J. Comb. Algebra 5 (2021), 93–122
DOI 10.4171/JCA/51

Words, permutations, and the nonsolvable length of a finite group

Alexander Bors* and Aner Shalev**

Abstract. We study the impact of certain identities and probabilistic identities on the structure of finite groups. More specifically, let w be a nontrivial word in d distinct variables and let G be a finite group for which the word map $w_G\colon G^d \to G$ has a fiber of size at least $\rho|G|^d$ for some fixed $\rho > 0$. We show that, for certain words w, this implies that G has a normal solvable subgroup of index bounded above in terms of w and ρ. We also show that, for a larger family of words w, this implies that the nonsolvable length of G is bounded above in terms of w and ρ, thus providing evidence in favor of a conjecture of Larsen.

Along the way we obtain results of independent interest on permutation groups; e.g. we show, roughly, that most elements of large finite permutation groups have large support.

Mathematics Subject Classification (2020). 20E10, 20P05; 20B05, 20D06, 20F22.

Keywords. Finite groups, nonsolvable length, probabilistic identities, identities, word maps.

1. Introduction

The purpose of this paper is twofold. The first is to obtain new results on permutation groups, with emphasis on the support of their elements. The second is the study of word maps and the impact of identities and probabilistic identities on the structure of finite groups. It will turn out that these two goals are related: our results on permutation groups, which are of independent interest, can be applied in the context of word maps and probabilistic identities.

For a permutation group $P \leq \mathrm{Sym}(\Omega)$ and $\sigma \in P$ we let $\mathrm{supp}(\sigma)$ denote the number of points moved by σ and $\mathrm{supp}(P)$ the number of points moved by some element of P. We also let $\mathrm{fix}(\sigma)$ denote the number of fixed points of σ, and $\deg(P) := |\Omega|$.

Theorem 1.1. *Let $P \leq \mathrm{Sym}(\Omega)$ be a permutation group (where P and Ω are not assumed to be finite). Let c be a positive integer, and suppose $\mathrm{supp}(\sigma) \leq c$ for all $\sigma \in P$. Then:*

*The first author was supported by the Austrian Science Fund (FWF), project J4072-N32 "Affine maps on finite groups".

**The second author acknowledges the support of ISF grant 686/17, BSF grant 2016072 and the Vinik chair of mathematics which he holds.

(1) $|P| \leq c!$, *with equality if and only if* $P \cong S_c$ *acting on* $\mathrm{supp}(P)$*;*

(2) $\mathrm{supp}(P) \leq 2(c-1)$.

Some remarks are in order. First, conclusions (1) and (2) above are immediate if P is transitive, since then P contains a fixed-point-free permutation σ, so

$$|\Omega| = \mathrm{supp}(\sigma) \leq c.$$

Secondly, if $c = 1$ then all the elements of P have support zero, so $P = 1$.

Finally, we claim that the bound in conclusion (2) above is also best possible at least when $c = 2^k$ for some $k \in \mathbb{N}^+$. To show this we use error correcting codes. Let $H_k \leq \mathbb{F}_2^{2^k}$ be the $[2^k, k, 2^{k-1}]_2$-Hadamard code (see e.g. [16, p. 248]). By its definition, it is clear that there is a unique coordinate where all elements of H_k are 0; we project H_k onto the $2^k - 1$ other coordinates, resulting in a subspace $\widetilde{H}_k \leq \mathbb{F}_2^{2^k-1}$ (which we regard as an additive group) with the following properties:

- every nonzero element of $\widetilde{H}_k$ has exactly 2^{k-1} nonzero entries (equal to 1);
- for each $i \in \{1, \ldots, 2^k - 1\}$, there is an element of $\widetilde{H}_k$ having entry 1 in the i-th coordinate.

Set

$$\Omega := \{1, \ldots, 2^k - 1\} \times \{0, 1\}.$$

Consider the function $f : \widetilde{H}_k \to \mathrm{Sym}(\Omega)$, where $f(x_1, \ldots, x_{2^k-1})$ is the product of the transpositions $((i, 0), (i, 1))$ for those $i \in \{1, \ldots, 2^k - 1\}$ where $x_i = 1$. Then, f is an injective group homomorphism, so the image $P := f[\widetilde{H}_k]$ is actually a subgroup of $\mathrm{Sym}(\Omega)$, and it satisfies

$$\mathrm{supp}(P) = |\Omega| = 2(2^k - 1)$$

and that all nontrivial elements of P have support size exactly $2 \cdot 2^{k-1} = 2^k$.

Let $C \in \mathbb{N}$, and let $P \leq \mathrm{Sym}(\Omega)$ be a permutation group. We denote by $\mathrm{SB}_C(P)$ the set of all $\sigma \in P$ whose support on Ω is of size at most C.

Theorem 1.2. *There is a function* $f : (0, 1] \times \mathbb{N} \to [1, \infty)$ *such that the following holds: Let* $\rho \in (0, 1]$, $C \in \mathbb{N}$, *and assume that* $P \leq \mathrm{Sym}(\Omega)$ *is a permutation group of finite degree such that* $|\mathrm{SB}_C(P)| \geq \rho |P|$. *Then,*

$$|P| \leq f(\rho, C).$$

Indeed, one may choose f *to be the following function:*

$$f(\rho, C) := \left(\lfloor \rho^{-1} + C + 1 \rfloor !\right)^{\lceil 8(C - \log \rho) \rceil}.$$

In other words, the above result shows that, if P is a finite permutation group, and ρ is the probability that a random permutation in P has support at most C, then the size of P is bounded above by $f(\rho, C)$.

The next result provides refined bounds on $|P|$ in terms of additional parameters. Let t denote the number of orbits of $P \leq \mathrm{Sym}(n)$, and let r denote the rank of P (namely the number of orbits on ordered pairs of points). Clearly $r \geq t^2$.

Proposition 1.3. *With the above notation we have:*

(1) *The probability that a random element* $\sigma \in P$ *satisfies* $\mathrm{supp}(\sigma) > (1-\epsilon)n$ *is at least* $1 - t/(\epsilon n)$ *for any* $0 < \epsilon < 1$. *Thus this probability tends to* 1 *as* $t = o(n)$.

(2) *The probability that a random element* $\sigma \in P$ *satisfies* $\mathrm{supp}(\sigma) > (1-\epsilon)n - t$ *is at least* $1 - (r - t^2)/(\epsilon^2 n^2)$ *for any* $0 < \epsilon < 1$. *Thus this probability tends to* 1 *as* $r - t^2 = o(n^2)$.

Note that statement (1) of Proposition 1.3 implies that

$$\deg(P) \leq t \cdot \mathbf{P}(\mathrm{supp}(\sigma) \leq C)^{-1} + C,$$

which, adopting the notation from Theorem 1.2 above, yields that $\deg(P) \leq t\rho^{-1} + C$, and so

$$|P| \leq \lfloor t\rho^{-1} + C \rfloor ! .$$

Similarly, statement (2) of Proposition 1.3 implies that, with the above notation, we have $\deg(P) \leq \sqrt{r - t^2}\rho^{-1} + t + C$, which yields

$$|P| \leq \lfloor \sqrt{r - t^2}\rho^{-1} + t + C \rfloor ! .$$

We now turn to our main results on word maps and probabilistic identities, some of which apply the above results on permutation groups.

The impact of identities on the structure of groups has been a central research topic for over a century. Major examples include the Burnside problems and their solutions, the theory of group varieties, as well as parts of combinatorial and geometric group theory.

In the realm of finite groups, Zelmanov's solution to the Restricted Burnside Problem bounds the order of a d-generator finite group satisfying the power identity $x^n \equiv 1$ in terms of d and n [17, 18]. The Hall–Higman reduction of this problem to p-groups involves bounding the p-length of solvable groups satisfying this identity for all primes p [4]. A recent related result of Segal bounds the generalized Fitting height of finite groups satisfying $x^n \equiv 1$ in terms of n [13, Theorem 10].

More generally, in recent years there has been extensive interest in probabilistic identities (defined below) of finite and residually finite groups. Finitely generated linear groups which satisfy a probabilistic identity were shown in [7] to be virtually solvable. Arbitrary residually finite groups satisfying a probabilistic identity were shown in [8] (using results from [2]) to have nonabelian upper composition factors of bounded size. Probabilistically nilpotent finite and infinite groups were recently studied in [14] and in [10]. See also [9] for additional results on probabilistic aspects of word maps.

It is easy to see that every finite group G has a normal series each of whose factors is either solvable or a direct product of nonabelian finite simple groups. The smallest number of nonsolvable factors in a shortest such series is defined by Khukhro and Shumyatsky in [6] to be the *nonsolvable length of* G, and is denoted by $\lambda(G)$ (see also Section 2 below for an alternative definition, which was also already given in [6, first paragraph of the introduction]); while this concept was explicitly introduced and studied in [6], the idea of writing a finite group G as an extension of two finite groups with smaller nonsolvable lengths for inductive purposes is already implicit in the Hall–Higman paper, see [4, proof of Theorem 4.4.1].

In this paper we present some ideas relating identities and probabilistic identities in finite groups with the nonsolvable length, and sometimes with the index of the solvable radical. We combine some machinery already developed by the first author in [3] (building on earlier work of Nikolov from [11]) with some new methods. Let us now explain this in some more detail.

For a positive integer d, denote by $\mathrm{F}(X_1, \ldots, X_d)$ the free group freely generated by $X_1, \ldots, X_d$. Elements of these groups are called *words*. For the definition of *probabilistic identity*, let $w \in \mathrm{F}(X_1, \ldots, X_d)$ be a nontrivial word. Then for every (not necessarily finite) group G, one has the word map $w_G \colon G^d \to G$, induced by substitution into w. If G is finite and $g \in G$, it makes sense to define

$$p_{w,G}(g) := \frac{1}{|G|^d} |\{(g_1, \ldots, g_d) \in G^d \mid w_G(g_1, \ldots, g_d) = g\}|,$$

the proportion in G^d of the fiber of g under w_G. For profinite groups G, $p_{w,G}(g)$ denotes the (normalized) Haar measure (in G^d) of the fiber $w_G^{-1}(g)$. We say that G *satisfies a probabilistic identity with respect to* w *and* $\rho \in (0, 1]$ if and only if there is an element $g \in G$ such that $p_{w,G}(g) \geq \rho$. A residually finite group is said to satisfy a probabilistic identity if its profinite completion satisfies a probabilistic identity.

In this paper, we will be interested in the following property of nontrivial words:

Definition 1.4. A nontrivial word $w \in \mathrm{F}(X_1, \ldots, X_d)$ is called *nonsolvable-length-bounding* (or NLB for short) if and only if there is a function $f_w \colon (0, 1] \to [0, \infty)$ such that for every $\rho \in (0, 1]$ and every finite group G, if G satisfies a probabilistic identity with respect to w and ρ, then $\lambda(G) \leq f_w(\rho)$.

We can now state the following.

Conjecture 1.5. *All nontrivial words are NLB.*

This conjecture, due to Michael Larsen (private communication), seems very challenging, in view of the fact that it is even unknown for $\rho = 1$, namely when w is an identity of G.

Conjecture 1.6. *The nonsolvable length of a finite group which satisfies a nontrivial identity* $w \equiv 1$ *is bounded above in terms of* w.

See also the last paragraph of [13] for a related problem, where the nonsolvable length is replaced by the generalized Fitting height.

Conjecture 1.6 is reduced to bounding the Fitting height $h(G)$ of finite solvable groups G satisfying a nontrivial identity w in terms of w alone; indeed this reduction follows from [6, Corollary 1.2].

It turns out that these two conjectures are related to problems regarding support of elements in finite permutation groups, discussed above.

In [3], the first author studied another property of nontrivial words w, that of being *multiplicity-bounding* (or *MB* for short), see [3, Definition 1.1.1]. This just means that if a finite group G satisfies a probabilistic identity with respect to w and $\rho \in (0, 1]$, then for each nonabelian finite simple group S, the multiplicity of S as a composition factor of G is bounded from above in terms of w, ρ and S. Several stronger and weaker properties than that of being MB were also studied in [3], such as the ones in the last two enumeration points of the following definition:

Definition 1.7. Let $w \in \mathrm{F}(X_1, \ldots, X_d)$ be a nontrivial word.

(1) A *variation of* w is a word obtained from w by "splitting variables", i.e. by adding, for each $i \in \{1, \ldots, d\}$, to each occurrence of $X_i^{\pm 1}$ in w some second index.

(2) For a nonabelian finite simple group S, we say that w *is a coset identity over* S if and only if there are $\alpha_1, \ldots, \alpha_d \in \mathrm{Aut}(S)$ such that $w_{\mathrm{Aut}(S)}$ is constant on the Cartesian product $\prod_{i=1}^{d} S\alpha_i$ of cosets of S in $\mathrm{Aut}(S)$.

(3) w is called *weakly multiplicity-bounding* (or *WMB* for short) if and only if w is not a coset identity over any nonabelian finite simple group.

(4) w is called *very strongly multiplicity-bounding* (or *VSMB* for short) if and only if every variation of w is WMB.

For example, the word $X_{1,5}^{-1} X_{2,17}^{-1} X_{1,5} X_{2,4}$ is a variation of the commutator word $X_1^{-1} X_2^{-1} X_1 X_2$. Note that our definition of a variation slightly differs from the one in [3, Definition 2.4(2)], which included a technical restriction on the second indices which one can assume w.l.o.g. anyway, but we will not need it here.

By a result of Larsen and the second author [8, Theorem 5.2], if a finite group G satisfies a probabilistic identity with respect to w and ρ, then the orders of the nonabelian composition factors of G are bounded from above in terms of w and ρ. Letting $\mathrm{Rad}(G)$ denote the solvable radical of a finite group G (namely the largest solvable normal subgroup of G), this implies the following.

Corollary 1.8. *A nontrivial word w is MB if and only if the assumption that a finite group G satisfies a probabilistic identity with respect to w and ρ implies that the radical index $[G : \mathrm{Rad}(G)]$ is bounded from above in terms of w and ρ. In particular, if w is MB then it is NLB.*

The proof of this result will be given in Subsection 4.2 for the reader's convenience.

Hence, [3, Theorem 1.1.2] provides us with some examples of NLB words. Also by [3, Theorem 1.1.2(1)], the shortest nontrivial words which are not MB are of the form x^8 where x is a variable. We will, however, be able to show that such words are NLB, and the crucial observation is that while these words are not MB, in particular not VSMB, they are "almost" VSMB, in the following exact sense:

Definition 1.9. Let $w \in \mathrm{F}(X_1, \ldots, X_d)$ be a nontrivial word. w is called *almost very strongly multiplicity-bounding* (or *almost VSMB* for short) if and only if every *proper* variation w' of w (i.e. such that the number of variables occurring in w' is strictly larger than the number of variables in w) is WMB.

Our first main result relates the concepts of almost VSMB and NLB words:

Theorem 1.10. *Almost VSMB words are NLB.*

Thus, almost VSMB words satisfy Conjecture 1.5. This theorem is proved using Theorem 1.2 above on permutation groups.

Using the above result, Corollary 1.8 and [3, Theorem 1.1.2(3)], the following is immediate:

Corollary 1.11. *Let $w \in \mathrm{F}(X_1, \ldots, X_d)$ be a nontrivial word of length at most 8. Then w is NLB.*

Theorem 1.10 and Corollary 1.11 provide evidence in favor of Larsen's conjecture mentioned above. We note that while X_1^{12} is also not MB, the authors cannot exclude the possibility that all words of lengths 9, 10 and 11 are VSMB, in particular NLB, thus possibly allowing to replace the constant 8 in Corollary 1.11 by 11. However, compared to studying words of lengths up to 8 as done by the first author in [3, Section 6], the computational cost of doing so even just for words of length 9 is considerable and would most likely require a medium- to large-scale parallel computation. Still, with some more theoretical machinery, we will at least be able to show the following:

Corollary 1.12. *Let $w \in \mathrm{F}(X_1, \ldots, X_d)$ be a nontrivial word of length at most 11. Then there is a constant $L_w \in \mathbb{N}$ such that if a finite group G satisfies the identity $w \equiv 1$, then $\lambda(G) \leq L_w$.*

Thus words of length at most 11 satisfy Conjecture 1.6. The proof of Corollary 1.12 is based on a result allowing one to infer, under certain assumptions on a nontrivial word w, that if a finite group H without nontrivial solvable normal subgroups satisfies the identity $w \equiv 1$, then the so-called *permutation part of H* (see Definition 4.1.1(1) below) satisfies a shorter identity. This result is formulated in detail in Subsection 5.1 as Theorem 5.1.4.

Apart from new techniques for relating (probabilistic) identities with the nonsolvable length, we will also give infinitely many new (i.e. not already implicit in [3, Theorem 1.1.2(1)]) examples of both MB and non-MB words. Recall that the power words x^e are MB for all odd e (as shown in the above reference). Answering [3, Question 7.1] we show the following.

Theorem 1.13. *Let x be a variable. Then the following hold:*

(1) *Let m be a positive integer such that every prime divisor l of m satisfies $l \equiv 1 \pmod{225}$. Then x^{2m} is MB.*

(2) *Let e be a positive integer with $e > 4$ and $4 \mid e$. Then x^e is not MB.*

Obtaining a better understanding for which positive integers e the word x^e is (or is not) MB is of intrinsic interest, but it also relates to bounding $\lambda(G)$ in terms of the group exponent $\exp(G)$, see Subsection 6.1. We note that Theorem 1.13(2) partially contradicts the first author's result [3, Theorem 1.1.2(1)]; more precisely, [3, Theorem 1.1.2(1)] wrongly states that x^{20} is MB, but it is not. However, as clarified in an erratum on [3] prepared by the first author, [3, Theorem 1.1.2(1)] does become true if one replaces the set $\{8, 12, 16, 18\}$ in its statement by $\{8, 12, 16, 18, 20\}$ (so 20 is the only exponent e for which the original version of [3, Theorem 1.1.2(1)] makes a wrong statement on the MB property status of x^e). Except for the paragraph at hand, whenever we cite [3, Theorem 1.1.2(1)] in our paper (as we already did above), we are actually always referring to the above mentioned corrected version of it.

Now that we have stated our main results, we give an overview of their dependencies on the classification of finite simple groups (CFSG) and consequences or precursor results thereof. In the following list of bullet points, the word "elementary" means "does not require the CFSG":

- Our results on permutation groups, Theorems 1.1 and 1.2 as well as Proposition 1.3, are elementary.

- Corollary 1.8 relies on [8, Theorem 5.2], which uses the CFSG.

- Theorem 1.10 depends on Lemmas 4.1.2 and 4.1.3, the former of which requires the Schreier Conjecture, whereas the latter is elementary.

- Corollary 1.11 relies on [3, Theorem 1.1.2(3)], which in turn is based on [11, Proposition 7], which uses the CFSG.

- Corollary 1.12 relies on Corollary 1.11, so it also depends on the CFSG. The auxiliary results Theorem 5.1.4 and Proposition 5.2.1, which are also used in the proof of Corollary 1.12, are elementary, though.

- The second statement in Theorem 1.13 is elementary, but the first depends on a reduction argument from [11, proof of Proposition 7], which uses the CFSG. As for auxiliary results, Lemma 6.1.1 requires the Feit–Thompson Theorem, and Proposition 6.1.4 depends on Lemma 6.1.1. Lemmas 6.1.3, 6.2.1 and 6.2.2 are elementary, although the fact that Lemma 6.1.3 implies Lemma 6.1.1 does use the Feit–Thompson Theorem.

We conclude the Introduction with a consequence of Conjecture 1.5, that for every word $w \neq 1$ and every $\rho \in (0, 1]$, the nonsolvable length of finite groups satisfying a probabilistic identity with respect to w and ρ is bounded above in terms of w and ρ. A

profinite group G is said to be *randomly free* if, for any $n \geq 1$ and randomly chosen elements $g_1, \ldots, g_n \in G$ (with respect to the normalized Haar measure on G), the probability that $g_1, \ldots, g_n$ freely generate a (discrete) free subgroup of rank n is 1. The nonsolvable length of a profinite group G is defined to be the supremal nonsolvable length of the finite quotients of G (with respect to open subgroups).

Conjecture 1.5 readily implies that *profinite groups of infinite nonsolvable length are randomly free*. Indeed, this implication follows from [8, Lemma 1.4]. The above (conditional) conclusion may be regarded as an extension of [8, Theorem 1.3], showing that profinite groups are randomly free provided the sizes of their nonabelian upper composition factors are unbounded.

This paper is organized as follows. In Section 2 we introduce some notation. Section 3 is devoted to permutation groups and the support of their elements. This is where results 1.1, 1.2 and 1.3 are proved. In Section 4 we study probabilistic identities and prove Theorem 1.10 and Corollary 1.11. Section 5 is devoted to identities and the proof of Corollary 1.12. Finally, in Section 6, we prove Theorem 1.13 as well as a few results on the impact of power word identities on the group structure. In particular we show there that the nonsolvable length of a finite group is bounded above by the exponent of its Sylow 2-subgroups.

2. Some notation and prerequisites

We first discuss an equivalent, but more explicit (though also more technical) definition of $\lambda(G)$.

Definition 2.1. Let G be a finite group.

(1) We denote by $\mathrm{Rad}(G)$ the *solvable radical of* G, the largest solvable normal subgroup of G.

(2) We denote by $\mathrm{Soc}(G)$ the *socle of* G, the subgroup of G generated by all the minimal normal subgroups of G.

(3) We define sequences $(G_k(G))_{k\geq 1}$, $(R_k(G))_{k\geq 1}$, $(H_k(G))_{k\geq 1}$, and $(T_k(G))_{k\geq 1}$ of characteristic sections of G recursively as follows:

   (a) $G_1(G) := G$.

   (b) For $k \geq 1$, $R_k(G) := \mathrm{Rad}(G_k(G))$.

   (c) For $k \geq 1$, $H_k(G) := G_k(G)/R_k(G)$.

   (d) For $k \geq 1$, $T_k(G) := \mathrm{Soc}(H_k(G))$.

   (e) For $k \geq 2$, $G_k(G) := H_{k-1}(G)/T_{k-1}(G)$.

We call a finite group H *semisimple* if and only if $\mathrm{Rad}(H)$ is trivial, i.e. if and only if H has no nontrivial solvable normal subgroups. For the basic structure theory of

finite semisimple groups (from which several of the subsequently listed facts follow), see [12, pp. 89ff].

For every finite group G, the groups $R_k(G)$ are by definition all solvable, the groups $H_k(G)$ are semisimple, and the groups $T_k(G)$ are direct products of nonabelian finite simple groups. Moreover, since $H_k(G)$ embeds into the automorphism group of $T_k(G)$, we have that $T_k(G)$ is trivial if and only if $H_k(G)$ is trivial, so there is a unique non-negative integer $\lambda'(G)$ such that $T_1(G), \ldots, T_{\lambda'(G)}(G)$ are all nontrivial and $T_k(G) = \{1\}$ for $k > \lambda'(G)$. Actually, $\lambda'(G) = \lambda(G)$, by [6, first paragraph in the introduction].

We now introduce some more notation and terminology that will be used throughout the paper. We denote by $\mathbb{N}$ the set of natural numbers (including 0) and by $\mathbb{N}^+$ the set of positive integers. When $f: X \to Y$ is a function and $M \subseteq X$, then $f_{|M}$ denotes the restriction of f to M, and $f[M]$ denotes the element-wise image of M under f. Euler's constant will be denoted by e (which is to be distinguished from the variable e). For $c > 1$, we denote by $\log_c$ the base c logarithm, and log denotes $\log_e$. For a set Ω, the symmetric group on Ω is denoted by $\mathrm{Sym}(\Omega)$, and for $n \in \mathbb{N}^+$, $\mathrm{Sym}(n)$ denotes the symmetric group on $\{1, \ldots, n\}$. The group of units of a field K is denoted by K^*, and the algebraic closure of K by $\bar{K}$. For a prime power q, the finite field with q elements is denoted by $\mathbb{F}_q$. For a subset M of a finite group G, we denote by $\exp(M)$ the least common multiple of the orders of the elements of M. Finally, for a nonabelian finite simple group S and a word $w \in \mathrm{F}(X_1, \ldots, X_d)$, a *coset word equation with respect to w over S* is an equation of the form $w(s_1\alpha_1, \ldots, s_d\alpha_d) = \beta$ where $\alpha_1, \ldots, \alpha_d, \beta$ are fixed automorphisms of S, and $s_1, \ldots, s_d$ are variables ranging over S (so that the solution set of such an equation is always a subset of S^d).

3. Permutation groups

In this section we prove our results on permutation groups, the support of their elements and its distribution.

We first prove Theorem 1.1.

Proof. We first assume Ω is finite, and then deduce the result without this assumption.

Set $n = |\Omega|$. We may assume P has no orbits of size 1 in its action on Ω, since we may delete these orbits from Ω, thereby obtaining a subset $\Omega' = \mathrm{supp}(P)$, and regard P as a permutation group on Ω'.

Suppose P has t orbits on Ω, of sizes $n_1, \ldots, n_t > 1$. Then,

$$|P| \leq n_1! \cdots n_t!\,.$$

Since $\mathrm{supp}(\sigma) \leq c$ for all $\sigma \in P$, we have $\mathrm{fix}(\sigma) \geq n - c$ for all $\sigma \in P$. Consider the random variable $X = \mathrm{fix}(\sigma)$, where $\sigma \in P$ is assumed to be chosen

uniformly at random. Then, by the Cauchy–Frobenius Lemma ("The lemma that is not Burnside's"),

$$\mathrm{E}(X) = t.$$

This yields $t \geq n - c$. In fact, since $\mathrm{fix}(1) = n$ we have $t > n - c$, hence

$$\sum_{i=1}^{t}(n_i - 1) = n - t \leq c - 1.$$

Since $n_i \geq 2$ we have $n_i \leq 2(n_i - 1)$, and so

$$\mathrm{supp}(P) = \sum_{i=1}^{t} n_i \leq 2\sum_{i=1}^{t}(n_i - 1) \leq 2(c-1).$$

This proves part (2).

To prove part (1) we claim that

$$\prod_{i=1}^{t} n_i! \leq \left(1 + \sum_{i=1}^{t}(n_i - 1)\right)!,$$

with equality if and only if $t = 1$. We prove the claim by induction on t. The case $t = 1$ is trivial, so suppose $t \geq 2$.

Induction hypothesis yields

$$\prod_{i=1}^{t-1} n_i! \leq \left(1 + \sum_{i=1}^{t-1}(n_i - 1)\right)!.$$

Set $d = 1 + \sum_{i=1}^{t-1}(n_i - 1)$. Then, since $d \geq 2g$ we have

$$n_t! < (d+1)(d+2)\cdots(d+n_t-1).$$

Hence,

$$\begin{aligned}\prod_{i=1}^{t} n_i! &\leq d!n_t! < d!(d+1)(d+2)\cdots(d+n_t-1)\\ &= (d+n_t-1)! = \left(1 + \sum_{i=1}^{t}(n_i-1)\right)!,\end{aligned}$$

proving the claim. We conclude that

$$|P| \leq \prod_{i=1}^{t} n_i! \leq \left(1 + \sum_{i=1}^{t}(n_i-1)\right)! \leq c!,$$

with equality if and only if $P = S_c$, proving part (1).

Suppose now Ω is infinite. Let Ω' be the support of P, as above. We claim that Ω' is finite, hence, regarding P as a permutation group on Ω', we reduce to the finite case.

To prove the claim, choose $\sigma_1 \in P$ and denote its support by B_1. If $B_1 = \Omega'$ then Ω' has size at most c and we are done. Otherwise there exists $\sigma_2 \in P$ with support B_2 which is not contained in B_1. If $B_1 \cup B_2 = \Omega'$ we are done. Otherwise we proceed so that in step i we choose $\sigma_i \in P$ with support B_i which is not contained in $\Omega_{i-1} := \cup_{j=1}^{i-1} B_j$. Let $P_i \leq P$ be the subgroup generated by $\sigma_1, \ldots, \sigma_i$ and let $\Omega_i = \cup_{j=1}^{i} B_j$. Then Ω_i is finite (of size at most ci) and $P_i \leq \mathrm{Sym}(\Omega_i)$. By the finite case we have

$$|\Omega_i| = \mathrm{supp}(P_i) \leq 2(c-1).$$

Since the sequence $|\Omega_j|$ is increasing the process must stop, which means that, for some i, $\Omega' = \Omega_i$ is finite. This completes the proof. □

Next, we prove Theorem 1.2.

Proof. This is clear if $C = 0$, since then $\mathrm{SB}_C(P) = \{\mathrm{id}_\Omega\}$, whence $|\mathrm{SB}_C(P)| \geq \rho|P|$ is equivalent to $|P| \leq \rho^{-1}$, and

$$\rho^{-1} \leq \lfloor \rho^{-1} + 1 \rfloor \leq \lfloor \rho^{-1} + 1 \rfloor ! \leq \left(\lfloor \rho^{-1} + 1 \rfloor ! \right)^{\lceil 8 \log \rho^{-1} \rceil}.$$

The assertion is also clear if $C \geq \deg(P)$. So we may henceforth assume that $1 \leq C < \deg(P)$. We first show the following claim:

"If P is transitive, then $\deg(P) \leq \rho^{-1} + C$."

To see that this claim holds true, consider the random variable $X = \mathrm{fix}(\sigma)$, where $\sigma \in P$ is assumed to be chosen uniformly at random. Then as noted in the proof of Theorem 1.1, by the Cauchy–Frobenius Lemma, $\mathrm{E}(X) = 1$.

Moreover, the Markov inequality (see for instance [1, p. 265]) shows that, for each positive integer k,

$$\mathbf{P}(X \geq k) \leq \frac{\mathrm{E}(X)}{k} = \frac{1}{k}.$$

Applied with $k := \deg(P) - C$, this yields

$$\mathbf{P}(\sigma \in \mathrm{SB}_C(P)) \leq \frac{1}{\deg(P) - C},$$

so that

$$\rho \leq \frac{1}{\deg(P) - C},$$

or equivalently, $\deg(P) \leq \rho^{-1} + C$. This concludes the proof of the above claim.

The claim yields in particular that the asserted upper bound on $|P|$ holds when P is transitive. Let us now give an argument for general P. Let $\Omega = \Omega_1 \sqcup \cdots \sqcup \Omega_t$ be

the partition of Ω into the orbits of P. For $i = 1, \ldots, t$, denote by $P_i \leq \mathrm{Sym}(\Omega_i)$ the (transitive) image of P under the restriction homomorphism

$$\pi_i \colon P \to \mathrm{Sym}(\Omega_i), \quad \sigma \mapsto \sigma_{|\Omega_i}.$$

Observe that $\pi_i[\mathrm{SB}_C(P)] \subseteq \mathrm{SB}_C(P_i)$, and so $|\mathrm{SB}_C(P_i)| \geq \rho|P_i|$ as well. Hence if, for any $i \in \{1, \ldots, t\}$, one has $|\Omega_i| > \rho^{-1} + C$, one gets a contradiction to the above claim. So we may assume that $|\Omega_i| \leq \rho^{-1} + C$ for each $i = 1, \ldots, t$; in particular,

$$|P_i| \leq \lfloor \rho^{-1} + C \rfloor !.$$

Aiming for a contradiction, assume now additionally that

$$|P| > \left(\lfloor \rho^{-1} + C + 1 \rfloor !\right)^{\lceil 8(C - \log \rho) \rceil}.$$

Then,

$$\frac{|P|}{\left(\lfloor \rho^{-1} + C \rfloor !\right)^j} > 1$$

for $j = 0, 1, \ldots, \lceil 8(C - \log \rho) \rceil$, allowing us to choose, for $s := \lceil 8(C - \log \rho) \rceil$, a length s sequence $(i_1, \ldots, i_s)$ of pairwise distinct indices from $\{1, \ldots, t\}$ such that for each $j \in \{1, \ldots, s\}$,

$$c_j := |\pi_{i_j}[\ker(\pi_{i_1}) \cap \cdots \cap \ker(\pi_{i_{j-1}})]| \geq 2.$$

What this means is that among all the elements of P, there occur $c_1 \geq 2$ distinct values in the i_1-th coordinate, and after fixing any of the c_1 many values in the i_1-th coordinate and considering only such elements of P, there still occur $c_2 \geq 2$ distinct values in the i_2-th coordinate, and after fixing both the i_1-th and i_2-th coordinate, there still occur $c_3 \geq 2$ distinct values in the i_3-th coordinate, and so on.

Now consider $\pi \colon P \to \prod_{j=1}^{s} P_{i_j}$, the projection of P to the coordinates number $i_1, \ldots, i_s$. The image $\pi[P]$ still satisfies that

$$|\mathrm{SB}_C\left(\pi[P]\right)| \geq \rho|\pi[P]|,$$

but on the other hand,

$$\mathrm{SB}_C\left(\pi[P]\right) \subseteq \Bigg\{(\sigma_{i_1}, \ldots, \sigma_{i_s}) \in \pi[P] \leq \prod_{j=1}^{s} P_{i_j} \mid$$
$$\exists M \subseteq \{1, \ldots, s\} : \left(|M| = C \text{ and } \sigma_{i_r} = \mathrm{id}_{\Omega_r} \text{ for all } r \in \{1, \ldots, s\} \setminus M\right)\Bigg\}.$$

Letting $d_1 \geq d_2 \geq \cdots \geq d_s$ be such that the multisets $\{c_1, \ldots, c_s\}$ and $\{d_1, \ldots, d_s\}$ are equal, this yields the following upper bound on the proportion of elements in $\pi[P]$

with support size at most C:

$$\frac{1}{|\pi[P]|}|\operatorname{SB}_C(\pi[P])| \le \frac{1}{d_1\cdots d_s}\binom{s}{C}d_1\cdots d_C$$
$$= \frac{\binom{s}{C}}{d_{C+1}\cdots d_s} \le \frac{\left(\frac{es}{C}\right)^C}{2^{s-C}} = \frac{\left(\frac{2es}{C}\right)^C}{2^s}.$$

We thus get the desired contradiction if we can argue that

$$\frac{\left(\frac{2es}{C}\right)^C}{2^s} < \rho. \tag{3.1}$$

Recall that $s = \lceil 8(C - \log\rho)\rceil$, and set $s' := \frac{s}{C}$, so that $s = C\cdot s'$. Then,

$$\frac{\left(\frac{2es}{C}\right)^C}{2^s} = \frac{(2es')^C}{2^{s'C}} = \left(\frac{2es'}{2^{s'}}\right)^C = \left(\frac{es'}{2^{s'-1}}\right)^C,$$

and that last expression is strictly smaller than ρ if and only if

$$s' - \log_2 s' - 1 > \frac{1-\frac{1}{C}\log\rho}{\log 2}.$$

Now by definition,

$$s' = \frac{s}{C} = \frac{\lceil 8(C-\log\rho)\rceil}{C} \ge 8\Big(1 - \frac{\log\rho}{C}\Big) \ge 8,$$

and so

$$s' - \log_2 s' - 1 \ge \frac{1}{2}s'.$$

Hence, Formula (3.1) is implied by

$$s' > 2\cdot\frac{1-\frac{1}{C}\log\rho}{\log 2} = \frac{2}{\log 2}\Big(1 - \frac{1}{C}\log\rho\Big),$$

which is clear by definition of s'. □

We now prove Proposition 1.3.

Proof. The Markov inequality applied in the proof of the above theorem shows that, for any fixed $\epsilon > 0$ we obtain (substituting $k = \epsilon n$),

$$\mathbf{P}\big(\operatorname{supp}(\sigma) > (1-\epsilon)n\big) \ge 1 - t/k = 1 - \frac{t}{\epsilon n},$$

which tends to 1 provided $t = o(n)$. Part (1) follows.

For part (2) we use the second moment method for the random variable $X = \mathrm{fix}(\sigma)$ ($\sigma \in P$). Then $\mathrm{E}(X) = t$, and as is well known, by applying the Cauchy–Frobenius Lemma to the action of P on $\{1, \ldots, n\}^2$, one also gets $\mathrm{E}(X^2) = r$. Therefore,

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = r - t^2.$$

By the Chebyshev inequality (see for instance [1, p. 267]) we have

$$\mathbf{P}\big(|X - \mathrm{E}(X)| \geq k\big) \leq \frac{\mathrm{Var}(X)}{k^2}.$$

Writing $k = \epsilon n$ we obtain

$$\mathbf{P}\big(|X - t| < \epsilon n\big) \geq 1 - \frac{r - t^2}{\epsilon^2 n^2}.$$

Clearly, $|X - t| < \epsilon n$ implies $\mathrm{fix}(\sigma) < t + \epsilon n$, which yields

$$\mathrm{supp}(\sigma) = n - \mathrm{fix}(\sigma) > (1 - \epsilon)n - t.$$

The result follows. □

We conclude this section with the following example, which shows that (in the notation used in Proposition 1.3(2)) $r - t^2$ is not always in $o(n^2)$:

Example 3.1. Let $P = \mathrm{D}_6 = \mathrm{Sym}(3)$ in its regular action on itself (hence on 6 points). Then P is sharply 1-transitive. For $m \in \mathbb{N}^+$, $m \geq 2$, let $\mathcal{G}_m$ be the set of length m sequences $\vec{\sigma} = (\sigma_1, \ldots, \sigma_m) \in P^m$ such that $\mathrm{ord}(\sigma_i) = 2$ for $i = 1, \ldots, m$ and $|\{\sigma_1, \ldots, \sigma_m\}| \geq 2$. Note that each such sequence $\vec{\sigma}$ is a generating sequence for P. Set $k_m := |\mathcal{G}_m|$, denote by $\pi_i^{(m)}$, for $i = 1, \ldots, m$, the projection $P^m \to P$ to the i-th coordinate, and let

$$P_m := \langle (\pi_1^{(m)}(\vec{\sigma}))_{\vec{\sigma} \in \mathcal{G}_m}, \ldots, (\pi_m^{(m)}(\vec{\sigma}))_{\vec{\sigma} \in \mathcal{G}_m} \rangle \leq P^{\mathcal{G}_m} \cong P^{k_m}.$$

Then P_m is a k_m-fold subdirect power of P; in particular, all orbits Ω_j, for $j = 1, \ldots, k_m$, of P_m are of length 6. Note also that the m listed generators of P_m are pairwise distinct, so that $|P_m| \geq m$. For each $j \in \{1, \ldots, k_m\}$ and each $\omega \in \Omega_j$, the point stabilizer $(P_m)_\omega$ consists only of even length products of the listed generators of P_m; in particular, for each $l \in \{1, \ldots, k_m\}$, the restriction of each element of $(P_m)_\omega$ to Ω_l is contained in the unique index 2 subgroup of the corresponding (sharply 1-transitive) action of $P = \mathrm{D}_6$ on Ω_l. Hence, $(P_m)_\omega$ is intransitive on each orbit Ω_l of P_m, whence each Cartesian product $\Omega_j \times \Omega_l$ of orbits of P splits into at least two distinct orbits under the component-wise action of P^2. In particular,

$$r(P_m) \geq 2t(P_m)^2,$$

and so

$$r(P_m) - t(P_m)^2 \geq t(P_m)^2 = k_m^2 = \left(\frac{6k_m}{6}\right)^2 = \left(\frac{\deg(P_m)}{6}\right)^2 = \frac{1}{36}\deg(P_m)^2.$$

4. Probabilistic identities

This section is concerned with the proofs of Theorem 1.10 and Corollary 1.11. We will introduce another word property, that of being *permutation-part-bounding* (see Definition 4.1.1(2)), which is stronger than the property of being NLB (see Lemma 4.1.2(2)). Using Theorem 1.2, we will be able to show that almost VSMB words are permutation-part-bounding (see Lemma 4.1.3). For the sake of completeness, we will also give a proof of Corollary 1.8 in Subsection 4.2.

4.1. Permutation-part-bounding words. We now introduce another word property that will be relevant for the proof of Theorem 1.10:

Definition 4.1.1. Consider the following notations and concepts:

(1) Let H be a nontrivial finite semisimple group, say

$$S_1^{n_1} \times \cdots \times S_r^{n_r} \le H \le \mathrm{Aut}(S_1^{n_1} \times \cdots \times S_r^{n_r}) = \big(\mathrm{Aut}(S_1) \wr \mathrm{Sym}(n_1)\big) \times \cdots \times \big(\mathrm{Aut}(S_r) \wr \mathrm{Sym}(n_r)\big),$$

where $S_1, \ldots, S_r$ are pairwise nonisomorphic nonabelian finite simple groups and $n_1, \ldots, n_r \in \mathbb{N}^+$. For $i = 1, \ldots, r$, denote by $\pi_i \colon H \to \mathrm{Aut}(S_i) \wr \mathrm{Sym}(n_i)$ the projection to the i-th coordinate, and let H_i be the image of H under π_i, which is again semisimple, with socle $S_i^{n_i}$. We introduce the following notations for isomorphism invariants of H:

(a) $P(H) := H/\big(H \cap (\mathrm{Aut}(S_1)^{n_1} \times \cdots \times \mathrm{Aut}(S_r^{n_r}))\big)$ for the so-called *permutation part of* H, which we can view naturally as a subgroup of $\mathrm{Sym}(n_1) \times \cdots \times \mathrm{Sym}(n_r)$.

(b) $\mathrm{Perm}(H)$ for the multiset $\{P(H_1), \ldots, P(H_r)\}$, and

(c) $\mathrm{MPO}(H)$ for the number $\max\{|P(H_i)| \mid i = 1, \ldots, r\}$.

(2) Let $w \in \mathrm{F}(X_1, \ldots, X_d)$ be a nontrivial word. We say that w is *permutation-part-bounding* (or *PPB* for short) if and only if there is a function $f_w \colon (0, 1] \to [1, \infty)$ such that for every nonsolvable finite group G satisfying a probabilistic identity with respect to w and ρ, $\mathrm{MPO}(H_1(G)) \le f_w(\rho)$.

Clearly, MB words are PPB. Moreover, we have the following:

Lemma 4.1.2. *The following hold:*

(1) *Let G be a finite group. Then $H_2(G)$ is a section of $\prod_{P \in \mathrm{Perm}(H_1(G))} P$.*

(2) *PPB words are NLB.*

Proof. For (1): By definition,

$$H_2(G) = G_2(G)/R_2(G) = \big(H_1(G)/\operatorname{Soc}(H_1(G))\big)/\operatorname{Rad}\big(H_1(G)/\operatorname{Soc}(H_1(G))\big).$$

It is thus sufficient to show that $G_2(G) = H_1(G)/\operatorname{Soc}(H_1(G))$ has a solvable normal subgroup N such that $G_2(G)/N$ is isomorphic to a subgroup of $\prod_{P \in \mathrm{Perm}(H_1(G))} P$.

Letting $\mathrm{Soc}(H_1(G)) \cong S_1^{n_1} \times \cdots \times S_r^{n_r}$, where $S_1, \ldots, S_r$ are pairwise nonisomorphic nonabelian finite simple groups and $n_1, \ldots, n_r \in \mathbb{N}^+$, we may view, up to natural isomorphism,

$$S_1^{n_1} \times \cdots \times S_r^{n_r} \leq H_1(G) \leq \mathrm{Aut}(S_1^{n_1} \times \cdots \times S_r^{n_r}).$$

We then find that

$$N := \big((\mathrm{Aut}(S_1)^{n_1} \times \cdots \times \mathrm{Aut}(S_r)^{n_r}) \cap H_1(G)\big)/(S_1^{n_1} \times \cdots \times S_r^{n_r}) \hookrightarrow \mathrm{Out}(S_1)^{n_1} \times \cdots \times \mathrm{Out}(S_r)^{n_r}$$

is a suitable choice.

For (2): Let $w \in \mathrm{F}(X_1, \ldots, X_d) \setminus \{1\}$ be PPB, and assume that G is a finite group that satisfies a probabilistic identity with respect to that word w and some given $\rho \in (0, 1]$. We want to bound $\lambda(G)$ in terms of w and ρ. If G is solvable, then $\lambda(G) = 0$, so assume that G is nonsolvable. Then,

$$|\,\mathrm{MPO}(G/\,\mathrm{Rad}(G))| = |\,\mathrm{MPO}(H_1(G))| \leq f_w(\rho),$$

where f_w is as in the definition of PPB words. In other words, $|P| \leq f_w(\rho)$ for each $P \in \mathrm{Perm}(G/\,\mathrm{Rad}(G))$. Moreover, by [8, Theorem 5.2], there is an $N_w(\rho) > 0$ such that all nonabelian composition factors of G have order at most $N_w(\rho)$. In particular, the number of nonisomorphic simple direct factors in $\mathrm{Soc}(G/\,\mathrm{Rad}(G))$ is bounded from above by $N_w(\rho)$ (because for each $k \geq 1$, the number of isomorphism types of nonabelian finite simple groups up to order k is at most k, as the orders of nonabelian finite simple groups are even and for each given order, there are at most two nonisomorphic nonabelian finite simple groups of that order). Using statement (1), it follows that

$$60^{\lambda(G)-1} \leq |H_2(G)| \leq \Big|\prod_{P \in \mathrm{Perm}(G/\,\mathrm{Rad}(G))} P\Big| \leq f_w(\rho)^{N_w(\rho)},$$

and thus,

$$\lambda(G) \leq \frac{1}{\log 60} N_w(\rho) \log f_w(\rho) + 1. \qquad \square$$

In particular, the proof of Theorem 1.10 is now reduced to the following, which we will show next:

Lemma 4.1.3. *Almost VSMB words are PPB.*

Proof. Let w be an almost VSMB word, let $\rho \in (0, 1]$, and assume that a finite nonsolvable group G satisfies a probabilistic identity with respect to w and ρ. Then every quotient of G also satisfies a probabilistic identity with respect to w and ρ; in particular, writing

$$\mathrm{Soc}(H_1(G)) = S_1^{n_1} \times \cdots \times S_r^{n_r},$$

where $S_1, \dots, S_r$ are pairwise nonisomorphic nonabelian finite simple groups and $n_1, \dots, n_r \in \mathbb{N}^+$, for $i = 1, \dots, r$, the group $H_{1,i}(G)$, defined as the projection of

$$H_1(G) \leq \mathrm{Aut}(S_1^{n_1}) \times \cdots \times \mathrm{Aut}(S_r^{n_r})$$

to the i-th coordinate, satisfies a probabilistic identity with respect to w and ρ. Note that up to isomorphism,

$$S_i^{n_i} \leq H_{1,i}(G) \leq \mathrm{Aut}(S_i^{n_i}) = \mathrm{Aut}(S_i) \wr \mathrm{Sym}(n_i),$$

and that when setting

$$P_{1,i}(G) := P(H_{1,i}(G)) \hookrightarrow \mathrm{Sym}(n_i),$$

one has by definition that

$$\mathrm{Perm}(H_1(G)) = \{P_{1,i}(G), \dots, P_{1,r}(G)\}.$$

Our goal is to find an upper bound in terms of w and ρ on $\max\{|P_{1,i}(G)| \mid i = 1, \dots, r\}$.

To that end, fix $i \in \{1, \dots, r\}$, and for notational simplicity, write S instead of S_i, n instead of n_i, H instead of $H_{1,i}(G)$, and P instead of $P_{1,i}(G)$. For $\sigma \in P$, denote by $\mathrm{Supp}(\sigma)$ the set of points moved by σ (so that, using the notation from Section 3, $\mathrm{supp}(\sigma) = |\mathrm{Supp}(\sigma)|$). Recall from above that H satisfies a probabilistic identity with respect to w and ρ, so we can fix an element $h = (\beta_1, \dots, \beta_n)\psi \in H$ such that $p_{w,H}(h) \geq \rho$.

Note: If w is a *repetition-free word*, i.e. if the maximum multiplicity of a variable in w is 1 (no variable occurs more than once in w), then the probabilistic identity implies that $|H| \leq \rho^{-1}$; in particular, $|P| \leq \rho^{-1}$ then, and we are done. So we may assume that w is *not* repetition-free.

Writing $w = x_1^{\epsilon_1} \cdots x_\ell^{\epsilon_\ell}$, where ℓ is the length of w, $\epsilon_1, \dots, \epsilon_\ell \in \{\pm 1\}$ and $x_1, \dots, x_\ell \in \{X_1, \dots, X_d\}$, we can find indices $j_1, j_2 \in \{1, \dots, \ell\}$ with $j_1 < j_2$ such that $x_{j_1} = x_{j_2}$, $x_j \neq x_{j_1}$ for all $j \in \{j_1 + 1, \dots, j_2 - 1\}$, and the (possibly empty) word segment $x_{j_1+1}^{\epsilon_{j_1+1}} \cdots x_{j_2-1}^{\epsilon_{j_2-1}}$ is repetition-free. Moreover, for $j = 1, \dots, \ell$, define the word

$$v_j := \begin{cases} x_1^{\epsilon_1} \cdots x_{j-1}^{\epsilon_{j-1}} & \text{if } \epsilon_j = 1, \\ x_1^{\epsilon_1} \cdots x_j^{\epsilon_j} & \text{if } \epsilon_j = -1, \end{cases}$$

see also [3, Lemma 2.7] and our Notation 5.1.1(1), and set $v := v_{j_1}^{-1} v_{j_2}$, see also Notation 5.1.1(2). Note that by choice of j_1 and j_2, v is a nonempty reduced word in which some variable occurs with multiplicity 1.

We bound the number of solutions to the equation $w(y_1, \dots, y_d) = h$, where $y_1, \dots, y_d$ are variables ranging over H, in a $\mathrm{Soc}(H)$-coset-wise counting argument. More precisely, fix first a d-tuple $(\sigma_1, \dots, \sigma_d) \in P^d$. There are two

fundamentally different cases in the counting argument, according to whether or not $v(\sigma_1,\dots,\sigma_d) \in \mathrm{SB}_{C(\rho)}(P)$, where

$$C(\rho) := \ell^2 \cdot \frac{\log(2/\rho)}{\log(1 + 1/(N_w(\rho)^\ell - 1))}$$

and $N_w(\rho)$ is chosen such that all nonabelian composition factors of a finite group that satisfies a probabilistic identity with respect to w and ρ have order at most $N_w(\rho)$.

(1) Assume first that $v(\sigma_1,\dots,\sigma_d) \notin \mathrm{SB}_{C(\rho)}(P)$, i.e. that $\mathrm{supp}(v(\sigma_1,\dots,\sigma_d)) > C(\rho)$. For each $k = 1,\dots,d$, fix one of the $[(H \cap \mathrm{Aut}(S)^n) : S^n]$ many cosets of S^n in H that have permutation part σ_k, say with coset representative $(\alpha_{k,1},\dots,\alpha_{k,n})\sigma_k$, and consider the equation

$$\begin{aligned} w\big((s_{1,1}\alpha_{1,1},\dots,s_{1,n}\alpha_{1,n})\sigma_1,\dots,(s_{d,1}\alpha_{d,1},\dots,s_{d,n}\alpha_{d,n})\sigma_d\big) &= h \\ &= (\beta_1,\dots,\beta_n)\psi, \end{aligned}$$

where the $s_{k,t}$, for $k = 1,\dots,d$ and $t = 1,\dots,n$, are variables ranging over S. As described in [3, Lemma 2.7], this equation can be rewritten into the conjunction of the single word equation $w(\sigma_1,\dots,\sigma_d) = \psi$ and the system of coset word equations over S with respect to some variations of w whose t-th equation, for $t \in \{1,\dots,n\}$, looks like this:

$$\big(s_{\iota(1),\chi_1^{-1}(t)}\alpha_{\iota(1),\chi_1^{-1}(t)}\big)^{\epsilon_1} \cdots \big(s_{\iota(\ell),\chi_\ell^{-1}(t)}\alpha_{\iota(\ell),\chi_\ell^{-1}(t)}\big)^{\epsilon_\ell} = \beta_t,$$

where ι is the unique function $\{1,\dots,\ell\} \to \{1,\dots,d\}$ such that $x_j = X_{\iota(j)}$ for $j = 1,\dots,\ell$, and $\chi_j = v_j(\sigma_1,\dots,\sigma_d)$ for $j = 1,\dots,\ell$.

Hence, for each $t \in \mathrm{Supp}(v(\sigma_1,\dots,\sigma_d))$, the underlying word of the $\chi_{j_1}(t)$-th coset word equation in the above equation system is a proper variation of w, as follows by considering the j_1-th and j_2-th factors in the product on the left-hand side: $\iota(j_1) = \iota(j_2)$ (i.e. w has the same variable, possibly with different exponents ± 1, in those positions), but

$$\chi_{j_1}^{-1}(\chi_{j_1}(t)) = t \neq (\chi_{j_2}^{-1}\chi_{j_1})(t) = \chi_{j_2}^{-1}(\chi_{j_1}(t))$$

(so the second indices of the variables at those positions in the $\chi_{j_1}(t)$-th coset word equation are different). As w is assumed to be almost VSMB, this implies that each coset word equation labeled by an index from $\chi_{j_1}[\mathrm{Supp}(v(\sigma_1,\dots,\sigma_d))]$ is not universally solvable; in particular, since $|S| \le N_w(\rho)$, its proportion of solutions (among the variables that occur in it) is at most $1 - (1/N_w(\rho)^l)$.

But as in [3, proof of Lemma 2.12], since $|\chi_{j_1}[\mathrm{Supp}(v(\sigma_1,\dots,\sigma_d))]| > C(\rho)$, we can find at least $\lceil C(\rho)/\ell^2 \rceil$ pairwise distinct indices in $\chi_{j_1}[\mathrm{Supp}(v(\sigma_1,\dots,\sigma_d))]$ such that the corresponding equations in the above system have pairwise disjoint occurring variable sets (i.e. they are "pairwise independent"), and this implies that

the proportion of solutions (in S^{nd}) of the entire system of equations is at most

$$\left(1-\frac{1}{N_w(\rho)^\ell}\right)^{\lceil C(\rho)/\ell^2\rceil} \le \left(1-\frac{1}{N_w(\rho)^\ell}\right)^{C(\rho)/\ell^2} = \frac{\rho}{2},$$

where the equality is by definition of $C(\rho)$.

(2) Assume now that $v(\sigma_1,\dots,\sigma_d) \in \mathrm{SB}_{C(\rho)}(P)$. Then we do not give a nontrivial upper bound on the number of solutions per d-tuple of socle cosets with permutation parts $(\sigma_1,\dots,\sigma_d)$, but we note that since v contains some variable with multiplicity 1, the proportion of such d-tuples $(\sigma_1,\dots,\sigma_d)$ in P^d is exactly $\frac{1}{|P|}|\mathrm{SB}_{C(\rho)}(P)|$.

In combination, this yields the following:

$$\rho \le p_{w,H}(h) \le \frac{\rho}{2} + \frac{1}{|P|}|\mathrm{SB}_{C(\rho)}(P)|,$$

and thus,

$$\frac{1}{|P|}|\mathrm{SB}_{C(\rho)}(P)| \ge \frac{\rho}{2},$$

so that an application of Theorem 1.2 shows that $|P|$ can indeed be bounded from above in terms of w and ρ, as required. □

4.2. Proof of Corollary 1.8. Let G be a finite group. Assume first that

$$[G : \mathrm{Rad}(G)] \le C$$

for some constant $C > 0$. Then since $\mathrm{Rad}(G)$ is solvable (i.e. it only has abelian composition factors), for each nonabelian finite simple group S, the multiplicities of S in G and $G/\mathrm{Rad}(G)$ are the same. It follows that $[G : \mathrm{Rad}(G)]$, and hence C, is an upper bound on the product of the orders of the nonabelian composition factors of G, counted with multiplicities. In particular, the maximum multiplicity of a nonabelian composition factor of G is at most $\log_{60}(C)$. This shows the implication "⇐" in the first sentence of Corollary 1.8.

Now assume that for each nonabelian finite simple group S, the multiplicity of S in G is at most C_S for some constant $C_S > 0$ that may depend on S. Assume also that the maximum order of a nonabelian composition factor of G is bounded from above by another constant $C > 0$. Then let D be the maximum value of C_S, where S ranges over the (finitely many) nonabelian finite simple groups of order at most C, so that any nonabelian composition factor of G occurs with multiplicity at most D. It follows that the socle $T_1(G)$ of $G/\mathrm{Rad}(G)$, which is of the form $S_1^{n_1} \times \cdots \times S_r^{n_r}$ where $S_1,\dots,S_r$ are pairwise nonisomorphic nonabelian finite simple groups and $n_1,\dots,n_r \in \mathbb{N}^+$, satisfies

$$|T_1(G)| \le C^{Dr} \le C^{CD},$$

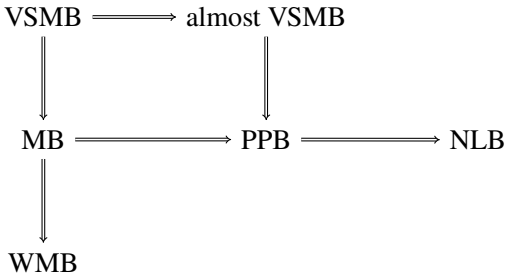
where the latter inequality uses that there are at most C distinct isomorphism types of nonabelian finite simple groups of order at most C (as was already observed in the proof of Lemma 4.1.2(2) above). This concludes the proof of the implication "$\Rightarrow$" in the first sentence of Corollary 1.8.

For the second sentence (the "In particular"), just observe that

$$[G : \mathrm{Rad}(G)] \geq \prod_{k=1}^{\infty} |T_k(G)| \geq 60^{\lambda(G)}.$$

This concludes the proof of Corollary 1.8.

We thus have the following implication diagram between the various word properties considered in this paper:



4.3. Proofs of Theorem 1.10 and Corollary 1.11. The proof of Theorem 1.10 is immediate by combining Lemmas 4.1.2 and 4.1.3. For Corollary 1.11, note that by [3, Theorem 1.1.2(3)], all nontrivial words w of length at most 8 are almost VSMB, so that we can conclude by an application of Theorem 1.10.

5. Identities

In this section, we are concerned with the proof of Corollary 1.12. This is based on a result stating that if each member of a certain subset of the variations of a given nontrivial word w is WMB, then there is a nontrivial word v of length strictly shorter than w (actually, v can be chosen to be some proper segment of w) such that if a finite semisimple group H satisfies the identity $w \equiv 1$, then the permutation part $P(H)$ satisifies the identity $v \equiv 1$. For the precise formulation of this result, see Theorem 5.1.4.

5.1. Segment identities. As noted in the Introduction, we will prove a result (Theorem 5.1.4 below) which will allow us to show that under certain assumptions, if a finite semisimple group H satisfies some identity $w \equiv 1$, then the permutation part $P(H)$ satisfies a shorter identity $v \equiv 1$, where v is some proper segment of w.

Let us first introduce some notations and terminology and then formulate and prove Theorem 5.1.4.

Notation 5.1.1. Let $w \in \mathrm{F}(X_1, \dots, X_d)$, say $w = x_1^{\epsilon_1} \cdots x_\ell^{\epsilon_\ell}$ where ℓ is the length of w, and for $i = 1, \dots, \ell$, $x_i \in \{X_1, \dots, X_d\}$ and $\epsilon_i \in \{\pm 1\}$.

(1) For $i = 1, \dots, \ell$, set

$$I_i(w) := \begin{cases} x_1^{\epsilon_1} \cdots x_{i-1}^{\epsilon_{i-1}} & \text{if } \epsilon_i = 1, \\ x_1^{\epsilon_1} \cdots x_i^{\epsilon_i} & \text{if } \epsilon_i = -1. \end{cases}$$

(2) For $1 \le i < j \le \ell$, set

$$\Delta_{i,j}(w) := I_i(w)^{-1} I_j(w).$$

Note the following two simple facts:

Remark 5.1.2. Using the notation from Notation 5.1.1, we note the following:

(1) The words $\Delta_{i,j}(w)$ are segments of w.

(2) $\Delta_{i,j}(w)$ is empty if and only if $j = i + 1$, $\epsilon_i = -1$ and $\epsilon_j = \epsilon_{i+1} = 1$. In particular, since w is reduced, $\Delta_{i,j}(w)$ is always nonempty if i and j are such that $x_i = x_j$.

(3) $\Delta_{i,j}(w) = w$ if and only if $i = 1$, $j = \ell$, $\epsilon_1 = 1$ and $\epsilon_\ell = -1$.

Definition 5.1.3. Let $w \in \mathrm{F}(X_1, \dots, X_d)$, with notation as in Notation 5.1.1. Moreover, let $w' = y_1^{\epsilon_1} \cdots y_\ell^{\epsilon_\ell}$ be a variation of w, and let $1 \le i < j \le \ell$. We say that w' is an (i, j)-*split variation of* w if and only if $x_i = x_j$ and $y_i \ne y_j$.

Theorem 5.1.4. *Let $w \in \mathrm{F}(X_1, \dots, X_d)$, with notation as in Notation 5.1.1. Also, assume that for some given $i, j \in \{1, \dots, \ell\}$ with $i < j$ and $x_i = x_j$, all (i, j)-split variations of w are WMB. Then, if a finite semisimple group H satisfies the identity $w \equiv 1$, the permutation part $P(H)$ satisfies the identity $\Delta_{i,j}(w) \equiv 1$. In particular, there is a nontrivial word $v \in \mathrm{F}(X_1, \dots, X_d)$ of length strictly smaller than ℓ such that $P(H)$ satisfies the identity $v \equiv 1$.*

Proof. The "In particular" follows from the main statement, as by Remark 5.1.2(1,2), $\Delta_{i,j}(w)$ is a nonempty segment of w, and so usually, one will simply choose $v := \Delta_{i,j}(w)$, unless $\Delta_{i,j}(w) = w$, which by Remark 5.1.2(3) can only happen if $w = xvx^{-1}$ with $v \in \mathrm{F}(X_1, \dots, X_d) \setminus \{1\}$ is not cyclically reduced, in which case H, and thus $P(H)$ satisfies the identity $v \equiv 1$. We thus focus on the proof of the main statement now.

Say, $\mathrm{Soc}(H) = S_1^{n_1} \times \cdots \times S_r^{n_r}$, where $S_1, \dots, S_r$ are pairwise nonisomorphic nonabelian finite simple groups and $n_1, \dots, n_r \in \mathbb{N}^+$. Then, H is a subdirect product of semisimple groups H_k, $k = 1, \dots, r$, such that $\mathrm{Soc}(H_k) = S_k^{n_k}$ for each k, and such that $P(H)$ is a subdirect product of the permutation parts $P(H_k)$,

for $k = 1, \dots, r$. Hence, it suffices to show that each $P(H_k)$ satisfies the identity $\Delta_{i,j}(w) \equiv 1$. This shows that we may assume w.l.o.g. that $\mathrm{Soc}(H) = S^n$ for some nonabelian finite simple group S and some $n \in \mathbb{N}^+$.

Aiming for a contradiction, we will also assume that $P(H)$ does *not* satisfy $\Delta_{i,j}(w) \equiv 1$. Then we can fix $\sigma_1, \dots, \sigma_d \in P(H)$ with $\Delta_{i,j}(w)(\sigma_1, \dots, \sigma_d) \neq \mathrm{id}$. Moreover, we fix $m_0 \in \{1, \dots, n\}$ with $\Delta_{i,j}(w)(\sigma_1, \dots, \sigma_d)(m_0) \neq m_0$, and set

$$m_1 := I_i(w)(\sigma_1, \dots, \sigma_d)(m_0).$$

Finally, we fix automorphism tuples

$$\vec{\alpha}_k = (\alpha_{k,1}, \dots, \alpha_{k,n}) \in \mathrm{Aut}(S)^n,$$

for $k = 1, \dots, d$, such that $\vec{\alpha}_k \sigma_k \in H$.

By assumption, we have that

$$w_H(S^n \vec{\alpha}_1 \sigma_1, \dots, S^n \vec{\alpha}_d \sigma_d) = \{1_H\}.$$

In particular, letting $s_{k,m}$, for $k = 1, \dots, d$ and $m = 1, \dots, n$, be variables ranging over S, then by [3, Lemma 2.7], we have that a certain system of n coset word equations over S in the variables $s_{k,m}$ is universally solvable, and setting

$$\chi_t := I_t(w)(\sigma_1, \dots, \sigma_d)$$

for $t = 1, \dots, l$ and denoting by ι the unique function $\{1, \dots, \ell\} \to \{1, \dots, d\}$ such that for $t = 1, \dots, \ell$, $x_t = X_{\iota(t)}$, one of the equations from the system is

$$\left(s_{\iota(1),\chi_1^{-1}(m_1)} \alpha_{\iota(1),\chi_1^{-1}(m_1)}\right)^{\epsilon_1} \cdots \left(s_{\iota(\ell),\chi_\ell^{-1}(m_1)} \alpha_{\iota(\ell),\chi_\ell^{-1}(m_1)}\right)^{\epsilon_\ell} = 1. \tag{5.1}$$

Note that by assumption, $\iota(i) = \iota(j)$, but also

$$\begin{aligned} \chi_j^{-1}(m_1) &= \left(I_j(w)(\sigma_1, \dots, \sigma_d)\right)^{-1}(m_1) \\ &= \left(\Delta_{i,j}(w)(\sigma_1, \dots, \sigma_d)^{-1} \cdot I_i(w)(\sigma_1, \dots, \sigma_d)^{-1}\right)(m_1) \\ &= \Delta_{i,j}(w)(\sigma_1, \dots, \sigma_d)^{-1}(m_0) \neq m_0 = \chi_i^{-1}(m_1). \end{aligned}$$

Hence, Equation (5.1) is a universally solvable coset word equation over S with respect to some (i, j)-split variation w' of w. But by assumption, S does not satisfy any coset identity with respect to w', which is the desired contradiction. □

5.2. A consequence of Theorem 5.1.4. Using Theorem 5.1.4, we can show the following, which will be used in the proof of Corollary 1.12:

Proposition 5.2.1. *Let $w \in \mathrm{F}(X_1, \dots, X_d)$, with notation as in Notation 5.1.1. Also, assume that for some given $k \in \{1, \dots, d\}$, $\mu_w(X_k) \leq 3$. Finally, let $i, j \in \{1, \dots, \ell\}$ with $i < j$ such that $x_i = x_j = X_k$. If a finite semisimple group H satisfies the identity $w \equiv 1$, then $P(H)$ satisfies $\Delta_{i,j}(w) \equiv 1$; in particular, $P(H)$ satisfies a nontrivial identity of length strictly shorter than ℓ.*

Proof. The proof of the "In particular" is as for Theorem 5.1.4. For the main statement: Since $\mu_w(X_k) \leq 3 < 2 \cdot 2$, in each (i, j)-split variation w' of w, there will be a variable that occurs with multiplicity exactly 1. Hence w' is VSMB, in particular WMB, by [3, Proposition 3.1(1)]. □

5.3. Proof of Corollary 1.12. By [3, Theorem 1.1.2(3)] and Corollary 1.11, it suffices to consider words w of lengths 9, 10 or 11. We start with the length 9 case. Then the existence of L_w (actually, with $L_w = 0$) is clear if w is a power of single variable. So we may also assume that w contains at least two distinct variables. But if the total number of variables occurring in w is at least 3, then since $9 < 3 \cdot 4$, there is a variable occurring with multiplicity at most 3 in w. Hence, by Proposition 5.2.1, $P(H_1(G))$ satisfies an identity $v \equiv 1$ for some word v of length at most 8. By Corollary 1.11, v is NLB, and so $P(H_1(G))$ satisfying $v \equiv 1$ entails that $\lambda(P(H_1(G)))$ (and thus $\lambda(G)$) is bounded from above by some constant, as required.

So we may henceforth assume that $w = w(x, y)$ is a two-variable word, and moreover (by an argument as in the previous paragraph, using Proposition 5.2.1), we may assume that each variable that occurs in w does so with multiplicity at least 4. Since $9 < 2 \cdot 5$, one of the two variables, say w.l.o.g. x, occurs with multiplicity exactly 4 in w. Using the notation of Notation 5.1.1 for w (with $l = 9$, of course), fix a pair (i, j) with $1 \leq i < j \leq 9$ and $x_i = x_j = x$.

We will now argue that each (i, j)-split variation w' of w is WMB. Since $\mu_w(x) = 4 < 3 \cdot 2$, at least one of the variables in w' derived from x, say x', must occur with multiplicity at most 2. If $\mu_{w'}(x') = 1$, w' is VSMB, in particular WMB, by [3, Proposition 3.1(1)]. So assume that $\mu_{w'}(x') = 2$. The segment between the two occurrences of $(x')^{\pm 1}$ in w' is of length at most 7, and thus it is VSMB by [3, Theorem 1.1.2(3)]. In view of this and [3, Proposition 3.1(2,3)], w' is VSMB, in particular WMB.

An application of Theorem 5.1.4 now yields that $P(H_1(G))$ satisfies an identity of the form $v \equiv 1$ where v is a word of length at most 8. Again, by Corollary 1.11, v is NLB, and so $\lambda(P(H_1(G)))$ is bounded from above by some constant.

The arguments for words of length $\ell \in \{10, 11\}$ are largely similar, so we only sketch them. The first paragraph of the above argument can almost literally be carried over, replacing 9 by ℓ, of course, and not only referring to Corollary 1.11 at the end, but also to the cases of length 9 resp. lengths 9 and 10 already done by then. In the two-variable case $w = w(x, y)$ with $\mu_w(x), \mu_w(y) \geq 4$, since $\ell < 2 \cdot 6$, we get that one of the two variables, say w.l.o.g. x, occurs with multiplicity 4 or 5 in w. When choosing the pair (i, j) with $1 \leq i < j \leq \ell$ with $x_i = x_j = x$, one must also choose it such that the difference $j - i$ is maximal among all such pairs. This way, in the third paragraph of the argument, it is ensured that the segment s between the two occurrences of x' in w' is of length at most $\ell - 3$ (not just $\ell - 2$, as in the argument for length 9 words). For $\ell = 10$, one can then conclude as in the length 9 case, and

for $\ell = 11$, one needs the additional observation that s cannot be an 8-th or (-8)-th power of a single variable, for then some variable (necessarily y) occurs in w with multiplicity at least 8, so that $\mu_w(x) \leq 3$, a contradiction.

6. Power words

The aim of this section is two-fold: Firstly, to discuss, in Subsection 6.1, a few results on the structure of finite groups G satisfying a power word identity $x^e \equiv 1$ (i.e. such that the group exponent $\exp(G)$ divides e). These results are closely related to observations, stated explicitly by Segal in [13, Theorem 10 and its proof] but already implicit in a paper by Hall and Higman [4, proof of Theorem 4.4.1], on the impact of power word identities on the generalized Fitting height of the group. Secondly, in Subsection 6.2, we will prove Theorem 1.13, which provides infinitely many new examples both of MB and of non-MB power words x^e.

6.1. Identities. It is clear by a result of Segal [13, Theorem 10] that for each positive integer e, if a finite group G satisfies the identity $x^e \equiv 1$ (in other words, if $\exp(G) \mid e$), then $\lambda(G)$ is bounded from above in terms of e (actually, Segal's result says that the same holds true if $\lambda(G)$ is replaced by the generalized Fitting height of G, which is an upper bound on $\lambda(G)$). Now Segal's proof uses the following, which is based on [4, proof of Theorem 4.4.1] and the Feit–Thompson theorem:

Lemma 6.1.1. *Let x be a variable, let $e \in \mathbb{N}^+$, and let H be a nontrivial finite semisimple group satisfying the identity $x^e \equiv 1$ (in particular, e is even). Then, $P(H)$ satisfies the identity $x^{e/2} \equiv 1$.* □

The aim of this subsection is two-fold: Firstly, to show a slightly stronger variant of Lemma 6.1.1 (see Lemma 6.1.3 below), and secondly, to use a Segal-like argument for gaining a simple explicit upper bound on the nonsolvable length $\lambda(G)$ in terms of $\exp(G)$ (see Proposition 6.1.4 below).

Let us start with the stronger version of Lemma 6.1.1, for which we introduce the following:

Definition 6.1.2. Let x be any fixed variable. Call a positive integer e *good* if and only if the word x^e is MB, and otherwise, call e *bad*. Moreover, for fixed $e \in \mathbb{N}^+$, denote by $\mathrm{BAD}(e)$ the set of all positive divisors of e that are bad.

Lemma 6.1.3. *Let x be a variable, let $e \in \mathbb{N}^+$, and let H be a nontrivial finite semisimple group satisfying the identity $x^e \equiv 1$ (in particular, e is bad). Then $P(H)$ satisfies the identity $x^{e/\gcd(\mathrm{BAD}(e))} \equiv 1$.*

Proof. We may w.l.o.g. assume that $S^n \leq H \leq \mathrm{Aut}(S^n)$ for some nonabelian finite simple group S and some $n \in \mathbb{N}^+$ (as H is, in general, a subdirect product of such groups, and likewise, $P(H)$ is a subdirect product of the permutation parts of those

groups). Fix $\sigma \in P(H)$. We will show that σ can only have cycles of lengths of the form $\frac{e}{d}$ where $d \in \mathrm{BAD}(e)$, and once we will have shown that, we will be done, as this implies that

$$\mathrm{ord}(\sigma) \mid \mathrm{lcm}_{d \in \mathrm{BAD}(e)} \frac{e}{d} = \frac{e}{\gcd(\mathrm{BAD}(e))}.$$

So let $\zeta = (i_1, \ldots, i_\ell)$ be a length ℓ cycle of σ. Note firstly that $\ell \mid e$, since $P(H)$, being a quotient of H, also satisfies the identity $x^e \equiv 1$. Now fix $(\alpha_1, \ldots, \alpha_n) \in \mathrm{Aut}(S)^n$ such that $(\alpha_1, \ldots, \alpha_n)\sigma \in H$. It follows that for all $s_1, \ldots, s_n \in S$,

$$\big((s_1\alpha_1, \ldots, s_n\alpha_n)\sigma\big)^e = 1,$$

and the expression on the left-hand side can be written as an element of $\mathrm{Aut}(S)^n$ whose i_ℓ-th entry is

$$\big(s_{i_\ell}\alpha_{i_\ell} s_{i_{\ell-1}}\alpha_{i_{\ell-1}} \cdots s_{i_1}\alpha_{i_1}\big)^{e/\ell},$$

which must in particular also be 1 for all choices of $s_{i_1}, \ldots, s_{i_\ell} \in S$. This shows that S satisfies a coset identity with respect to $x^{e/\ell}$, and so e/ℓ is bad by [3, Proposition 2.9(3)], i.e. $\ell = e/d$ for some $d \in \mathrm{BAD}(e)$, as required. □

Note that by [3, Corollary 5.2], all bad positive integers are even, and so in Lemma 6.1.3,

$$\gcd(\mathrm{BAD}(e)) \geq 2,$$

whence Lemma 6.1.3 does imply Lemma 6.1.1, as asserted above. While it is true that the greatest common divisor of all bad positive integers is 2 (since, for example, 8 and 18 are bad by [3, Theorem 1.1.2(1)]), and thus that Lemma 6.1.3 does not always provide strictly stronger information than Lemma 6.1.1, in some cases, it is better. As a somewhat extreme example, note that

$$\mathrm{BAD}(30) = \{30\}$$

by [3, Theorem 1.1.2(1)], and so by Lemma 6.1.3, H satisfying the identity $x^{30} \equiv 1$ implies that $P(H)$ is trivial (as opposed to it just satisfying the identity $x^{15} \equiv 1$, which is what Lemma 6.1.1 gives).

Using the bound from Lemma 6.1.1, we will now show, similarly to Segal's proof of [13, Theorem 10]:

Proposition 6.1.4. *For every finite group G, $\lambda(G) \leq \nu_2(\exp(G))$.*

Proof. By induction on $v := \nu_2(\exp(G))$. If $v = 0$, then G is solvable by the Feit–Thompson Theorem, so $\lambda(G) = 0$, and the bound is clear in that case. Now assume that $v \geq 1$, and also assume that G is nonsolvable (otherwise, again, $\lambda(G) = 0$ and the bound is clear). Then since G satisfies the identity $x^{\exp(G)} \equiv 1$, so does

$$H_1(G) = G/\mathrm{Rad}(G).$$

By Lemma 6.1.1, it follows that $P(H_1(G))$ satisfies the identity $x^{\exp(G)/2} \equiv 1$, and thus, by the induction hypothesis,

$$\lambda(G) - 1 = \lambda(P(H_1(G))) \leq \nu_2(\exp(G)/2) = \nu_2(\exp(G)) - 1,$$

which yields the desired bound, $\lambda(G) \leq \nu_2(\exp(G))$. □

6.2. New examples of MB and non-MB power words. In this subsection, we are concerned with the proof of Theorem 1.13. It relies on the following two lemmas of some independent interest:

Lemma 6.2.1. *Let $w \in \mathrm{F}(X_1, \ldots, X_d)$, let S be a nonabelian finite simple group, and let $\alpha_1, \ldots, \alpha_d \in \mathrm{Aut}(S)$. The following are equivalent:*

(1) $w(S\alpha_1, \ldots, S\alpha_d) \neq \{w(\alpha_1, \ldots, \alpha_d)\}$.

(2) $w(S\alpha_1, \ldots, S\alpha_d) \neq \{1\}$.

We note that by Lemma 6.2.1, [3, Corollary 5.2] may be viewed as a direct consequence of Nikolov's earlier result [11, Proposition 10].

Lemma 6.2.2. *Let $f \in \mathbb{N}^+$ be odd, let $S = \mathrm{PSL}_2(3^f)$, and let α be an automorphism of S with nontrivial diagonal part and whose field part is of order f. Then,* $\exp(S\alpha) = 4f$.

Computer calculations show that the statement of Lemma 6.2.2 is also true for $f \in \{2, 4\}$, so it might actually hold for all $f \in \mathbb{N}^+$. Let us now prove these two lemmas before proceeding with the proof of Theorem 1.13.

Proof of Lemma 6.2.1. For "1 ⇒ 2": We will show the contraposition: Assume that

$$w(S\alpha_1, \ldots, S\alpha_d) = \{1\}.$$

Then, since $w(\alpha_1, \ldots, \alpha_d) \in w(S\alpha_1, \ldots, S\alpha_d)$, it follows that $w(\alpha_1, \ldots, \alpha_d) = 1$, and so

$$w(S\alpha_1, \ldots, S\alpha_d) = \{1\} = \{w(\alpha_1, \ldots, \alpha_d)\},$$

as required.

For "2 ⇒ 1": Let $\beta \in w(S\alpha_1, \ldots, S\alpha_d) \setminus \{1\} \subseteq \mathrm{Aut}(S) \setminus \{1\}$. Then there is an $s \in S$ with $\beta^s \neq \beta$. But,

$$\beta^s \in w(S\alpha_1, \ldots, S\alpha_d)^s = w((S\alpha_1)^s, \ldots, (S\alpha_d)^s) = w(S\alpha_1, \ldots, S\alpha_d).$$

It follows that $|w(S\alpha_1, \ldots, S\alpha_d)| \geq 2$. In particular, $w(S\alpha_1, \ldots, S\alpha_d)$ cannot be equal to the singleton $\{w(\alpha_1, \ldots, \alpha_d)\}$, as required. □

Proof of Lemma 6.2.2. We view $S = \mathrm{PSL}_2(3^f)$ as the subgroup of $\mathrm{PGL}_2(3^f)$ consisting of the images under the canonical projection $\mathrm{GL}_2(3^f) \to \mathrm{PGL}_2(3^f)$ of all matrices in $\mathrm{GL}_2(3^f)$ whose determinant is a square in $\mathbb{F}_{3^f}$. Note that the order

of every element of $S\alpha$ is divisible by f and that by [5, Proposition 4.1], $(S\alpha)^f$ lies in some copy of $\mathrm{PGL}_2(3) \cong \mathrm{Sym}(4)$ inside the simple Chevalley group $A_1(\overline{\mathbb{F}_3})$ containing S. In particular, each element order in $(S\alpha)^f$ lies in $\{1, 2, 3, 4\}$, so we are done if we can show the following two statements:

- For all $s \in S$, $\mathrm{ord}((s\alpha)^f) \notin \{1, 3\}$.
- There is an $s \in S$ with $\mathrm{ord}((s\alpha)^f) = 4$.

Let us start with the first statement. Write $\alpha = s'\delta\phi$ where $s' \in S$, δ is any fixed element of $\mathrm{PGL}_2(3^f) \setminus \mathrm{PSL}_2(3^f)$, and ϕ is a field automorphism of order f (not necessarily the entry-wise Frobenius automorphism $x \mapsto x^3$). Then for each $s \in S$,

$$(s\alpha)^f = (ss'\delta\phi)^f = (ss'\delta)(ss'\delta)^{\phi} \cdots (ss'\delta)^{\phi^{f-1}},$$

and so, since $ss'\delta \in \mathrm{PGL}_2(3^f) \setminus \mathrm{PSL}_2(3^f)$ and f is odd, it follows that the order of $(s\alpha)^f$ is even. This concludes the proof of the first statement.

For the second statement, denote again by ϕ the common field part of the elements of $S\alpha$. Since $\mathrm{PGL}_2(3) \cong \mathrm{Sym}(4)$, we have that

$$\mathrm{PGL}_2(3) \setminus \mathrm{PSL}_2(3) = \mathrm{PGL}_2(3) \setminus \mathrm{PGL}_2(3)'$$

contains an element $g\zeta \,\mathrm{GL}_2(3)$ of order 4. Observe that the lift $g \in \mathrm{GL}_2(3)$ of $g\zeta \,\mathrm{GL}_2(3)$ must have determinant -1, for its determinant must be a non-square in $\mathbb{F}_3$. But since f is odd, -1 is also a non-square in $\mathbb{F}_{3^f}$, so $\beta := g\zeta \,\mathrm{GL}_2(3^f)$ lies in $\mathrm{PGL}_2(3^f) \setminus \mathrm{PSL}_2(3^f)$ and also has order 4. Since β is centralized by ϕ and $\gcd(4, f) = 1$, it follows that $\beta\phi \in S\alpha$ has order $4f$, as required. □

We are now ready for the:

Proof of Theorem 1.13. Let us start with the proof of statement (2), because it is shorter and easier. Firstly, note that since x^8 is not MB by [3, Theorem 1.1.2(1)], we also have that x^{8k} is not MB for any $k \in \mathbb{N}^+$ (if a finite group satisfies a probabilistic identity with respect to x^8 and ρ, it also satisfies one with respect to x^{8k} and ρ). We may thus assume that $e \equiv 4 \pmod 8$; in other words, $e = 4f$ for some odd $f \in \mathbb{N}^+$ with $f > 1$. But by Lemma 6.2.2, if $\alpha \in \mathrm{Aut}(\mathrm{PSL}_2(3^f))$ is as in the formulation of Lemma 6.2.2, then

$$\left(\mathrm{PSL}_2(3^f)\alpha\right)^e = \{1\},$$

whence e is bad by [3, Proposition 2.9(3)].

We now give the proof of statement (1). First, note that the assumption implies that each prime divisor of m is larger than 226. We need to show that for every nonabelian finite simple group S and all $\alpha \in \mathrm{Aut}(S)$, $(S\alpha)^{2m} \neq \{\alpha^{2m}\}$. By [3, Theorem 5.1], it suffices to show this for S of one of the two forms $\mathrm{PSL}_2(p^f)$ or $\mathrm{Suz}(2^{2k+1})$. In what follows, we denote by Φ_S the field automorphism group of S, which is cyclic, generated by ϕ, the entry-wise Frobenius automorphism $a \mapsto a^p$ (for this to make

sense in the Suzuki case, view $\mathrm{Suz}(2^{2k+1})$ as a subgroup of $\mathrm{GL}_4(2^{2k+1})$ as in [15]). As in the proof of Lemma 6.2.2 above, we view $\mathrm{PSL}_2(p^f)$ as a subgroup of

$$\mathrm{PGL}_2(p^f) = \mathrm{Inndiag}(\mathrm{PSL}_2(p^f)),$$

and we denote the image of a matrix $M \in \mathrm{GL}_2(p^f)$ under the canonical projection $\mathrm{GL}_2(p^f) \to \mathrm{PGL}_2(p^f)$ by $\bar{M}$.

(1) Case: $S = \mathrm{Suz}(2^{2k+1})$. Then,

$$\mathrm{Out}(S) = \Phi_S = \langle \phi \rangle,$$

so we may choose $\alpha = \phi^t$ for some $t \in \{0, \ldots, 2k\}$. Then α centralizes

$$\mathrm{Frob}(20) \cong \mathrm{Suz}(2) \leq S.$$

In particular, there is an $s \in S$ of order $5 \nmid 2m$ centralized by α. It follows that

$$(s\alpha)^{2m} = s^{2m}\alpha^{2m} \neq \alpha^{2m}.$$

(2) Case: $S = \mathrm{PSL}_2(p^f)$. We make a subcase distinction:

(a) Subcase: $p = 2$. Then,

$$\mathrm{Out}(S) = \Phi_S = \langle \phi \rangle,$$

so we may choose $\alpha = \phi^t$ for some $t \in \{0, \ldots, f-1\}$. Then α centralizes

$$\mathrm{Sym}(3) \cong \mathrm{PSL}_2(2) \leq S.$$

In particular, there is an $s \in S$ of order $3 \nmid 2m$ centralized by α. It follows that

$$(s\alpha)^{2m} = s^{2m}\alpha^{2m} \neq \alpha^{2m}.$$

(b) Subcase: $p > 2$. Then,

$$\mathrm{Out}(S) = \mathrm{Outdiag}(S).\Phi_S = \left\langle \overline{\left(\begin{smallmatrix} \xi & 0 \\ 0 & 1 \end{smallmatrix}\right)} S \right\rangle.\langle \phi \rangle,$$

where ξ is some fixed generator of $\mathbb{F}_{p^f}^*$. We may thus choose $\alpha = \overline{\left(\begin{smallmatrix} \xi & 0 \\ 0 & 1 \end{smallmatrix}\right)}^{\epsilon} \phi^t$ for some $\epsilon \in \{0, 1\}$ and some $t \in \{0, \ldots, f-1\}$. If $\epsilon = 0$, then we can conclude as in Subcase (1), using that $\mathrm{PSL}_2(p)$ contains an element of order 3. So assume that $\epsilon = 1$. We make a subsubcase distinction:

(i) Subsubcase: $p \geq 7$ or $\gcd(f, t) > 1$. Note that the centralizer of α in $\mathrm{Inndiag}(S) = \mathrm{PGL}_2(p^f)$ contains the element

$$\alpha^{\mathrm{ord}(\phi^t)} = \alpha^{f/\gcd(f,t)} = \overline{\begin{pmatrix} \prod_{k=0}^{f/(\gcd(f,t)-1)} \xi^{p^{kt}} & 0 \\ 0 & 1 \end{pmatrix}}$$

$$= \overline{\begin{pmatrix} \prod_{k=0}^{f/(\gcd(f,t)-1)} \xi^{p^{k\gcd(f,t)}} & 0 \\ 0 & 1 \end{pmatrix}} = \overline{\begin{pmatrix} \xi^{(p^f-1)/(p^{\gcd(f,t)}-1)} & 0 \\ 0 & 1 \end{pmatrix}},$$

whose order is $p^{\gcd(f,t)} - 1$. In particular, since

$$[\mathrm{Inndiag}(S) : S] = 2,$$

there is an $s \in S$ of order $(p^{\gcd(f,t)} - 1)/2$ centralized by α. We will now argue that $(p^{\gcd(f,t)} - 1)/2$ does not divide $2m$, then we can conclude as in Subcase (1). To that end, note that by the subsubcase assumption, $(p^{\gcd(f,t)} - 1)/2 > 2$, so it suffices to show that $(p^{\gcd(f,t)} - 1)/2$ is not of the form n or $2n$ for some $n > 1$ that is odd and satisfies the congruence $n \equiv 1 \pmod{225}$.

- If $(p^{\gcd(f,t)} - 1)/2 = n$. Then,

  $$2n + 1 = p^{\gcd(f,t)}.$$

  By assumption, $n \equiv 1 \pmod 3$, so that $3 \mid 2n+1$ and thus $p = 3$. But $2n+1 > 3$, so one would need to have $9 \mid 2n + 1$, which is impossible since $n \equiv 1 \pmod 9$ by assumption.

- If $(p^{\gcd(f,t)-1})/2 = 2n$. Then,

  $$4n + 1 = p^{\gcd(f,t)}.$$

  By assumption, $n \equiv 1 \pmod 5$, so that $5 \mid 4n+1$ and thus $p = 5$. But $4n+1 > 5$, so one would need to have $25 \mid 4n + 1$, which is impossible since $n \equiv 1 \pmod{25}$ by assumption.

(ii) Subsubcase: $p \in \{3, 5\}$ and $\gcd(f, t) = 1$ (i.e. ϕ^t is a generator of Φ_S). By Lemma 6.2.1, it suffices to show that

$$(S\alpha)^{2m} \neq \{1\}.$$

Since $f = \mathrm{ord}(\phi^t) \mid \mathrm{ord}(s\alpha)$ for all $s \in S$, this is clear if $f \nmid 2m$, so assume that $f \mid 2m$. Note that by the argument from the previous subsubcase, we always have that $(p-1)f \mid \mathrm{ord}(\alpha)$. In particular, if $p = 5$, or if $p = 3$ and f is even, then $4 \mid \mathrm{ord}(\alpha)$, so that $\mathrm{ord}(\alpha) \nmid 2m$ and we are done. So we may assume that $p = 3$ and f is odd. But then Lemma 6.2.2 yields that some element in $S\alpha$ has order divisible by 4; in particular, the $(2m)$-th power of that element is nontrivial, as required. □

References

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Nonlinear programming*. Second edition, Athena Scientific Optimization and Computation Series, Athena Scientific, Belmont, MA, 1999. Zbl 1015.90077 MR 3444832

[2] A. Bors, Fibers of automorphic word maps and an application to composition factors, *J. Group Theory*, **20** (2017), no. 6, 1103–1133. Zbl 1401.20021 MR 3719319

[3] A. Bors, Fibers of word maps and the multiplicities of non-abelian composition factors, *Internat. J. Algebra Comput.*, **27** (2017), no. 8, 1121–1148. Zbl 06826303 MR 3741006

[4] P. Hall and G. Higman, On the p-length of p-soluble groups and reduction theorems for Burnside's problem, *Proc. London Math. Soc. (3)*, **6** (1956), 1–42. Zbl 0073.25503 MR 72872

[5] B. Hartley, A general Brauer–Fowler theorem and centralizers in locally finite groups, *Pacific J. Math.*, **152** (1992), no. 1, 101–117. Zbl 0713.20023 MR 1139975

[6] E.I. Khukhro and P. Shumyatsky, Nonsoluble and non-p-soluble length of finite groups, *Israel J. Math.*, **207** (2015), no. 2, 507–525. Zbl 1333.20023 MR 3359710

[7] M. Larsen and A. Shalev, A probabilistic Tits alternative and probabilistic identities, *Algebra Number Theory*, **10** (2016), no. 6, 1359–1371. Zbl 1356.20030 MR 3544299

[8] M. Larsen and A. Shalev, Words, Hausdorff dimension and randomly free groups, *Math. Ann.*, **371** (2018), no. 3-4, 1409–1427. Zbl 06923807 MR 3831276

[9] M. Larsen, A. Shalev, and P. H. Tiep, Probabilistic Waring problems for finite simple groups, *Ann. of Math. (2)*, **190** (2019), no. 2, 561–608. Zbl 1448.20063 MR 3997129

[10] A. Martino, M. C. H. Tointon, M. Valunis, and E. Ventura, Probabilistic nilpotence in infinite groups. arXiv:1805.11520

[11] N. Nikolov, Verbal width in anabelian groups, *Israel J. Math.*, **216** (2016), no. 2, 847–876. Zbl 1398.20040 MR 3557468

[12] D. J. S. Robinson, *A course in the theory of groups*, Graduate Texts in Mathematics, 80, Springer-Verlag, New York-Berlin, 1982. Zbl 0483.20001 MR 648604

[13] D. Segal, Remarks on profinite groups having few open subgroups, *J. Comb. Algebra*, **2** (2018), no. 1, 87–101. Zbl 06857321 MR 3763908

[14] A. Shalev, Probabilistically nilpotent groups, *Proc. Amer. Math. Soc.*, **146** (2018), no. 4, 1529–1536. Zbl 1427.20036 MR 3754339

[15] M. Suzuki, A new type of simple groups of finite order, *Proc. Nat. Acad. Sci. U.S.A.*, **46** (1960), 868–870. Zbl 0093.02301 MR 120283

[16] L. R. Vermani, *Elements of algebraic coding theory*, Chapman and Hall Mathematics Series, Chapman and Hall, Ltd., London, 1996. Zbl 0861.20007 MR 1406568

[17] E. I. Zelmanov, Solution of the restricted Burnside problem for groups of odd exponent (Russian), *Izv. Akad. Nauk SSSR Ser. Mat.*, **54** (1990), no. 1, 42–59, 221; translation in *Math. USSR-Izv.*, **36** (1991), no. 1, 41–60. Zbl 0709.20020 MR 1044047

[18] E. I. Zelmanov, Solution of the restricted Burnside problem for 2-groups (Russian), *Mat. Sb.*, **182** (1991), no. 4, 568–592; translation in *Math. USSR-Sb.*, **72** (1992), no. 2, 543–565. Zbl 0782.20038 MR 1119009

Received 08 April, 2019

A. Bors, School of Mathematics and Statistics, Carleton University,
1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

E-mail: alexanderbors@cunet.carleton.ca

A. Shalev, Einstein Institute of Mathematics, Hebrew University,
Jerusalem 91904, Israel

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