WORDS, HAUSDORFF DIMENSION AND RANDOMLY FREE GROUPS

MICHAEL LARSEN AND ANER SHALEV

Abstract. We bound the size of fibers of word maps in finite and residually finite groups, and derive various applications. Our main result shows that, for any word $1 \neq w \in F_d$ there exists $\epsilon > 0$ such that if Γ is a residually finite group with infinitely many non-isomorphic non-abelian upper composition factors, then all fibers of the word map $w : \Gamma^d \to \Gamma$ have Hausdorff dimension at most $d - \epsilon$.

We conclude that profinite groups $G := \hat{\Gamma}$, Γ as above, satisfy no probabilistic identity, and therefore they are *randomly free*, namely, for any $d \geq 1$, the probability that randomly chosen elements $g_1, \ldots, g_d \in G$ freely generate a free subgroup (isomorphic to F_d) is 1. This solves an open problem from [DPSS].

Additional applications and related results are also established. For example, combining our results with recent results of Bors, we conclude that a profinite group in which the set of elements of finite odd order has positive measure has an open prosolvable subgroup. This may be regarded as a probabilistic version of the Feit-Thompson theorem.

2010 *Mathematics Subject Classification.* Primary 20E26, 20P06; Secondary 20D06, 20G40.

ML was partially supported by NSF grant DMS-1401419. AS was partially supported by ERC advanced grant 247034, BSF grant 2008194, ISF grant 1117/13 and the Vinik Chair of mathematics which he holds.

1. Introduction

In the past few decades there has been considerable interest in the theory of word maps, see for instance [Bo, LiS, La, Sh, LS1, LST, GT, N2], as well as Segal's monograph [Se] and the references therein. Many of these works focus on the image of word maps on finite simple groups and on a related Waring type problem. There is also increasing interest in fibers of word maps and related problems, see [LS2, LS3, B1, B2].

In this paper we prove various results showing that fibers of word maps are small, not just in a measure-theoretic sense, but also in the stronger sense of Hausdorff dimension. Our results apply for a wide family of finite and infinite groups, well beyond the family of finite simple groups.

The main result of this paper is the following.

Theorem 1.1. *Let Γ be a residually finite group with infinitely many non-isomorphic non-abelian upper composition factors. Let $d \geq 1$ and let $w \in F_d$ be a non-trivial word. Then there exists $\epsilon > 0$ depending only on w such that the Hausdorff dimension of any fiber of the associated word map $w : \Gamma^d \to \Gamma$ is at most $d - \epsilon$.*

Here F_d denotes a free group of rank d freely generated by $x_1, \dots x_d$. A word $w = w(x_1, \dots, x_d) \in F_d$ (which we write in reduced form) gives rise to a word map $w : \Gamma^d \to \Gamma$ on any group Γ, which is induced by substitution.

By an *upper composition factor* of Γ we mean a composition factor of some finite quotient Γ/Δ of Γ, where Δ is a normal subgroup of Γ of finite index (and we assume Δ is open if Γ is a profinite group).

For a residually finite group Γ and $d \geq 1$, we define the *Hausdorff dimension* of a subset $S \subseteq \Gamma^d$ by

$$\mathrm{Hdim}(S) = \liminf_{\Delta} \frac{\log |S\Delta^d/\Delta^d|}{\log |\Gamma/\Delta|},$$

where Δ ranges over the finite index normal subgroups of Γ (and again if Γ is profinite we also assume that Δ is open). Thus $\mathrm{Hdim}(\Gamma^d) = d$ and Theorem 1.1 states that, under the assumptions of the theorem we have

$$\mathrm{Hdim}(w^{-1}(g)) \leq d - \epsilon$$

for every $g \in \Gamma$.

In fact the proof of Theorem 1.1 gives a bit more; see Theorem 5.2 for an effective finitary version.

These results may be regarded as a far-reaching extension of the main result of [LS2], stating that, for w as above there exists $\epsilon = \epsilon(w) > 0$ such that, if T is a large enough finite simple group then the fibers of $w : T^d \to T$ have size at most $|T|^{d-\epsilon}$. Here and throughout this paper, by a finite simple group we mean a non-abelian finite simple group.

We now list several consequences of Theorem 1.1. The first one deals with linear groups and strengthens results from [LS3].

It is easy to see, using strong approximation (see [N1, Pi, We]), that a finitely generated linear group which is not virtually solvable has infinitely many finite simple groups of Lie type as upper composition factors. Applying Theorem 1.1, we deduce the following.

Theorem 1.2. *Let Γ be a finitely generated linear group over any field. Suppose Γ is not virtually solvable. Then the fibers of any word map on Γ^d induced by a non-trivial word $w \in F_d$ have Hausdorff dimension less then d.*

The next consequence of our main result deals with probabilistic identities. We need more notation. Given a word $w \in F_d$ and a finite group G, let $p_{w,G}$ denote the associated probability distribution on G. Thus, for $g \in G$ we have $p_{w,G}(g) = |w^{-1}(g)|/|G|^d$.

Recall that a word $1 \neq w \in F_d$ is said to be *a probabilistic identity* of a residually finite group Γ if there exists $\delta > 0$ such that, in any finite quotient H of Γ, we have $p_{w,H}(1) \geq \delta$. This amounts to saying that, in the profinite completion G of Γ, the probability (with respect to the normalized Haar measure on G) that $w(g_1, \ldots, g_d) = 1$ where $g_1, \ldots, g_d \in G$ are random elements is positive.

For example, $w = x_1^2$ is a probabilistic identity of the infinite dihedral group $\Gamma = D_\infty$.

It follows from [Ma] that a residually finite group which satisfies the probabilistic identity x_1^2 is finite-by-abelian-by-finite. A similar conclusion holds for the probabilistic identity $[x_1, x_2]$, as follows from the earlier paper [Ne], which is applied in [Ma]. However, very little is known about groups satisfying more general probabilistic identities.

Now, let Γ be a residually finite group with infinitely many non-isomorphic non-abelian upper composition factors, and let $1 \neq w \in F_d$. By Theorem 1.1 we see that Γ has arbitrarily large finite quotients H such that $p_{w,H}(1) \leq |H|^{-\epsilon}$, which tends to 0. This implies the following new result.

Theorem 1.3. *Let Γ be a residually finite group with infinitely many non-isomorphic non-abelian upper composition factors. Then Γ does not satisfy any probabilistic identity.*

This result generalizes Theorem 1.2 in [LS3], showing that a finitely generated linear group which satisfies a probabilistic identity is virtually solvable. It also enables us to solve an open problem regarding randomly free groups.

We say that a profinite group G is *randomly free* if, for every $d \geq 1$, the probability that a d-tuple of elements of G freely generates a free subgroup is 1. We use *freely generate* in the sense of abstract group theory, i.e., no non-trivial word evaluated at the chosen elements should give the identity. A residually finite group is said to be randomly free if its profinite completion is randomly free.

See [Ep] and [Sz] for earlier results on groups in which almost all subgroups are free.

We need the following straightforward observation.

Lemma 1.4. *A residually finite group Γ is randomly free if and only if it does not satisfy any probabilistic identity.*

To show this, let G be the profinite completion of Γ. Note that $g_1, \dots, g_d \in G$ freely generate a free subgroup of G if and only if $w(g_1, \dots, g_d) \neq 1$ for every $1 \neq w \in F_d$. Now, suppose Γ does not satisfy any probabilistic identity. Then the probability that $w(g_1, \dots, g_d) = 1$ ($g_i \in G$) is 0 for any such word w. As Haar measure is σ-additive, the probability that there exists $w \neq 1$ such that $w(g_1, \dots, g_d) = 1$ is also 0. Thus, $g_1, \dots, g_d \in G$ freely generate a free subgroup with probability 1, proving that Γ is randomly free. The reverse implication is trivial.

Combining the above lemma with Theorem 1.3 we obtain the following.

Theorem 1.5. *Let Γ be a residually finite group with infinitely many non-isomorphic non-abelian upper composition factors. Then Γ is randomly free.*

This immediately implies (using strong approximation) the following probabilistic Tits alternative, which is the main result of [LS3].

Corollary 1.6. *A finitely generated linear group is either virtually solvable or randomly free.*

Theorem 1.5 (applied to profinite groups G) solves a problem raised in 2003 in [DPSS] (see Problem 7 there). A partial solution, assuming G has infinitely many alternating groups as upper composition factors, was given in 2005 by Abért, see [Ab, Theorem 1.7]. Hence it remains to prove the theorem assuming G has infinitely many simple groups of Lie type as upper composition factors.

A recent work of Bors [B1] can be used to handle the case where G has classical groups of unbounded rank as composition factors. Indeed Theorem 1.1.2 there implies Theorem 1.3 in this case. Hence it remains to handle the case where G has infinitely many groups of Lie type of bounded rank as upper composition factors.

After discussing various consequences of Theorem 1.1, let us now state some results of independent interest on finite groups, which play a crucial role in its proof. The first result deals with almost simple groups of Lie type of bounded rank.

Theorem 1.7. *For any non-trivial word $w = w(x_1, \dots, x_d)$ and any positive integer r, there exist $N, \epsilon > 0$ depending only on w and r such that, if T is a finite simple group of Lie type of order $\geq N$ and rank $\leq r$, then for all $g_1, \dots, g_d, g \in \mathrm{Aut}\,(T)$,*

$$\left| \{(t_1, \dots, t_d) \in T^d \mid w(t_1 g_1, \dots, t_d g_d) = g\} \right| \leq |T|^{d-\epsilon}.$$

Combining this with results from [LS2] and [B1], we show that, for any $1 \neq w \in F_d$ there exist $N, \epsilon > 0$ such that for any almost simple group G of size at least N and any $g \in G$ we have $p_{w,G}(g) \leq |G|^{-\epsilon}$—see Corollary 4.6 below. This extends the main result of [LS2] from simple groups to almost

simple groups. This extension is by no means routine, and it occupies a large part of our paper.

Combining the result above with other tools we deduce the following more general theorem on finite groups which are semisimple in the sense of Fitting.

Theorem 1.8. *Let G be a finite group such that $T^k \leq G \leq \mathrm{Aut}\,(T^k)$ for some $k \geq 1$ and a finite simple group T. Suppose $w \neq 1$ is a word. Then there exist constants $N = N(w), \epsilon = \epsilon(w) > 0$ depending only on w such that, if $|T| \geq N$, then for any $g \in G$ we have $p_{w,G}(g) \leq |T^k|^{-\epsilon}$.*

Note that, if G is any finite group with a chief factor $G_1/G_2 \cong T^k$, then G has a semisimple quotient $G/C_G(G_1/G_2)$ (the section centralizer) lying between T^k and its automorphism group, to which Theorem 1.8 may be applied.

The next result enables us to pass from any finite group with a large non-abelian composition factor to a large semisimple quotient K as above, in which the chief factor T^k is almost as large as K.

Proposition 1.9. *For any $\delta > 0$ there exists $f = f(\delta) > 0$ such that if G is a finite group with a non-abelian composition factor of order $\geq f$, then G has a quotient K with a composition factor T of order $\geq f$ such that $T^k \leq K \leq \mathrm{Aut}\,(T^k)$ and $|T^k| \geq |K|^{1-\delta}$.*

Results 1.8 and 1.9 easily imply our main result, namely Theorem 1.1.

Finally, combining Theorem 1.8 with a new result of Bors [B2, 1.1.2] we obtain a result which may be regarded as a probabilistic version of the Odd Order Theorem of Feit and Thompson [FT]. For a finite group G denote by $\mathrm{Rad}\,(G)$ the solvable radical of G, namely the maximal solvable normal subgroup of G.

Theorem 1.10. (i) *Let k be an odd integer and let $w = x^k$. Then for any $\epsilon > 0$ there is a number $M = M(k, \epsilon)$ such that, if G is any finite group satisfying $p_{w,G}(g) \geq \epsilon$ for some $g \in G$, then*

$$|G/\mathrm{Rad}\,(G)| \leq M.$$

(ii) *Let G be a profinite group and suppose that the set of elements of G of finite odd order has positive Haar measure. Then G has a prosolvable open normal subgroup.*

Therefore profinite groups as in part (ii) above are virtually prosolvable. Part (i) above shows that, for odd k, if the probability that $g^k = 1$ in G is bounded away from zero, then G has a solvable normal subgroup of bounded index.

However, it is not true that if the probability that $g \in G$ has odd order is at least $\epsilon > 0$ (or even at least $1 - \epsilon$), then G is solvable-by-bounded. Indeed, simple groups of Lie type in characteristic 2 provide counterexamples (since most of their elements are semisimple, hence of odd order).

A result similar to part (i) of Theorem 1.10 with $w = [x_1, \dots, x_d]$, a left normed commutator, also follows by combining Theorem 1.1.2 of Bors [B2] with Theorem 1.8 above.

Let us now discuss the strategy of the proof of Theorem 1.7, which is the basis of most of our other results. The main idea is, roughly, to convert it to a problem in algebraic geometry. If, to simplify slightly, $T = \underline{G}(\mathbb{F}_q)$ for some algebraic group $\underline{G}$ of bounded rank, then the parameter q must go to infinity as $|T| \to \infty$. Any non-trivial word w defines a non-constant morphism $\underline{G}^d \to \underline{G}$. The fibers are therefore of dimension $\leq d \dim \underline{G} - 1$, and as q goes to infinity and g varies, the "complexity" of the fibers remains bounded, so we can deduce that

$$|w^{-1}(g)| = O(|\underline{G}(\mathbb{F}_q)|^{d-1/\dim \underline{G}})$$

from standard point counting results for varieties over finite fields. This idea is not new to this paper (see, e.g., [DPSS, LS2]). However, there are technical difficulties in implementing it when we must take outer automorphisms of T into account. For field automorphisms, when $\mathbb{F}_q$ is large but its characteristic p is small, this requires a new idea, namely finding big gaps between consecutive powers of Frobenius appearing in any specified d-tuple of cosets of T (see the proof of Theorem 1.7, below). The Suzuki and Ree groups also pose a technical challenge, since we are no longer counting points on varieties over finite fields but rather taking fixed points of maps which are square roots of ordinary Frobenius maps.

In §2, we develop a simple formalism for making precise the idea of bounded complexity mentioned above. In §3, we present upper bounds for certain point counting problems. There are two main variants, one aimed at proving estimates which are uniform in characteristic and one aimed at dealing with the special difficulties of the Suzuki-Ree case. In §4 we prove Theorem 1.7 and extend it to all almost simple groups.

Finally, in §5 we study finite semisimple groups and finite groups in general. This is where results 1.8, 1.9, 1.1 and 1.10 are established. If $T^k \leq G \leq \mathrm{Aut}\,(T^k) \cong \mathrm{Aut}\,(T) \wr S_k$ as in Theorem 1.8, then G induces a transitive permutation group P on the k copies of T, and tools from the theory of permutation groups become relevant. Theorem 1.8 is proved, roughly, by using the case $k = 1$, which has already been established, and by finding a large set of independent equations induced by the given word equation on G.

We then prove Proposition 1.9 by bounding the order of P above, and deduce a finitary version—Theorem 5.2—of Theorem 1.1, which readily implies it.

We would like to thank the referee for pointing out several deficiencies in the original version of this paper.

2. Degree Bounds

Definition 2.1. Let R be a commutative ring. By a *generated commutative algebra* over R (GCA for short), we mean a pair (A, S) consisting of a commutative R-algebra A and a finite set S of generators of A over R. An isomorphism $(A, S) \to (B, T)$ of GCAs is an isomorphism $A \to B$ of R-algebras which maps S onto T.

Thus, every GCA is isomorphic to one of the form

$$(R[x_1, \ldots, x_N]/I, \{\overline{x}_1, \ldots, \overline{x}_N\}).$$

For $a \in A$ we define $\deg_S a$ to be the minimum integer m such that a can be realized as a degree m polynomial in the elements S with coefficients in R. If (A, S) and (B, T) are GCAs over R and $\phi\colon A \to B$ is an R-algebra homomorphism, we define

$$\deg \phi = \max_{s \in S} \deg_T \phi(s). \tag{1}$$

The following lemma is obvious:

Lemma 2.2. *With notation as above,*

$$\deg_T \phi(a) \leq \deg \phi \deg_S a.$$

If (C, U) is also a GCA over R, and $\psi\colon B \to C$ is an R-algebra homomorphism,

$$\deg \psi \circ \phi \leq \deg \phi \deg \psi.$$

If R' is a commutative R-algebra and (A', S') denotes the base change of (A, S) to R' (i.e., $A' = A \otimes_R R'$, and $S' = \{s \otimes 1 \mid s \in S\}$), then

$$\deg_{S'} a \otimes 1 \leq \deg_S a.$$

The following lemma, asserting that every q-Frobenius map has degree at most q, is likewise immediate from the definitions:

Lemma 2.3. *If (A, S) is a GCA over $R = \mathbb{F}_q$, and $\phi\colon A \to A$ denote the q-Frobenius map, then* $\deg \phi \leq q$.

We say generating sets S and T of A are *equivalent* if they generate the same R-submodule of A. In this case, $\deg_S a = \deg_T a$ for all $a \in A$. Thus, every surjective R-algebra homomorphism $\phi\colon R[x_1, \ldots, x_N] \to A$ determines a well-defined equivalence class of generating sets in A represented by $\{\phi(x_1), \ldots, \phi(x_N)\}$. Geometrically, a closed immersion of Spec A into a vector space scheme over R determines a generating set, and two closed immersions which are the same up to an invertible linear transformation determine equivalent generating sets.

Lemma 2.4. *Let R be a field, $\underline{G} =$ Spec A an adjoint semisimple algebraic group over R, α an automorphism of $\underline{G}$ as algebraic group over R, and $\rho\colon \underline{G} \to \mathbb{A}^{\dim \underline{G}}$ the adjoint representation with respect to some basis of the Lie algebra of $\underline{G}$. Up to equivalence, ρ and $\rho \circ \alpha$ determine the same equivalence class of generating sets of A.*

Proof. The homomorphisms ρ and $\rho \circ \alpha$ give the same representation, which means they are conjugate, so they define the same equivalence class of generating sets. □

In general, given an adjoint semisimple algebraic group $\underline{G}$ over a field $\mathbb{F}_q$, we regard its coordinate ring A as a GCA endowed with a generating set S belonging to this equivalence class.

Proposition 2.5. *Let $\underline{G} := \operatorname{Spec} A$ be an adjoint semisimple algebraic group over $\mathbb{F}_p$ with root system Φ. Let F_p denote the p-Frobenius endomorphism of $\underline{G}$. Let $\beta_1, \dots, \beta_l$ be automorphisms of $\underline{G}$ defined over $\mathbb{F}_p$, $j_1, \dots, j_l$ non-negative integers, and $n_1, \dots, n_l \in \{1, \dots, d\}$. Then, for $\underline{G}$ as above, the morphism $\underline{G}^d \to \underline{G}$ given by*

$$(g_1, \dots, g_d) \mapsto F_p^{j_1} \beta_1(g_{n_1})^{\pm 1} \cdots F_p^{j_l} \beta_l(g_{n_l})^{\pm 1} \tag{2}$$

has degree $O(p^j)$, where $j := \max_i j_i$ and where degree is defined by (1) with the coordinate rings of $\underline{G}$ and $\underline{G}^d$ equipped with generating sets S and $S \coprod \cdots \coprod S$ respectively. (The implicit constant depends on Φ and l but not on p or j.)

Proof. As the multiplication and inversion morphisms on $\underline{G}$ have degrees which are bounded independent of p, by Lemma 2.2 and induction on l, it suffices to prove the claim for $g \mapsto F_p^j \beta(g)$. The proposition follows from Lemma 2.2, Lemma 2.3, and Lemma 2.4.

□

3. Point Bounds

It is well known that a degree s hypersurface in $\mathbb{A}^N$ has at most sq^{N-1} points over $\mathbb{F}_q$. We present several variants of this observation.

Proposition 3.1. *Let p be prime, q a power of p, $\underline{X} \subset \mathbb{A}^N$ over $\overline{\mathbb{F}}_p$ be the closed subscheme defined by the ideal $(f_1, \dots, f_n)$, where $f_i \in \overline{\mathbb{F}}_p[x_1, \dots, x_N]$ have degrees $d_1, \dots, d_n$. If*

$$d_1 \cdots d_n \le Aq,$$

then

$$|\underline{X}(\overline{\mathbb{F}}_p) \cap \mathbb{A}^N(\mathbb{F}_q)| \le (2A+1)^N d_1 \cdots d_n q^{\dim \underline{X}}.$$

Proof. Let

$$\phi_i(x_0, \dots, x_N) := x_0^{d_i} f_i(x_1/x_0, \dots, x_N/x_0).$$

The intersection of the hypersurfaces in $\mathbb{P}^N$ defined by the ϕ_i defines a closed subscheme $\overline{\underline{X}}$ of $\mathbb{P}^N$, and the intersection of this subscheme with $\mathbb{A}^N$ is $\underline{X}$. Therefore, the number of irreducible components of $\underline{X}$ over $\overline{\mathbb{F}}_p$ is less than or equal to the number of components in $\overline{\underline{X}}$, which by a suitable version of Bézout's theorem (see [Fu, 12.3.1]), implies that the number of

such components is less than or equal to $d_1 \cdots d_n$. This implies the theorem for $\dim \underline{X} = 0$.

We proceed by double induction, first on $\dim \underline{X}$ and then on N. The base case for the inner induction is $N = \dim \underline{X}$, and the proposition holds in this case trivially. For the induction step, we count pairs (x, H) consisting of a point $x \in \underline{X}(\overline{\mathbb{F}}_p) \cap \mathbb{A}^N(\mathbb{F}_q)$ and H an $\mathbb{F}_q$-rational hyperplane in $\mathbb{A}^N$ containing x (but not necessarily 0). The number of such pairs for a given $x \in \underline{X}(\overline{\mathbb{F}}_p) \cap \mathbb{A}^N(\mathbb{F}_q)$ is $|\mathbb{P}^{N-1}(\mathbb{F}_q)|$, so the total number is

$$(1 + q + \cdots + q^{N-1})|\underline{X}(\overline{\mathbb{F}}_p) \cap \mathbb{A}^N(\mathbb{F}_q)|.$$

We say that H is *bad* for $\underline{X}$ if it contains an irreducible component of $\underline{X}$ of dimension $\dim \underline{X}$. The number of hyperplanes containing a given top-dimensional component is at most

$$1 + q + \cdots + q^{N-\dim \underline{X}-1},$$

so the number of bad hyperplanes is less than

$$2q^{N-\dim \underline{X}-1} d_1 \cdots d_n.$$

By the induction hypothesis for N, the number of pairs (x, H) where H is bad is less than

$$2q^{N-\dim \underline{X}-1} d_1 \cdots d_n (2A+1)^{N-1} d_1 \cdots d_n q^{\dim \underline{X}} \leq 2A(2A+1)^{N-1} q^N d_1 \cdots d_n.$$

By the induction hypothesis for $\dim \underline{X}$, the number of pairs (x, H) where H is good is less than

$$(q + q^2 + \cdots + q^N)(2A+1)^{N-1} d_1 \cdots d_n q^{\dim \underline{X}-1}.$$

Thus, the total number of pairs is less than

$$(q + q^2 + \cdots + q^N)(2A+1)^N d_1 \cdots d_n q^{\dim \underline{X}},$$

so

$$|\underline{X}(\overline{\mathbb{F}}_p) \cap \mathbb{A}^N(\mathbb{F}_q)| \leq (2A+1)^N d_1 \cdots d_n q^{\dim \underline{X}},$$

and the proposition holds by induction. □

Proposition 3.2. *Let (A, S) be a GCA over $\mathbb{Z}$. Let (A_q, S_q) be a GCA over $\mathbb{F}_q$ such that there exists an isomorphism $\iota\colon A \otimes \overline{\mathbb{F}}_q \to A_q \otimes \overline{\mathbb{F}}_q$ with respect to which $S \otimes 1$ and $S_q \otimes 1$ are equivalent, and let $\underline{Y} = \operatorname{Spec} A_q$. If $f \in A \otimes \overline{\mathbb{F}}_q$ is not a zero divisor, then*

$$|\{y \in \underline{Y}(\mathbb{F}_q) \mid f(y) = 0\}| = O((\deg_{S\otimes 1} f) q^{\dim \underline{Y}-1}).$$

Here the implicit constant depends only on (A, S) (in particular, not on q, A_q, S_q, or ι).

Proof. Let $S = \{s_1, \ldots, s_N\}$ and $S_q = \{t_1, \ldots, t_M\}$, and consider the homomorphisms $\phi\colon \mathbb{Z}[x_1, \ldots, x_N] \to A$ and $\phi_q\colon \mathbb{F}_q[y_1, \ldots, y_M] \to A_q$ mapping x_i

to $s_i \otimes 1$ and y_i to $t_i \otimes 1$ respectively. The equivalence of generating sets guarantees the existence of a commutative diagram

$$\begin{array}{ccc} \overline{\mathbb{F}}_q[x_1, \dots, x_N] & \longrightarrow & \overline{\mathbb{F}}_q[y_1, \dots, y_M] \\ \downarrow{\scriptstyle \phi \otimes 1} & & \downarrow{\scriptstyle \phi_q \otimes 1} \\ A \otimes \overline{\mathbb{F}}_q & \xrightarrow{\quad \iota \quad} & A_q \otimes \overline{\mathbb{F}}_q \end{array}$$

where the top arrow maps each x_i to a linear combination of the y_j. If $f_1, \dots, f_n$ is a generating set for $\ker \phi$, then we have an isomorphism of $\mathbb{F}_q$-algebras

$$A_q \otimes \overline{\mathbb{F}}_q/(\iota(f \otimes 1)) \cong \overline{\mathbb{F}}_q[y_1, \dots, y_N]/(f_1 \otimes 1, \dots, f_n \otimes 1, f).$$

The proposition now follows by applying Proposition 3.1 to

$$\underline{X} = \operatorname{Spec} \overline{\mathbb{F}}_q[x_1, \dots, x_N]/(f_1 \otimes 1, \dots, f_n \otimes 1, f).$$

□

Proposition 3.3. *Let $q = p^f$, (A, S) a GCA defined over $\mathbb{F}_p$, and $\underline{X} =$ Spec A. For all $a \in A$, we let $\underline{X}_a$ denote* Spec $A/(a)$ *regarded as a closed subscheme of $\underline{X}$. Let k be a positive integer prime to f, $F_q \colon \underline{X} \to \underline{X}$ the q-Frobenius morphism, and $F \colon \underline{X} \to \underline{X}$ an endomorphism such that $F^k = F_q$. If $\dim \underline{X}_a < \dim \underline{X}$ and t is a positive integer, then*

$$|\underline{X}_a(\overline{\mathbb{F}}_p) \cap \underline{X}(\overline{\mathbb{F}}_p)^{F^t}| \leq O((\deg_S a) p^{(tf/k)(\dim \underline{X} - 1)}),$$

where the implicit constant does not depend on t or a.

Proof. Let g denote the g.c.d. of t and k. Replacing F by F^g and t and k by t/g and k/g respectively, we may assume without loss of generality that t and k are relatively prime.

Every endomorphism of a commutative ring maps the nilradical to itself, so without loss of generality, we may assume that $\underline{X}$ is reduced. We may also assume that the irreducible components $\underline{X}_1, \dots, \underline{X}_h$ of $\underline{X}$ are permuted transitively by powers of F.

By induction, we may assume that the proposition holds for all affine schemes of dimension less than $\dim \underline{X}$. Thus, if $\underline{U}$ is any F-stable dense open affine subvariety of $\underline{X}$, it suffices to prove the proposition for $\underline{U}$. If $\underline{U}$ is any dense open affine subvariety of $\underline{X}$, the orbit of $\underline{U}$ under $\langle F \rangle$ is finite, so we may replace $\underline{X}$ by an F-stable dense open affine subvariety of $\underline{U}$. Since every quasi-affine variety contains a dense affine subvariety, it suffices to prove the desired estimate after replacing $\underline{X}$ by our choice of dense quasi-affine $\underline{U} \subseteq \underline{X}$.

Applying this observation to the complement of

$$\bigcup_{1 \leq i < j \leq h} \underline{X}_i \cap \underline{X}_j$$

we may assume that the connected components of $\underline{X}$ coincide with the irreducible components $\underline{X}_i$, so the proposition is trivial unless $t = hs$ for some integer s. We may therefore assume that h is relatively prime to k. Replacing F by F^h and t and f by s and fh respectively, we may assume without loss of generality that F fixes every component of $\underline{X}$, so we may reduce to the case that $\underline{X}$ itself is irreducible.

If K denotes the fraction field of A, then F^* defines an inclusion $K \hookrightarrow K$, and $(F^k)^* = (F_q)^*$. As K is finitely generated, the latter inclusion is of degree $q^{\dim \underline{X}} = p^{f \dim \underline{X}}$. By hypothesis, k is relatively prime to f, so k divides $\dim \underline{X}$.

Let $\dim \underline{X} = mk$. We claim that there exists a transcendence basis $b_{11}, \dots, b_{1k}, b_{21}, \dots, b_{mk}$ for K in A such that

$$F^* b_{ij} = \begin{cases} b_{i\,j+1} & \text{if } j < k, \\ b_{i1}^q & \text{otherwise.} \end{cases}$$

Indeed, by the same reasoning that shows that the transcendence degree of K is a multiple of k, we see that every F^*-stable subfield of K has transcendence degree a multiple of k. Thus, we can construct the b_{i1} iteratively, defining $b_{q+1\,1}$ to be any element of A not algebraic over the (algebraically independent) set $\{b_{ij} \mid i \le q\}$ and setting $b_{i\,j+1} = (F^j)^* b_{i1}$.

Endowing $\mathbb{A}^{mk} = \operatorname{Spec} \mathbb{F}_p[x_{ij}]$ $(1 \le i \le m,\ 1 \le j \le k)$ with the F-action given by

$$F^* x_{ij} = \begin{cases} x_{i\,j+1} & \text{if } j < k, \\ x_{i1}^q & \text{otherwise,} \end{cases} \tag{3}$$

we see that the map $x_{ij} \mapsto b_{ij}$ is a generically finite F-equivariant morphism $\underline{X}_1 \to \mathbb{A}^{mk}$. Passing to a sufficiently small F-stable affine open subset $\operatorname{Spec} A[1/h]$ of the base, we may assume that $A[1/h]$ is a finitely generated free $\mathbb{F}_p[x_{ij}][1/h]$-module of some rank ρ. We can therefore realize $A[1/h]$ as a commutative subring of $M_\rho(\mathbb{F}_p[x_{ij}][1/h])$. In particular, all elements of S can be realized in this matrix ring, and it follows that any polynomial of degree $\deg_S a$ in the elements of S can be realized as a matrix whose entries are polynomials of degree $O(\deg_S a)$ in the x_{ij} and $1/h$. The determinant of this matrix is a polynomial of degree $O(\deg_S a)$ in x_{ij} and $1/h$. Every point in the zero locus $\underline{X}_a$ maps to a point in the zero locus of this polynomial. Since the map $\operatorname{Spec} A[1/h] \to \operatorname{Spec} \mathbb{F}_p[x_{ij}][1/h]$ is of degree ρ, it suffices to bound the number of F^t-points on the zero locus of this determinant. Expressing the determinant as a rational function in the x_{ij}, the numerator is a polynomial of degree $O(\deg_S a)$, and it suffices to bound the number of F^t-points of the zero locus of the numerator, viewed as a hypersurface in $\operatorname{Spec} \mathbb{F}_p[x_{ij}]$. Thus, we can reduce the general problem to the case $\underline{X} = \operatorname{Spec} \mathbb{F}_p[x_{ij}]$, with the F-action given by (3).

More explicitly, it suffices to prove that if Q is a polynomial in the x_{ij},

$$|\{(a_{ij}) \in \overline{\mathbb{F}}_p^{mk} \mid Q(a_{ij}) = 0\}^{F^t}| = O(\deg Q(t/k) p^{mk-1}).$$

We project $\mathbb{A}^{mk}$ to $\mathbb{A}^k$ by

$$\pi_m \colon (a_{ij}) \mapsto (a_{m1}, \dots, a_{mk}).$$

Thus π_m is F-equivariant, where

$$F(c_1, \dots, c_k) = (c_2, c_3, \dots, c_k, c_1^q),$$

so

$$\begin{aligned} F^t(c_1, \dots, c_k) &= \left(c_{t+1}^{q^{\lfloor t/k \rfloor}}, c_{t+2}^{q^{\lfloor (t+1)/k \rfloor}}, \dots, c_{t+k-1}^{q^{\lfloor (t+k-1)/k \rfloor}}\right) \\ &= \left(c_{t+1}^{q^{\lfloor t/k \rfloor}}, \dots, c_k^{q^{\lfloor t/k \rfloor}}, c_1^{q^{\lfloor t/k \rfloor + 1}}, \dots, c_t^{q^{\lfloor t/k \rfloor + 1}}\right) \end{aligned}$$

where the c_i are numbered cyclically. Since s and k are relatively prime, any F^t-fixed $(c_1, \dots, c_k)$ is determined by c_1 satisfying $c_1^{p^t} = c_1$. The number of possibilities for $c_1 \in \overline{\mathbb{F}}_q$ is therefore p^t. Any F^t-fixed point (a_{ij}) projects under π_m to a F^t-fixed point.

There are at most $\deg Q$ factors of Q over $\overline{\mathbb{F}}_p[x_{ij}]$ of the form $x_{m1} - c_1$, and for any $c_1 \in \overline{\mathbb{F}}_p$ such that $x_{m1} - c_1$ is not such a factor and $(c_1, \dots, c_k)$ is fixed by F^t, the fiber $\pi_m^{-1}(c_1, \dots, c_k)$ is a hypersurface in $\mathbb{A}^{(m-1)k}$ of degree $\leq \deg Q$. If $(c_1, \dots, c_m)$ is fixed by F^t and $x_{m1} - c_1$ divides Q, then the number of F^t fixed points in $\pi_m^{-1}(c_1, \dots, c_m)$ is the number of F^t fixed points of $\mathbb{A}^{(m-1)k}$, i.e. $q^{t(m-1)}$. By induction on m, it suffices to consider the base case, $m = 1$.

For $m = 1$, if $(c_1, \dots, c_m)$ is a fixed point of F^t, we have

$$c_1 = c_{t+1}^{q^{\lfloor t/k \rfloor}} = \cdots = c_{(k-1)t+1}^{q^{t - \lceil t/k \rceil}},$$

so $Q(c_1, \dots, c_k)$ can be expressed as a polynomial in $c_{(k-1)t+1}$ of degree at most

$$\begin{aligned} (\dim Q) q^{t - \lceil t/k \rceil} &= (\dim Q) p^{tf - \lceil t/k \rceil f} \\ &\leq (\dim Q) p^{tf - tf/k} = (\dim Q) p^{(tf/k)(\dim \underline{X} - 1)}, \end{aligned}$$

which proves the proposition.

□

4. Almost simple groups

In this section we bound the size of fibers of word maps for almost simple groups. Most of the section is devoted to the proof of Theorem 1.7 on almost simple groups of Lie type of bounded rank.

We make essential use of the following result of Nikolov [N2, Corollary 8], ruling out a given coset identity in large almost simple groups.

Proposition 4.1. *For every word $1 \neq w \in F_d$ there exists $c_0 = c_0(w)$ such that if T is a finite simple group of order $\geq c_0$ and G is an almost simple group with socle T, then for every $g_1, \dots, g_d \in G$ we have $w(Tg_1, \dots, Tg_d) \neq \{1\}$.*

This easily yields the following.

Corollary 4.2. *With notation as above, there exists $c_1 = c_1(w)$ such that if $|T| \geq c_1$ then $|w(Tg_1, \ldots, Tg_d)| > 1$.*

Proof. Define $v(x_1, \ldots, x_{2d}) = w(x_1, \ldots, x_d)w(x_{d+1}, \ldots, x_{2d})^{-1}$.
Setting $c_1(w) = c_0(v)$ the result follows. □

Now let $w = x_{n_1}^{e_1} \cdots x_{n_l}^{e_l}$ be a reduced word of length $l \geq 1$ in F_d. This means that $n_1, \ldots, n_l \in \{1, 2 \ldots, d\}$, $e_i = \pm 1$, and $n_i = n_{i+1}$ implies $e_i = e_{i+1}$.

Let G be a group, and let $\alpha_1, \ldots, \alpha_l \in \mathrm{Aut}\,(G)$. The map $G^d \to G$ given by

$$(g_1, \ldots, g_d) \mapsto \alpha_1(g_{n_1})^{e_1} \cdots \alpha_l(g_{n_l})^{e_l}$$

is called a *generalized word function* (allowing w to be non-reduced) or an *automorphic word map* on G^d based on w (see [Se, §1.3] and [B1, 1.2.1]).

Let $T \lhd G$ and consider the word map $w : G^d \to G$. Its restriction to $Tg_1 \times \cdots \times Tg_d$ (where $g_i \in G$) can be regarded as a map $T^d \to G$ whose image lies in a T-coset of G. Indeed, more is true, namely:

Lemma 4.3. *With the above notation, the map*

$$(t_1, \ldots, t_d) \mapsto w(g_1, \ldots, g_d)^{-1} w(t_1 g_1, \ldots, t_d g_d)$$

is an automorphic word map $T^d \to T$ based on w.

This well known observation is easily proved by induction on l.

Lemma 4.4. *Given integers $j_1, \ldots, j_l$, there exist nonnegative integers $j_1', \ldots, j_l' \leq m - \lceil m/l \rceil$ such that*

$$j_1 - j_1' \equiv \cdots \equiv j_l - j_l' \pmod{m}.$$

Proof. Without loss of generality, we may assume $0 \leq j_1 \leq \cdots \leq j_l < m$. Setting $j_{l+1} = m + j_1$, we have $j_{i+1} - j_i \geq 0$ for $i = 1, 2, \ldots, l$, so $j_{r+1} - j_r \geq m/l$ for some $1 \leq r \leq l$, so setting

$$j_i' = \begin{cases} m - j_{r+1} + j_i & \text{if } 1 \leq i \leq r \\ j_i - j_{r+1} & \text{if } r+1 \leq i \leq m, \end{cases}$$

the lemma follows. □

We now prove Theorem 1.7.

Proof. For every group T as above, there exists a prime p, an adjoint split simple algebraic group $\underline{G}_0$ over $\mathbb{F}_p$ with root system Φ, and a generalized Frobenius endomorphism F of $\underline{G}_0$ such that T is the derived group of $\underline{G}_0(\overline{\mathbb{F}_p})^F$. Let S denote a fixed set of generators of the coordinate ring A of the adjoint Chevalley group scheme with root system Φ over $\mathbb{Z}$. Let F_p denote the p-Frobenius map of $\underline{G}_0$. The condition that F is a generalized Frobenius endomorphism means that some positive power of F is a positive power of F_p.

Fix cosets of $Tg_1, \dots, Tg_d$ of T in $\mathrm{Aut}\,(T)$ and let $w \in F_d$ be as above. By Lemma 4.3 there exist $\alpha_1, \dots, \alpha_l \in \mathrm{Aut}\,(T)$ such that for $(t_1, \dots, t_d) \in T^d$ the map

$$(t_1, \dots, t_d) \mapsto w(g_1, \dots, g_d)^{-1} w(t_1 g_1, \dots, t_d g_d) = \alpha_1(t_{n_1})^{e_1} \cdots \alpha_l(t_{n_l})^{e_l} \tag{4}$$

is an automorphic word map $T^d \to T$ based on w.

Each α_i can be written $F_p^{j_i} \beta_i$, where β_i is a product of a diagonal automorphism and a graph automorphism. By Corollary 4.2, the map (4) is not constant, assuming T is large enough given w. We can define (4) as the restriction to $T \subset \underline{G}_0(\overline{\mathbb{F}}_p)$ of a morphism $\underline{G}_0^d \to \underline{G}_0$ defined over $\overline{\mathbb{F}}_p$ in terms of the comultiplication map on the coordinate ring $A \otimes \mathbb{F}_p$ of $\underline{G}_0$, the inverse map on $A \otimes \mathbb{F}_p$, and the various endomorphism maps on $A \otimes \overline{\mathbb{F}}_p$ giving rise to the α_i. This morphism cannot be constant, and by Proposition 2.5, its degree (with respect to $S_p \coprod \cdots \coprod S_p$) is $O(p^j)$, where $j = \max_{1 \le i < d} j_i$. There exists $s \in S_p$ such that the composition of s and $\underline{G}_0^d \to \underline{G}_0$ gives a non-constant morphism

$$\underline{G}_0^d \to \mathbb{A}^1 \tag{5}$$

defined over $\overline{\mathbb{F}}_p$, and the degree of this morphism is $O(p^j)$.

At this point, we divide the problem into the Chevalley-Steinberg case and the Suzuki-Ree case. In the former, we can identify $\underline{G}_0(\overline{\mathbb{F}}_p)^F$ with $\underline{G}(\mathbb{F}_q)$, where $\underline{G}$ is an adjoint simple algebraic group over $\mathbb{F}_q$, $q = p^m$, which becomes isomorphic to $\underline{G}_0$ after extension of scalars from $\mathbb{F}_q$ to $\mathbb{F}_{q^k}$ and $\mathbb{F}_p$ to $\mathbb{F}_{q^k}$ respectively. As F_p^m acts as a (possibly trivial) graph automorphism of T, we may assume $0 \le j_i < m$ for all i. Applying a suitable power of F_p to $\alpha_1(t_{n_1})^{e_1} \cdots \alpha_l(t_{n_l})^{e_l}$ and using Lemma 4.4, we may assume without loss of generality that $0 \le j_i \le m - \lceil m/l \rceil$ for all i. Thus, the $S_p \coprod \cdots \coprod S_p$-degree of the element of $A^{\otimes d} \otimes \overline{\mathbb{F}}_p$ defining (5) is $O(p^{m - \lceil m/l \rceil})$, where by Proposition 2.5, the implicit constant does not depend on p.

We define a generating set for the coordinate ring of $\underline{G}$ by choosing a basis of the adjoint representation of $\underline{G}$ and letting the generators correspond to matrix entries in the adjoint representation with respect to this basis. Extending to $\overline{\mathbb{F}}_p$, this generating set may not be the same as that obtained by extending S_p, but it is equivalent. By Proposition 3.2, the theorem holds for Chevalley and Steinberg groups.

For the Suzuki and Ree cases, we can again use Lemma 4.4 to assume that the non-constant morphism (5) has degree $O(p^{m - \lceil m/l \rceil})$. Applying Proposition 3.3 to this map, we deduce the theorem in these cases.

□

We now deduce the following more general result.

Theorem 4.5. *For any non-trivial word $w = w(x_1, \dots, x_d)$ there exist $N, \epsilon > 0$ depending only on w such that, if T is any finite simple group of order $\ge N$, then for all $g_1, \dots, g_d, g \in \mathrm{Aut}\,(T)$,*

$$\left| \{(t_1, \dots, t_d) \in T^d \mid w(t_1 g_1, \dots, t_d g_d) = g\} \right| \le |T|^{d - \epsilon}.$$

Proof. If T is an alternating group of degree larger than some function $f(w)$ of w, or a classical group of rank larger than $f(w)$, then the conclusion follows from Theorem 3.1.2 in [B1]. Indeed, the latter result shows that, under our assumption on T, there is $\epsilon = \epsilon(w) > 0$ such that all fibers of all automorphic word maps on T^d based on w have size $\leq |T|^{d-\epsilon}$, which implies the required conclusion.

Otherwise T is a simple group of Lie type of rank $\leq f(w)$, and the result follows by Theorem 1.7.

□

Theorem 4.5 strengthens Proposition 4.1 of Nikolov: not only there is no fixed coset identity w in large almost simple groups, the probability of w attaining any fixed value g on each subset $Tg_1 \times \ldots \times Tg_d$ tends to zero very fast as $|T| \to \infty$.

We conclude with the following probabilistic consequence.

Corollary 4.6. *For any non-trivial word w there exist $N, \epsilon > 0$ depending only on w such that, for any almost simple group G of order at least N and any element $g \in G$ we have*

$$p_{w,G}(g) \leq |G|^{-\epsilon}.$$

Proof. Since T is generated by two elements we trivially have $|G| \leq |\mathrm{Aut}\,(T)| \leq |T|^2$ (much better upper bounds hold, of course). Therefore it suffices to show that, for some $N, \epsilon > 0$ depending on w, if $|T| \geq N^{1/2}$ then $p_{w,G}(g) \leq |T|^{-2\epsilon}$.

This follows from Theorem 4.5 above, by summing up the probabilities over all subsets $Tg_1 \times \ldots \times Tg_d$ of G^d. □

From this, we can immediately deduce Theorem 1.8 in the case of almost simple groups.

5. Semisimple groups and proof of main theorem

In this section we prove Theorem 1.8 and Proposition 1.9 and then use them to deduce Theorems 1.1 and 1.10.

As in §4, we let $w = x_{n_1}^{e_1} \cdots x_{n_l}^{e_l}$ be a reduced word of length l in F_d.

Proof of Theorem 1.8. It is well known that $\mathrm{Aut}\,(T^k) = \mathrm{Aut}\,(T) \wr S_k$. We fix a d-tuple of cosets of T^k in G: $T^k g_1, \ldots, T^k g_d$ and $g_i = (h_{i1}, \ldots, h_{ik}).\sigma_i$ where $h_{ij} \in \mathrm{Aut}\,(T)$ and $\sigma_i \in S_k$.

Let $w = x_{n_1}^{e_1} \ldots x_{n_l}^{e_l}$, where $1 \leq n_i \leq d$, $e_i = \pm 1$. Note that $n_i = n_{i+1}$ implies $e_i = e_{i+1}$, since w is reduced. For $i = 1, \ldots, d$ let $s_i = (t_{i1}, \ldots, t_{ik}) \in T^k$.

Fix $g \in G$. If $w(s_1 g_1, \ldots, s_d g_d) = g$ then

$$(s_{n_1} g_{n_1})^{e_1} \cdots (s_{n_l} g_{n_l})^{e_l} = g,$$

so

$$((t_{n_1 1}h_{n_1 1}, \dots, t_{n_1 k}h_{n_1 k})\sigma_{n_1})^{e_1} \cdots ((t_{n_l 1}h_{n_l 1}, \dots, t_{n_l k}h_{n_l k})\sigma_{n_l})^{e_l} = g.$$

Moving the σ_{n_i} terms all the way to the right of the LHS we can express this equation in the form

$$\tau_1((t_{n_1 1}h_{n_1 1}, \dots, t_{n_1 k}h_{n_1 k}))^{e_1} \cdots \tau_l((t_{n_l 1}h_{n_l 1}, \dots, t_{n_l k}h_{n_l k}))^{e_l}\sigma = g, \tag{6}$$

where $\tau_i, \sigma \in S_k$ do not depend on the t_{ij}. Note that $\tau_i((t_{n_i 1}h_{n_i 1}, \dots, t_{n_i k}h_{n_i k}))$ is simply

$$(t_{n_i\tau_i(1)}h_{n_i\tau_i(1)}, \dots, t_{n_i\tau_i(k)}h_{n_i\tau_i(k)}).$$

Let $H \lhd G$ be the kernel of the permutation action of G on the k factors of T. Then $G/H \leq S_k$ and $T^k \leq H \leq \mathrm{Aut}\,(T)^k$.

The equation (6) above has no solutions unless σ is the image of g in G/H. In the latter case it can be written as a system of k equations. We claim that each of these is a reduced equation of length l applied to a d-tuple of cosets of T. Indeed, if $(n_i, \tau_i(j)) = (n_{i+1}, \tau_{i+1}(j))$ then $n_i = n_{i+1}$ so $e_i = e_{i+1}$. Taking all of these equations together, each variable t_{ij} appears at most l times.

We obtain k subsets of $\{t_{ij}\}$, each of size at most l, such that each variable t_{ij} belongs to at most l subsets. Therefore each subset intersects at most $l(l-1)$ other subsets. Starting with one subset, deleting the other subsets it intersects, choosing a new remaining subset (if there is one), deleting the other subsets it intersects, choosing a new remaining one (if possible), and so on, we construct at least $m := \left\lceil \frac{k}{l^2-l+1} \right\rceil$ pairwise disjoint subsets. These subsets correspond to m equations in disjoint sets of variables, which are clearly independent.

We now apply Theorem 4.5 which implies the case $k = 1$ of Theorem 1.8. We conclude that there exist $N, \delta > 0$ depending only on w such that, if $|T| \geq N$, then any one of the k equations discussed above holds with probability $\leq |T|^{-\delta}$. Since our system of equations contains at least m independent equations, we obtain

$$p_{w,G}(g) \leq (|T|^{-\delta})^m \leq |T^k|^{-\delta/(l^2-l+1)}.$$

Setting $\epsilon = \delta/(l^2 - l + 1) > 0$ we complete the proof of Theorem 1.8.

□

The following result, which is Lemma 2.2 of [BCP], will play a key role in the proof of Proposition 1.9 and Theorem 1.1.

Lemma 5.1. *Fix $c \geq 6$. A permutation group of degree k without composition factors isomorphic to A_i with $i > c$ has order at most c^{k-1}.*

Proof of Proposition 1.9. Fix a positive integer f. Among all chief factors G_1/G_2 of G (where $G_i \lhd G$) corresponding to non-abelian composition factors of order $\geq f$ choose one of minimal index (namely $|G : G_1|$ is minimal).

Write $G_1/G_2 = T^k$ for $k \geq 1$ and a finite simple group T. Let C be the centralizer in G of G_1/G_2 and let $K = G/C$. Then $T^k \leq K \leq \mathrm{Aut}\,(T^k) = \mathrm{Aut}\,(T) \wr S_k$. Let H be the kernel of the permutation action of G on the k copies of T. Then $G_1 \leq H \lhd G$ and $G/H \leq S_k$. By the minimality of $|G : G_1|$ it follows that all non-abelian composition factors of G/H have order $< f \leq |T|$.

Let $c \geq 6$ be minimal such that G/H does not have a composition factor A_i with $i > c$. Then $|G/H| \leq c^{k-1}$ by Lemma 5.1. If A_i $(i \geq 5)$ is a composition factor of G/H then $2^i < |A_i| < f \leq |T|$. This shows (assuming $f \geq 64$ as we may) that $c \leq \log |T|$ (where logarithms are to the base 2).

It is well known that $|\mathrm{Out}\,(T)| \leq \log |T|$ for all finite simple groups T. We conclude that

$$|K| \leq |\mathrm{Aut}\,(T)|^k |G/H| \leq |T|^k |\mathrm{Out}\,(T)|^k (\log |T|)^k \leq |T^k| (\log |T|)^{2k}.$$

Now, given $\delta > 0$ choose $f = f(\delta)$ such that $\log t \leq t^{\delta/2}$ for all $t \geq f$. Since $|T| \geq f$ we obtain

$$|K| \leq |T^k|^{1+\delta}.$$

Therefore $|T^k| \geq |K|^{1-\delta}$, completing the proof.

□

Combining results 1.9 and 1.8 we obtain the following.

Theorem 5.2. *For every word $1 \neq w \in F_d$ there exist constants $N, \epsilon > 0$ depending only on w such that, if G is a finite group with a non-abelian composition factor S with $|S| \geq N$, then G has a quotient K with $|K| \geq |S|$ such that $p_{w,K}(g) \leq |K|^{-\epsilon}$ for all $g \in K$.*

Proof. Let $N(w), \epsilon(w)$ be as in Theorem 1.8. Fix ϵ with $0 < \epsilon < \epsilon(w)$. Define $\delta = 1 - \epsilon/\epsilon(w)$ and let $f = f(\delta)$ be as in Proposition 1.9. Set $N = \max(f, N(w))$.

Now let G, S be as in the theorem. Since $|S| \geq f$, Proposition 1.9 (applied with $|S|$ in the role of f) shows that there is a quotient K of G and a finite simple group T with $|T| \geq |S|$ such that $T^k \leq K \leq \mathrm{Aut}\,(T^k)$ and $|T^k| \geq |K|^{1-\delta}$.

By Theorem 1.8 we have

$$p_{w,K}(g) \leq |T^k|^{-\epsilon(w)} \leq |K|^{-(1-\delta)\epsilon(w)} = |K|^{-\epsilon}$$

for every $g \in G$, as required. □

We now prove Theorem 1.1.

Proof. This follows immediately from Theorem 5.2 applied to finite quotients of Γ with non-abelian composition factors of orders tending to infinity.

□

Proof of Theorem 1.10. We first prove part (i).

Theorem 1.1.2 of [B2] shows that, for certain words w, including x^k (k odd) and $[x_1, \dots, x_d]$, if G is a finite group, $p_{w,G}(g) \geq \epsilon > 0$ for some $g \in G$, and T is a finite simple group, then the multiplicity of T as a composition factor of G is bounded above by some function $f_1(T, w, \epsilon)$ of T, w and ϵ only.

Now, by Theorem 1.8 (applied to a suitable quotient of G), if $p_{w,G}(g) \geq \epsilon > 0$ and T is a non-abelian composition factor of G, then $|T| \leq f_2(w, \epsilon)$ for a suitable function f_2. This implies that the product of the orders of all non-abelian composition factors of G is bounded above by

$$\prod_{T, |T| \leq f_2(w,\epsilon)} |T|^{f_1(T,w,\epsilon)} \leq f_3(w, \epsilon),$$

for a suitable function f_3.

Now $\mathrm{Soc}\,(G/\mathrm{Rad}\,(G))$ has the form $\prod_i T_i^{n_i}$ for (non-abelian) simple groups T_i, hence $|\mathrm{Soc}\,(G/\mathrm{Rad}\,(G))| \leq f_3(w, \epsilon)$. Since $G/\mathrm{Rad}\,(G)$ is embedded in $\mathrm{Aut}\,(\mathrm{Soc}\,(G/\mathrm{Rad}\,(G)))$ we obtain

$$|G/\mathrm{Rad}\,(G)| \leq M,$$

where $M = f_4(w, \epsilon) = f_3(w, \epsilon)^{\log f_3(w,\epsilon)}$. This proves part (i).

For part (ii), note that, since the Haar measure is σ-additive, there exists an odd integer $k > 0$ such that the measure of the elements g of the profinite group G satisfying $g^k = 1$ is positive. The required conclusion now follows from part (i).

□

References

[Ab] M. Abért, Group laws and free subgroups in topological groups, *Bull. London Math. Soc.* **37** (2005), 525–534.

[BCP] L. Babai, P.J. Cameron and P.P. Pálfy, On the orders of primitive groups with restricted nonabelian composition factors, *J. Algebra* **79** (1982), 161–168.

[Bo] A. Borel, On free subgroups of semisimple groups, *Enseign. Math.* **29** (1983), 151–164.

[B1] A. Bors, Fibers of automorphic word maps and an application to composition factors, arXiv math:1608.00131.

[B2] A. Bors, Fibers of word maps and the multiplicities of nonabelian composition factors, arXiv math:1703.00408.

[DPSS] J.D. Dixon, L. Pyber, Á. Seress, A. Shalev, Residual properties of free groups and probabilistic methods, *J. reine angew. Math. (Crelle's)* **556** (2003), 159–172.

[Ep] D.B.A. Epstein, Almost all subgroups of a Lie group are free, *J. Algebra* **19** (1971), 261–262.

[SGA3] M. Demazure M. and A. Grothendieck, *Schémas en Groupes III*, Lecture Notes in Math. **153**, Springer Verlag, Berlin, 1970.

[FT] W. Feit and J.G. Thompson, Solvability of groups of odd order, *Pacific J. Math.* **13** (1963), 775–1029.

[Fu] W. Fulton, Intersection theory, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Springer-Verlag, Berlin, 1984.

[GT] R.M. Guralnick and Ph. H. Tiep, Effective results on the Waring problem for finite simple groups, *Amer. J. Math.* **137** (2015), no. 5, 1401–1430.

[La] M. Larsen, Word maps have large image, *Israel J. Math.* **139** (2004), 149–156.

[LS1] M. Larsen and A. Shalev, Word maps and Waring type problems, *J. Amer. Math. Soc.* **22** (2009), 437–466.

[LS2] M. Larsen and A. Shalev, Fibers of word maps and some applications, *J. Algebra* **354** (2012), 36–48.

[LS3] M. Larsen and A. Shalev, Probabilistic identities and a probabilistic Tits alternative, *Algebra and Number Theory* **10** (2016), 1359–1371.

[LST] M. Larsen, A. Shalev and Pham Huu Tiep, The Waring problem for finite simple groups, *Annals of Math.* **174** (2011), 1885–1950.

[LiS] M.W. Liebeck and A. Shalev, Diameters of finite simple groups: sharp bounds and applications, *Annals of Math.* **154** (2001), 383–406.

[LOST] M.W. Liebeck, E.A. O'Brien, A. Shalev, and Pham Huu Tiep, The Ore conjecture, *J. Eur. Math. Soc.* **12** (2010), 939–1008.

[Ma] A. Mann, Finite groups containing many involutions, *Proc. Amer. Math. Soc.* **122** (1994), 383–385.

[Ne] P.M. Neumann, Two combinatorial problems in group theory, *Bull. London Math. Soc.* **21** (1989), 456–458.

[N1] N. Nikolov, Strong approximation methods, *Lectures on profinite topics in group theory*, London Math. Soc. Stud. Texts **77**, Cambridge Univ. Press, Cambridge 2011, pp. 63–97.

[N2] N. Nikolov, Verbal width in anabelian groups, *Israel J. Math.*, **216** (2016), 847-876.

[Pi] R. Pink, Strong approximation for Zariski dense subgroups over arbitrary global fields, *Comment. Math. Helv.* **75** (2000), 608–643.

[Se] D. Segal, *Words: Notes on Verbal Width in Groups*, London Math. Soc. Lecture Note Series **361**, Cambridge University Press, Cambridge, 2009.

[Sh] A. Shalev, Word maps, conjugacy classes, and a noncommutative Waring-type theorem, *Annals of Math.* **170** (2009), 1383–1416.

[Sz] B. Szegedy, Almost all finitely generated subgroups of the Nottingham group are free, *Bull. London Math. Soc.* **37** (2005), 75–79.

[We] B. Weisfeiler, Strong approximation for Zariski-dense subgroups of semisimple algebraic groups, *Annals of Math.* **120** (1984), 271–315.

E-mail address: mjlarsen@indiana.edu

Department of Mathematics, Indiana University, Bloomington, IN 47405, U.S.A.

E-mail address: shalev@math.huji.ac.il

Einstein Institute of Mathematics, Hebrew University, Givat Ram, Jerusalem 91904, Israel