Annals of Mathematics **179** (2014), 405–429
http://dx.doi.org/10.4007/annals.2014.179.2.1

A product theorem in free groups

By Alexander A. Razborov

Abstract

If A is a finite subset of a free group with at least two noncommuting elements, then $|A\cdot A\cdot A| \geq \frac{|A|^2}{(\log|A|)^{O(1)}}$. More generally, the same conclusion holds in an arbitrary virtually free group, unless A generates a virtually cyclic subgroup. The central part of the proof of this result is carried on by estimating the number of collisions in multiple products $A_1 \cdot \ldots \cdot A_k$. We include a few simple observations showing that in this "statistical" context the analogue of the fundamental Plünnecke-Ruzsa theory looks particularly simple and appealing.

1. Introduction

Let G be a group, and let A be its finite subset. Assume that for some fixed $k \geq 2$, $|\underbrace{A \cdot A \cdot \ldots \cdot A}_{k \text{ times}}|$ (where the *product set* $\underbrace{A \cdot A \cdot \ldots \cdot A}_{k \text{ times}}$ is defined as $\{\, b \in G \mid (\exists a_1, \ldots, a_k \in A)(b = a_1 a_2 \cdots a_k)\,\}$) is much smaller than $|A|^k$. What can be said about the internal structure of A?

Questions of this (and similar) sort are known in arithmetic combinatorics as *inverse problems.* (Most of the material briefly surveyed in this section can be found in comprehensive monographs [21], [25].) Originally they were studied for $G = \mathbb{Z}$. (The case $G = \mathbb{Z}^n$ is easily seen to be "essentially equivalent" to this one.) One of the deepest and hardest results in the area is Freiman's theorem [8] that provides a complete characterization of sets $A \subseteq \mathbb{Z}$ with $|A + A + \cdots + A| \leq O(|A|)$.

For many applications, however, it is highly desirable to be able to infer at least something intelligent about the structure of A from the weaker assumption

$$|\underbrace{A + A + \cdots + A}_{k \text{ times}}| \leq |A|^{1+o(1)}. \tag{1}$$

Supported by the NSF grant ITR-0324906 and by the Russian Foundation for Basic Research.

For the case of abelian groups, this is a widely open problem (perhaps, *the* central problem in the whole area). This state of the art is particularly embarrassing given the amount of useful information one *can* extract from (1) with the help of powerful *Plünnecke-Ruzsa theory.* As one of the most cited corollaries, let us just mention that the conditions (1) are equivalent for all (fixed) $k \geq 2$ and, moreover, this equivalence still holds if some pluses are replaced by minuses. Further, (1) follows from $|A+B| \leq |A|^{o(1)}|B|$ for an *arbitrary* set B with $|B| \leq |A|$. Unfortunately, these powerful conclusions tell us very little about the *internal* structure of A.

Somewhat surprisingly, inverse problems have turned out to be simpler for more complicated algebraic structures. For example, sum-product estimates in commutative rings by Bourgain, Katz and Tao [5] do give strong inverse results in the range (1) if we append the analogous restriction $|A \cdot A \cdot \ldots \cdot A| \leq |A|^{1+o(1)}$ for product sets.

In this paper we are interested in another class of algebraic structures that has recently sparkled a considerable attention, the class of *non-abelian* groups [11], [24], [7]. One of the reasons for this interest lies in the motivations of the pioneering papers by Helfgott [11] and Bourgain and Gamburd [4] that linked this kind of question to estimating the diameter of Cayley graphs in certain finite groups and, via this, to difficult open problems about explicit constructions of expanders. But before reviewing these latest developments, it is worth mentioning that for groups equipped with a length function, very similar problems were studied long before, in quite a different context and in a different community. Specifically, the *Rapid Decay Property* [10], [13] implies that any set A satisfying (1) (or, in fact, the weaker assumption $|A \cdot A| \leq A^{2-\Omega(1)}$) cannot be positioned within a small ball and must necessarily contain elements of length $|A|^{\Omega(1)}$. Among others, this property is known for free groups [10], groups of polynomial growth and hyperbolic groups [13].

An easy example shows that the Plünnecke-Ruzsa theory does not literally transfer to the non-abelian case: $|A \cdot A|$ can be small, whereas already $|A \cdot A \cdot A|$ is large. Tao [24] and Helfgott [11], however, proved that this theory catches up already at the next level: say, the statements (1) become equivalent for $k = 3, 4, \ldots$. For this reason, in the non-abelian case it does make sense to concentrate on the study of sets A with *small tripling* (that is, $k = 3$) as opposed to sets with *small doubling* in the abelian case. Helfgott [11] indeed proved a strong inverse result for tripling in the range (1) when $G = \mathrm{SL}_2(\mathbb{Z}_p)$. Chang [7] proved a similar theorem for $G = \mathrm{SL}_2(\mathbb{C})$ and made a very substantial step toward obtaining an analogous result for $G = \mathrm{SL}_3(\mathbb{Z})$.

Chang's former result (for $\mathrm{SL}_2(\mathbb{C})$) in fact looks rather intriguing since it exhibits the following "threshold behavior." There exists a *fixed* constant $\delta > 0$ such that the structural conclusion she gets from $|A \cdot A \cdot A| \leq |A|^{1+\delta}$

is *exactly* the same as the conclusion one gets from much stronger bound (1): A generates a virtually abelian subgroup. (This reduces the inverse problem for $\mathrm{SL}_2(\mathbb{C})$ to the same problem for abelian groups — the best we can hope for without actually solving the latter!) This is very unusual for arithmetic combinatorics, where the conclusion usually depends on things like $|A \cdot A|$ or $|A \cdot A \cdot A|$ *numerically* and *smoothly*. Chang also remarked that the same conclusion holds (via any known embedding of F_m into $\mathrm{SL}_2(\mathbb{C})$) for free groups F_m,[1]. She asked for a purely combinatorial proof of this fact and for any estimates of the threshold constant δ.

The main result of our paper provides an answer to her question, and we show that in fact $\delta = 1$ (which is clearly optimal). More precisely, we prove the following:

Main Theorem. *Let A be a finite subset of a free group F_m with at least two noncommuting elements. Then*

$$|A \cdot A \cdot A| \geq \frac{|A|^2}{(\log |A|)^{O(1)}}.$$

More generally, the same conclusion holds for any finite subset A of an arbitrary fixed virtually free group, unless the subgroup generated by A is virtually cyclic. In particular, this is true for the modular group $\mathrm{PSL}_2(\mathbb{Z})$, as well as for $\mathrm{SL}_2(\mathbb{Z})$ and $\mathrm{GL}_2(\mathbb{Z})$, and this makes an improvement over [7, Th. 5.1]. (The latter gave the bound $|A \cdot A \cdot A| \geq |A|^{1+\delta}$ for $\mathrm{SL}_2(\mathbb{Z})$ and for an unspecified constant $\delta > 0$.)

Our proof is heavily based on the machinery of combinatorial group theory and, more specifically, its part known as the theory of (highly) periodic words. It is worth noting that this theory lies in the heart of two of the deepest (and *extremely* involved technically) achievements in that area: the work on Burnside problem [1], and the work on equations in free groups [18], [6], [20], [22] that has recently culminated in independent solutions of Tarski's problem given by Kharlampovich-Miasnikov [14], [15] and Sela [23].

Instead of lower bounds on the cardinalities of sum/product sets, it is often more convenient to go after *upper* bounds on the dual quantities[2] defined like

$$\mathbf{c}(A, B) \stackrel{\text{def}}{=} \left| \left\{ (a, b, a', b') \in (A \times B)^2 \mid ab = a'b' \right\} \right|.$$

---

[1]Breuillard (personal communication) observed that this can be derived already from the work of Helfgott [11].

[2]Accordingly, they appeared in the literature under many different names, e.g., *quadruples* [9] or *additive energy* [24], [25]. In order to stress our purely combinatorial treatment, we prefer to follow the lead of [2] and call them *collision numbers* or, after appropriate normalization, *collision probabilities*.

These collision numbers are related to the cardinalities of sum/product sets via a simple Cauchy-Schwartz by

$$|A \cdot B| \geq \frac{|A|^2|B|^2}{\mathbf{c}(A,B)},$$

but they display much more analytical (and in many cases more convenient) behavior than $|A \cdot B|$. The Balog-Szemerédi-Gowers theorem shows how to go in the opposite direction (from large $\mathbf{c}(A,B)$ to large subsets $A_0 \subseteq A,\ B_0 \subseteq B$ with small $|A_0 \cdot B_0|$) without losing too much. But we would also like to note that one of the most striking recent applications of arithmetic combinatorics [2], [3] actually *needs* upper bounds on collision numbers/probabilities rather than lower bounds on the size of sum/product sets.

The most crucial part of our argument (contained in Section 5) also works entirely in this framework (that we, following [2] once more, will call *statistical*) and essentially utilizes all its versatility. This has motivated us to wonder how far we can get in the world in which all quantities like $|A_1 \cdot A_2 \cdot \ldots \cdot A_k|$ are *systematically* replaced by their statistical counterparts $\mathbf{c}(A_1, \ldots, A_k)$. We contribute to this a few simple remarks showing that the statistical version of Plünnecke-Ruzsa theory looks particularly simple and appealing, without ever mentioning cardinalities $|A_1 \cdot A_2 \cdot \ldots \cdot A_k|$, Menger's theorem or Ruzsa's covering lemma inherent to its "classical" versions.

These remarks are given in the concluding Section 6, and all the preceding part of the paper is entirely devoted to the proof of the Main Theorem. In Section 2 we give the necessary background, mostly from combinatorial group theory. In Section 3 we get rid of cancellations and also show that when lower bounding $|A \cdot B \cdot C|$ in a free semi-group, we can assume without loss of generality that A is a prefix chain, and C is a suffix chain. In Section 4 we further reduce our problem to the case when the triple (A, B, C) has "enough aperiodicity" in it. Then in Section 5 comes the central part of our proof: we upper bound the collision numbers $\mathbf{c}(A, B, C)$, ruling out the only unpleasant case with the help of "aperiodicity constraints" enforced in the previous Section 4.

Added in proof. The present paper has been circulated in preprint form since 2007. Since then, developments on triple-product theorems and their applications have succeeded each other rapidly. See, for example, the survey papers by Kowalski [16], Lubotzky [17] and Helfgott [12].

2. **Background**

All the material in this section related to the combinatorial group theory can be found, e.g., in [19], [1].

We let $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$. Let F_m be the free group with the basis $\{x_1, \ldots, x_m\}$. A word w in the alphabet $\{x_1, x_1^{-1}, \ldots, x_m, x_m^{-1}\}$ is *reduced* if for any $i \in [m]$, x_i and x_i^{-1} never appear in w as adjacent letters. The elements of F_m are in one-to-one correspondence with the set of reduced words, and we will always represent them in this form. The unit element is the empty word, denoted by Λ. $|w|$ is the length of the word w.

The notation $\eqcirc$ stands for *graphical* (or *letter-for-letter*) equality: for $u_1, \ldots, u_r, v_1, \ldots, v_s \in F_m$, $u_1 u_2 \cdots u_r \eqcirc v_1 v_2 \cdots v_s$ by definition means that $u_1 u_2 \cdots u_r = v_1 v_2 \cdots v_s$ in F_m *and* both words $u_1 u_2 \cdots u_r$, $v_1 v_2 \cdots v_s$ are reduced. In the opposite direction, $u = vw$ in F_m if and only if there exist (uniquely defined) $v', c, w' \in F_m$ such that $v \eqcirc v'c$, $w \eqcirc c^{-1}w'$ and $u \eqcirc v'w'$. In this case we say that the word c is the *cancellation* (or *gets canceled*) in the product vw. If $c = v$ [$c = w^{-1}$], then we say that v [w, respectively] *gets completely canceled* in this product. And if $c = \Lambda$, then we say that *there is no cancellation* in vw, or that vw is *reduced*.

A word v is a *sub-word* of u, denoted $v \subseteq u$, if there exist words L, R such that $u \eqcirc LvR$. Any such representation is called an *occurrence* of v into u, and L, R are called *wings* of this occurrence. If $L = \Lambda$ [$R = \Lambda$], then we say that u *begins with* v, or that v *is a prefix of* u [u *ends with* v/v *is a suffix of* u, respectively]. A prefix or a suffix v of u is *proper* if $v \neq u$. We let $a \leq b$ denote that a is a prefix of b. This is a partial ordering on the set of all reduced words called the *prefix order*. Let $a \leq^* b$ be the dual *suffix order*.

A reduced word w is *cyclically reduced* if w^2 (and, hence, also all higher powers w^s) is reduced. Two cyclically reduced words u, v are *cyclic shifts* of each other, denoted $u \sim v$, if for some w_1, w_2, we have

$$u \eqcirc w_1 w_2, \ v \eqcirc w_2 w_1. \tag{2}$$

Thus, $u \sim v$ if and only if cyclically reduced words u, v are conjugated (in the ordinary sense) in F_m, and $\sim$ is an equivalence relation on the set of cyclically reduced words. A *cyclic word* is an equivalence class of this relation. That is, a cyclic word is a cyclically reduced word considered up to cyclic shifts. Cyclic words are in one-to-one correspondence with conjugacy classes of F_m. $u \sim v$ implies $|u| = |v|$, therefore the length of a cyclic word is well defined.

A cyclically reduced word w is *simple* if it cannot be represented in the form $w \eqcirc v^s$, $s > 1$. (Thus, simple words are nonempty.) Simple (cyclically reduced) words will be also called *periods*[3] and will be denoted by capital letters P, Q. If P is a period, u is a cyclically reduced word and $P \sim u$, then u is a period, too. Different cyclic shifts of a period are also different as words. That is, if in (2) u (and, hence, also v) is a period and both w_1, w_2 are nonempty, then $u \neq v$. Cyclic words consisting of periods will be called *cyclic periods* and

[3]This is a slight deviation from the notation of [1] where periods are not required to be simple.

denoted by the letters $\mathfrak{p}, \mathfrak{q}$. Thus, cyclic periods are periods considered up to cyclic shifts. It is worth noting that if we further identify $\mathfrak{p}$ with $\mathfrak{p}^{-1}$, then these will be in one-to-one correspondence with maximal cyclic subgroups of F_m considered up to conjugacy.

Let P be a period. A reduced word u is *P-periodic* if $u \subseteq P^s$ for some $s > 0$ and $|u| \geq 2|P|$. We denote by $\mathrm{Per}(P)$ the set of all P-periodic words. u is *periodic* if it is P-periodic for some period P and *aperiodic* otherwise.

Clearly, u is P-periodic if and only if it is representable in the form Q^sQ', where $Q \sim P,\ s \geq 2$ and Q' is a proper prefix of Q. (We will see soon that such a representation is unique.) In particular, if $P \sim Q$ and u is P-periodic, then it is also Q-periodic. Therefore, for every *cyclic* period $\mathfrak{p}$, we have the well-defined notion $\mathrm{Per}(\mathfrak{p})$ of *$\mathfrak{p}$-periodic words.*

In order to go any further, we need the following simple but very fundamental *Overlapping Lemmas* (see, e.g., [1, §1.2]):

LEMMA 2.1 (First Overlapping Lemma). *Let P, Q be two periods, and let u, v, w be reduced words such that*

$$uv = P'P^s,\ vw = Q^tQ', \tag{3}$$

where $s, t \geq 0$, P' is a proper suffix of P and Q' is a proper prefix of Q. Assume further that

$$|v| \geq |P| + |Q|.$$

Then $P \sim Q$. Moreover, the two representations (3) *are* **compatible in phase** *in the following sense: if $v = P''P^{s'}$, where P'' is a (possibly another) suffix of P, $P = P^{(3)}P''$, then $Q = P''P^{(3)}$.*

For the sake of completeness, we include a sketch of its proof. It is based on the following description of commuting elements in a free semi-group:

LEMMA 2.2. *If u and v are (reduced) words such that $uv = vu$, then there exists another reduced word w such that $u = w^k,\ v = w^\ell$ for some integers $k, \ell > 0$.*

Proof. By induction on $|u| + |v|$. If $u = \Lambda$ or $v = \Lambda$, the statement is obvious. Otherwise, assume without loss of generality that $|u| \geq |v|$. Then $uv = vu$ implies that v is a prefix of u; that is, $u = vu'$ for some reduced u'. Thus, $vu'v = v^2u'$, which implies $u'v = vu'$. Now we apply the inductive assumption to the pair (u', v). $\square$

Proof of Lemma 2.1. As suggested by the second part of the statement, let $R \sim P$ be a period such that $v = R^{s'}P''$ for some prefix P'' of R. From (3) we also know that $v = Q^{t'}Q''$, where Q'' is a prefix of Q. In this terminology, the first part of Lemma 2.1 claims that $R \sim Q$ while the second part refines this by stating that in fact $R = Q$.

Assume without loss of generality that $|R| \geq |Q|$. Then (since R is a prefix of v) $R \doteq Q^h Q^{(3)}$, where $h \geq 1$ and $Q \doteq Q^{(3)} Q^{(4)}$. Since $|v| \geq |R| + |Q|$, the first representation $v \doteq R^{s'} P''$ implies that v begins with $RQ^{(3)}Q^{(4)}$ while $v \doteq Q^{t'}Q''$ implies that v begins with $RQ^{(4)}Q^{(3)}$. Hence $Q^{(3)}Q^{(4)} \doteq Q^{(4)}Q^{(3)}$, and by Lemma 2.2 we have $Q^{(3)} \doteq w^k$, $Q^{(4)} \doteq w^\ell$ for some reduced word w and $k, \ell \geq 0$. Since Q is simple, we conclude that actually $w \doteq Q$, and since now we know that $R \doteq Q^h Q^{(3)} \doteq Q^{h+k}$ and R is also simple, we infer that $R \doteq Q$. □

Applying Lemma 2.1 in the case when the wings u, w are empty, we find that $\mathrm{Per}(\mathfrak{p}) \cap \mathrm{Per}(\mathfrak{q}) = \emptyset$ for any two different cyclic periods $\mathfrak{p}, \mathfrak{q}$.[4] The *left period* of $u \in \mathrm{Per}(\mathfrak{p})$ is defined as that particular $P \in \mathfrak{p}$ for which $u \doteq P^s P'$ $(s \geq 2)$, and *right periods* are defined symmetrically. Then the second part of Lemma 2.1 implies that left and right periods of periodic words are *uniquely defined*. Also, if we know left and right periods of $u \in \mathrm{Per}(\mathfrak{p})$, and also know $|u|$ within additive error $C \cdot |\mathfrak{p}|$, then u itself is completely determined up to $(2C+1)$ possibilities. (This simple remark will play a crucial role in Section 5.)

Let us note another important implication of Lemma 2.1 that we will be using extensively (and often implicitly). An occurrence $u \doteq LvR$ of a $\mathfrak{p}$-periodic word is *maximal* if there does not exist any strictly larger occurrence[5] $u \doteq L'v'R'$ of a $\mathfrak{p}$-periodic word into the same word u.

LEMMA 2.3. *Let $\mathfrak{p}$ be a period. Then two different maximal occurrences of $\mathfrak{p}$-periodic words into the same word intersect in a word of length $< 2|\mathfrak{p}|$. In particular, every occurrence of a $\mathfrak{p}$-periodic word has a* **unique** *extension to a maximal occurrence of a $\mathfrak{p}$-periodic word into the same word.*

Proof. Assume the contrary. Then by Lemma 2.1 the union of these two occurrences would also be a $\mathfrak{p}$-periodic word, in contradiction to the assumption of maximality. □

The First Overlapping Lemma basically says that occurrences of sufficiently periodic words cannot overlap "accidentally," and this is what one needs for the problems where the periodical structure is given to us *a priori* (which is the case, e.g., for the Burnside problem). On the contrary, the Second Overlapping Lemma tells us how to *extract* such structure from any two occurrences of an *arbitrary* word, provided they are close enough. This lemma lies in the heart of the research on equations in free groups cited in the introduction.

[4]Note that $|v| \geq 2|P|$ and $|v| \geq 2|Q|$ imply $|v| \geq |P| + |Q|$.

[5]This means $|L'| \leq |L|$, $|R'| \leq |R|$ and at least one of these inequalities is strict.

Lemma 2.4 (Second Overlapping Lemma). *Let $u \doteq LvR,\ u \doteq L'vR'$ be two different occurrences of the same word v into u. Assume that*

$$||L'| - |L|| \leq \frac{1}{3}|v|. \tag{4}$$

Then $v \in \mathrm{Per}(\mathfrak{p})$ for some cyclic period $\mathfrak{p}$ and, moreover, these two occurrences of v into u have the same maximal $\mathfrak{p}$-periodic extension.

Proof (*sketch*). Assume without loss of generality that $|L'| \geq |L|$ and let, say, $L' \doteq LP^h$ for some period P. Applying the same inductive process as in the proof of Lemma 2.2 and condition (4), we see that $v \doteq P^s P'$, where $s \geq 3h$ and P' is a prefix of P, which already implies the first part of the lemma. The second part follows from Lemma 2.3 as the intersection $P^{s-h}P'$ of the two occurrences of v into u has length $\geq 2|P|$. $\square$

If G is a group and $A_1, \ldots, A_k \subseteq G$, then

$$A_1 \cdot \ldots \cdot A_k \stackrel{\text{def}}{=} \{\, b \in G \,|\, (\exists (a_1, \ldots, a_k) \in A_1 \times \cdots \times A_k)(b = a_1 a_2 \cdots a_k) \,\} \,.$$

Throughout the paper we use the asymptotic notation $O, \Omega, \widetilde{O}, \widetilde{\Omega}$ quite customary in Combinatorics and Theoretical Computer Science. Thus,[6] $f \leq O(g)$ $[f \geq \Omega(g)]$ means "there exists an absolute constant $C > 0$ $[\varepsilon > 0]$ such that $f \leq Cg$ $[f \geq C\varepsilon$, respectively$]$ for all possible values of parameters assumed in f, g explicitly or implicitly." Its "soft" version $f \leq \widetilde{O}(g)$ and $f \geq \widetilde{\Omega}(g)$ can be used when all parameters $n_1, \ldots, n_t$ to f, g are integer and given explicitly (or, at least, are clear from the context). $f(n_1, \ldots, n_t) \leq \widetilde{O}(g(n_1, \ldots, n_t))$ $[f(n_1, \ldots, n_t) \geq \widetilde{\Omega}(g(n_1, \ldots, n_t))]$ means there exist absolute constants $C, k > 0$ $[\varepsilon, k > 0]$ such that $\forall n_1, \ldots, n_t (f(n_1, \ldots, n_t) \leq C \cdot \log^k(n_1 + \cdots + n_t) g(n_1, \ldots, n_t))$ $[\forall n_1, \ldots, n_t (f(n_1, \ldots, n_t) \geq \varepsilon g(n_1, \ldots, n_t) / \log^k(n_1 + \cdots + n_t))$, respectively$]$. Thus, in this notation our main result looks as follows:

Theorem 2.5. *Let $A \subseteq F_m$ be a finite subset of the free group F_m with at least two noncommuting elements. Then $|A \cdot A \cdot A| \geq \widetilde{\Omega}(|A|^2)$.*

Remark 1. In one place of our proof (namely, Lemma 3.5) constants assumed in the asymptotic notation do become dependent on the number of generators m. But this dependence can be eliminated by considering any fixed embedding $\phi : F_m \longrightarrow F_2$ and applying Theorem 2.5 to $\phi(A)$ (instead of applying it to the original $A \subseteq F_m$).

In fact, our main Lemma 3.2 readily implies a more general result. Recall that a group G is *virtually free* [*virtually cyclic*] if it contains a free [cyclic, respectively] subgroup of finite index.

[6]Most people would have used the equality sign here, but we find the combination of this notation with $\leq, \geq$ particularly expressive and instructive.

Theorem 2.6. *Let G be any fixed virtually free group, and let $A \subseteq G$ be its finite subset such that the subgroup generated by A is* **not** *virtually cyclic. Then* $|A \cdot A \cdot A| \geq \widetilde{\Omega}(|A|^2)$.

In particular, it is well known that the modular group $\mathrm{PSL}_2(\mathbb{Z}) \approx \mathbb{Z}_2 * \mathbb{Z}_3$ is virtually free (e.g., because its commutant is torsion-free, therefore it is a free subgroup (of index 6) by the Kurosh subgroup theorem [19, Th. IV.1.10]). The same is true for $\mathrm{SL}_2(\mathbb{Z})$ (every *free* subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ can be lifted to $\mathrm{SL}_2(\mathbb{Z})$), as well as for $\mathrm{GL}_2(\mathbb{Z})$. Therefore, Theorem 2.6 improves upon [7, Th. 5.1] (which, under the same assumptions, stated the bound $|A \cdot A \cdot A| \geq |A|^{1+\delta}$ for $\mathrm{SL}_2(\mathbb{Z})$ and for an unspecified constant $\delta > 0$).

3. Reduction: combinatorial part

This and the next two sections are entirely devoted to the proof of Theorems 2.5, 2.6. Our overall strategy is to analyze a potential counterexample by exhibiting in it "sufficiently large" subsets with "sufficiently rich" structure. (Accordingly, most of the proof is written in the distinct "top-down" style.) As stated, Theorem 2.5 turns out to be very inconvenient for this purpose. Our first task is to replace it with a stronger (and much clumsier) statement specifically designed with several types of reduction in mind.

Definition 3.1. For a finite subset $A \subseteq F_m$, $\Delta(A)$ is the maximal possible size of the intersection $A \cap C$, where C runs over all cosets of maximal cyclic subgroups[7] in F_m.

Note that Δ is monotone ($\Delta(A) \leq \Delta(B)$ if $A \subseteq B$) and invariant under left and right shifts ($\Delta(A) = \Delta(uA) = \Delta(Au)$).

Lemma 3.2 (Main Lemma). *Let $A, B, C \subseteq F_m$ be finite subsets, and assume that*

$$|A|, |C| \leq O(|B|).$$

Then one of the following two is true:

(a) $|A \cdot B \cdot C| \geq \widetilde{\Omega}(|A| \cdot |C|)$,
(b) $\Delta(B) \geq \Omega(|B|)$.

For the benefit of the reader who may feel uncomfortable with that much of asymptotic notation, we provide a translation of this statement to the ε/δ-language. (In all analogous places below, the translation is quite similar.)

[7] Since this class of subgroups is invariant under conjugacy, it does not matter whether we consider left or right cosets in this definition.

LEMMA 3.3 (Main Lemma, ε/δ-version). *For every* $D > 0$, *there exist* $\varepsilon, K > 0$ *such that the following is true*: *For all finite* $A, B, C \subseteq F_m$ *with* $|A|, |C| \leq D \cdot |B|$, *either* $|A \cdot B \cdot C| \geq \varepsilon \cdot \frac{|A| \cdot |C|}{\log^K(|A|+|B|+|C|)}$ *or* $\Delta(B) \geq \varepsilon \cdot |B|$ *holds.*

Proof of Theorems 2.5, 2.6 *from Lemma* 3.2. Since every virtually cyclic subgroup of a free group is cyclic, Theorem 2.6 implies Theorem 2.5, and we only have to prove the former.

Let G be a virtually free group, and let $F \leq G$ be a free subgroup of finite index; without loss of generality, we can assume that F is normal. Let $A \subseteq G$ be finite; represent it as $A = \dot{\bigcup}_{u \in U} (uA_u)$, where U is an arbitrary set of representatives for cosets of F and $A_u \subseteq F$. Choose that $u \in U$ for which $|A_u|$ is maximal (thus, $|A_u| \geq \Omega(|A|)$), and note that $(uA_u)(uA_u)(uA_u) = u^2(u^{-1}A_u u)A_u(uA_u u^{-1})u$. We apply Lemma 3.2 with $A := u^{-1}A_u u$, $B := A_u$, $C := uA_u u^{-1}$. If conclusion (a) holds, we are done. If $\Delta(A_u) \geq \Omega(|A_u|) \geq \Omega(|A|)$, there exists a maximal cyclic subgroup $C \leq F$ and $v \in F$ such that $|A_u \cap (vC)| \geq \Omega(|A|)$. Denoting $w = uv$, we conclude that $|A \cap (wC)| \geq \Omega(|A|)$. Let $N \leq G$ be the normalizer of C.

If $w \notin N$, we are done: since C and (wCw^{-1}) are *different* maximal cyclic subgroups in F, they have empty intersection. Therefore, all products $c_1 c_2$ $(c_1, c_2 \in (wC))$ are pairwise distinct and $|A \cdot A \cdot A| \geq |A \cdot A| \geq |A \cap (wC)|^2 \geq \Omega(|A|^2)$.

Assume $w \in N$. Since $N \cap F = C$, C has a finite index in N and, therefore, N is virtually cyclic. Since A does not generate a virtually cyclic subgroup, $A \not\subseteq N$; fix arbitrarily $a \in A \setminus N$. Now we are done by the same argument as above, applied to the product $(wC)a(wC)$. □

Remark 2. The statement of Lemma 3.2 allows the following three types of reductions that we are going to use:

- Let $u, v \in F_m$, $A_0 \stackrel{\text{def}}{=} Au^{-1}$, $B_0 \stackrel{\text{def}}{=} uBv$ and $C_0 \stackrel{\text{def}}{=} v^{-1}C$. Then the validity of Lemma 3.2 for the triple (A_0, B_0, C_0) implies its validity for the original (A, B, C).
- The same conclusion holds if $A_0 \subseteq A$, $B_0 \subseteq B, C_0 \subseteq C$ are arbitrary subsets with the only restriction $|A_0| \geq \widetilde{\Omega}(|A|)$, $|B_0| \geq \Omega(|B|)$, $|C_0| \geq \widetilde{\Omega}(|C|)$.
- Assume that $A = A_1 \stackrel{.}{\cup} \cdots \stackrel{.}{\cup} A_{\ell_A}$ and $C = C_1 \stackrel{.}{\cup} \cdots \stackrel{.}{\cup} C_{\ell_C}$ are decompositions of A and C into disjoint unions of subsets, and further assume that all $\ell_A \ell_C$ sets $A_i B C_j$ $(i \in [\ell_A],\ j \in [\ell_C])$ are pairwise disjoint. Then the validity of Lemma 3.2 for all triples (A_i, B, C_j) implies its validity for (A, B, C).

In the reduction of the last type we of course require the *uniform* dependence of assumed constants (that is, ε, K on D in the notation of Lemma 3.3).

After this preparatory work, we begin the real proof by getting rid of cancellations.

LEMMA 3.4. *For any finite $A \subseteq F_m$, there exists $u \in F_m$ such that for any letter $y \in \{x_1, x_1^{-1}, \dots, x_m, x_m^{-1}\}$, at least $\frac{1}{4m}|A|$ words in Au^{-1} do* **not** *end with y.*

Proof. Let us call $u \in F_m$ *populated* if it is a suffix of at least $\frac{1}{4|m|}|A|$ words in A. Λ is populated whereas sufficiently long words are not. Choose the longest populated word u; we claim that it has the required property.

Indeed, every one of the words yu ($y \in \{x_1, x_1^{-1}, \dots, x_m, x_m^{-1}\}$, u does not begin with y^{-1}) is not populated and therefore may appear as a suffix in $\leq \frac{1}{4m}|A|$ words from A. Hence u is a suffix of at most $\frac{1}{2}|A|$ words in A. (On the other hand, it is a suffix of at least $\frac{1}{4m}|A|$ words since u itself is populated.) It only remains to note that if u is not a suffix of $a \in A$, then au^{-1} ends with the same letter as u^{-1}, and if it is its suffix, then au^{-1} ends with a different letter, unless it is empty. □

LEMMA 3.5. *For any finite $A, B, C \subseteq F_m$ with $|B| \geq 2$, there exist $u, v \in F_m$ and $A_0 \subseteq Au^{-1}$, $B_0 \subseteq uBv$, $C_0 \subseteq v^{-1}C$ such that $|A_0| \geq \Omega(|A|)$, $|B_0| \geq \Omega(|B|)$, $|C_0| \geq \Omega(|C|)$ and all products abc ($a \in A_0$, $b \in B_0$, $c \in C_0$) are reduced.*

Proof. Apply Lemma 3.4 to A, and apply its dual version to C; let u, v be the resulting elements. Removing from uBv the empty word (if it is there), we find a subset $B_0 \subseteq uBv$ with $|B_0| \geq \frac{1}{4m^2}(|B| - 1)$ such that all words in B_0 begin with the same letter y and end with the same letter z. Finally, let $A_0 \subseteq Au^{-1}$ consist of all those words that do not end with y^{-1}, and similarly for $C_0 \subseteq v^{-1}C$. Since $|A_0| \geq \Omega(|A|)$ and $|C_0| \geq \Omega(|C|)$ hold by Lemma 3.4, this completes the proof. □

From this point on, cancellations will never appear again, and the reader may freely assume that we are working in a free *semi-group*. Note that if $abc = a'b'c'$ is a collision in the product $A \cdot B \cdot C$, then a, a' are comparable in the prefix order and c, c' are comparable in the suffix order. This suggests that the most difficult case should be when the elements of A form a *prefix chain* (defined as a set of words mutually comparable in the prefix order), and C forms a *suffix chain*. The following lemma makes this intuition precise.

Definition 3.6. Two prefix [suffix] chains A_1, A_2 are *incomparable* if any two $a_1 \in A_1$, $a_2 \in A_2$ are incomparable in the prefix [suffix, respectively] order.

In particular, incomparable prefix/suffix chains are necessarily disjoint. Also, two prefix chains A_1, A_2 are incomparable if and only if their *minimal* elements are incomparable.

LEMMA 3.7. *Every finite set of words A contains a collection $A_1, \ldots, A_\ell \subseteq A$ of mutually incomparable prefix chains such that*

$$|A_1 \cup \cdots \cup A_\ell| = \sum_{i=1}^{\ell} |A_i| \geq \widetilde{\Omega}(|A|), \tag{5}$$

and a similar statement holds for suffix chains.

Proof. Consider the restriction of the prefix order $\leq$ onto A. For $a \in A$, let $h(a)$ be its *height* defined as the maximal possible length of a prefix chain having a as its minimal element (and entirely contained in A). All elements of the same height h are mutually incomparable; let ℓ_h be their number. Then

$$|A| = \sum_{h=1}^{|A|} \ell_h,$$

and also for every h there exist ℓ_h mutually incomparable prefix chains of length h each. (For every element a of height h, include an arbitrarily chosen prefix chain of height h with the minimal element a.)

Thus, if t is the maximal possible value of $|A_1 \cup \cdots \cup A_\ell|$ in (5), then $t \geq h\ell_h$ for each h, which implies

$$|A| \leq t \cdot \sum_{h=1}^{|A|} \frac{1}{h} \leq O(t \log |A|)$$

and, therefore, $t \geq \widetilde{\Omega}(|A|)$. □

Now, by Lemma 3.5 we may assume in Lemma 3.2 that all products abc ($a \in A, b \in B, c \in C$) are reduced. By Lemma 3.7 we may also assume that A [C] can be decomposed as a union of mutually incomparable prefix [suffix, respectively] chains; say, $A = A_1 \dot\cup \cdots \dot\cup A_{\ell_A}$, $C = C_1 \dot\cup \cdots \dot\cup C_{\ell_C}$. But if $i \neq i' \in [\ell_A]$, then A_iBC and $A_{i'}BC$ are disjoint (since A_i and $A_{i'}$ are incomparable in the prefix order), and similarly for $j \neq j' \in [\ell_C]$. This means that we can apply the reduction of the third type from Remark 2.

Summarizing what we have achieved so far, *in Lemma* 3.2 *we can assume without loss of generality that all products abc* ($a \in A, b \in B, c \in C$) *are reduced and that, moreover, A is a prefix chain and C is a suffix chain.*

4. **Reduction: finding aperiodicity**

At this point we bring into the analysis periodic words, and the rest of the proof is split into two almost independent parts. Namely (thinking in terms of a hypothetical counterexample to Lemma 3.2), we want to show that

- if $|A \cdot B \cdot C|$ is small, there is enough "periodical structure" in A, B, C;
- if $\Delta(B)$ is small, then some large subsets A_0, B_0, C_0 display enough "aperiodicity" in them.

These two conclusions will contradict each other. Of these two, the first task is much more difficult, interesting and natural to start with. But for technical reasons, we have to begin with the second.

Definition 4.1. Let $a, b \in F_m$, and assume that the product ab is reduced. We say that ab is *left regular* if b is periodic, and a ends with P^2, where P is the left period of b (equivalently, $b \in \mathrm{Per}(\mathfrak{p})$ for some cyclic period $\mathfrak{p}$, and its maximal $\mathfrak{p}$-periodic extension in ab has length $\geq |b| + 2|\mathfrak{p}|$). In all other cases ab is *left singular*. *Right regular* and *right singular* products bc are defined by symmetry.

Definition 4.2. Let P be a period, and let $A \subseteq F_m$ be a finite set. We define $\Delta_{\ell,P}(A)$ as the maximal possible size of the intersection $A \cap C$, where C runs over all sets of the form

$$\left\{ LP^t \mid t \geq 0 \right\} \ (L \in F_m,\ LP \text{ reduced}).$$

$\Delta_{r,P}(A)$ is defined by symmetry.

Clearly, $\Delta_{\ell,P}(A), \Delta_{r,P}(A) \leq \Delta(A)$.

LEMMA 4.3. *Let $A, B, C \subseteq F_m$ be finite sets, and assume that all products abc ($a \in A,\ b \in B,\ c \in C$) are reduced. Then either*

$$\Delta(B) \geq \Omega(|B|), \tag{6}$$

or there exist $A_0 \subseteq A$, $B_0 \subseteq B$, $C_0 \subseteq C$ with $|A_0| \geq \Omega(|A|)$, $|B_0| \geq \Omega(|B|)$, $|C_0| \geq \Omega(|C|)$ such that at least one of the following three is true:

(a) *At least $\frac{1}{2}|A_0||B_0|$ products ab ($a \in A_0,\ b \in B_0$) are left singular.*
(b) *At least $\frac{1}{2}|B_0||C_0|$ products bc ($b \in B_0,\ c \in C_0$) are right singular.*
(c) *For every period P that is the left period of at least one periodic word in B_0, $\Delta_{\ell,P}(A_0) \leq O(1)$, and the dual conclusion holds for right periods.*

Proof. Either at least half of all words in B are aperiodic, or at least half of them are periodic. In the first case both (a) and (b) hold trivially. Removing from B all aperiodic words in the second case, we may assume without loss of generality that all words in B are periodic.

Consider now any individual cyclic period $\mathfrak{p}$ for which $B_{\mathfrak{p}} \stackrel{\text{def}}{=} B \cap \mathrm{Per}(\mathfrak{p})$ is nonempty. If there exists $P \in \mathfrak{p}$ that appears as either the left period in at least half of all words from $B_{\mathfrak{p}}$ or the right period in at least half of them, remove from $B_{\mathfrak{p}}$ all words violating this. Repeating this procedure once more if necessary, we will find $B'_{\mathfrak{p}} \subseteq B_{\mathfrak{p}}$ with $|B'_{\mathfrak{p}}| \geq \Omega(B_{\mathfrak{p}})$ and such that one of the following is true:

(a) Every period $P \in \mathfrak{p}$ appears as the left period in $\leq \frac{1}{2}|B'_{\mathfrak{p}}|$ words from $B'_{\mathfrak{p}}$.
(b) Every period $P \in \mathfrak{p}$ appears as the right period in $\leq \frac{1}{2}|B'_{\mathfrak{p}}|$ words from $B'_{\mathfrak{p}}$.
(c) All words in $B'_{\mathfrak{p}}$ have the same left and right periods.

Let $B' \stackrel{\text{def}}{=} \bigcup_{\mathfrak{p}} B'_{\mathfrak{p}}$. At the expense of decreasing $|B'|$ by at most a factor of three, we may assume that one and the same of these three alternatives holds for *every* cyclic period $\mathfrak{p}$ for which $B'_{\mathfrak{p}}$ is nonempty.

Alternatives (a) and (b) (along with Lemma 2.1) immediately apply the corresponding conclusions in the statement of Lemma 4.3 (with $A_0 := A,\ B_0 := B',\ C_0 := C$) since then in (say) case (a), for every $a \in A$ and every cyclic period $\mathfrak{p}$, there would be at most $\leq \frac{1}{2}|B'_{\mathfrak{p}}|$ words $b \in B'_{\mathfrak{p}}$ for which ab is left regular. So, we are left with the case when for every $\mathfrak{p}$, all words in $B'_{\mathfrak{p}}$ have the same left and right periods. Note that in this case $B'_{\mathfrak{p}}$ is a subset of the coset $\{\, P'P^tP'' \mid t \in \mathbb{Z} \,\}$ of a cyclic subgroup and, therefore,

$$|B'_{\mathfrak{p}}| \leq |\Delta(B)|. \tag{7}$$

Let $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_d\}$ be the enumeration of all cyclic periods $\mathfrak{p}$ for which $B'_{\mathfrak{p}} \neq \emptyset$ in the order of nondecreasing length:

$$|\mathfrak{p}_1| \leq |\mathfrak{p}_2| \leq \cdots \leq |\mathfrak{p}_d|. \tag{8}$$

Choose the minimal ℓ for which $\sum_{i=1}^{\ell} |B'_{\mathfrak{p}_i}| \geq \frac{1}{3}|B'|$ (therefore, $\sum_{i=1}^{\ell-1} |B'_{\mathfrak{p}_i}| < \frac{1}{3}|B'|$). If $\sum_{i=1}^{\ell} |B'_{\mathfrak{p}_i}| \geq \frac{2}{3}|B'|$, then $|B'_{\mathfrak{p}_\ell}| \geq \frac{1}{3}|B'|$, and hence (7) implies (6).

Otherwise, $\sum_{i=\ell+1}^{d} |B'_{\mathfrak{p}_i}| \geq \frac{1}{3}|B'|$, and we first try out the set $\bigcup_{i=\ell+1}^{d} B'_{\mathfrak{p}_i}$ as B_0. If at least $\frac{1}{2}|A||B_0|$ products ab ($a \in A,\ b \in B_0$) are left singular, or at least $\frac{1}{2}|B_0||C|$ products bc ($b \in B_0,\ c \in C$) are right singular, we are done.

Otherwise, there exist *fixed* $b_\ell, b_r \in \bigcup_{i=\ell+1}^{d} B'_{\mathfrak{p}_i}$ such that for at least half of all $a \in A$ the product ab_ℓ is left regular, and for at least half of all $c \in C$, $b_r c$ is right regular. We remove from A and C all elements violating these properties, and we let A_0, C_0 be the result of this removal. Set also

$$B_0 \stackrel{\text{def}}{=} \bigcup_{i=1}^{\ell} B_{\mathfrak{p}_i}.$$

We finally claim that A_0, B_0, C_0 satisfy alternative (c) in Lemma 4.3 and, by symmetry, it is sufficient to check this only on the left side.

Indeed, all words in A_0 end with Q^2, where Q is the left period of b_ℓ (and hence $Q \in \mathfrak{p}_j$ for some $j \geq \ell + 1$). If a period P appears as the left period of some word in B_0, then $P \in \mathfrak{p}_i$ for some $i \leq \ell$. In particular, $P \not\sim Q$ and, by (8),

$$|P| \leq |Q|. \tag{9}$$

According to Definition 4.2, consider any fixed word L such that LP is reduced. If $LP^t \in A_0$, then LP^t ends with Q^2. The word Q^2, however, is not $\mathfrak{p}_i$-periodic; therefore, due to (9), it cannot be a sub-word of P^s for any s, which means that P^t is a suffix of Q^2. Moreover, if $t \geq 2$, then the maximal $\mathfrak{p}_i$-periodic extension of P^t in LP^t is a proper suffix of Q^2 and, therefore, has the same length as its maximal $\mathfrak{p}_i$-periodic extension in Q^2. In particular, *this* *extension does not depend on* t. This implies that there can be at most one value $t \geq 2$ for which LP^t ends with Q^2, which shows that $\Delta_{\ell,P}(A_0) \leq 3$ and completes the proof of Lemma 4.3. $\square$

To summarize, so far we have reduced Lemma 3.2 to its partial case described as follows. (Alternative (b) in the statement of Lemma 3.2 has already been used up in (6), and we do not need to carry it any longer.)

LEMMA 4.4. *Let* $A, B, C \subseteq F_m$ *be finite sets such that*

$$|A|, |C| \leq O(|B|).$$

Assume that all products abc $(a \in A,\ b \in B,\ c \in C)$ *are reduced, that* A *is a prefix chain and that* C *is a suffix chain. Moreover, assume that one of the following three is true*:

(a) *At least* $\frac{1}{2}|A||B|$ *products* ab $(a \in A,\ b \in B)$ *are left singular.*
(b) *At least* $\frac{1}{2}|B||C|$ *products* bc $(b \in B,\ c \in C)$ *are right singular.*
(c) *For every period* P *that is the left period of at least one periodic word in* B, $\Delta_{\ell,P}(A) \leq O(1)$, *and the symmetric conclusion holds for the right periods.*

Then $|A \cdot B \cdot C| \geq \widetilde{\Omega}(|A| \cdot |C|)$.

5. Finding periodicity with collision numbers

In this section we prove Lemma 4.4, thereby completing the proof of our main result. Fix $A, B, C \subseteq F_m$ satisfying all the premises of Lemma 4.4. Define $T \subseteq A \times B \times C$ as follows. If one of the alternatives (a) or (b) holds, T consists of those triplets (a, b, c) for which *either* ab is left singular *or* bc is right singular. In the remaining case (c), we simply let $T := A \times B \times C$. Note that in any case

$$|T| \geq \Omega(|A| \cdot |B| \cdot |C|). \tag{10}$$

We define the *collision number* $\mathbf{c}_T(A, B, C)$ as

$$\mathbf{c}_T(A, B, C) \stackrel{\text{def}}{=} \left| \left\{ ((a, b, c), (a', b', c')) \in T^2 \mid abc \doteq a'b'c' \right\} \right|.$$

For $u \in F_m$, let

$$n(u) \stackrel{\text{def}}{=} |\{ (a, b, c) \in T \mid abc \doteq u \}|.$$

Then by Cauchy-Schwartz and (10),

$$\begin{aligned} \mathbf{c}_T(A, B, C) &= \sum_{u \in A \cdot B \cdot C} n(u)^2 \geq \frac{1}{|A \cdot B \cdot C|} \left(\sum_{u \in A \cdot B \cdot C} n(u) \right)^2 \\ &= \frac{|T|^2}{|A \cdot B \cdot C|} \geq \Omega\left(\frac{|A|^2 |B|^2 |C|^2}{|A \cdot B \cdot C|} \right). \end{aligned} \tag{11}$$

Thus, in order to complete our proof, we only have to show that

$$\mathbf{c}_T(A,B,C) \leq \widetilde{O}(|A||B|^2|C|). \tag{12}$$

Our next task is to set the stage for the Second Overlapping Lemma 2.4, and for this, we need one more reduction (this time in terms of collision numbers). But now the reduction is slightly more subtle than those based on Remark 2 seen in previous sections. For this reason, we prefer to change the gears, and we first *formulate* the statement we are reducing to.

LEMMA 5.1. *Let $A, B, C \subseteq F_m$ be finite sets such that*

$$|A|, |C| \leq O(|B|). \tag{13}$$

Assume that all products abc ($a \in A,\ b \in B,\ c \in C$) are reduced, that A is a prefix chain of even length and that C is a suffix chain of even length: $A = \{a_1, \ldots, a_{2n_A}\},\ C = \{c_1, \ldots, c_{2n_C}\}$, where $a_1 < a_2 < \cdots < a_{2n_A}$ and $c_1 <^* c_2 <^* \cdots <^* c_{2n_C}$. Let $T \subseteq A \times B \times C$ be such that either (A,B,C) satisfies property (c) *in the statement of Lemma* 4.4, *or for every $(a,b,c) \in T$, either ab is left singular or bc is right singular. Then*

$$\begin{aligned} |\{((a_i, b, c_j), (a_{i'}, b', c_{j'})) \in T^2 \mid\ & a_i b c_j = a_{i'} b' c_{j'}, \\ & |\{i, i'\} \cap \{1, 2, \ldots, n_A\}| = 1, \\ & |\{j, j'\}| \cap \{1, 2, \ldots, n_C\}| = 1\} \\ \leq O(|A||B|^2|C|). & \end{aligned} \tag{14}$$

Thus, the only difference in the conclusion from (12) is that we additionally require that the "middle" prefix a_{n_A} of a_{2n_A} separates i from i', and the same holds for j, j'.

Proof of (12) *from Lemma* 5.1. Let (A,B,C) satisfy the assumptions of Lemma 4.4, and let T be defined as in the beginning of this section. Assume for simplicity that $|A|$ and $|C|$ are powers of 2, and represent A and C as in the statement of Lemma 5.1: $A = \{a_1, \ldots, a_{n_A}\},\ C = \{c_1, \ldots, c_{n_C}\}$, where $a_1 < a_2 < \cdots < a_{n_A}$ and $c_1 <^* c_2 <^* \cdots <^* c_{n_C}$. For $d \leq \log_2 n_A,\ d^* \leq \log_2 n_C$ and integers α, γ, let

$$\begin{aligned} A^d_\alpha &\stackrel{\text{def}}{=} \left\{ a_i \in A \,\middle|\, \lfloor i/2^d \rfloor = \alpha \right\}, \\ C^{d^*}_\gamma &\stackrel{\text{def}}{=} \left\{ c_j \in C \,\middle|\, \lfloor j/2^{d^*} \rfloor = \gamma \right\}. \end{aligned}$$

For any fixed values of d, d^*, α, γ, we can apply Lemma 5.1 to the triple $(A^d_\alpha, B, C^{d^*}_\gamma)$ letting $T := T \cap (A^d_\alpha \times B \times C^{d^*}_\gamma)$. Summing up the right-hand sides of the resulting estimates (14), we get

$$O\left(\sum_{d,d^*}\sum_{\alpha,\gamma}|A^d_\alpha||B|^2|C^{d^*}_\gamma|\right) = O\left(\sum_{d,d^*}|A||B|^2|C|\right) \le \widetilde{O}(|A||B|^2|C|),$$

as d, d^* take on only logarithmically many values.

On the other hand, the sets in the left-hand sides of (14) give a partition of all those tuples $((a_i, b, c_j), (a_{i'}, b', c_{j'})) \in T^2$ for which $a_i b c_j = a_{i'} b' c_{j'}$ *and* $i \neq i'$, $j \neq j'$. Namely, such a tuple is counted in that $(A^d_\alpha, B, C^{d^*}_\gamma)$, where d is the most significant bit in which binary representations of i and i' differ, d^* is defined in the same way from j, j' and $\alpha = \lfloor i/2^d \rfloor (= \lfloor i'/2^d \rfloor)$, $\gamma = \lfloor j/2^{d^*} \rfloor$.

Since there are at most $2|A||B|^2|C|$ tuples $((a_i, b, c_j), (a_{i'}, b', c_{j'}))$ with $a_i b c_j = a_{i'} b' c_{j'}$ for which either $i = i'$ or $j = j'$, we are done. □

Now we prove Lemma 5.1, and at this point we have to break the symmetry by assuming (without loss of generality) that

$$|C| \le |A|. \tag{15}$$

For two words a, a' comparable in the prefix order, we let $\delta(a, a')$ denote their difference. (That is, $a = a'\delta(a, a')$ or $a' = a\delta(a, a')$, depending on which of the two is longer.) Let $\mathcal{P}$ be the set of all those cyclic periods $\mathfrak{p}$ for which there exists an occurrence

$$a_{2n_A} = L_\mathfrak{p} u_\mathfrak{p} R_\mathfrak{p} \tag{16}$$

of a $\mathfrak{p}$-periodical word $u_\mathfrak{p}$ in a_{2n_A} that is "nontrivially cut" by a_{n_A} in the following sense:

$$a_{n_A} = L_\mathfrak{p} v,\ v \text{ is a prefix of } u_\mathfrak{p} \text{ with } |v| \ge 2|\mathfrak{p}| \text{ and } |\delta(v, u_\mathfrak{p})| \ge 2|\mathfrak{p}|. \tag{17}$$

It follows from Lemma 2.3 that for any fixed $\mathfrak{p}$, maximal $\mathfrak{p}$-periodic extensions of all such occurrences coincide, and we choose (16) to be this maximal (and uniquely defined) occurrence.

Next, let $A_\mathfrak{p}$ be the set of all $a_i \in A$ for which we, like in (17), still have $a_i = L_\mathfrak{p} v$, where v is a prefix of $u_\mathfrak{p}$ with $|v| \ge 2|\mathfrak{p}|$ and $|\delta(v, u_\mathfrak{p})| \ge 2|\mathfrak{p}|$, but now *we also additionally require that* $|\delta(a_i, a_{n_A})| \ge 2|\mathfrak{p}|$. This new condition implies, in particular, that $a_{n_A} \notin A_\mathfrak{p}$. In fact, it implies that for $a_i \in A_\mathfrak{p}$, the word $\delta(a_i, a_{n_A})$ is $\mathfrak{p}$-periodic; therefore, $A_\mathfrak{p} \cap A_\mathfrak{q} = \emptyset$ *for every two different cyclic periods* $\mathfrak{p}, \mathfrak{q}$.

After this setup, we begin proving the bound (14). First we drop from circulation the condition $|\{j, j'\} \cap \{1, 2, \ldots, n_C\}| = 1$ and simplify the dual one by insisting that $i \le n_A < i'$. That is, we will prove (14) in the form

$$
\begin{aligned}
(18) \qquad & |\{((a_i, b, c_j), (a_{i'}, b', c_{j'})) \in T^2 \mid a_i b c_j = a_{i'} b' c_{j'}, \\
& \qquad i \le n_A,\ i' \ge n_A + 1\}| \\
& \le O(|A||B|^2|C|).
\end{aligned}
$$

We do it by case analysis according to the structural properties of a tuple $((a_i, b, c_j), (a_{i'}, b', c_{j'}))$ contributing to the left-hand side. In every of the four cases our strategy will be the same: we will show that four out of six elements of the tuple $((a_i, b, c_j), (a_{i'}, b', c_{j'}))$ already determine it up to $O(1)$ possibilities. But the exact choice of these four entries will depend on the case.

Case 1: *There is no cyclic period* $\mathfrak{p}$ *such that* $\{a_i, a_{i'}\} \subseteq A_{\mathfrak{p}}$.

Let us call such pairs $(a_i, a_{i'})$ *singular*. First we claim that every fixed $d \in F_m$ can be realized in the form $\delta(a_i, a_{i'})$ for at most 12 singular pairs $(a_i, a_{i'})$. Indeed, any such realization $a_{i'} = a_i d$ defines the occurrence $a_{2n_A} = a_i d \delta(a_{i'}, a_{2n_A})$ of d into a_{2n_A} and, moreover, $|a_{n_A}| - |d| \le |a_i| \le |a_{n_A}|$. Suppose for the sake of contradiction that d possesses ≥ 13 realizations. Then, by the pigeon-hole principle, we could choose five of them $d = \delta(a_{i_1}, a_{i'_1}) = \cdots = \delta(a_{i_5}, a_{i'_5})$ $(i_1 \le \cdots \le i_5)$ such that $||a_{i_\alpha}| - |a_{i_\beta}|| \le |d|/3$ for all $\alpha, \beta \in [5]$. Therefore, we could apply Lemma 2.4 and conclude that $d \in \mathrm{Per}(\mathfrak{p})$ for some cyclic period $\mathfrak{p}$ and, moreover, all five selected occurrences of d into a_{2n_A} would be contained in the same maximal occurrence of a $\mathfrak{p}$-periodic word in a_{2n_A}. Further, they would be compatible in phase (in the sense of Lemma 2.1); that is, all $||a_{i_\alpha}| - |a_{i_\beta}||$ would be multiples of $|\mathfrak{p}|$. This would readily imply that this maximal occurrence would necessarily be the occurrence (16) and that $\{a_{i_3}, a_{i'_3}\} \subseteq A_{\mathfrak{p}}$, a contradiction.

Now we only have to observe that $a_i b c_j = a_{i'} b' c_{j'}$ implies $\delta(a_i, a_{i'}) = \delta(b c_j, b' c_{j'})$; that is, $b, c_j, b', c_{j'}$ determine $\delta(a_i, a_{i'})$. Therefore, they also determine $a_i, a_{i'}$ up to ≤ 12 possibilities, and hence the contribution of Case 1 to (18) is estimated as $O(|B|^2|C|^2)$, which is $O(|A||B|^2|C|)$ by (15).

Case 2: $\{a_i, a_{i'}\} \subseteq A_{\mathfrak{p}}$ *for some cyclic period* $\mathfrak{p}$ *and* $|b| \le 2|\mathfrak{p}|$.

In this case we claim that the tuple can be retrieved (again, up to $O(1)$ possibilities) from $a_i, c_j, a_{i'}, b'$. Indeed, since $a_i, a_{i'} \in A_{\mathfrak{p}}$, we have $|\delta(a_i, a_{i'})| = |\delta(a_i, a_{n_A})| + |\delta(a_{i'}, a_{n_A})| \ge 4|\mathfrak{p}|$. This implies that $|a_{i'}| - |a_i b| \ge 2|\mathfrak{p}|$ and hence $\delta(a_i b, a_{i'})$ is $\mathfrak{p}$-periodic. Its left period is completely determined by c_j (as $\delta(a_i b, a_{i'}) \le c_j$), and its right period is determined by $a_{i'}$ (as $\delta(a_i b, a_{i'}) \le^* a_{i'}$). Finally, since $|b| \le 2|\mathfrak{p}|$, we can estimate its length as $|a_{i'}| - |a_i| - 2|\mathfrak{p}| \le |\delta(a_i b, a_{i'})| \le |a_{i'}| - |a_i|$. Thus, given $a_i, c_j, a_{i'}$, there are at most three possibilities for $\delta(a_i b, a_{i'})$, and once we know it, we also know b and then $c'_j = \delta(a_i b c_j, a_{i'} b')$. Thus, Case 2 contributes at most $O(|A|^2|B||C|)$, which is $O(|A||B|^2|C|)$ by (13).

Case 3: $\{a_i, a_{i'}\} \subseteq A_{\mathfrak{p}}$ *for some cyclic period* $\mathfrak{p}$, $|b| \geq 2|\mathfrak{p}|$ *but either* $b \notin \mathrm{Per}(\mathfrak{p})$ *or* $b \in \mathrm{Per}(\mathfrak{p})$ *and the product* bc_j *is right singular.*

This time the tuple is determined by $b, a_{i'}, b', c_{j'}$ (as always, up to $O(1)$ possibilities). Indeed, from these four entries we know $u = a_i' b' c_{j'} = a_i b c_j$, as well as the occurrence

$$u = a_{n_A}\delta(a_{n_A}, a_{i'})(b'c_{j'}) \tag{19}$$

of the $\mathfrak{p}$-periodic word $\delta(a_{n_A}, a_{i'})$ into it. The prefix v of b of length $2|\mathfrak{p}|$ is a prefix of $\delta(a_i, a_{i'})$ and thus $\mathfrak{p}$-periodic; let $b = vw$ and $R \stackrel{\text{def}}{=} wc_j$. Now consider its (yet unknown!) occurrence

$$u = a_i v R \tag{20}$$

into u. These two occurrences of $\mathfrak{p}$-periodic words into u possess a common (also unknown) $\mathfrak{p}$-periodic extension $u = a_i\delta(a_i, a_{i'})(b'c_{j'})$. Therefore, by Lemma 2.2 the *maximal* $\mathfrak{p}$-periodic extension $u = \tilde{a}_i\hat{v}R'$ of (20) is the same as the maximal $\mathfrak{p}$-periodic extension of the known occurrence (19) and hence *is also determined by* $(a_{i'}, b', c_{j'})$. Further, if $\hat{v}_1$ is the maximal $\mathfrak{p}$-periodic extension of the prefix v in the word bc_j, then it should have the same "right wing" R': $bc_j = \hat{v}_1 R'$. The assumptions of Case 3 imply that $|\hat{v}_1|$ (and hence also $|c_j|$ since b and R' are already known) is determined within accuracy $2|\mathfrak{p}|$ *by the word* b *only*. Namely, it cannot exceed by more than $2|\mathfrak{p}|$ the length of the maximal $\mathfrak{p}$-periodic extension of v in b. Therefore, $|a_i|$ and then $\delta(a_i, a_{n_A})$ are also determined within that accuracy. But the left and right periods of the latter words are known (it is a suffix of a_{n_A} and has v as its prefix), hence this word (and then a_i) is determined up to $O(1)$ possibilities.

Case 4: $\{a_i, a_{i'}\} \subseteq A_{\mathfrak{p}}$ *for some cyclic period* $\mathfrak{p}$, $b \in \mathrm{Per}(\mathfrak{p})$ *and the product* bc_j *is right regular.*

In this final case we also claim that the information can be retrieved from $b, a_{i'}, b', c_{j'}$ (but for entirely different reasons). Namely, recalling the definition (16), the word $\delta(L_{\mathfrak{p}}, a_i)b$ is $\mathfrak{p}$-periodic and $|\delta(L_{\mathfrak{p}}, a_i)| \geq 2|\mathfrak{p}|$. Hence, the product $a_i b$ is left regular. Since $(a_i, b, c_j) \in T$, this implies (recall the statement of Lemma 5.1) that (A, B, C) must necessarily satisfy property (c) in the statement of Lemma 4.4. In particular, $\Delta_{\ell,P}(A) \leq O(1)$, where P is the left period of b. Let $L'_{\mathfrak{p}}$ be the prefix of $L_{\mathfrak{p}}u_{\mathfrak{p}}$ in (16) with $|L_{\mathfrak{p}}| \leq |L'_{\mathfrak{p}}| \leq |L_{\mathfrak{p}}| + |\mathfrak{p}|$ and such that the left period of $\delta(L'_{\mathfrak{p}}, L_{\mathfrak{p}}u_{\mathfrak{p}})$ is equal to P. Then a_i must necessarily have the form $L'_{\mathfrak{p}}P^t$ for some integer t. Now the condition $\Delta_{\ell,P}(A) \leq O(1)$ again pinpoints it down to $O(1)$ possibilities.

We have shown that every one of four logically possible cases contributes at most $O(|A||B|^2|C|)$ to the left-hand side of (18). This completes the proof of Lemma 5.1, (12), Lemmas 4.4, 3.2 and Theorems 2.5, 2.6.

6. Statistical version of Plünnecke-Ruzsa inequalities

In this section G will be an abelian group. For its finite subsets $A_1, \ldots, A_k$, define the *collision number* $\mathbf{c}(A_1, \ldots, A_k)$ as

$$\mathbf{c}(A_1, \ldots, A_k) \stackrel{\text{def}}{=} |\{((a_1, \ldots, a_k), (a_1', \ldots, a_k')) \in (A_1 \times \cdots \times A_k)^2 | \\ a_1 + \cdots + a_k = a_1' + \cdots + a_k'\}|.$$

These qualities were extensively used in additive combinatorics, mostly for the case $k = 2$. In the previous section we saw their application (in the non-abelian case) for $k = 3$. Here we observe how extremely natural and appealing the Plünnecke-Ruzsa theory looks in this setting.

By "the setting" we mean the following. By Cauchy-Schwartz (cf. (11)),

$$\mathbf{c}(A_1, \ldots, A_k) \geq \frac{|A_1|^2 \cdots |A_k|^2}{|A_1 \cdot \ldots \cdot A_k|},$$

so we have the lower bound

$$|A_1 \cdot \ldots \cdot A_k| \geq \frac{|A_1|^2 \cdots |A_k|^2}{\mathbf{c}(A_1, \ldots, A_k)}. \tag{21}$$

Assuming we are willing to accept the right-hand side as a "good enough" substitute for $|A_1 \cdot \ldots \cdot A_k|$, we can infer Plünnecke-Ruzsa inequalities as follows.

Lemma 6.1.

$$\mathbf{c}(B_1, \ldots, B_k, A, A) \geq \frac{\mathbf{c}(B_1, \ldots, B_k, A)^2}{|B_1| \cdot (|B_2| \cdot \ldots \cdot |B_k|)^2}.$$

Proof. For $\vec{b} = (b_1, \ldots, b_k) \in B_1 \times \cdots \times B_k$, let $n(\vec{b})$ be the number of tuples $(\vec{b}', a, a')$ such that $b_1 + \cdots + b_k + a = b_1' + \cdots + b_k' + a'$; thus,

$$\mathbf{c}(B_1, \ldots, B_k, A) = \sum_{\vec{b}} n(\vec{b}).$$

On the other hand, for any fixed $\vec{b}$, every couple of tuples $(\vec{b}^{(1)}, a_1, a_1')$, $(\vec{b}^{(2)}, a_2, a_2')$ contributing to $n(\vec{b})$ as

$$b_1 + \cdots + b_k + a_1 = b_1^{(1)} + \cdots + b_k^{(1)} + a_1',$$
$$b_1 + \cdots + b_k + a_2 = b_1^{(2)} + \cdots + b_k^{(2)} + a_2'$$

also contributes to $\mathbf{c}(B_1, \ldots, B_k, A, A)$ as

$$b_1^{(1)} + \cdots + b_k^{(1)} + a_1' + a_2 = b_1^{(2)} + \cdots + b_k^{(2)} + a_2' + a_1.$$

Every such contribution is counted at most $|B_2| \cdot \ldots \cdot |B_k|$ times (as this is an upper bound on the number of tuples $\vec{b}$ for which $b_1 + \cdots + b_k$ takes on the prescribed value $b_1^{(1)} + \cdots + b_k^{(1)} + a_1' - a_1$). This implies that

$$\mathbf{c}(B_1, \ldots, B_k, A, A) \geq \frac{1}{|B_2| \cdot \ldots \cdot |B_k|} \cdot \sum_{\vec{b}} n(\vec{b})^2$$

and makes our lemma the result of yet another application of Cauchy-Schwartz. $\square$

Lemma 6.2. $\mathbf{c}(A_1, \ldots, A_k) \geq \frac{\mathbf{c}(B, A_1, \ldots, A_k)}{|B|^2}$.

Proof. Applying the union bound to all possible choices of b, b',

$$\mathbf{c}(B, A_1, \ldots, A_k) \leq |B|^2 \cdot \max_{d \in G} \mathbf{c}_d(A_1, \ldots, A_k),$$

where $\mathbf{c}_d(A_1, \ldots, A_k)$ is the "shifted" version of $\mathbf{c}(A_1, \ldots, A_k)$:

$$\mathbf{c}_d(A_1, \ldots, A_k) \stackrel{\text{def}}{=} |\{((a_1, \ldots, a_k), (a_1', \ldots, a_k')) \in (A_1 \times \cdots \times A_k)^2 | \\ a_1 + \cdots + a_k + d = a_1' + \cdots + a_k'\}|.$$

But

$$\mathbf{c}(A_1, \ldots, A_k) \geq \mathbf{c}_d(A_1, \ldots, A_k) \tag{22}$$

is easy (and well known). Namely, if $n(e)$ is the number of representations of $e \in G$ in the form $a_1 + \cdots + a_k$, then

$$\mathbf{c}(A_1, \ldots, A_k) = \sum_e n(e)^2,$$
$$\mathbf{c}_d(A_1, \ldots, A_k) = \sum_e n(e) n(e + d),$$

and since the vectors $(n(e) \mid e \in G), (n(e + d) \mid e \in G)$ have the same ℓ_2 norm, (22) follows by Cauchy-Schwartz. $\square$

Theorem 6.3.

$$\mathbf{c}(\underbrace{\pm A, \pm A, \ldots, \pm A}_{k \ \textit{times}}) \geq \frac{\mathbf{c}(B, A)^{2^{k-1}}}{|B|^{(2^{k-1}+1)} |A|^{2^k - 2k}}.$$

Proof. $\mathbf{c}(A_1, \ldots, A_k)$ is clearly invariant under negating components, so we may assume that all signs are actually plus signs. Applying Lemma 6.1 to $B_1 := B, \ B_2 := \cdots := B_k := A$, we find

$$\mathbf{c}(B, \underbrace{A, \ldots, A}_{k \text{ times}}) \geq \frac{1}{|B| \cdot |A|^{2(k-2)}} \cdot \mathbf{c}(B, \underbrace{A, \ldots, A}_{k-1}) \ (k \geq 2).$$

By induction on k,

$$\mathbf{c}(B, \underbrace{A, \ldots, A}_{k \text{ times}}) \geq \frac{\mathbf{c}(B, A)^{2^{k-1}}}{|B|^{(2^{k-1}-1)} |A|^{2^k - 2k}}.$$

Applying Lemma 6.2 finishes the proof. $\square$

In order to interpret this result, recall that the standard doubling constant $K_{A,B}$ given by

$$|A \cdot B| = K_{A,B}|B|$$

in our framework corresponds, via (21), to

$$\mathbf{c}(A, B) = \varepsilon_{A,B}|A|^2|B| \ (\varepsilon_{A,B}$$

In this notation, Theorem 6.3 can be re-written as

$$\mathbf{c}(\underbrace{\pm A, \ldots, \pm A}_{k \text{ times}}) \geq \varepsilon_{A,B}^{2^{k-1}} \cdot \frac{|A|^{2k}}{|B|},$$

which (again, via (21)) corresponds exactly to the "classical" conclusion $| \pm A \pm A \pm \cdots \pm A| \leq K_{A,B}^{O(1)}|B|$.

The material in this section can be readily generalized to convolutions of discrete probability measures (replacing uniform distributions on $A_1, \ldots, A_k$). Namely, the *collision probability* $\mathbf{cp}(\mu)$ of a discrete probability measure μ is defined as

$$\mathbf{cp}(\mu) \stackrel{\text{def}}{=} \mathbf{P}[\boldsymbol{a} = \boldsymbol{a'}],$$

where $\boldsymbol{a}, \boldsymbol{a'}$ are two random variables picked independently at random according to μ. We also let

$$\ell_\infty(\mu) \stackrel{\text{def}}{=} \max_{a \in \text{Sup}(\mu)} \mu(\{a\}).$$

(Thus, the *min-entropy* $H^\infty(\mu)$ is equal to $-\log_2 \ell_\infty(\mu)$.) If A is the support of μ, then clearly

$$\ell_\infty(\mu) \geq \mathbf{cp}(\mu) \geq \frac{1}{|A|}.$$

For probability measures $\mu_1, \ldots, \mu_k$ on an abelian group G, we denote by $\mu_1 + \cdots + \mu_k$ their *convolution*, that is the measure corresponding to the random variable $\boldsymbol{a_1} + \cdots + \boldsymbol{a_k}$, where $\boldsymbol{a_1}, \ldots, \boldsymbol{a_k}$ are picked uniformly at random according to the measures $\mu_1, \ldots, \mu_k$.

In this notation, the proof of Theorem 6.3 can be easily generalized to give the inequality

$$\frac{1}{\ell_\infty(\eta)} \cdot \mathbf{cp}(\underbrace{\pm\mu \pm \mu \pm \cdots \pm \mu}_{k \text{ times}}) \geq \left(\frac{1}{\ell_\infty(\eta)} \cdot \mathbf{cp}(\mu + \eta) \right)^{2^{k-1}}$$

for any two discrete probability measures μ, η on G.

A further generalization is apparently possible in the continuous setting of Tao [24]. It is not clear, however, whether any interesting analogue of this exists in the non-abelian case.

Acknowledgments. I am greatly indebted to Jean Bourgain for posing this problem and for his constant encouragement afterward. I am also grateful to Avi Wigderson for very useful conversations and for finding a gap in the first version of the argument. My thanks are due to Akshay Venkatesh for pointing out the references [10], [13] and to Lisa Carbone for her question that inspired the extension of the main result to virtually free groups. I am grateful to an anonymous referee for many useful remarks.

References

[1] S. I. Adian, *The Burnside Problem and Identities in Groups*, *Ergeb. Math. Grenzgeb.* **95**, Springer-Verlag, New York, 1979, translated from the Russian by John Lennox and James Wiegold. MR 0537580.

[2] B. Barak, R. Impagliazzo, and A. Wigderson, Extracting randomness using few independent sources, *SIAM J. Comput.* **36** (2006), 1095–1118. MR 2272272. Zbl 1127.68030. http://dx.doi.org/10.1137/S0097539705447141.

[3] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson, 2-source dispersers for sub-polynomial entropy and Ramsey graphs beating the Frankl-Wilson construction, in *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, ACM, New York, 2006, pp. 671–680. MR 2277192. Zbl 1122.68300. http://dx.doi.org/10.1145/1132516.1132611.

[4] J. Bourgain and A. Gamburd, Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$, *Ann. of Math.* **167** (2008), 625–642. MR 2415383. Zbl 1216.20042. http://dx.doi.org/10.4007/annals.2008.167.625.

[5] J. Bourgain, N. Katz, and T. Tao, A sum-product estimate in finite fields, and applications, *Geom. Funct. Anal.* **14** (2004), 27–57. MR 2053599. Zbl 1145.11306. http://dx.doi.org/10.1007/s00039-004-0451-1.

[6] V. K. Bulitko, Equations and inequalities in a free group and a free semigroup, *Tul. Gos. Ped. Inst. Učen. Zap. Mat. Kaf.* (1970), 242–252. MR 0393235.

[7] M.-C. Chang, Product theorems in SL_2 and SL_3, *J. Inst. Math. Jussieu* **7** (2008), 1–25. MR 2398145. Zbl 1167.20328. http://dx.doi.org/10.1017/S1474748007000126.

[8] G. A. Freĭman, *Foundations of a Structural Theory of Set Addition*, Amer. Math. Soc., Providence, R. I., 1973, translated from the Russian, *Transl. Math. Monogr.* **37**. MR 0360496. Zbl 0271.10044.

[9] W. T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* **8** (1998), 529–551. MR 1631259. Zbl 0907.11005. http://dx.doi.org/10.1007/s000390050065.

[10] U. Haagerup, An example of a non nuclear C^*-algebra, which has the metric approximation property, *Invent. Math.* **50** (1978/79), 279–293. MR 0520930. Zbl 0408.46046. http://dx.doi.org/10.1007/BF01410082.

[11] H. A. Helfgott, Growth and generation in $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, *Ann. of Math.* **167** (2008), 601–623. MR 2415382. Zbl 1213.20045. http://dx.doi.org/10.4007/annals.2008.167.601.

[12] H. A. Helfgott, Growth in groups: ideas and perspectives, 2013. arXiv 1303.0239.

[13] P. Jolissaint, Rapidly decreasing functions in reduced C^*-algebras of groups, *Trans. Amer. Math. Soc.* **317** (1990), 167–196. MR 0943303. Zbl 0711.46054. http://dx.doi.org/10.2307/2001458.

[14] O. Kharlampovich and A. Myasnikov, Implicit function theorem over free groups, *J. Algebra* **290** (2005), 1–203. MR 2154989. Zbl 1094.20016. http://dx.doi.org/10.1016/j.jalgebra.2005.04.001.

[15] O. Kharlampovich and A. Myasnikov, Elementary theory of free non-abelian groups, *J. Algebra* **302** (2006), 451–552. MR 2293770. Zbl 1110.03020. http://dx.doi.org/10.1016/j.jalgebra.2006.03.033.

[16] E. Kowalski, Sieve in expansion, Technical Report 1028, Séminaire Bourbaki, 63éme année, 2010–2011.

[17] A. Lubotzky, Expander graphs in pure and applied mathematics, *Bull. Amer. Math. Soc.* **49** (2012), 113–162. MR 2869010. Zbl 1232.05194. http://dx.doi.org/10.1090/S0273-0979-2011-01359-3.

[18] R. C. Lyndon, Equations in free groups, *Trans. Amer. Math. Soc.* **96** (1960), 445–457. MR 0151503. Zbl 0108.02301. http://dx.doi.org/10.2307/1993533.

[19] R. C. Lyndon and P. E. Schupp, *Combinatorial Group Theory*, *Ergeb. Math. Grenzgeb.* **89**, Springer-Verlag, New York, 1977. MR 0577064. Zbl 0368.20023.

[20] G. S. Makanin, Equations in a free group, *Izv. Akad. Nauk SSSR Ser. Mat.* **46** (1982), 1199–1273, 1344, in Russian; translated in *Math. USSR-Izvestiya* **21** (1983), 483–546. MR 0682490. Zbl 0527.20018. http://dx.doi.org/10.1070/IM1983v021n03ABEH001803.

[21] M. B. Nathanson, *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, *Grad. Texts in Math.* **165**, Springer-Verlag, New York, 1996. MR 1477155. Zbl 0859.11003.

[22] A. A. Razborov, On systems of equations in a free group, *Izv. Akad. Nauk SSSR Ser. Mat.* **48** (1984), 779–832, in Russian; translated in *Math. USSR-Izvestiya* **25** (1985), 115–162. MR 0755958. Zbl 0579.20019. http://dx.doi.org/10.1070/IM1985v025n01ABEH001272.

[23] Z. Sela, Diophantine geometry over groups. VI. The elementary theory of a free group, *Geom. Funct. Anal.* **16** (2006), 707–730. MR 2238945. Zbl 1118.20035. http://dx.doi.org/10.1007/s00039-006-0565-8.

[24] T. Tao, Product set estimates for non-commutative groups, *Combinatorica* **28** (2008), 547–594. MR 2501249. Zbl 1254.11017. http://dx.doi.org/10.1007/s00493-008-2271-7.

[25] T. Tao and V. H. Vu, *Additive Combinatorics*, *Cambridge Stud. Adv. Math.* no. 105, Cambridge Univ. Press, Cambridge, 2006. MR 2289012. Zbl 1127.11002. http://dx.doi.org/10.1017/CBO9780511755149.

(Received: June 18, 2007)
(Revised: September 20, 2013)

Steklov Mathematical Institute, Moscow, Russia
Institute for Advanced Study, Princeton, NJ
Current address: University of Chicago, Chicago, IL
E-mail: razborov@cs.uchicago.edu