hexagon:2611.00943v1

Expansion in powers of Suzuki groups

Alexander Lubotzky* Yotam Shomroni*

November 23, 2026

Abstract

In a companion paper, we proved that every d-generated direct power G^m of a nonabelian finite simple group G, other than an alternating group or a Suzuki group, has a symmetric generating multiset of at most Cd elements whose normalized spectral gap is at least $C^{-1}d^{-\log_2 3}$, for an absolute constant C. Here, in an AI-generated addendum to that paper, the same is proved for the Suzuki groups $\mathrm{Sz}(q)$. The main step is that, for $m \le q^{1/100}$ and with probability tending to one as q grows, a random pair of elements generates $\mathrm{Sz}(q)^m$ and gives a Cayley graph whose spectral gap is bounded below independently of q and m. For $m = 1$ this is a theorem of Breuillard, Green and Tao. The proof follows the method of Bourgain and Gamburd, with a growth theorem for $\mathrm{Sz}(q)^s$ whose constants do not depend on s.

1 Introduction

For a finite group K, a finite multiset in K is **symmetric** if every element and its inverse have the same multiplicity, and **generating** if its support generates K. The normalized spectral gap $\mathrm{gap}_K(S)$ of a symmetric multiset S in K is defined in Theorem 1.5. In [LS26, Theorem 1.2], we proved the analogue of the following theorem for every nonabelian finite simple group other than the alternating groups and the Suzuki groups. Here we prove it for the Suzuki groups $\mathrm{Sz}(q)$, defined in Theorem 1.4.

Theorem 1.1 (Main theorem)**.** *There is an absolute constant $C > 0$ with the following property. Let $q = 2^f$ with $f \ge 3$ odd, let $G = \mathrm{Sz}(q)$, let $d \ge 2$ and $m \ge 1$, and suppose that G^m is generated by d elements. Then G^m has a symmetric generating multiset S such that*

$$|S| \le Cd, \qquad \mathrm{gap}_{G^m}(S) \ge C^{-1}d^{-\log_2 3}.$$

For groups of Lie type of bounded rank, the proof in [LS26] uses a theorem of Liebeck, Nikolov and Shalev [LNS11], by which such a group over $\mathbb{F}_q$, other than $\mathrm{Sz}(q)$, is a product of a bounded number of subgroups isomorphic to $\mathrm{SL}_2(\mathbb{F}_q)$ or $\mathrm{PSL}_2(\mathbb{F}_q)$. The orders of $\mathrm{SL}_2(\mathbb{F}_q)$ and $\mathrm{PSL}_2(\mathbb{F}_q)$ are divisible by 3, while, by [DJ14, Section 2.2(1)], the order of $\mathrm{Sz}(q)$ is $q^2(q^2+1)(q-1)$, which is not, as $q \equiv 2 \pmod 3$; hence $\mathrm{Sz}(q)$ has no such subgroups. As in [LS26], we use the following consequence of the amplification theorem [LS26, Theorem 1.6].

Theorem 1.2 ([LS26, Corollary 3.9])**.** *Let G be a nonabelian finite simple group, let $\lambda \ge 1$ be a real number, and let $m_0 \ge |G|^{1/\lambda}$ be an integer. Suppose that G^{m_0} has a symmetric generating multiset of size at most k and gap at least $\delta > 0$. Then every d-generated power G^m has a symmetric generating multiset S with*

$$|S| \le 2\lambda k\, d, \qquad \mathrm{gap}_{G^m}(S) \ge \delta\,(2\lambda d)^{-\log_2 3}.$$

*Weizmann Institute of Science, Rehovot. alex.lubotzky@mail.huji.ac.il, yotam.shomroni@weizmann.ac.il.

Breuillard, Green and Tao [BGT11, Theorem 1.2 and the first remark after it][1] proved that, with probability tending to one as q grows, a random pair of elements of $\mathrm{Sz}(q)$ generates it, and the Cayley graphs so obtained form a family of expanders. Breuillard, Green, Guralnick and Tao [BGGT15, Theorem 8.3] extended this to products of finite simple groups of Lie type, with constants that depend on the rank of the product, and hence, for $\mathrm{Sz}(q)^m$, on m. We prove the following extension to powers, with a spectral gap and with constants that do not depend on m.

Theorem 1.3. *There are absolute constants $\varepsilon, C_1, q_* > 0$ with the following property. Let $q = 2^f \geq q_*$ with $f \geq 3$ odd, let $1 \leq m \leq q^{1/100}$, choose $a, b \in \mathrm{Sz}(q)^m$ independently and uniformly, and let*

$$\mu := \tfrac{1}{4}(\mathbf{1}_{\{a\}} + \mathbf{1}_{\{a^{-1}\}} + \mathbf{1}_{\{b\}} + \mathbf{1}_{\{b^{-1}\}}),$$

a symmetric probability measure on $\mathrm{Sz}(q)^m$. Then, with probability at least $1 - C_1 q^{-1/10}$, the elements a, a^{-1}, b, b^{-1} are distinct, the set $S = \{a, a^{-1}, b, b^{-1}\}$ generates $\mathrm{Sz}(q)^m$, and μ, the uniform measure on S, has absolute spectral gap at least ε, defined in Theorem 1.5.

We deduce Theorem 1.1 from Theorems 1.2 and 1.3 at the end of this section. The proof of Theorem 1.3 has three parts. In Theorem 3.5 we show that, for every nonempty set I of coordinates, after about $20|I| \log q$ steps the random walk on $\mathrm{Sz}(q)^I$ whose steps are the projections of $a^{\pm 1}$ and $b^{\pm 1}$ is in a given proper subgroup H with probability at most a fixed negative power of $[\mathrm{Sz}(q)^I : H]$. This part follows [BGT11, Theorem 2.3 and Section 4], with the word lengths kept as variables, so that all $2^m - 1$ nonempty sets of coordinates are handled at once. Following [LS20, proof of Proposition 33], in Section 4 we prove a growth theorem for $\mathrm{Sz}(q)^s$ whose constants do not depend on s. In Section 5, following the method of Bourgain and Gamburd [BG08], we show that the ℓ^2 norm of $\mu^{(n)}$ becomes small for n of order $\log|\mathrm{Sz}(q)^m|$, and deduce the spectral gap from the lower bound $c_{\mathrm{rep}} q^{3/2}$ of Theorem 2.3(iv) on the dimensions of the nontrivial irreducible representations of $\mathrm{Sz}(q)$.

All logarithms are natural, and $[r] := \{1, \dots, r\}$ for $r \in \mathbb{Z}_{\geq 1}$. Each lower bound on q below is absolute unless a dependence is stated.

Fix an odd integer $f \geq 3$ and denote $q := 2^f$. The automorphisms of $\mathbb{F}_q$ are the maps $\sigma_j \colon z \mapsto z^{2^j}$ with $0 \leq j < f$, and its subfields are the fields $\mathbb{F}_{2^d} = \{z \in \mathbb{F}_q : z^{2^d} = z\}$, one for each divisor d of f.

Definition 1.4 (Suzuki group; [BGT11, Definition 3.1])**.** Denote $\theta := 2^{(f+1)/2}$. For $\alpha, \beta \in \mathbb{F}_q$ and $\gamma \in \mathbb{F}_q \setminus \{0\}$, denote

$$U(\alpha,\beta) := \begin{pmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ \alpha^{\theta+1} + \beta & \alpha^\theta & 1 & 0 \\ \alpha^{\theta+2} + \alpha\beta + \beta^\theta & \beta & \alpha & 1 \end{pmatrix}, \qquad \Omega := \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

$$D(\gamma) := \mathrm{diag}(\gamma^{\theta+1}, \gamma, \gamma^{-1}, \gamma^{-\theta-1}),$$

$$Q := \{U(\alpha,\beta) : \alpha, \beta \in \mathbb{F}_q\}, \qquad B := \{U(\alpha,\beta)D(\gamma) : \alpha, \beta \in \mathbb{F}_q,\ \gamma \neq 0\}.$$

The **Suzuki group** is the group of matrices

$$G = \mathrm{Sz}(q) := B \ \cup\ \{U(\alpha,\beta)D(\gamma)\Omega U(\alpha',\beta') : \alpha, \beta, \alpha', \beta' \in \mathbb{F}_q,\ \gamma \neq 0\}.$$

From now on f, q, θ, G, Q, B have the meanings just fixed.

For a group K and a finite set I, the direct power K^I has coordinates indexed by I, and $K^r := K^{[r]}$. A subscript I on an element of $K^{[m]}$, or on a probability measure on $K^{[m]}$, denotes

[1] We cite the numbering of the author's version of [BGT11], linked in the bibliography.

its projection to K^I. For $E \subseteq K$, denote by $\langle E \rangle$ the subgroup generated by E and by $\mathbf{1}_E$ the indicator function of E, and denote $E^j := \{e_1 \cdots e_j : e_i \in E\}$ and $E^{-1} := \{e^{-1} : e \in E\}$.

For a probability measure μ on a finite group K and $E \subseteq K$, denote $\mu(E) := \sum_{x \in E} \mu(x)$ and $\operatorname{supp} \mu := \{x : \mu(x) > 0\}$; the measure μ is **symmetric** if $\mu(x^{-1}) = \mu(x)$ for every $x \in K$. For probability measures μ and ν on K, denote

$$(\mu * \nu)(x) := \sum_{y \in K} \mu(y)\nu(y^{-1}x), \qquad \widetilde{\mu}(x) := \mu(x^{-1}),$$

let $\mu^{(n)}$ be the convolution of n copies of μ, with $\mu^{(0)} = \mathbf{1}_{\{1\}}$, and let u_A be the uniform measure on a nonempty finite $A \subseteq K$. The inner product on $\ell^2(K)$ is $\langle v, w \rangle = \sum_x v(x)\overline{w(x)}$, with the norm $\|v\|_2 = \langle v, v \rangle^{1/2}$, and $\ell^2_0(K)$ is the subspace of functions with sum zero. Denote by $\|T\|_{\mathrm{op}}$ the operator norm of a linear map T, and by $\mathbb{P}$ and $\mathbb{E}$ probability and expectation.

Definition 1.5 (Spectral gap)**.** Let K be a finite group and let μ be a symmetric probability measure on K. Its **averaging operator** is $M_\mu v(x) := \sum_{y \in K} \mu(y)v(y^{-1}x)$, and μ has **absolute spectral gap at least ε** if

$$\|M_\mu v\|_2 \le (1 - \varepsilon)\|v\|_2 \qquad \text{for every } v \in \ell^2_0(K).$$

For a finite symmetric multiset S in K, let u_S be the uniform measure on S, counted with multiplicity, and denote, as in [LS26, equation (4)],

$$\operatorname{gap}_K(S) := \inf_{0 \neq v \in \ell^2_0(K)} \frac{\langle v - M_{u_S} v, v \rangle}{\|v\|_2^2}. \tag{1}$$

Proof of Theorem 1.1. Let ε, C_1 and q_* be from Theorem 1.3, and let $q_2 \ge \max(q_*, 2^{200})$ satisfy $C_1 q_2^{-1/10} < 1$. Let $q \ge q_2$, and denote $G = \mathrm{Sz}(q)$ and $m_0 := \lfloor q^{1/100} \rfloor$. By Theorem 1.3, there are $a, b \in G^{m_0}$ such that the set $S_0 = \{a, a^{-1}, b, b^{-1}\}$ generates G^{m_0} and u_{S_0} has absolute spectral gap at least ε. Hence $\operatorname{gap}_{G^{m_0}}(S_0) \ge \varepsilon$, because, by the Cauchy–Schwarz inequality, every $v \in \ell^2_0(G^{m_0})$ satisfies $\langle v - M_{u_{S_0}} v, v \rangle \ge \|v\|_2^2 - \|M_{u_{S_0}} v\|_2 \|v\|_2 \ge \varepsilon \|v\|_2^2$. The group G is nonabelian simple and $|G| < q^5$ (see Theorem 2.3(i)). Hence, as $\lfloor x \rfloor \ge x/2$ for $x \ge 1$ and as $q^{1/200} \ge 2$,

$$m_0 \ge \frac{q^{1/100}}{2} \ge q^{1/200} > |G|^{1/1000} > |G|^{1/1024}.$$

By Theorem 1.2 with $\lambda = 2^{10}$, $k = 4$ and $\delta = \varepsilon$, every d-generated power G^m has a symmetric generating multiset S with $|S| \le 2^{13} d$ and $\operatorname{gap}_{G^m}(S) \ge \varepsilon\,(2^{11} d)^{-\log_2 3} = 3^{-11} \varepsilon\, d^{-\log_2 3}$. The groups $\mathrm{Sz}(q)$ with $q < q_2$ form a finite set, so, by [LS26, Corollary 6.1], there is $C' > 0$ such that, for each of them, every d-generated power G^m has a symmetric generating multiset S with $|S| \le C' d$ and $\operatorname{gap}_{G^m}(S) \ge C'^{-1} d^{-\log_2 3}$. Therefore Theorem 1.1 holds with $C := \max\{2^{13}, 3^{11}/\varepsilon, C'\}$. □

Acknowledgements and AI disclosure

The authors were partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 882751, TeStability). This paper is an addendum to our paper [LS26]. It is entirely AI-generated: it was written by OpenAI's GPT-6 and then edited with Anthropic's Claude Opus, and we have not yet verified it.

2 Preliminaries

2.1 Walk inequalities

Lemma 2.1. *Let K be a finite group, μ a symmetric probability measure, $H \le K$, and $x \in K$. For integers $n \ge n' \ge 0$,*

$$\mu^{(2n)}(H) \le \mu^{(2n')}(H), \qquad \mu^{(n')}(xH)^2 \le \mu^{(2n')}(H).$$

If a probability measure ν satisfies $\nu(xH) \le b_H$ for every $x \in K$, then $(\xi * \nu)(xH) \le b_H$ for every probability measure ξ and every x. Also $\|\nu\|_2^2 = (\widetilde{\nu} * \nu)(1)$ and $\|\xi * \nu\|_2 \le \|\nu\|_2$.

Proof. We repeat the short proofs of these estimates from [BGT11, proof of Theorem 2.3, second paragraph of Section 4] and [Var12, proofs of Lemma 15 and of Theorem 3]. Let K/H be the set of left cosets of H. On functions on K/H, let $P_g v(zH) = v(g^{-1}zH)$ and $P = \sum_g \mu(g) P_g$. Each P_g preserves the ℓ^2 norm, so $\|P\|_{\mathrm{op}} \le 1$. Its adjoint is $\sum_g \mu(g) P_{g^{-1}} = P$, by symmetry. With $e_H = \mathbf{1}_{\{H\}}$, we have

$$\mu^{(n')}(xH) = \langle P^{n'} e_H, e_{xH} \rangle, \qquad \mu^{(2n')}(H) = \|P^{n'} e_H\|_2^2.$$

The first conclusion follows from $\|P\|_{\mathrm{op}} \le 1$; the second follows from the Cauchy–Schwarz inequality and $\|e_{xH}\|_2 = 1$. For convolution, sum first over the coset:

$$(\xi * \nu)(xH) = \sum_g \xi(g) \nu(g^{-1}xH) \le \sum_g \xi(g) b_H = b_H.$$

Moreover, $(\widetilde{\nu} * \nu)(1) = \sum_g \nu(g^{-1})\nu(g^{-1}) = \|\nu\|_2^2$. Finally, $\|\xi * \nu\|_2 \le \|\nu\|_2$ is Young's inequality. □

2.2 Words

Let F be the free group on x, y, let $|w|$ be the length of the reduced word $w \in F$, and let $w(a, b)$ be the result of substituting a pair a, b of group elements for x, y. Let W_n be the product of n independent uniform choices from x, x^{-1}, y, y^{-1}, reduced in F, so that $W_n(a, b)$ has law $\mu^{(n)}$ for the pair a, b and the measure μ of Theorem 1.3, for every $n \ge 0$. Denote $\rho := \sqrt{3}/2$.

Denote $[g, h] := g^{-1}h^{-1}gh$, and define words in 2^j variables by $\psi_0(z) = z$ and

$$\psi_{j+1}(z_1, \dots, z_{2^{j+1}}) = [\psi_j(z_1, \dots, z_{2^j}), \psi_j(z_{2^j+1}, \dots, z_{2^{j+1}})].$$

The derived series of a group K is $K^{(0)} = K$ and $K^{(j+1)} = \langle [g, h] : g, h \in K^{(j)} \rangle$, and K has **derived length at most** j if $K^{(j)} = \{1\}$. By induction on j, every value of ψ_j on K is in $K^{(j)}$. In particular, ψ_3 takes the value 1 on every eight elements of a group of derived length at most three.

Proposition 2.2. *Let F and W_n be as above.*

(i) By a theorem of Kesten [Kes59, Theorem 3], in the form [CV96, Theorem 1.1(b)], the averaging operator M_F of the uniform measure on $\{x, x^{-1}, y, y^{-1}\}$ on $\ell^2(F)$ has $\|M_F\|_{\mathrm{op}} = \rho$.

(ii) By [BGT11, Lemma 4.1], there is an absolute constant C_0 such that the following holds. Let $t \ge 1$, and let $E \subseteq \{w \in F : |w| \le t\}$. If $\psi_3(w_1, \dots, w_8) = 1$ for all $w_1, \dots, w_8 \in E$, then $|E| \le C_0 t^8$.

Part (ii) weakens the exponent of the cited lemma to eight. By the Cauchy–Schwarz inequality and Theorem 2.2(i), for every $w \in F$ and $n \ge 0$,

$$\mathbb{P}(W_n = w) = \langle M_F^n \mathbf{1}_{\{1\}}, \mathbf{1}_{\{w\}} \rangle \le \|M_F\|_{\mathrm{op}}^n \|\mathbf{1}_{\{1\}}\|_2 \|\mathbf{1}_{\{w\}}\|_2 = \rho^n. \tag{2}$$

2.3 Inputs on the Suzuki groups

Proposition 2.3. *Let $G = \mathrm{Sz}(q)$ be as in Theorem 1.4.*

(i) By [DJ14, Section 2.2(1)], the group G is simple, of order $|G| = q^2(q^2+1)(q-1)$, and, by [DJ14, Section 2.1], $|B| = q^2(q-1)$. Hence $q^4 < |G| < q^5$, as $(q^2+1)(q-1) = q^3 - q^2 + q - 1$, and G is nonabelian, because its order is not prime. By [BGT11, Definition 3.1 and the paragraph after it], the parameterizations in Theorem 1.4 are one-to-one. Every element of G has determinant one, because $\det U(\alpha,\beta) = \det \Omega = 1$ and $\det D(\gamma) = \gamma^{\theta+1}\gamma\gamma^{-1}\gamma^{-\theta-1} = 1$.

(ii) By [BGT11, Lemma 3.2], every proper subgroup of G is either solvable of derived length at most three or conjugate to $\mathrm{Sz}(2^d)$ for a proper divisor d of f; here $d \geq 3$, because, by [DJ14, Section 2.3.3], $\mathrm{Sz}(2) \cong \mathrm{AGL}_1(5)$, which is solvable of derived length two.

(iii) By [BGT11, Lemma 3.4], there is an absolute $c_0 > 0$ such that every nontrivial $w \in F$ with $|w| < c_0\sqrt{q}$ has $w(a,b) \neq 1$ for some $a, b \in G \setminus B$.

(iv) By [BGT11, Theorem 2.2], there is an absolute $c_{\mathrm{rep}} > 0$ such that every nontrivial irreducible complex representation π of G has $\dim \pi \geq c_{\mathrm{rep}} q^{3/2}$.

(v) By [PS11, Theorem 77], applied to the Suzuki groups, there is an absolute $\varepsilon_0 > 0$ such that every generating set $A \subseteq G$ satisfies either $A^3 = G$ or $|A^3| \geq |A|^{1+\varepsilon_0}$.

(vi) By [DJ14, Section 2.2(2)], every automorphism of G has the form $g \mapsto x\sigma_j(g)x^{-1}$ for some $x \in G$ and some $0 \leq j < f$, where σ_j acts on matrix entries.

In part (iii) the hypothesis that w be nontrivial is ours: the cited lemma omits it, although its proof uses it. For $g \in G$, denote by $p_g(X) \coloneqq \det(XI_4 - g)$ its characteristic polynomial, where I_4 is the identity matrix. Its constant term is $\det g = 1$, by Theorem 2.3(i), so

$$p_g(X) = X^4 + c_1(g)X^3 + c_2(g)X^2 + c_3(g)X + 1 \qquad \text{for some } c_1(g), c_2(g), c_3(g) \in \mathbb{F}_q.$$

For a subfield $\mathbb{F}_{2^d} \leq \mathbb{F}_q$, the notation $p_g \in \mathbb{F}_{2^d}[X]$ means that $c_1(g), c_2(g), c_3(g) \in \mathbb{F}_{2^d}$.

Proposition 2.4 ([BGT11, Lemmas 4.2–4.3])**.** *Let q, θ, G be as in Theorem 1.4, with $f \geq 5$.*

(i) Let P be a polynomial in $2k$ variables over $\mathbb{F}_q$, of degree at most e in each variable. If the function $P(z_1, \dots, z_k, z_1^\theta, \dots, z_k^\theta)$ is not identically zero on $\mathbb{F}_q^k$, then its zero set has at most $C_2 k e q^{k-1/2}$ elements, where C_2 is absolute.

(ii) Let $w \in F$ have length $L \geq 1$, let $j \in \{1,2,3\}$, and let $z \in \mathbb{F}_q \setminus \{0\}$. For independent uniform $a, b \in G$, the probability that $c_j(w(a,b)) = z$ is at most $C_3 L q^{-1/2}$, where C_3 is absolute.

(iii) By [BGT11, proof of Theorem 2.3, the paragraph after Corollary 4.4], if $g \in G$ has $c_1(g) = c_2(g) = c_3(g) = 0$, then $g^4 = 1$.

3 Simultaneous subgroup escape

For a polynomial p over $\mathbb{F}_q$, let $\sigma_j(p)$ be the polynomial obtained by applying σ_j to each coefficient of p. Denote

$$\mathcal{A} \coloneqq \{g \in G : p_g \in \mathbb{F}_{2^d}[X] \text{ for some proper divisor } d \text{ of } f\},$$
$$\mathcal{D} \coloneqq \{(g,h) \in G^2 : p_h = \sigma_j(p_g) \text{ for some } 0 \leq j < f\}.$$

Lemma 3.1. *There is an absolute C_4 with the following property. Let $f \geq 5$, and let $1 \neq w \in F$ have length $L < c_0\sqrt{q}/4$, where c_0 is from Theorem 2.3(iii). Let (a, b) and (a', b') be independent uniform pairs in G^2. Then*

$$\begin{aligned}
\mathbb{P}(w(a,b) = 1) &\leq C_4 L q^{-1/2}, \\
\sup_{p \in \mathbb{F}_q[X]} \mathbb{P}(p_{w(a,b)} = p) &\leq C_4 L q^{-1/2}, \\
\mathbb{P}(w(a,b) \in \mathcal{A}) &\leq C_4 L q^{-1/6}, \\
\mathbb{P}((w(a,b), w(a',b')) \in \mathcal{D}) &\leq C_4 f L q^{-1/2}.
\end{aligned}$$

Proof. We follow [BGT11, the proof of Lemma 3.5 in Appendix B] and [BGT11, the proofs of Corollary 4.4 and of equation (4.7)], keeping the length L as a variable.

Step 1. Relations. The probability that a or b belongs to B is at most $2|B|/|G| = 2/(q^2+1)$, so it is enough to bound the probability of $w(a,b) = 1$ on the event $a, b \in G \setminus B$ and then to add $2/(q^2+1)$.

Write $a = U(\alpha_1, \beta_1)D(\gamma_1)\Omega U(\alpha_1', \beta_1')$ and $b = U(\alpha_2, \beta_2)D(\gamma_2)\Omega U(\alpha_2', \beta_2')$, so that the pair is determined by ten parameters $z_1, \dots, z_{10} \in \mathbb{F}_q$, five for each of a and b, of which γ_1 and γ_2 must be nonzero. We shall apply Theorem 2.4(i) with $k = 10$, its $2k$ variables being $z_1, \dots, z_{10}$ and $z_1^\theta, \dots, z_{10}^\theta$.

Each entry of $U(\alpha, \beta)$ is a polynomial of degree at most two in each of $\alpha, \beta, \alpha^\theta, \beta^\theta$, and Ω is constant, while $D(\gamma)$ contributes only the two variables γ, γ^θ, with exponents between -1 and 1. Each parameter occurs in exactly one factor of $UD\Omega U'$, so every entry of a positive letter has degree at most two in each variable, with diagonal exponents between -1 and 1. A negative letter is the inverse of such a matrix, of determinant one, so each of its entries is a three-by-three minor of it: degree at most six, with diagonal exponents between -3 and 3. Multiplying L letters adds degrees, so every entry of $w(a,b)$ has degree at most $6L$ in each variable, with diagonal exponents between $-3L$ and $3L$.

The sixteen equations $w(a,b)_{ij} = (I_4)_{ij}$ therefore become polynomial equations after multiplication by $(\gamma_1\gamma_1^\theta\gamma_2\gamma_2^\theta)^{3L}$, of degree at most $e = 6L$ in each variable. By Theorem 2.3(iii), at least one of the sixteen is not identically zero on $(G \setminus B)^2$, hence not identically zero as a function of the ten variables. By Theorem 2.4(i), its zero set has at most $C_2 \cdot 10 \cdot 6L\, q^{10-1/2}$ elements, out of q^{10} tuples, so its probability under unrestricted parameters is at most $60C_2Lq^{-1/2}$. Conditioning γ_1, γ_2 to be nonzero multiplies this by $(1-1/q)^{-2} \leq 2$, and adding $2/(q^2+1)$ gives a bound $C_5Lq^{-1/2}$ with C_5 absolute. This argument needs only $L < c_0\sqrt{q}$.

Step 2. A fixed characteristic polynomial. Every p_g is monic of degree four with constant term one, so $\mathbb{P}(p_{w(a,b)} = p) = 0$ for every other p. If p is of that form and has a nonzero coefficient of X, X^2 or X^3, then, by Theorem 2.4(ii) applied to that coefficient, $\mathbb{P}(p_{w(a,b)} = p) \leq C_3Lq^{-1/2}$. The one p left is X^4+1, and then, by Theorem 2.4(iii), $w(a,b)^4 = 1$. The word w^4 is nontrivial, as F is torsion-free, and $|w^4| \leq 4L < c_0\sqrt{q}$, so Step 1 applies to w^4. The second estimate holds with $C_6 = \max(C_3, 4C_5)$.

Step 3. Proper subfields. Every proper divisor of the odd integer f is at most $f/3$. Thus the union of the proper subfields has size at most

$$\sum_{d=1}^{\lfloor f/3 \rfloor} 2^d = 2^{\lfloor f/3 \rfloor + 1} - 2 < 2q^{1/3}.$$

If $p_{w(a,b)}$ is defined over a proper subfield, then either an intermediate coefficient is nonzero and belongs to this union, or all three coefficients vanish. Hence, by a union bound, Theorem 2.4(ii) and Step 2,

$$\mathbb{P}(w(a,b) \in \mathcal{A}) \leq 3(2q^{1/3})C_3Lq^{-1/2} + C_6Lq^{-1/2} \leq (6C_3 + C_6)Lq^{-1/6}.$$

Step 4. A pair of characteristic polynomials. Condition on (a,b). There are at most f polynomials $\sigma_j(p_{w(a,b)})$, and, by Step 2, $p_{w(a',b')}$ equals each of them with probability at most $C_6Lq^{-1/2}$. Hence $\mathbb{P}((w(a,b),w(a',b')) \in \mathcal{D}) \le fC_6Lq^{-1/2}$. Take C_4 to be the maximum of the four absolute constants just obtained. □

Lemma 3.2. *Let $1 \le m \le q^{1/100}$, choose $a,b \in G^m$ independently and uniformly, and denote $\mu := \frac{1}{4}(\mathbf{1}_{\{a\}} + \mathbf{1}_{\{a^{-1}\}} + \mathbf{1}_{\{b\}} + \mathbf{1}_{\{b^{-1}\}})$, as in Theorem 1.3. For $r \ge 1$, denote*

$$n_r := 2\lceil 10r \log q \rceil, \qquad g_r := \lfloor (r/10) \log q \rfloor,$$

so that $n_r \le 20r \log q + 2 \le 21r \log q$ once $r \log q \ge 2$, which holds for $q \ge 8$. For all sufficiently large q, with probability at least $1 - C_7 q^{-1/10}$, the following hold at once.

(i) If $I \subseteq [m]$ has size r, no nontrivial word of length at most g_r equals 1 at (a_I, b_I).

(ii) If $I \subseteq [m]$ has size r, then $\mu_I^{(n_r)}(\mathcal{A}^I) \le q^{-r/100}$, where $\mathcal{A}^I$ is the set of elements of G^I all of whose coordinates are in $\mathcal{A}$.

(iii) A matching is a set of pairs (i,j) with $i < j$, in which no coordinate occurs in two pairs. For a matching M, denote

$$E_M := \{z \in G^m : (z_i, z_j) \in \mathcal{D} \text{ for every } (i,j) \in M\}.$$

If M has r pairs, then $\mu^{(n_r)}(E_M) \le q^{-r/100}$.

The constant C_7 is absolute.

Proof. We follow the word-counting and averaging argument of [BGT11, the proof of Lemma 3.5 in Appendix B, and the proof of Theorem 2.3], applied to independent sets of coordinates.

Step 1. Uniform bounds for all words used. For large q we have $n_m \le 21m \log q \le 21q^{1/100} \log q < c_0\sqrt{q}/4$. Therefore, by Theorem 3.1, for every nontrivial word w of length at most n_m, and for all $i \ne j$ in $[m]$, as (a_i, b_i) and (a_j, b_j) are independent uniform pairs in G^2, we have

$$\begin{aligned} \mathbb{P}(w(a_i,b_i) = 1) &\le q^{-1/3}, \\ \mathbb{P}(w(a_i,b_i) \in \mathcal{A}) &\le q^{-1/7}, \\ \mathbb{P}((w(a_i,b_i), w(a_j,b_j)) \in \mathcal{D}) &\le q^{-1/3} \end{aligned} \tag{3}$$

for all large q. Indeed, $L \le n_m \le 21q^{1/100} \log q$, so, by Theorem 3.1, the middle probability is at most $21C_4 q^{1/100} \log q \cdot q^{-1/6}$, which is at most $q^{-1/7}$ as soon as

$$21C_4 \log q \le q^{1/6 - 1/100 - 1/7} = q^{29/2100}.$$

In the same way, the first probability is at most $21C_4 q^{1/100} \log q \cdot q^{-1/2}$ and the third is at most that times $f = (\log q)/\log 2$, and both are at most $q^{-1/3}$ as soon as

$$\frac{21C_4 (\log q)^2}{\log 2} \le q^{1/2 - 1/100 - 1/3} = q^{47/300}.$$

Step 2. Relations on subsets. There are $4 \cdot 3^{j-1}$ reduced words of length $j \ge 1$: the first letter has four choices, and each later letter has three. Thus there are $2(3^{g_r} - 1) < 2 \cdot 3^{g_r}$ nontrivial words of length at most g_r. For a fixed such word and a fixed r-element set I, by the first inequality of (3) and the independence of the coordinates, the probability of a relation at every

coordinate in I is at most $q^{-r/3}$. There are at most $m^r \le q^{r/100}$ such sets, and $3^{g_r} \le q^{r(\log 3)/10}$. Hence the probability that (i) fails is at most

$$2\sum_{r=1}^{m} q^{-r(1/3-1/100-(\log 3)/10)}.$$

Step 3. Arithmetic constraints. As $n_r \ge 20r\log q$ and $20\log(1/\rho) = 10\log(4/3) > 1/3$, by (2) we have $\mathbb{P}(W_{n_r} = 1) \le \rho^{n_r} \le q^{-r/3}$. For a fixed nontrivial reduced word w of length at most n_r, by the second inequality of (3) and the independence of the coordinates, $\mathbb{P}(w(a_I, b_I) \in \mathcal{A}^I) \le q^{-r/7}$. Hence

$$\mathbb{E}_{a,b}\,\mu_I^{(n_r)}(\mathcal{A}^I) = \sum_w \mathbb{P}(W_{n_r} = w)\,\mathbb{P}(w(a_I, b_I) \in \mathcal{A}^I) \le q^{-r/3} + q^{-r/7} \le 2q^{-r/7}.$$

By Markov's inequality, $\mu_I^{(n_r)}(\mathcal{A}^I) > q^{-r/100}$ with probability at most $2q^{-r/7+r/100}$. Summing over the at most $m^r \le q^{r/100}$ sets I of size r, and over r, the probability that (ii) fails is at most $2\sum_{r=1}^{m} q^{-r(1/7-2/100)}$.

Step 4. Paired constraints. For a fixed matching of size r, its pairs use disjoint random coordinates, so, for a fixed nontrivial word of length at most n_r, by the third inequality of (3) and independence, the probability that every pair satisfies its constraint is at most $q^{-r/3}$. The trivial word has probability $\mathbb{P}(W_{n_r} = 1) \le q^{-r/3}$. Hence, as in Step 3, by Markov's inequality and a sum over the at most $m^{2r} \le q^{2r/100}$ matchings, the probability that (iii) fails is at most $2\sum_{r=1}^{\lfloor m/2 \rfloor} q^{-r(1/3-3/100)}$.

The exponents $1/3 - 1/100 - (\log 3)/10$, $1/7 - 2/100$ and $1/3 - 3/100$ of the three series are greater than $1/10$, and $\sum_{r\ge 1} x^r \le 2x$ for $0 \le x \le 1/2$, so, for $q \ge 2^{10}$, each of the three failure probabilities is at most $4q^{-1/10}$, and $C_7 = 12$ works. □

Lemma 3.3. *There is an absolute d_0 with $0 < d_0 < 1/100$ with the following property. Assume q is sufficiently large, let (a, b) satisfy Theorem 3.2(i), and let $I \subseteq [m]$ have size $r \ge 1$. Then $\mu_I^{(n)}(H) \le q^{-d_0 r}$ for every subgroup $H \le G^I$ of derived length at most three and every even $n \ge n_r$.*

Proof. We follow [BGT11, proof of Theorem 2.3, Section 4, algebraic case]. Denote $t_r := 2\lfloor g_r/128 \rfloor$. If $r\log q \ge 2580$, then, by the definitions of the floors,

$$\frac{r\log q}{1280} \le t_r \le \frac{r\log q}{640}, \qquad 64t_r \le g_r.$$

For the lower bound, $t_r \ge (r\log q)/640 - 2 - 1/64$, which is at least $(r\log q)/1280$ as $r\log q \ge 2580$. Let $E = \{w \in F : |w| \le t_r,\ w(a_I, b_I) \in H\}$. The word $\psi_3(w_1, \dots, w_8)$ has length at most $4^3 t_r = 64t_r$, and it evaluates to 1 whenever $w_1, \dots, w_8 \in E$. By Theorem 3.2(i), it is therefore the trivial free word. Every w with $W_{t_r} = w$ and $w(a_I, b_I) \in H$ is in E, so, by (2) and Theorem 2.2(ii),

$$\mu_I^{(t_r)}(H) \le |E|\rho^{t_r} \le C_0 t_r^8 \rho^{t_r}.$$

Choose a fixed t_0 so that $C_0 t^8 \le \rho^{-t/2}$ for all $t \ge t_0$. This is possible because $\log C_0 + 8\log t \le (t/2)\log(1/\rho)$ for all large t. If $\log q \ge \max(2580, 1280t_0)$, then

$$\mu_I^{(t_r)}(H) \le \rho^{t_r/2} \le q^{-r\log(1/\rho)/2560}.$$

Take $d_0 = \log(1/\rho)/2560$. Both t_r and n are even and $n \ge n_r \ge t_r$, so, by Theorem 2.1, $\mu_I^{(n)}(H) \le \mu_I^{(t_r)}(H) \le q^{-d_0 r}$. □

A subgroup of a direct product is **subdirect** if its projection to each factor is onto.

Proposition 3.4 (Scott's lemma; [DS19, Lemma 2.3(1) and Definition 2.2])**.** *Let K be a nonabelian simple group and let $P \le K^{[t]}$ be subdirect, with $t \ge 1$. There are a partition of $[t]$ into nonempty sets $I_1, \dots, I_k$ and automorphisms $\alpha_1, \dots, \alpha_t$ of K such that, denoting by $j(i)$ the index with $i \in I_{j(i)}$,*

$$P = \{(\alpha_i(z_{j(i)}))_{i\in[t]} : z_1, \dots, z_k \in K\}.$$

In particular, $|P| = |K|^k$.

Proposition 3.5. *There is an absolute $\eta > 0$ with the following property. Assume q is sufficiently large, and let (a, b) satisfy all three parts of Theorem 3.2. Then, for every $I \subseteq [m]$ of size $s \ge 1$ and every proper subgroup $H < G^I$,*

$$\mu_I^{(n_s)}(H) \le [G^I : H]^{-\eta}, \qquad \mu_I^{(n_s)}(xH) \le [G^I : H]^{-\eta/2} \quad \text{for every } x \in G^I.$$

Moreover, a_I and b_I generate G^I.

Proof. Step 1. Count the constraints on H. Split I into three sets: I_{sol}, the coordinates at which the projection of H is proper and solvable; I_{sub}, those at which it is proper and not solvable; and I_{full}, those at which it is G. Denote their sizes by u, v, t, so that $u + v + t = s$. By Theorem 2.3(ii), each projection at a coordinate in I_{sol} has derived length at most three, and each projection at a coordinate in I_{sub} is a conjugate of a subfield group $\mathrm{Sz}(2^d)$ for a proper divisor d of f. For $t > 0$, apply Theorem 3.4 to the projection of H to $G^{I_{\mathrm{full}}}$, which is subdirect, and let k be its number of blocks. In a block of size ℓ, choose $\lfloor \ell/2 \rfloor$ disjoint pairs. Since $\lfloor \ell/2 \rfloor \ge (\ell - 1)/2$, the resulting matching M has size $e \ge (t - k)/2$. If $t = 0$, set $k = e = 0$.

Step 2. Bound the walk on each type of constraint. If $u > 0$, denote by H_i the projection of H to the coordinate i. The product $\prod_{i\in I_{\mathrm{sol}}} H_i$ has derived length at most three, since commutators are computed coordinatewise, and it contains the projection of H to $G^{I_{\mathrm{sol}}}$. Hence, by Theorem 3.3 with $r = u$ and $n = n_s$, which is even and at least n_u,

$$\mu_I^{(n_s)}(H) \le \mu_{I_{\mathrm{sol}}}^{(n_s)}\Big(\prod_{i\in I_{\mathrm{sol}}} H_i\Big) \le q^{-d_0 u}.$$

If $v > 0$, let Γ be the product of the v conjugate subfield groups, a subgroup of $G^{I_{\mathrm{sub}}}$. It is contained in $\mathcal{A}^{I_{\mathrm{sub}}}$: the entries of an element g of $\mathrm{Sz}(2^d)$ are in $\mathbb{F}_{2^d}$, so $p_g \in \mathbb{F}_{2^d}[X]$, and conjugation does not change the characteristic polynomial. By Theorem 3.2(ii), $\mu_{I_{\mathrm{sub}}}^{(n_v)}(\Gamma) \le q^{-v/100}$, and, by Theorem 2.1, as Γ is a subgroup, the same bound holds for $\mu_{I_{\mathrm{sub}}}^{(n_s)}(\Gamma)$. The projection of H to $G^{I_{\mathrm{sub}}}$ is contained in Γ, so $\mu_I^{(n_s)}(H) \le \mu_{I_{\mathrm{sub}}}^{(n_s)}(\Gamma) \le q^{-v/100}$.

If $e > 0$, then, by Theorem 3.4, for every pair $(i, i') \in M$, every $z \in H$ satisfies $z_{i'} = \alpha_{ii'}(z_i)$, where $\alpha_{ii'} := \alpha_{i'}\alpha_i^{-1} \in \mathrm{Aut}(G)$. By Theorem 2.3(vi), there are $\kappa \in G$ and $0 \le j < f$ with $\alpha_{ii'}(g) = \kappa\sigma_j(g)\kappa^{-1}$ for every $g \in G$, so $p_{\alpha_{ii'}(g)} = p_{\sigma_j(g)}$, as conjugation does not change the characteristic polynomial. The determinant is a polynomial in the matrix entries with coefficients in the prime field, so σ_j commutes with it, and $p_{\sigma_j(g)} = \sigma_j(p_g)$. Hence $\Gamma_M := \{z \in G^I : z_{i'} = \alpha_{ii'}(z_i) \text{ for every } (i, i') \in M\}$ is a subgroup of G^I containing H, and its preimage in G^m is contained in the set E_M of Theorem 3.2(iii). Therefore $\mu_I^{(n_e)}(\Gamma_M) \le \mu^{(n_e)}(E_M) \le q^{-e/100}$, and, by Theorem 2.1, as Γ_M is a subgroup,

$$\mu_I^{(n_s)}(H) \le \mu_I^{(n_s)}(\Gamma_M) \le \mu_I^{(n_e)}(\Gamma_M) \le q^{-e/100}.$$

Since $d_0 < 1/100$, these three bounds imply

$$\mu_I^{(n_s)}(H) \le q^{-d_0 \max(u,v,e)}.$$

Step 3. Compare the constraint count with the index. The projection of H to $G^{I_{\mathrm{full}}}$ has order $|G|^k$, so $|H| \geq |G|^k$. Also

$$s - k = u + v + t - k \leq u + v + 2e \leq 4\max(u, v, e).$$

By Theorem 2.3(i), $|G| < q^5$; consequently,

$$[G^I : H] \leq |G|^{s-k} \leq q^{20\max(u,v,e)}, \qquad q^{-d_0\max(u,v,e)} \leq [G^I : H]^{-d_0/20}.$$

Take $\eta = d_0/20$. By Theorem 2.1,

$$\mu_I^{(n_s)}(xH)^2 \leq \mu_I^{(2n_s)}(H) \leq \mu_I^{(n_s)}(H) \leq [G^I : H]^{-\eta}.$$

This proves the coset bound. If a_I, b_I generated a proper subgroup $H < G^I$, then $\mu_I^{(n_s)}(H) = 1$, whereas, by the first bound, $\mu_I^{(n_s)}(H) < 1$, since $[G^I : H] > 1$. Hence $\langle a_I, b_I\rangle = G^I$. □

4 Growth in a direct power

In the growth theorem we need, Theorem 4.11 below, all direct factors are equal. Its proof follows the scheme of [LS20, proof of Proposition 33]; we cannot apply that proposition itself, because its hypotheses include that the factors be pairwise nonisomorphic.

4.1 Subgroups and representations

For a group K, denote by $Z(K)$ its centre, by $C_K(x)$ the centralizer of $x \in K$, and by $N_K(H)$ the normalizer of $H \leq K$.

Denote $L_0 := 16$. By Theorem 2.3(iv), for all sufficiently large q,

$$\dim \pi > |G|^{1/L_0} \qquad \text{for every nontrivial irreducible representation } \pi \text{ of } G. \tag{4}$$

Indeed, by Theorem 2.3(i), $|G| < q^5$, and $c_{\mathrm{rep}} q^{3/2} > q^{5/16}$ for all large q. Consequently,

$$[G : H] > |G|^{1/L_0} \qquad \text{for every } H < G. \tag{5}$$

Indeed, G acts on the functions with sum zero on G/H, a nonzero space of dimension $[G : H] - 1$ with no nonzero invariant vector, as the action is transitive; so, by (4), $[G : H] - 1 > |G|^{1/L_0}$.

Denote $A_0 := \{D(\gamma) : \gamma \in \mathbb{F}_q \setminus \{0\}\}$ (in [DJ14, Section 2.1], $D(\gamma)$ is denoted a_{γ^θ}). As $D(\gamma)D(\gamma') = D(\gamma\gamma')$, the group A_0 is cyclic of order $q - 1$; also $B = QA_0$ and $Q \cap A_0 = \{1\}$, by Theorem 1.4.

Proposition 4.1. *Let $G = \mathrm{Sz}(q)$, with Q, B from Theorem 1.4, and let $\mathcal{T}$ be the set of those cyclic subgroups of odd order of G that are not contained in any larger one; we call them* ***tori****.*

(i) By [DJ14, Section 2.1], Q is a group of order q^2. By [DJ14, Section 2.2(3)], distinct conjugates of Q intersect in $\{1\}$, and the intersection of two distinct conjugates of B is a conjugate of A_0. By [DJ14, Section 2.2(4)], every element of Q has fourth power 1, and each element of order two in Q is in $Z(Q)$. By [DJ14, Section 2.2(7)], distinct members of $\mathcal{T}$ intersect in $\{1\}$, and every nonidentity element of G either is in a conjugate of Q or has odd order.

(ii) By [DJ14, Section 2.2(5)], the group A_0 acts regularly by conjugation on the nonidentity elements of $Z(Q)$; in particular, conjugation by a nonidentity element of A_0 fixes none of them. By [DJ14, Section 3.4.2], every nontrivial subgroup of A_0 has centralizer A_0. By [DJ14, Theorem 4.1(2)], $N_G(B) = B$, and $N_G(Q) = B$, because B normalizes Q

and, by [DJ14, Section 2.1 and Theorem 4.1(3)], $|N_G(Q)| = q^2(q-1) = |B|$. By [DJ14, Section 2.2(7), Section 2.3.2 and Theorem 4.1(6)], for every $A \in \mathcal{T}$ there is an element c of order two or four with $N_G(A) = A\langle c\rangle$ and $A \cap \langle c\rangle = \{1\}$, and conjugation by each nonidentity power of c fixes no nonidentity element of A.

(iii) By [DJ14, Proposition 3.4, Section 2.3.3, Theorem 4.1(1)], every proper subgroup of G is contained in a conjugate of B, contained in some $N_G(A)$ with $A \in \mathcal{T}$, or is a conjugate of $\mathrm{Sz}(2^d)$ for some proper divisor $d \geq 3$ of f, and $N_G(\mathrm{Sz}(2^d)) = \mathrm{Sz}(2^d)$.

Two consequences of Theorem 4.1 are used repeatedly. First, every element $z \neq 1$ of odd order generates a cyclic subgroup of odd order, which is contained in a maximal one, because G is finite; hence z is in a member of $\mathcal{T}$, and, by Theorem 4.1(i), every $z \neq 1$ is in a conjugate of Q or in a member of $\mathcal{T}$. Second, $A_0 \in \mathcal{T}$: it is cyclic of odd order $q-1$, and if $A_0 \leq A$ with A cyclic, then A is abelian, so $A \leq C_G(A_0) = A_0$ and $A = A_0$. Since $\mathcal{T}$ is closed under conjugation, every conjugate of A_0 is a torus as well, so two distinct conjugates of B meet in a member of $\mathcal{T}$.

The following subgroup condition is [LS20, assumption $(V3)_L$, preceding Proposition 33], with two levels (the families $\mathcal{H}_1$ and $\mathcal{H}_2$ below), for the group G; as G is nonabelian simple, it has trivial centre and is almost simple, which is [LS20, assumption $(V1)_L$]. For subgroups $H, J \leq G$, denote $H \preceq J$ if $[H : H \cap J] < L_0$, and denote

$$\mathcal{H}_0 \coloneqq \{\{1\}\}, \quad \mathcal{H}_1 \coloneqq \{gQg^{-1} : g \in G\} \cup \mathcal{T}, \quad \mathcal{H}_2 \coloneqq \{gBg^{-1} : g \in G\}.$$

For each proper divisor $d \geq 3$ of f, let $\mathcal{H}'_d$ be the set of conjugates of $\mathrm{Sz}(2^d)$.

Lemma 4.2. *The families just defined have the following properties.*

(i) They are closed under conjugation. There are at most $L_0 \log|G|$ families $\mathcal{H}'_d$. Every proper subgroup H satisfies $H \preceq J$ for some J in one of the listed families.

(ii) For $i = 1, 2$, distinct $H, H' \in \mathcal{H}_i$ satisfy $H \cap H' \preceq J$ for some $J \in \mathcal{H}_j$ with $j < i$. Every $H \in \mathcal{H}_i$ satisfies $N_G(H) \preceq J$ for some $J \in \mathcal{H}_1 \cup \mathcal{H}_2$.

(iii) Distinct $H, H' \in \mathcal{H}'_d$ have $H \cap H' \preceq J$ for some $J \in \mathcal{H}_1$. Also $[N_G(H) : H] \leq L_0$ for $H \in \mathcal{H}'_d$.

Proof. Step 1. The first two assertions. Conjugation preserves each definition. The number of proper divisors of f is at most $f = (\log q)/\log 2$, and, by Theorem 2.3(i), $|G| > q^4$, so $f < \log|G| \leq L_0 \log|G|$. A subgroup H of a torus normalizer $N_G(A)$ has $[H : H \cap A] \leq [N_G(A) : A] \leq 4 < L_0$, that is, $H \preceq A \in \mathcal{H}_1$; in all the other cases of Theorem 4.1(iii), H is contained in a listed group. This proves (i). Distinct groups in $\mathcal{H}_1$ meet in $\{1\} \in \mathcal{H}_0$: for two conjugates of Q and for two tori this is Theorem 4.1(i), and a group of two-power order meets a group of odd order trivially. Two distinct groups in $\mathcal{H}_2$ meet in a member of $\mathcal{T}$, as shown after Theorem 4.1, and $\mathcal{T} \subseteq \mathcal{H}_1$. The normalizer assertions in (ii) follow from Theorem 4.1(ii): the indices needed there are one, two or four.

Step 2. Centralizers of nonidentity elements. We claim that the centralizer $C_G(z)$ of every $z \neq 1$ is contained in a member of $\mathcal{H}_1$. Suppose first that z is in a conjugate Q' of Q, which is then the only one, since distinct conjugates of Q meet in $\{1\}$. If x centralizes z, then $xQ'x^{-1}$ is a conjugate of Q containing $xzx^{-1} = z$, hence equals Q'; thus x normalizes Q', and therefore x is in the conjugate $B' = N_G(Q')$ of B. If z has order two, it belongs to $Z(Q')$. Every element of B' equals ut for some $u \in Q'$ and some t in the corresponding conjugate of A_0. The element u commutes with z. If ut also commutes with z, then t does so, and, by the fixed-point assertion of Theorem 4.1(ii), $t = 1$. Thus $C_G(z) \leq Q'$. If z has order four, then z^2 is a nonidentity element of order two and $C_G(z) \leq C_G(z^2) \leq Q'$. If z is instead in some $A \in \mathcal{T}$,

this A is unique, since distinct members of $\mathcal{T}$ meet in $\{1\}$. If x centralizes z, then xAx^{-1} is a member of $\mathcal{T}$ containing z, hence equals A; thus $C_G(z) \le N_G(A) = A\langle c\rangle$. The group A is abelian, and, by the fixed-point assertion for c, $C_G(z) \le A$. These cases prove the claim.

Step 3. Two conjugate subfield groups. Let $K_d = \mathrm{Sz}(2^d)$, and let σ apply σ_d to every matrix entry. The elements of G fixed by σ are exactly K_d. Indeed, σ raises each parameter in Theorem 1.4 to its 2^d-th power, so the parameters of a fixed matrix are fixed, hence are in $\mathbb{F}_{2^d}$; conversely, those parameter choices give fixed matrices. On $\mathbb{F}_{2^d}$ the exponent θ agrees with the Suzuki exponent $2^{(d+1)/2}$ of that subfield: writing $f = kd$ with k odd, the difference $(f+1)/2 - (d+1)/2 = d(k-1)/2$ of the powers of 2 in θ and in $2^{(d+1)/2}$ is a multiple of d, and $z \mapsto z^{2^d}$ is the identity on $\mathbb{F}_{2^d}$. Suppose $K_d \ne gK_dg^{-1}$ and $h \in K_d \cap gK_dg^{-1}$. Then $h = gkg^{-1}$ for some $k \in K_d$. As σ is an automorphism of G fixing h and k,

$$h = \sigma(h) = \sigma(g)k\sigma(g)^{-1},$$

so $\sigma(g)g^{-1}$ commutes with h. This element is not 1: equality would give $g \in K_d$ and hence $gK_dg^{-1} = K_d$. The intersection is therefore contained in $C_G(\sigma(g)g^{-1})$, which, by Step 2, is contained in a member of $\mathcal{H}_1$. Conjugating this argument handles any two members of $\mathcal{H}'_d$. The last normalizer assertion follows from Theorem 4.1(iii). □

Let ε_0 be from Theorem 2.3(v), let $0 < a < 1$, and let $A \subseteq G$ generate G with $|A| < |G|^{1-a}$. Then

$$|A^3| \ge |A|^{1+\min(a,\varepsilon_0)}. \tag{6}$$

Indeed, by Theorem 2.3(v), either $|A^3| \ge |A|^{1+\varepsilon_0}$, which gives (6) at once, or $A^3 = G$, and then $|A|^{1+\min(a,\varepsilon_0)} \le |A|^{1+a} < |G|^{(1-a)(1+a)} = |G|^{1-a^2} < |G| = |A^3|$.

4.2 Three analytic inputs

For a random variable X on a finite set, denote by $p_X(x) \coloneqq \mathbb{P}(X = x)$ its law. The three entropies used here are

$$\mathrm{Ent}(X) = -\sum_x p_X(x)\log p_X(x), \qquad \mathrm{Ent}_2(X) = -\log\sum_x p_X(x)^2,$$
$$\mathrm{Ent}_\infty(X) = -\log\max_x p_X(x),$$

where $0\log 0 = 0$. As in [LS20, Definition 35 and equation (31)], a conditional entropy is the average of the entropies of the conditional laws:

$$\mathrm{Ent}_2(X \mid Y) = \sum_{y:\mathbb{P}(Y=y)>0} \mathbb{P}(Y=y)\left[-\log\sum_x \mathbb{P}(X=x \mid Y=y)^2\right],$$

and $\mathrm{Ent}(X \mid Y)$ is defined in the same way, with Ent in place of Ent_2.

Proposition 4.3 ([LS20, Lemmas 36–37])**.** *Let $X, Y, X_1, \dots, X_s$ be finite-valued random variables, and let ϕ be a function. Then*

$$\mathrm{Ent}(X_1,\dots,X_s) = \sum_{i=1}^{s}\mathrm{Ent}(X_i \mid X_1,\dots,X_{i-1}), \qquad \mathrm{Ent}(X \mid \phi(Y)) \ge \mathrm{Ent}(X \mid Y),$$
$$\log|\operatorname{supp} p_X| \ge \mathrm{Ent}(X) \ge \mathrm{Ent}_2(X) \ge \mathrm{Ent}_\infty(X).$$

If X, Y are independent and group-valued, each of $\mathrm{Ent}, \mathrm{Ent}_2, \mathrm{Ent}_\infty$ of XY is at least the corresponding entropy of either factor. Multiplication by a fixed group element preserves all three entropies.

Proposition 4.4 ([LS20, Lemma 34]). *There is an absolute $R \geq 1$ with the following property. Let ν, ξ be probability measures on a finite group K, and let $\Lambda > 2$. If*

$$\|\nu * \xi\|_2 > \Lambda^{-1} \|\nu\|_2^{1/2} \|\xi\|_2^{1/2},$$

then there is a nonempty symmetric $A \subseteq K$ such that

$$\Lambda^{-R} \|\nu\|_2^{-2} \leq |A| \leq \Lambda^R \|\nu\|_2^{-2}, \quad |A^3| \leq \Lambda^R |A|, \quad (\widetilde{\nu} * \nu)(a) \geq \Lambda^{-R} |A|^{-1} \quad \text{for every } a \in A.$$

Proposition 4.5 ([LS20, Proposition 41]). *Take q large enough for (4). Fix $a' > 0$ and $\beta > 0$. Let $X_1, \dots, X_8$ be independent G-valued random variables with $\mathrm{Ent}_\infty(X_i) \geq a' \log |G|$. Let $Y_1, \dots, Y_7$ be independent G-valued random variables with a common law, independent of all the X_i. Suppose that*

$$\mathbb{P}(Y_1 \in gH) \leq [G : H]^{-\beta} \qquad \text{for all } g \in G \text{ and } H < G.$$

For a fixed tuple $y = (y_1, \dots, y_7)$, denote $Z_y := X_1 y_1 X_2 y_2 \cdots y_7 X_8$ and

$$\beta' := 8^{-3} \min(\beta/(5L_0), a'/2).$$

If q exceeds a bound depending only on a', β, L_0, then, except for a set of y of probability at most $|G|^{-\beta/(4L_0)}$,

$$\mathbb{P}(Z_y \in gH) \leq [G : H]^{-\beta'/2} \qquad \text{for all } g \in G \text{ and } H < G.$$

The hypotheses of [LS20, Proposition 41] on the group are the bound (4) on the dimensions of the nontrivial irreducible representations of G, the triviality of the centre of G, and the subgroup condition of Theorem 4.2. The hypothesis of [LS20, Proposition 41] on the law of the Y_i is that, for some $0 < \alpha \leq \beta/4$, the coset bound $[G : H]^{-\beta}$ holds for every $H < G$ with $|H| \geq |G|^\alpha$ and for every H in $\mathcal{H}_1 \cup \mathcal{H}_2$. Our coset bound holds for every proper subgroup, so any α in that range serves; take $\alpha = \beta/8$. In [LS20, Proposition 41], with the two levels of Theorem 4.2, there are $2^{2+1} = 8$ variables X_i, and the exponent contains the factor 8^{-3}.

4.3 An entropy gain proportional to the initial entropy

The next lemma and its proof follow [LS20, Lemma 39 and Proposition 42]; we give the proof because we need γ independent of a.

Lemma 4.6. *Fix $\beta > 0$ and denote $b_0 := 1/(3L_0)$. There is $\gamma > 0$, depending only on $\beta, L_0, \varepsilon_0$, with the following property. Fix $h > 0$. Let q be sufficiently large in terms of these constants and h. Let $D = |G|^a$ be an integer, where $h \leq a \leq 1 - b_0$. Let $X_1, \dots, X_9$ be independent and uniform on sets of size D in G. Let $Y_1, \dots, Y_7$ be independent G-valued random variables with a common law, independent of all nine X_i, and suppose that $\mathbb{P}(Y_1 \in gH) \leq [G : H]^{-\beta}$ for all $g \in G$ and $H < G$. Then*

$$\mathrm{Ent}_2(X_1 Y_1 X_2 Y_2 \cdots Y_7 X_8 X_9 \mid Y_1, \dots, Y_7) \geq (1 + \gamma) \log D.$$

Proof. Step 1. Choose constants before h. Let R be from Theorem 4.4, and denote

$$k := \frac{1}{2 \cdot 8^3} \min\left(\frac{\beta}{5L_0}, \frac{1}{4}\right), \quad \Delta := \min(b_0/4, \varepsilon_0), \quad \gamma := \min\left(\frac{1}{4R}, \frac{b_0}{8R}, \frac{k}{2RL_0}, \frac{\Delta}{4R}\right).$$

Step 2. Obtain coset bounds. Let a' be the largest member of $\{h2^j : j \geq 0\}$ that is at most a. It exists because $h \leq a$, and $a' > a/2$, since otherwise $2a'$ would be a larger member of that set still at most a. Also $a' \leq a \leq 1$. Each X_i has $\mathrm{Ent}_\infty(X_i) = a \log |G| \geq a' \log |G|$.

Once h is fixed there are only finitely many possible values of a', so one lower bound on q suffices in Theorem 4.5 for every allowed a. Its β' satisfies $\beta'/2 \geq ka$: indeed $a \leq 1$ gives $\beta/(5L_0) \geq a\beta/(5L_0)$, and $a' > a/2$ gives $a'/2 > a/4$, so

$$\frac{\beta'}{2} = \frac{1}{2 \cdot 8^3} \min\left(\frac{\beta}{5L_0}, \frac{a'}{2}\right) \geq \frac{a}{2 \cdot 8^3} \min\left(\frac{\beta}{5L_0}, \frac{1}{4}\right) = ka.$$

Call y good when the conclusion of Theorem 4.5 holds for y. For good y, the law ν of Z_y satisfies $\nu(gH) \leq [G : H]^{-ka}$ for all $g \in G$ and $H < G$.

Step 3. Gain entropy for a good tuple. Let ξ be the law of X_9 and denote $t := \mathrm{Ent}_2(Z_y)$. By Theorem 4.3, $t \geq a \log|G|$. If $t > (1 - b_0/2)\log|G|$, then

$$\mathrm{Ent}_2(Z_y X_9) \geq t > a\log|G| + (b_0/2)\log|G| \geq a\log|G| + 2\gamma a \log|G|,$$

where the middle inequality uses $a \leq 1 - b_0$, and the last one uses $a \leq 1$ and $2\gamma \leq b_0/2$.

Assume instead that $t \leq (1 - b_0/2)\log|G|$. Suppose, for a contradiction, that

$$\mathrm{Ent}_2(Z_y X_9) < \frac{t + a\log|G|}{2} + 2\gamma a \log|G|.$$

Every finite-valued U has $\mathrm{Ent}_2(U) = -2\log\|p_U\|_2$, and here $\mathrm{Ent}_2(X_9) = \log D = a\log|G|$, so this inequality reads

$$-2\log\|\nu * \xi\|_2 < \frac{-2\log\|\nu\|_2 - 2\log\|\xi\|_2}{2} + 2\gamma a \log|G|,$$

that is,

$$\|\nu * \xi\|_2 > |G|^{-\gamma a}\|\nu\|_2^{1/2}\|\xi\|_2^{1/2}.$$

For large q we have $|G|^{\gamma a} \geq |G|^{\gamma h} > 2$, so we may apply Theorem 4.4 with $\Lambda = |G|^{\gamma a}$. The resulting symmetric set A satisfies

$$|G|^{a/2} \leq |A| < |G|^{1-b_0/4}, \quad |A^3| \leq |G|^{R\gamma a}|A|, \quad (\widetilde{\nu} * \nu)(A) \geq |G|^{-R\gamma a}.$$

For the lower size bound, use $t \geq a\log|G|$ and $R\gamma \leq 1/4$. For the upper size bound, use $t \leq (1 - b_0/2)\log|G|$ and $R\gamma a \leq b_0/8 < b_0/4$. The third bound is the pointwise bound of Theorem 4.4, summed over A.

Let $H = \langle A \rangle$. If $H < G$, then, for independent Z, Z' with law ν, by (5),

$$(\widetilde{\nu} * \nu)(H) = \mathbb{P}(Z^{-1}Z' \in H) = \sum_{gH \in G/H} \nu(gH)^2 \leq \max_{gH} \nu(gH) \leq |G|^{-ka/L_0},$$

which contradicts $A \subseteq H$ and $R\gamma < k/L_0$. Thus A generates G. By (6) with $b_0/4$ in place of a, as $|A| < |G|^{1-b_0/4}$,

$$\frac{|A^3|}{|A|} \geq |A|^{\Delta} \geq |G|^{\Delta a/2} > |G|^{R\gamma a},$$

again a contradiction. Hence the assumed strict inequality fails, and, as $t \geq a\log|G|$,

$$\mathrm{Ent}_2(Z_y X_9) \geq \frac{t + a\log|G|}{2} + 2\gamma a\log|G| \geq a\log|G| + 2\gamma a\log|G|.$$

Step 4. Average. For every tuple y, good or bad, by Theorem 4.3, $\mathrm{Ent}_2(Z_y X_9) \geq a\log|G|$. For large q, the probability p of a bad tuple, which is at most $|G|^{-\beta/(4L_0)}$, is at most $1/2$, so

$$\mathrm{Ent}_2(X_1 Y_1 \cdots Y_7 X_8 X_9 \mid Y_1, \ldots, Y_7) \geq (1-p)(1+2\gamma)a\log|G| + p\,a\log|G| \geq (1+\gamma)\log D. \quad \square$$

4.4 Regular sets and auxiliary variables

For $z = (z_1, \dots, z_s)$, a prefix of length i is $(z_1, \dots, z_i)$; the prefix of length zero is empty. Following [LS20, Definition 44], a nonempty $A \subseteq G^s$ is **regular with degrees $D_1, \dots, D_s$** if every prefix of length $i-1$ occurring in A has exactly D_i possible next coordinates in A.

Lemma 4.7 ([LS20, Section 4.3, the passage after Lemma 45]; see also [Var12, Section 3.2, equation (13)])**.** *Fix $h > 0$. For all sufficiently large $|G|$, in terms of h only, every $s \geq 1$ and every nonempty $S \subseteq G^s$ satisfy the following: S contains a regular subset A such that*

$$|A| \geq |S||G|^{-2hs}, \qquad D_i = 1 \text{ or } D_i \geq |G|^h \quad \text{for } 1 \leq i \leq s.$$

Moreover, $|A| = \prod_i D_i$, since exactly $\prod_{j>i} D_j$ elements of A have a given occurring prefix of length i; and if X is uniform on A, then, conditional on any occurring prefix of length $i-1$, its next coordinate is uniform on a set of size D_i.

Proposition 4.8 ([LS20, Lemma 55], after [Var12, Lemma 17])**.** *Take q large enough for (4). Let $s \geq 1$, denote $K := G^s$ and $N := |G|^s$, and take $0 < \varepsilon_1 < 1$ and $0 < \delta < \varepsilon_1/(8L_0)$. Let $S \subseteq K$ be nonempty and satisfy*

$$u_S(xH) \leq [K : H]^{-\varepsilon_1} N^{\delta} \qquad \text{for all } x \in K \text{ and } H < K.$$

There are a nonempty subset $S_1 \subseteq S$ and a set $J \subseteq [s]$ with the following properties. Denote $\beta := \varepsilon_1/(2L_0)$ and let Y be uniform on S_1. For $i \in J$,

$$\mathbb{P}(Y_i \in gH) \leq [G : H]^{-\beta} \qquad \text{for all } g \in G \text{ and } H < G.$$

Moreover, $|G|^{|[s] \setminus J|} \leq N^{\delta/\beta}$.

4.5 Growth from coordinates with large degrees

The next proposition is [LS20, Proposition 51] for G^s.

Proposition 4.9. *Let $0 < \varepsilon_1 < 1$. There are $\kappa > 0$ and $\delta_0 > 0$, depending only on ε_1 and L_0, with the following property. Let q be large enough for (4) and (5), let $s \geq 1$, denote $K := G^s$ and $N := |G|^s$, and assume that $0 < \delta < \delta_0$ and $|G| \geq 2^{2/\varepsilon_1}$. Let $S \subseteq K$ be nonempty and symmetric, with $u_S(xH) \leq [K : H]^{-\varepsilon_1} N^{\delta}$ for all $x \in K$ and $H < K$, and let $A \subseteq S$ be regular with degrees D_i. Denote $I_{\text{large}} := \{i : D_i > |G|^{1-1/(3L_0)}\}$ and $I_{\text{small}} := [s] \setminus I_{\text{large}}$. Then either $|S^3| > N^{1-\varepsilon_1+\delta}$ or*

$$|S^{14}| \geq N^{\kappa} |G|^{|I_{\text{large}}|}.$$

In [LS20], a finite group K is L^{-1}-quasi-random if $\dim \pi > |K|^{1/L}$ for every nontrivial irreducible representation π of K; for $G = \mathrm{Sz}(q)$ and $L = L_0$ this is (4). Theorem 4.9 is proved there for a direct product of pairwise nonisomorphic finite simple groups, which G^s is not when $s \geq 2$. Its proof uses this hypothesis only in [LS20, Lemma 43], to obtain [LS20, inequality (43)], which [LS20, proof of Lemma 49] uses only in the form $2^{|I_{\text{small}}|} \leq |G|^{\varepsilon_1 |I_{\text{small}}|/2}$. Under our hypothesis $|G| \geq 2^{2/\varepsilon_1}$ we have $2 \leq |G|^{\varepsilon_1/2}$, and hence that inequality, so the proof applies verbatim to G^s. Its two constants are implicit there, and named here.

4.6 The growth theorem

We need one more set inequality, due to Helfgott.

Proposition 4.10 ([Hel08, proof of Lemma 2.2], quoted as [Var12, equation (14)])**.** *Let S be a nonempty finite symmetric subset of a group. For every integer $k \geq 3$,*

$$\frac{|S^k|}{|S|} \leq \left(\frac{|S^3|}{|S|} \right)^{k-2}.$$

Theorem 4.11. *For each $0 < \varepsilon_1 < 1$, there are $\delta > 0$ and q_0 depending only on ε_1 such that the following holds. Let $q \geq q_0$, let $s \geq 1$, and denote $K := \mathrm{Sz}(q)^s$ and $N := |K|$. Let $S \subseteq K$ be nonempty and symmetric. Suppose that*

$$|S| < N^{1-\varepsilon_1}, \qquad u_S(xH) \leq [K : H]^{-\varepsilon_1} N^{\delta} \quad \text{for all } x \in K \text{ and } H < K.$$

Then $|S^3| \geq |S|^{1+\delta}$.

Proof. Step 1. Fix the constants in order. Denote $\beta := \varepsilon_1/(2L_0)$, and take κ and δ_0 from Theorem 4.9 and γ from Theorem 4.6 for this β. Choose

$$0 < h < \min\left(1, \frac{\gamma\kappa}{8(1+\gamma)}\right).$$

Choose $\delta > 0$ so that

$$\delta < \min\left(\delta_0, \frac{\varepsilon_1}{8L_0}, \frac{\beta\kappa}{4}, \frac{\gamma\kappa}{28(1+\gamma)}\right).$$

All choices depend only on ε_1 and absolute constants. Take q_0 so large that Theorem 4.7 (for this h), Theorem 4.6 (for this β and h), Theorem 4.8 and Theorem 4.9 apply; none of these thresholds depends on s. Denote $G := \mathrm{Sz}(q)$, so that $N = |G|^s$. By Theorem 4.7, there is a regular $A \subseteq S$ with $|A| \geq |S|N^{-2h}$, with degrees D_i. Take $S_1 \subseteq S$ and $J \subseteq [s]$ from Theorem 4.8, and denote $I_{\mathrm{large}} := \{i : D_i > |G|^{1-1/(3L_0)}\}$ and $I_{\mathrm{small}} := [s] \setminus I_{\mathrm{large}}$, as in Theorem 4.9.

Step 2. The case $|S^3| > N^{1-\varepsilon_1+\delta}$. If $|S^3| > N^{1-\varepsilon_1+\delta}$, then

$$|S|^{1+\delta} < N^{(1-\varepsilon_1)(1+\delta)} = N^{1-\varepsilon_1+\delta-\varepsilon_1\delta} \leq N^{1-\varepsilon_1+\delta} < |S^3|.$$

Otherwise, by Theorem 4.9, and as $D_i \leq |G|$,

$$\log|S^{14}| \geq \kappa \log N + \sum_{i \in I_{\mathrm{large}}} \log D_i. \tag{7}$$

Step 3. Sum the weighted entropy gain. Choose $X_1, \ldots, X_9$ uniform on A and $Y_1, \ldots, Y_7$ uniform on S_1, with all sixteen variables independent. For a fixed tuple $y = (y_1, \ldots, y_7) \in S_1^7$, denote $Z_y := X_1 y_1 X_2 y_2 \cdots y_7 X_8 X_9$, a random element of S^{16}. Denote by $Z_y^{(i)}$ its coordinate i, and by P_i the tuple of the nine prefixes of length $i-1$ of the X_j. The first $i-1$ coordinates of Z_y are a function of P_i, so, by the chain rule, the conditioning inequality, and $\mathrm{Ent} \geq \mathrm{Ent}_2$ applied to each conditional law, all from Theorem 4.3,

$$\mathrm{Ent}(Z_y) = \sum_{i=1}^{s} \mathrm{Ent}(Z_y^{(i)} \mid Z_y^{(1)}, \ldots, Z_y^{(i-1)}) \geq \sum_{i=1}^{s} \mathrm{Ent}(Z_y^{(i)} \mid P_i) \geq \sum_{i=1}^{s} \mathrm{Ent}_2(Z_y^{(i)} \mid P_i).$$

Conditional on a fixed occurring value of P_i, the nine next coordinates are independent, and each is uniform on a set of size D_i, by Theorem 4.7. The independence holds because the nine variables X_j are independent and each prefix condition concerns only one of them. Conditioning on P_i does not change the joint law of the Y_j, which stay independent of the X_j.

For $i \in I_{\mathrm{small}} \cap J$ with $D_i > 1$, denote $a_i := (\log D_i)/\log|G|$. Denote by $\mathbb{E}_y$ the average over y with the law of $(Y_1, \ldots, Y_7)$, and by $\mathrm{Ent}_2(\cdot \mid P_i = \vec{p})$ the Ent_2 of the conditional law given $P_i = \vec{p}$. Then $h \leq a_i$ by Theorem 4.7, and $a_i \leq 1 - 1/(3L_0)$ as $i \in I_{\mathrm{small}}$. For each fixed occurring prefix tuple $\vec{p}$, the nine next coordinates are independent and uniform on sets of size $D_i = |G|^{a_i}$, and, as $i \in J$, the $Y_j^{(i)}$ satisfy the coset bound of Theorem 4.8, so, by Theorem 4.6 with $a = a_i$,

$$\mathbb{E}_y \mathrm{Ent}_2(Z_y^{(i)} \mid P_i = \vec{p}) \geq (1+\gamma) \log D_i.$$

Hence, by the definition of conditional entropy,

$$\mathbb{E}_y \operatorname{Ent}_2(Z_y^{(i)} \mid P_i) = \sum_{\vec{p}} \mathbb{P}(P_i = \vec{p})\, \mathbb{E}_y \operatorname{Ent}_2(Z_y^{(i)} \mid P_i = \vec{p}) \geq (1+\gamma) \log D_i.$$

For $D_i = 1$, this lower bound is 0, and it holds as entropy is nonnegative. For every other coordinate i, by Theorem 4.3, $\operatorname{Ent}_2(Z_y^{(i)} \mid P_i) \geq \log D_i$ for each fixed tuple y. Summing these bounds over i,

$$\begin{aligned} \log |S^{16}| &\geq \mathbb{E}_y \operatorname{Ent}(Z_y) \\ &\geq \sum_i \log D_i + \gamma \sum_{i \in I_{\text{small}} \cap J} \log D_i \\ &\geq \log |S| + \gamma \sum_{i \in I_{\text{small}}} \log D_i - (2h + \gamma\delta/\beta) \log N. \end{aligned} \tag{$*$}$$

For the last line we used $\sum_i \log D_i = \log |A| \geq \log |S| - 2h \log N$ and, by Theorem 4.8,

$$\sum_{i \notin J} \log D_i \leq |[s] \setminus J| \log |G| \leq (\delta/\beta) \log N.$$

Step 4. Combine the two bounds. Adding γ times (7) to ($*$), and using $|S^{14}| \leq |S^{16}|$,

$$(1+\gamma) \log |S^{16}| \geq \log |S| + \gamma\Big(\sum_{i \in I_{\text{large}}} \log D_i + \sum_{i \in I_{\text{small}}} \log D_i \Big) + \gamma\kappa \log N - \Big(2h + \frac{\gamma\delta}{\beta}\Big) \log N,$$

in which the two sums over coordinates add up to $\sum_{i=1}^{s} \log D_i = \log |A|$. Using $\log |A| \geq \log |S| - 2h \log N$ a second time,

$$(1+\gamma) \log |S^{16}| \geq (1+\gamma) \log |S| + \left(\gamma\kappa - 2h(1+\gamma) - \frac{\gamma\delta}{\beta} \right) \log N.$$

By the choices of h and δ in Step 1, $2h(1+\gamma) \leq \gamma\kappa/4$ and $\gamma\delta/\beta \leq \gamma\kappa/4$, so the parenthesis is at least $\gamma\kappa/2$. By Theorem 4.10 with $k = 16$,

$$14 \log \frac{|S^3|}{|S|} \geq \log \frac{|S^{16}|}{|S|} \geq \frac{\gamma\kappa}{2(1+\gamma)} \log N \geq \frac{\gamma\kappa}{2(1+\gamma)} \log |S|.$$

As $\delta \leq \gamma\kappa/[28(1+\gamma)]$, it follows that $|S^3| \geq |S|^{1+\delta}$. □

5 From escape to a spectral gap

In this section we apply the method of Bourgain and Gamburd [BG08]: we bound the ℓ^2 norms of convolution powers of the projections μ_I of μ (Theorem 5.1; see [Var12, Theorem 3 and Section 3.1]), and deduce a spectral gap by bounding a trace, as in [Var12, Section 5, p. 303].

Lemma 5.1. *Let η be from Theorem 3.5. There are absolute $j_0 \geq 1$ and q_1 with the following property. Let $q \geq q_1$, $s \geq 1$, $K := \operatorname{Sz}(q)^s$, and $N := |K|$. Let ν be a symmetric probability measure on K such that*

$$\|\nu\|_2^2 \leq N^{-\eta}, \qquad \nu(xH) \leq [K : H]^{-\eta/2} \quad \text{for all } x \in K \text{ and } H < K.$$

Then $\|\nu^{(2^{j_0})}\|_2^2 \leq N^{-7/8}$.

Proof. Step 1. A bound on $\|\xi * \xi\|_2$. Denote $\alpha := \min(\eta/4, 1/32)$, and take $\delta > 0$ from Theorem 4.11 for $\varepsilon_1 = \alpha$. Let R be from Theorem 4.4. Choose a positive absolute ζ with $R\zeta < \min(\eta/2, 1/16, \delta/2, \delta\eta/2)$. Take q_1 large enough that $|\operatorname{Sz}(q)|^\zeta > 2$ for every $q \geq q_1$, and that Theorem 4.11 applies.

Let ξ be a convolution power of ν and suppose

$$N^{-7/8} < \|\xi\|_2^2 \leq N^{-\eta}.$$

We claim that $\|\xi * \xi\|_2 \leq N^{-\zeta}\|\xi\|_2$. If not, apply Theorem 4.4 with both measures equal to ξ and $\Lambda = N^\zeta$. The measure ξ is symmetric, and, by Theorem 2.1, every convolution power $\nu^{(n)}$ with $n \geq 1$ satisfies the coset bound of ν. Denote by τ the number with $\|\xi\|_2^{-2} = N^\tau$; then $\eta \leq \tau < 7/8$. The resulting symmetric set A satisfies

$$N^{\tau - R\zeta} \leq |A| \leq N^{\tau + R\zeta}, \qquad |A^3| \leq N^{R\zeta}|A|.$$

For all $x \in K$ and $H < K$, as $\widetilde{\xi} = \xi$, by the pointwise bound of Theorem 4.4 and by the coset bound of $\xi * \xi$,

$$N^{-R\zeta} u_A(xH) = N^{-R\zeta}\frac{|A \cap xH|}{|A|} \leq \sum_{z \in A \cap xH} (\xi * \xi)(z) \leq (\xi * \xi)(xH) \leq [K : H]^{-\eta/2}.$$

As $R\zeta < \delta$ and $\alpha < \eta/2$, it follows that $u_A(xH) \leq N^{R\zeta}[K : H]^{-\eta/2} \leq N^\delta [K : H]^{-\alpha}$. As $\eta \leq \tau < 7/8$, $R\zeta < \min(\eta/2, 1/16)$ and $\alpha \leq 1/32$,

$$|A| \geq N^{\tau - R\zeta} > N^{\eta/2}, \qquad |A| \leq N^{\tau + R\zeta} < N^{7/8 + 1/16} = N^{15/16} < N^{1-\alpha}.$$

Hence, by Theorem 4.11,

$$N^{R\zeta} \geq \frac{|A^3|}{|A|} \geq |A|^\delta \geq N^{\delta\eta/2},$$

which contradicts $R\zeta < \delta\eta/2$ and proves the claim.

Step 2. Iterate. Choose an integer $j_0 \geq 1$ with $2j_0\zeta \geq 7/8$, and denote $\xi_j := \nu^{(2^j)}$ for $0 \leq j \leq j_0$. By Theorem 2.1, $\|\xi_{j+1}\|_2 \leq \|\xi_j\|_2$ for $j < j_0$, so, if $\|\xi_j\|_2^2 \leq N^{-7/8}$ for some $j < j_0$, then $\|\xi_{j_0}\|_2^2 \leq N^{-7/8}$. Otherwise, every ξ_j with $j < j_0$ satisfies $N^{-7/8} < \|\xi_j\|_2^2 \leq \|\nu\|_2^2 \leq N^{-\eta}$, so, by Step 1, $\|\xi_{j+1}\|_2^2 = \|\xi_j * \xi_j\|_2^2 \leq N^{-2\zeta}\|\xi_j\|_2^2$. Hence, as $2j_0\zeta \geq 7/8$,

$$\|\xi_{j_0}\|_2^2 \leq N^{-2j_0\zeta}\|\nu\|_2^2 \leq N^{-7/8-\eta} < N^{-7/8}. \qquad \square$$

Proof of Theorem 1.3. Work on the event that the three parts of Theorem 3.2 hold, whose probability is at least $1 - C_7 q^{-1/10}$; take $C_1 = C_7$, and take q_* larger than every absolute threshold used below.

Step 1. Apply Theorem 5.1 to every nonempty set of coordinates. Let $I \subseteq [m]$ be nonempty, let $s := |I|$, and denote $K := G^I$, $N := |G|^s$ and $\nu := \mu_I^{(n_s)}$. By Theorem 3.5, $\nu(xH) \leq [K : H]^{-\eta/2}$ for all $x \in K$ and $H < K$. As ν is symmetric and n_s is even, by Theorem 2.1 with $n = n_s$, $n' = n_s/2$ and $H = \{1\}$, and by Theorem 3.5 with $H = \{1\}$,

$$\|\nu\|_2^2 = (\nu * \nu)(1) = \mu_I^{(2n_s)}(1) \leq \mu_I^{(n_s)}(1) \leq [G^I : \{1\}]^{-\eta} = N^{-\eta}.$$

Denote $T := 2^{j_0} n_s$, so that $\nu^{(2^{j_0})} = \mu_I^{(T)}$. By Theorem 5.1,

$$\|\mu_I^{(T)}\|_2^2 \leq N^{-7/8}. \tag{8}$$

Moreover, $T \leq C_T \log N$, where $C_T := 21 \cdot 2^{j_0}/4$ is absolute. Indeed, by Theorem 2.3(i), $|G| > q^4$, so $\log N = s \log |G| \geq 4s \log q$, and, by Theorem 3.2, $n_s \leq 21 s \log q \leq (21/4) \log N$.

Step 2. Bound the eigenvalues on representations nontrivial at every coordinate of I. For all sufficiently large q we have $c_{\mathrm{rep}} q^{3/2} \geq q^{5/4} > |G|^{1/4}$. Let $\pi\colon K \to \mathrm{GL}(V)$ be an irreducible unitary representation that is the tensor product of nontrivial irreducible representations π_i of the factors, for $i \in I$. By Theorem 2.3(iv),

$$\dim \pi = \prod_{i \in I} \dim \pi_i \geq |G|^{s/4} = N^{1/4}.$$

Denote by $\pi(\mu_I) \coloneqq \sum_{g \in K} \mu_I(g)\pi(g)$ the image of μ_I in π, and similarly for other representations and measures. The operator $\pi(\mu_I)$ is self-adjoint, because $\pi(g)^* = \pi(g^{-1})$ and μ_I is symmetric; let λ be any of its eigenvalues.

For $k \geq 1$, the measure $\mu_I^{(k)}$ is symmetric, and, for $v \in \ell^2(K)$ and $x \in K$,

$$M_{\mu_I} M_{\mu_I^{(k)}} v(x) = \sum_{g,h} \mu_I(g)\mu_I^{(k)}(h) v((gh)^{-1}x) = M_{\mu_I^{(k+1)}} v(x),$$

so $M_{\mu_I}^{2T} = M_{\mu_I^{(2T)}}$. The operator M_{μ_I} on $\ell^2(K)$ is the image of μ_I in the regular representation of K, which contains $\dim \pi$ copies of the irreducible representation π, so λ is an eigenvalue of M_{μ_I} of multiplicity at least $\dim \pi$. Like $\pi(\mu_I)$, the operator M_{μ_I} is self-adjoint, so its eigenvalues are real, and their $2T$-th powers are nonnegative. Hence $\dim \pi \cdot |\lambda|^{2T} \leq \mathrm{Tr}(M_{\mu_I}^{2T})$. In the basis $\{\mathbf{1}_{\{x\}} : x \in K\}$, each diagonal entry of $M_{\mu_I}^{2T}$ is $\mu_I^{(2T)}(1)$, so, by Theorem 2.1 and (8),

$$\mathrm{Tr}(M_{\mu_I}^{2T}) = N\mu_I^{(2T)}(1) = N\|\mu_I^{(T)}\|_2^2 \leq N^{1/8}.$$

Therefore $|\lambda|^{2T} \leq N^{1/8}/\dim \pi \leq N^{-1/8}$, and

$$|\lambda| \leq \exp\left(-\frac{\log N}{16T}\right) \leq \exp\left(-\frac{1}{16C_T}\right).$$

Step 3. Treat every nontrivial representation. Every irreducible representation of G^m is a tensor product of irreducible representations of the m factors. If it is nontrivial, the set I of coordinates at which its factor is nontrivial is nonempty, and the factors at the other coordinates are trivial, so the image of μ in it equals the image of μ_I in the tensor product of its factors at the coordinates in I. Fix a decomposition of $\ell^2(G^m)$ into pairwise orthogonal irreducible subrepresentations. Hence, by Step 2, every eigenvalue of M_μ on every nontrivial summand of this decomposition has absolute value at most $\exp(-1/(16C_T))$. The constant functions form the only trivial summand, so $\ell_0^2(G^m)$ is the orthogonal sum of the nontrivial summands, and M_μ preserves each of them. As M_μ is self-adjoint, it follows that $\|M_\mu|_{\ell_0^2(G^m)}\|_{\mathrm{op}} \leq \exp(-1/(16C_T))$. Take $\varepsilon = 1 - \exp(-1/(16C_T)) > 0$.

Step 4. Check the generating set. By Theorem 3.5 with $I = [m]$, the elements a and b generate G^m. We may take q_* so large that $g_1 \geq 2$. By Theorem 3.2(i), applied to one coordinate, none of the words x^2, y^2, xy and xy^{-1}, which are nontrivial of length two, equals 1 at (a, b). Each equality between two of a, a^{-1}, b, b^{-1} is equivalent to one of these four relations, so the four elements are distinct. The multiset $\{a, a^{-1}, b, b^{-1}\}$ is therefore a symmetric set S of four elements with $u_S = \mu$. □

References

[BG08] J. Bourgain and A. Gamburd, *Uniform expansion bounds for Cayley graphs of* $\mathrm{SL}_2(\mathbb{F}_p)$, Ann. of Math. (2) **167** (2008), 625–642.

[BGGT15] E. Breuillard, B. Green, R. Guralnick, and T. Tao, *Expansion in finite simple groups of Lie type*, J. Eur. Math. Soc. **17** (2015), 1367–1434. Publisher's PDF.

[BGT11] E. Breuillard, B. Green, and T. Tao, *Suzuki groups as expanders*, Groups Geom. Dyn. **5** (2011), 281–299. Author's PDF.

[CV96] P.-A. Cherix and A. Valette, *On spectra of simple random walks on one-relator groups*, with an appendix by P. Jolissaint, Pacific J. Math. **175** (1996), 417–438. Publisher's PDF.

[DJ14] M. Downs and G. A. Jones, *The Möbius function of the Suzuki groups, with applications to enumeration*, arXiv:1404.5470v1 (2014). Cited version.

[DS19] P. H. P. Daldegan and C. Schneider, *Transitive characteristically simple subgroups of finite quasiprimitive permutation groups*, arXiv:1901.07285v3 (2019). Cited version.

[Hel08] H. A. Helfgott, *Growth and generation in* $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, Ann. of Math. (2) **167** (2008), 601–623.

[Kes59] H. Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354.

[LNS11] M. W. Liebeck, N. Nikolov, and A. Shalev, *Groups of Lie type as products of* SL_2 *subgroups*, J. Algebra **326** (2011), 201–207.

[LS20] B. Longo and A. Salehi Golsefidy, *Towards super-approximation in positive characteristic*, arXiv:1908.07014v2 (2020). Cited version.

[LS26] A. Lubotzky and Y. Shomroni, *Expansion in powers of simple groups of Lie type*, arXiv:2611.21492v1 (2026). Cited version.

[PS11] L. Pyber and E. Szabó, *Growth in finite simple groups of Lie type of bounded rank*, arXiv:1005.1858v2 (2011). Cited version.

[Var12] P. P. Varjú, *Expansion in* $\mathrm{SL}_d(\mathcal{O}_K/I)$*, I square-free*, J. Eur. Math. Soc. **14** (2012), 273–305. Publisher's PDF.