g^{-1}h)v(g^{-1}h)^{-1}g^{-1}$ and that the pair $(g, g^{-1}h)$ is uniformly distributed over G^2. Therefore, another way of describing the distribution is that we choose v randomly, then take a random conjugate of v, then multiply it on the left by u (which is equal to xv^{-1}) and take a random conjugate again. For each given u and v this is precisely the distribution $T_uT_v\delta_e$, where T_u and T_v are as defined just before the statement of Theorem 2.4.

By Lemma 3.3, with probability $1 - O(q^{-1})$ we have that $T_u T_v \delta_e$ is within $O(q^{-1/2})$ of uniform.

Since for an arbitrary v we also know that $T_u T_v \delta_e$ is within 2 of uniform, we obtain that

$$\mathbb{E}_{x,x'}|\mathbb{E}_z \Delta(x,z)\Delta(x',z) - |G|^{-4}| = O(|G|^{-4}q^{-1/2}).$$

For the same reason, we know that $\mathbb{E}_z \Delta(x,z)\Delta(x',z) = O(|G|^{-4})$ for every x, x'. It follows that

$$\mathbb{E}_{x,x'}(\mathbb{E}_z \Delta(x,z)\Delta(x',z) - |G|^{-4})^2 = O(|G|^{-8}q^{-1/2}).$$

This proves that $\|\Delta\|_\square^4 \leq |G|^{-8}(1 + O(|G|^{-1/6}))$. Combining this with our previous calculations, we have the following lemma.

Lemma 5.3. *Let G be the group $SL(2,q)$ and let $\Gamma : G^2 \times G^2 \to \mathbb{R}$ be defined by setting $\Gamma(a,b) = 1$ if $a_1 b_1 a_2 b_2 = e$ and 0 otherwise. Let $f(a,b) = \Gamma(a,b) - |G|^{-1}$ and let $g(a,a') = \mathbb{E}_b f(a,b) f(a',b)$. Then $\|g\|_\square = O(|G|^{-2-1/24})$.*

Proof. At the end of the previous section we proved that $\|g\|_\square^4 = \|\Delta\|_\square^4 - \delta^8$, where $\delta = \mathbb{E}_{a,b}\Gamma(a,b)$, which in this case is $|G|^{-1}$. The result therefore follows from the estimate we have just given for $\|\Delta\|_\square^4$. □

Applying this result and Lemma 5.1 with u and v the characteristic functions of sets A and B, we obtain the following result.

Theorem 5.4. *Let G be the group $SL(2,q)$, and let $A, B \subset G^2$ be subsets of density α and β, respectively. Then for every $g \in G$,*

$$|\mathbb{E}_{a_1 b_1 a_2 b_2 = g} A(a)B(b) - \alpha\beta| = O(|G|^{-1/48})(\alpha\beta)^{1/2}.$$

As we noted for the general case, we can use Bayes's theorem to reformulate this statement in a useful way. We have

$$\begin{aligned}\mathbb{E}_{a_1 b_1 a_2 b_2 = g} A(a)B(b) &= \mathbb{P}[a \in A, b \in B \mid a_1 b_1 a_2 b_2 = g] \\ &= \alpha\beta|G|\,\mathbb{P}[a_1 b_1 a_2 b_2 = g \mid a \in A, b \in B].\end{aligned}$$

Therefore, Theorem 5.4 is telling us that

$$\mathbb{P}[a_1 b_1 a_2 b_2 = g \mid a \in A, b \in B] - |G|^{-1}| \leq (\alpha\beta)^{-1/2}|G|^{-49/48}.$$

5.3 A generalization to arbitrary distributions

In this section we state and quickly prove the following generalization of Theorem 5.4.

Theorem 5.5. *Let u and v be two probability distributions on G^2 and let U be the uniform distribution on G^2. Then*

$$\sum_{a_1 b_1 a_2 b_2 = e} u(a_1,a_2)v(b_1,b_2) = n^{-1} + O(n^{1-1/48})\|u\|_2\|v\|_2.$$

Proof. From Lemmas 5.1 and 5.3 we obtain the result that

$$|\mathbb{E}_{a,b}(\Gamma(a,b) - n^{-1})u(a)v(b)| = O(n^{-1-1/48})(\mathbb{E}_a u(a)^2)^{1/2}(\mathbb{E}_b v(b)^2)^{1/2},$$

where $\Gamma(a,b) = 1$ if $a_1 b_1 a_2 b_2 = e$ and 0 otherwise. Multiplying both sides by n^4, we deduce that

$$|\sum_{a_1 b_1 a_2 b_2 = e} u(a_1, a_2)v(b_1, b_2) - n^{-1}\|u\|_1\|v\|_1| = O(n^{1-1/48})\|u\|_2\|v\|_2,$$

which is what we wanted. □

We remark that the value of $\mathbb{E}_{a,b}(\Gamma(a,b) - n^{-1})u(a)v(b)$ is unchanged if a constant is added to either u or v, so we can if we wish improve the bound to $O(n^{1-1/48})\|u - U\|_2\|v - U\|_2$.

We sketch an alternative proof of Theorem 5.5 that relies only on Theorem 5.4. First, note that any distribution u is a convex combination of distributions u_i which are uniform on at least $1/(2\|u\|_2)$ points, if $\|u\|_2$ is at most $1/2$. This is because the ℓ_2 upper bound implies an ℓ_∞ upper bound, and the distributions u_i are the vertexes of the polytopes of distributions with that ℓ_∞ bound. The extra factor of two on the number of points accounts for the possibility that $\|u\|_2$ is not an integer. Given this, we can fix distributions u_i and v_i that maximize the probability that we are trying to bound, and invoke Theorem 5.4.

5.4 Proof that G is quasirandom

An immediate consequence of Theorem 5.4 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 5.6. *Let G be the group $SL(2,q)$, and let $A, B, C, D \subset G$ 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

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

Proof. For the first statement we simply apply Theorem 5.4 with A replaced by $A \times C$ and B replaced by $B \times D$. For the second, we apply the equivalent version stated just after the proof of Theorem 5.4. □

It turns out that from this result for four sets follows the same result for three sets. This is of some interest, because it gives the first proof that G is quasirandom, in the sense of [Gow08], that does not use representation theory.

Corollary 5.7. *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

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

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 5.6. 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 $\mathbb{E}_{abc=g}A(a)B(b)C(c) = \mathbb{P}[a \in A, b \in B, c \in C \mid abc = g]$. □

Acknowledgments. 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

[ABH12] Edith Adan-Bante and John M. Harris. On conjugacy classes of $\mathrm{SL}(2, q)$. *Revista Colombiana de Matemticas*, 46(2):97–111, 2012.

[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.

[AH85] Z. Arad and M. Herzog, editors. *Products of conjugacy classes in groups*, volume 1112 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1985.

[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.

[CGW89] Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[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.

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

[ESY84] Shimon Even, Alan L. Selman, and Yacov Yacobi. The complexity of promise problems with applications to public-key cryptography. *Information and Control*, 61(2):159–173, 1984.

[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.

[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.

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

[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.

[LS01] Martin W. Liebeck and Aner Shalev. Diameters of finite simple groups: sharp bounds and applications. *Annals of Mathematics*, (154):383–406, 2001.

[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.

[Sch07] Issai Schur. Untersuchungen über die darstellung der endlichen gruppen durch gebrochene lineare substitutionen. *Journal für die reine und angewandte Mathematik*, (132):85–137, 1907.

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

[Tho87] Andrew Thomason. Pseudo-random graphs. *Annals of Discrete Mathematics*, 33:307–331, 1987.

[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.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de