THE LARGEST INDEPENDENT SUBSET OF A SPHERE OR A BALL IN A FREE GROUP: EXACT VALUES, AND THE COINCIDENCE WITH THE MAXIMUM RANK

KOREA SUPERINTELLIGENCE LABS

Abstract. Let F_r be the free group of rank r, let S_n be its sphere of radius n — the freely reduced words of length exactly n — and let B_n be the corresponding ball. A subset X of a group is *independent* when no element of X lies in the subgroup generated by the others. Koch-Hyde and Olive (arXiv:2609.00382) determine the largest rank $\rho(r,n)$ of a subgroup of F_r generated by elements of S_n, or of B_n, answering a problem of Dotsenko, and close by asking for the largest *independent* subset of S_n and of B_n. They leave that question open, recording only that the answer is $\Theta\big(\sqrt{2r-1}^{\,n}\big)$ and that $\rho(r,n)+\lfloor\log_2 v(r,n)\rfloor$ is an upper bound, $v(r,n)$ being their bound on the number of vertices of a Stallings core graph. We determine the answer in four cells. For every rank r, the largest independent subset of S_1 and of B_1 has exactly r elements. In F_2 the largest independent subset of S_2 and of B_2 has exactly 3 elements and that of S_3 has exactly 6; in F_3 the largest independent subset of S_2 and of B_2 has exactly 5. Every one of these numbers is $\rho(r,n)$ exactly, so in each settled cell the logarithmic slack in the stated bound is unused, and in the cells of radius at least two both bounds recorded in the source are strict. The coincidence is not formal: $\{a^2,a^3\}$ is independent of size two and generates a subgroup of rank one. Whether it persists we leave as a question. The upper bounds rest on exhaustive finite computation — for each cell, a list of certified word identities among the inverse classes of the sphere or the ball, together with an enumeration of all index sets avoiding them. Every theorem, proposition and lemma below is machine-checked in Lean 4 against Mathlib.

1. Introduction

The objects. Fix $r \geq 1$ and let F_r be the free group on $x_1,\dots,x_r$; for $r=2$ and $r=3$ we write $a=x_1$, $b=x_2$, $c=x_3$. Every element of F_r has a unique freely reduced expression, and $|w|$ denotes its length. Following Koch-Hyde and Olive [5] we set, in their words,

> $S_n \subset F_r$ is the set of all freely reduced words of length exactly n, and $B_n \subset F_r$ is the set of all freely reduced words of length at most n,

the sphere and the ball of radius n around the identity in the word metric, and we use their definition of independence, again verbatim:

> A subset X of a group G is *independent* when for all $x \in X$, $x \notin \langle X - \{x\}\rangle$.

Since S_n and B_n are finite, so is every subset of them, and we may write

$$I_S(r,n) \;=\; \max\{\,|X| : X \subseteq S_n \text{ independent}\,\}, \qquad I_B(r,n) \;=\; \max\{\,|X| : X \subseteq B_n \text{ independent}\,\}$$

for the two quantities in question.

The source and its open problem. The paper [5] solves two problems. Its Theorem `thm:max rank` answers Dotsenko's Problem F42 of the Kapovich–Myasnikov–Shpilrain collection [4, 1]: the greatest rank of a subgroup of F_r generated by elements of S_n, and likewise of B_n, is

$$\rho(r,n) \;=\; \begin{cases} (2r-1)^{n/2}, & n \text{ even},\\ r\,(2r-1)^{(n-1)/2}, & n \text{ odd}. \end{cases} \tag{1}$$

(The environments of [5] are auto-numbered, so we quote its internal labels rather than numbers throughout.) Its Theorem `thm:independent set limit` answers Grigorchuk's Problem 10.13 of

the Kourovka Notebook [6] by showing that every independent $S \subseteq B_n \subseteq F_r$ satisfies $|S| \in O\big(\sqrt{2r-1}^{\,n}\big)$; a companion theorem in its §4 makes the constant explicit, bounding $|S|$ by the number of edges of a Stallings core graph:

$$|S| \;\le\; \beta(r,n) \;=\; \begin{cases} r\dfrac{\sqrt{2r-1}^{\,n}-1}{r-1}, & n \text{ even}, \\[2ex] r\dfrac{r\sqrt{2r-1}^{\,n-1}-1}{r-1}, & n \text{ odd} \end{cases} \qquad \Big(\text{that is, } \beta(r,n) = \tfrac{r}{r-1}\big(\rho(r,n)-1\big)\Big). \tag{2}$$

The paper then closes with three problems, the last of which reads, verbatim:

> **Problem.** What is the largest independent subset of: (a) $S_n \subset F_r$? (b) $B_n \subset F_r$?

followed by the commentary

> Here we know that the answer lies in $\Theta\left(\sqrt{2r-1}^{\,n}\right)$ but a finer characterization would be interesting. Although we will not prove it, it is not hard to see using the results provided herein that the following is an upper bound: $\rho(r,n)+\lfloor\log_2(v(r,n))\rfloor$,

where

$$v(r,n) \;=\; \frac{r(2r-1)^{\lfloor n/2\rfloor}-1}{r-1} \tag{3}$$

is their bound on the number of vertices of the core graph. No exact value of I_S or I_B appears anywhere in [5], at any (r,n).

Two elementary inequalities frame the question. A basis of a free subgroup is independent, and $S_n \subseteq B_n$, so

$$\rho(r,n) \;\le\; I_S(r,n) \;\le\; I_B(r,n). \tag{4}$$

In the other direction the two bounds above give $I_B(r,n) \le \min\big\{\rho(r,n) + \lfloor\log_2 v(r,n)\rfloor,\ \beta(r,n)\big\}$, the first term being the bound [5] states without proof. Independence is *strictly* weaker than being a basis — $\{a^2, a^3\}$ has two elements and generates the rank-one subgroup $\langle a\rangle$ — so nothing formal forces the left inequality of (4) to be an equality, and the $\lfloor\log_2 v\rfloor$ term is exactly the room that leaves.

Results. We compute I_S and I_B in four cells. Everything in the table below is proved in this paper except the parenthesised entry, which is a computation outside the formal development (Section 7).

cell	$\rho(r,n)$	$I_S(r,n)$	$I_B(r,n)$	$\rho + \lfloor\log_2 v\rfloor$	$\beta(r,n)$
$n=1$, every r	r	r	r	r	r
$r=2$, $n=2$	3	3	3	5	4
$r=3$, $n=2$	5	5	5	7	6
$r=2$, $n=3$	6	6	≥ 6 $(=6)$	8	10

The last two columns are the two bounds of [5], evaluated in Section 2; both carry a factor $1/(r-1)$, so the first row of those columns is read for $r \ge 2$. The radius-one line (Theorem 3.1) is not new content, and we say so plainly: at $n=1$ the bound $\rho(r,n) + \lfloor\log_2 v(r,n)\rfloor$ that [5] states without proof evaluates to $r + \lfloor\log_2 1\rfloor = r$, the free basis realises r, and the value follows. What is offered here is a proof of it, uniform in r. The three remaining cells (Theorems 4.1, 4.2 and 4.3) sit strictly below both bounds of [5], so no published estimate implies them.

The uniform reading of the table is Theorem 5.1: in every cell settled here the answer to the problem of [5] is exactly $\rho(r,n)$, the answer to Dotsenko's problem. The logarithmic slack is never used; and passing from "freely generates a subgroup" to "is independent" — Dotsenko's hypothesis to that of [5] — costs nothing in these cells. Proposition 5.2 records that this is a statement and not a tautology. Whether the pattern persists is Question 5.3, which we pose and do not answer.

What is not claimed. We prove no general formula and no new asymptotic bound; the values above are the settled cells of a finite computation, not a theorem about all (r, n). We claim no error in [5]: every value we compute is consistent with both bounds recorded there. Section 8 records two observations about the text of [5] that a reimplementer may find useful, neither of which affects the correctness of any statement in that paper. Sections 6 and 7 say exactly which claims rest on exhaustive computation, what finite searches were run, and which cells were priced and abandoned.

Related work. The notion of independence used here — no element in the subgroup generated by the others — is the one fixed in [5], and is weaker than the "normal independence" of Grigorchuk's Problem 10.13, in which $\langle\cdot\rangle$ is replaced by the normal closure. It also differs from a usage common in the literature on free groups, in which an "independent subset" means a free basis. The closest published occurrence of the present notion known to us is Problem 16.45 of the Kourovka Notebook [6], proposed by P. J. Cameron, which introduces the maximum size $\mu'(G)$ of an independent set in a group G; that problem concerns finite groups and imposes no length bound. For the folding technique that the searches of Section 6 use — never to certify, only to find — see Stallings [10] and Kapovich–Myasnikov [3].

2. Preliminaries

Independence. Throughout, G is a group and $X \subseteq G$ is independent when $x \notin \langle X \setminus \{x\}\rangle$ for every $x \in X$. Two facts are used constantly.

Lemma 2.1. *Let $X \subseteq G$ be independent. Then $1 \notin X$; and if $x \in X$ with $x^{-1} \neq x$ then $x^{-1} \notin X$. Every subset of X is independent.*

Proof. The identity lies in every subgroup, so $1 \in X$ would contradict independence at 1. If both x and x^{-1} lay in X and $x^{-1} \neq x$, then $x^{-1} \in X \setminus \{x\}$ and hence $x \in \langle X \setminus \{x\}\rangle$. For the last claim, $Y \subseteq X$ gives $Y \setminus \{y\} \subseteq X \setminus \{y\}$ and closure is monotone. □

In a free group $x^{-1} = x$ only for $x = 1$, so an independent subset of F_r contains at most one member of each pair $\{x, x^{-1}\}$. We call such a pair an *inverse class*. The sphere S_n $(n \geq 1)$ is a union of $|S_n|/2 = r(2r-1)^{n-1}$ inverse classes, and B_n is a union of $(|B_n| - 1)/2$ inverse classes together with $\{1\}$. Replacing an element of an independent set by its inverse leaves the set independent, since a subgroup contains g if and only if it contains g^{-1}; this halving is what makes the searches of Section 6 feasible, and it is also what the upper-bound tool below exploits.

Certifying a lower bound.

Lemma 2.2. *Let $W \subseteq G$ be finite. Suppose that for every $x \in W$ there are a finite set Ω, a homomorphism $\pi\colon G \to \mathrm{Sym}(\Omega)$ and a point $p \in \Omega$ such that $\pi(s)\,p = p$ for all $s \in W$ with $s \neq x$, while $\pi(x)\,p \neq p$. Then W is independent.*

Proof. Fix x and take π, p as given. The stabiliser $K = \{g \in \mathrm{Sym}(\Omega) : g\,p = p\}$ is a subgroup, and π carries $W \setminus \{x\}$ into K, hence carries $\langle W \setminus \{x\}\rangle$ into K; but $\pi(x) \notin K$. So $x \notin \langle W \setminus \{x\}\rangle$. □

Such witnesses always exist for a finite independent subset W of a free group F, so the method is not a restriction. Theorem 5.1 of M. Hall [2] separates finitely many elements from a finitely generated subgroup that omits them: since $x \notin \langle W \setminus \{x\}\rangle$, it gives a finite-index $H \leq F$ with $W \setminus \{x\} \subseteq H$ and $x \notin H$, and the action of F on the cosets of H, with $p = H$, is a witness of the required kind. No proof below depends on that fact: each witness is exhibited explicitly, and the required permutation identities are decided by evaluation.

Certifying an upper bound. The upper bounds are certified by explicit dependences among the inverse classes.

Definition 2.3. Let $e\colon \{1,\dots,N\} \to G$. A *circuit* for e is a datum $c = \big(t;\ (j_1,\varepsilon_1),\dots,(j_m,\varepsilon_m)\big)$ with $t, j_1, \dots, j_m \in \{1,\dots,N\}$, $\varepsilon_l \in \{\pm 1\}$, subject to

$$t \notin \{j_1,\dots,j_m\} \qquad \text{and} \qquad e(t) \;=\; e(j_1)^{\varepsilon_1} e(j_2)^{\varepsilon_2} \cdots e(j_m)^{\varepsilon_m}.$$

Its *support* is $\operatorname{supp}(c) = \{t, j_1, \dots, j_m\}$. A set $M \subseteq \{1,\dots,N\}$ is *circuit-free* for a list $\mathcal{C}$ of circuits when $\operatorname{supp}(c) \not\subseteq M$ for every $c \in \mathcal{C}$.

Lemma 2.4. *Let $e\colon \{1,\dots,N\} \to G$, let $\mathcal{C}$ be a finite list of circuits for e, and let k be such that every circuit-free $M \subseteq \{1,\dots,N\}$ has $|M| \le k$. Then every independent $X \subseteq \{\, e(i)^{\pm 1} : 1 \le i \le N \,\}$ has $|X| \le k$.*

Proof. Choose $f(x) \in \{1,\dots,N\}$ with $x = e(f(x))^{\pm 1}$ for each $x \in X$. The map f is injective on X: if $f(x) = f(y) = i$ and $x \neq y$, then $\{x, y\} = \{e(i), e(i)^{-1}\}$, so $y = x^{-1}$, and then $x = y^{-1} \in \langle X \setminus \{x\} \rangle$, contradicting independence. Put $M = f(X)$, so $|M| = |X|$.

Suppose some $c \in \mathcal{C}$ had $\operatorname{supp}(c) \subseteq M$. Pick $x_0 \in X$ with $f(x_0) = t$. For each l there is $y_l \in X$ with $f(y_l) = j_l$, and $y_l \neq x_0$ because $j_l \neq t$ and f is injective on X; since $e(j_l) = y_l^{\pm 1}$, we get $e(j_l) \in \langle X \setminus \{x_0\} \rangle$. The defining identity of c then places $e(t) \in \langle X \setminus \{x_0\} \rangle$, hence also $x_0 = e(t)^{\pm 1}$, contradicting independence at x_0. So M is circuit-free and $|X| = |M| \le k$. □

In the applications, e enumerates one representative of each inverse class of S_n — or of B_n, in which case one extra index is spent on the identity, entered as the circuit $(t; \varnothing)$ with empty product, valid because 1 lies in every subgroup. The hypothesis of Lemma 2.4 then holds for every independent subset of the sphere or the ball, by a finite check that each element of S_n (respectively B_n) is $e(i)^{\pm 1}$ for some i: one enumerates the $(2r)^n$ letter sequences of length n (respectively of length at most n), discards the unreduced ones and matches the rest against the list.

The two quantities of [5]**.** We will use ρ, v and β as defined in (1), (3) and (2). Both v and β carry a factor $1/(r-1)$ and so are read for $r \ge 2$; ρ itself is defined for every r. For the reader's convenience: $v(r,1) = 1$, $v(2,2) = v(2,3) = 5$ and $v(3,2) = 7$, so that $\rho + \lfloor \log_2 v \rfloor$ equals r, 5, 7 and 8 in the four cells of the table, in that order, while β equals r, 4, 6 and 10.

3. Radius one, every rank

Theorem 3.1. *For every $r \ge 1$,*

$$I_S(r,1) \;=\; I_B(r,1) \;=\; r.$$

Proof. Lower bound. The free basis $\{x_1,\dots,x_r\}$ lies in $S_1 \subseteq B_1$, has r elements, and is independent by Lemma 2.2: for x_i take Ω of size two, π the homomorphism sending x_i to the transposition and every other generator to the identity, and p either point. Then π kills x_j for $j \neq i$, so it fixes p there, and moves p at x_i.

Upper bound. A word of length one is x_i or x_i^{-1} for a unique i. Let $X \subseteq S_1$ be independent and let $\lambda(w) \in \{1,\dots,r\}$ be the index of the letter of w. If $\lambda(x) = \lambda(y)$ for $x, y \in X$ with $x \neq y$ then $y = x^{-1}$, which Lemma 2.1 forbids; so λ is injective on X and $|X| \le r$. For the ball, an independent $X \subseteq B_1$ cannot contain 1 (Lemma 2.1), so $X \subseteq S_1$ and the same bound applies. □

As noted in Section 1, the value r is what the bound stated without proof in [5] gives at $n = 1$, together with the free basis; the content here is a proof, and one uniform in the rank.

4. Three exact cells

Each of the next three theorems has the same shape: an explicit independent set of the stated size, certified by one finite permutation quotient per element, and a matching upper bound from Lemma 2.4 applied to an explicit list of circuits. The lower bounds are the easy half and are essentially those of [5]: in each case the witness is one of that paper's own maximal bases, up to replacing individual elements by their inverses, which changes neither independence nor the property of being a basis. The content is the upper bound.

Theorem 4.1. *In F_2,*

$$I_S(2,2) \;=\; I_B(2,2) \;=\; 3.$$

Proof sketch. Lower bound. The set $W = \{a^{-2},\ ab,\ ab^{-1}\}$ lies in $S_2 \subseteq B_2$ and is independent: for each of its three elements we exhibit a homomorphism $F_2 \to \mathrm{Sym}(\{1,2,3\})$ and a point fixed by the other two elements and moved by that one, and apply Lemma 2.2. (This is the set $Q_2(1)$ of [5] with one element inverted.)

Upper bound. The sphere $S_2 \subset F_2$ has 12 elements in 6 inverse classes; we fix the representatives

$$e = \big(a^{-2},\ ab,\ ab^{-1},\ a^{-1}b,\ a^{-1}b^{-1},\ b^{-2}\big)$$

and exhibit 7 circuits for e, of support size at most 4, each expressing its target as a product of at most 3 of the marked classes and their inverses; a representative one is $a^{-2} = (a^{-1}b)(ab)^{-1}$. Each defining identity is checked by reducing both sides to normal form. Every element of S_2 is $e(i)^{\pm 1}$ for some i, verified by enumerating the $4^2 = 16$ letter sequences of length two and matching the 12 reduced ones. Finally, all $2^6 = 64$ subsets of $\{1,\dots,6\}$ are enumerated, and every circuit-free one is found to have at most 3 elements; Lemma 2.4 gives $I_S(2,2) \le 3$.

For the ball, $B_2 \subset F_2$ has 17 elements: 8 inverse classes and the identity, which is given a ninth index. The same scheme runs with 18 circuits — 17 genuine ones plus the empty-product circuit that rules the identity out of any independent set — and an enumeration of all $2^9 = 512$ subsets, giving $I_B(2,2) \le 3$. Both bounds are matched by W. □

Theorem 4.2. *In F_3,*

$$I_S(3,2) \;=\; I_B(3,2) \;=\; 5.$$

Proof sketch. Lower bound. The set $W = \{a^{-2},\ ab,\ ab^{-1},\ ac,\ ac^{-1}\}$ lies in $S_2 \subseteq B_2$; it is $Q_3(1)$ of [5] with one element inverted, and independence is certified by five homomorphisms into $\mathrm{Sym}(\{1,2,3\})$ as in Lemma 2.2.

Upper bound. $S_2 \subset F_3$ has 30 elements in 15 inverse classes. We exhibit 197 circuits for the corresponding marking, of support size at most 6 and expressing each target as a product of at most 5 marked classes and their inverses, and enumerate all $2^{15} = 32\,768$ index sets: every circuit-free one has at most 5 elements. Covering — that every element of S_2 is $e(i)^{\pm 1}$ for some i — is checked over the $6^2 = 36$ letter sequences of length two. For the ball, $B_2 \subset F_3$ has 37 elements: 18 inverse classes and the identity, so 19 indices, 289 circuits (288 plus the identity circuit), and an enumeration of all $2^{19} = 524\,288$ index sets, every circuit-free one of size at most 5. □

Theorem 4.3. *In F_2,*

$$I_S(2,3) \;=\; 6.$$

Proof sketch. Lower bound. The six words

$$W = \big\{\, a^{-3},\ (aba)^{-1},\ a^{-1}ba^{-1},\ (bab)^{-1},\ b^{-1}ab^{-1},\ b^{-3} \,\big\}$$

all have length three; they form the set $P_2(1)$ of [5] with four of its six elements inverted. Independence is certified elementwise by Lemma 2.2, with permutation quotients of degree 4 for the two elements a^{-3}, b^{-3} and degree 6 for the other four.

Upper bound. $S_3 \subset F_2$ has 36 elements in 18 inverse classes. We exhibit 246 circuits for the marking; their supports have at most 4 indices, but the expressing words are longer than in the

previous cells, being products of up to 23 of the marked classes and their inverses. Covering is checked over the $4^3 = 64$ letter sequences of length three. All $2^{18} = 262\,144$ index sets are then enumerated, and every circuit-free one has at most 6 elements, so $I_S(2,3) \le 6$ by Lemma 2.4. □

Corollary 4.4. *$I_B(2,3) \ge 6$: the six words of Theorem 4.3 form an independent subset of B_3.*

Proof. $S_3 \subseteq B_3$. □

The matching upper bound for the ball at $(2,3)$ is not proved here; see Section 7.

5. The coincidence with the maximum rank

Reading the four cells against (1) gives the uniform statement.

Theorem 5.1. *In each of the cells settled above — $(r,1)$ for every r, and $(2,2)$, $(3,2)$, $(2,3)$ — the largest independent subset of the sphere has exactly $\rho(r,n)$ elements:*

$$I_S(r,1) = \rho(r,1), \quad I_S(2,2) = \rho(2,2), \quad I_S(3,2) = \rho(3,2), \quad I_S(2,3) = \rho(2,3).$$

The same holds for the ball at $(r,1)$, $(2,2)$ and $(3,2)$, and at $(2,3)$ the value $\rho(2,3) = 6$ is realised by an independent subset of B_3. In particular, in each of these cells the term $\lfloor \log_2 v(r,n) \rfloor$ of the upper bound stated in [5] *is entirely unused, and both bounds recorded there are strict for $n \ge 2$.*

Proof. Immediate from Theorems 3.1, 4.1, 4.2, 4.3, Corollary 4.4 and the evaluations $\rho(r,1) = r$, $\rho(2,2) = 3$, $\rho(3,2) = 5$, $\rho(2,3) = 6$ of (1). The final sentence is the comparison $\rho < \rho + \lfloor \log_2 v \rfloor$ and $\rho < \beta$, valid because $\lfloor \log_2 v(r,n) \rfloor = 2$ and $\beta(r,n) > \rho(r,n)$ for the three cells with $n \ge 2$ listed in the table of Section 1. □

The statement is not a tautology, because independence does not bound a set by the rank of what it generates.

Proposition 5.2. *In F_2 the set $\{a^2, a^3\}$ is independent and $\langle a^2, a^3 \rangle = \langle a \rangle$, a subgroup of rank one. Hence an independent set can be strictly larger than the rank of the subgroup it generates. Moreover, independence is a genuine restriction on subsets of a sphere: the three elements $a^{-2}, ab, a^{-1}b$ all lie in $S_2 \subset F_2$ and are not independent, since $a^{-2} = (a^{-1}b)(ab)^{-1}$.*

Proof. Independence of $\{a^2, a^3\}$ is Lemma 2.2 with two witnesses: a homomorphism onto a 3-cycle sends a^3 to the identity and a^2 to a point-moving permutation, and a homomorphism onto a transposition sends a^2 to the identity and a^3 to a point-moving permutation. For the closure, $a^2, a^3 \in \langle a \rangle$ gives one inclusion and $a = a^3(a^2)^{-1}$ the other. The last claim is the displayed identity. □

Dotsenko's Problem F42, in the wording of the authors' own online copy of the collection [1], asks for "the largest number of elements from S_n that freely generate a subgroup of F_r", which is the problem of finding the largest *basis* inside S_n; [5] renders it as the highest rank of a subgroup generated by elements of S_n. Its Theorem `thm:max rank` answers both readings at once with the value $\rho(r,n)$, by bounding the rank of $\langle A \rangle$ for any $A \subseteq B_n$ and exhibiting bases of size $\rho(r,n)$ inside S_n. Theorem 5.1 says that in the settled cells, weakening the hypothesis on the set from "is a basis" to "is independent" does not increase the answer.

Question 5.3. Is $I_S(r,n) = I_B(r,n) = \rho(r,n)$ for all $r \ge 2$ and all $n \ge 1$?

The evidence is the four cells of Theorem 5.1, one of them uniform in the rank; we regard that as evidence and not as a proof, and we do not conjecture an answer. A proof would settle the problem of [5] outright. The restriction $r \ge 2$ is necessary.

Remark 5.4. In $F_1 = \mathbb{Z}$ the pattern fails. Reading (1) at $r = 1$ gives $\rho(1,n) = 1$ for every n, while $\{6, 10, 15\} \subseteq B_{15}$ is independent: $6 \notin \langle 10, 15 \rangle = 5\mathbb{Z}$, $10 \notin \langle 6, 15 \rangle = 3\mathbb{Z}$, and $15 \notin \langle 6, 10 \rangle = 2\mathbb{Z}$. So $I_B(1,15) \ge 3$, and any proof of the pattern must use $r \ge 2$ somewhere. This observation is elementary and is stated outside the formal development.

Remark 5.5. The natural route to Question 5.3 runs through the argument of [5] itself. Its Lemma `lem:monotone` yields $|S| \leq E$, the number of edges of the core graph of a maximal basis $K \subseteq S$, and by the Stallings rank formula $E = \operatorname{rank}\langle K\rangle + V - 1$. Closing the gap between E and ρ is the whole content of Question 5.3; the cells computed here say that in those cases the $V - 1$ vertices contribute nothing.

6. How the certificates were found

Nothing in the proofs depends on the programs described in this section: every certificate they produce is re-verified from scratch, and the verification is a finite check on explicit data. We record the searches because the theorems of Section 4 rest on exhaustive computation and a reader is entitled to know exactly which.

The exact values, before certification. For each cell, the maximum independent subset was first computed by depth-first search over the inverse classes of the sphere or the ball, using a subgroup-membership oracle built from Stallings foldings [10, 3] written for this purpose: to test $x \in \langle Y\rangle$, fold the wedge of loops labelled by Y and read x along the resulting core graph from the base point. The reduction to inverse classes is justified in Section 2. These searches are exhaustive over the reduced space — at most 2^N subsets, where $N = 6, 8, 15, 18, 18$ is the number of inverse classes in the five certified cells — and finish in seconds.

Lower-bound certificates. For each element x of the witness W, a homomorphism into a symmetric group of degree at most 6 separating x from $\langle W \setminus \{x\}\rangle$ was found by brute-force enumeration over $\mathrm{Sym}(m)^r$ for $m \leq 5$, with a completion from the Stallings graph of $\langle W \setminus \{x\}\rangle$ as a fallback. The degrees actually used are 2 at $n = 1$, 3 at $(2, 2)$ and $(3, 2)$, and 4 and 6 at $(2, 3)$.

Upper-bound certificates. The circuit lists were assembled in four stages: (i) all minimal dependent sets of at most three or four classes were taken as seeds, the expressing words coming from Nielsen reduction with expression tracking, with a length-capped breadth-first product search as a fallback; (ii) each seed was closed under the $2^r r!$ length-preserving automorphisms of F_r that permute the generators and invert some of them — these carry circuits to circuits with a computable sign correction, and the effect is substantial: on the $(3, 2)$ sphere the loop of the next stage needed 8 rounds with the orbit closure and 177 rounds without; (iii) a cutting-plane loop added circuits until the largest circuit-free index set matched the true maximum found by the search; (iv) a greedy pass deleted every circuit the bound does not need. Every expressing word was evaluated and checked before being recorded.

What is exhaustive in the proofs. The upper bound in each cell consists of two finite checks and nothing else. First, the validity of each circuit: an identity between reduced words in a free group, decided by normal form; the longest such identity in this paper is one of the 246 circuits of Theorem 4.3, whose right-hand side is a product of 23 of the marked classes and their inverses. Second, the combinatorial statement that every circuit-free index set has at most k elements, decided by enumerating all 2^N subsets of the index set, with $N = 6, 9, 15, 19, 18$ in the five cells. The covering checks (that every sphere or ball element is a marked class or its inverse) enumerate 4^2, 4^2+4+1, 6^2, 6^2+6+1 and 4^3 letter sequences respectively. The combinatorial checks are not vacuous: replacing k by $k - 1$ makes them false, as recorded for $(2, 2)$ in the formal development.

7. What is not proved, and what it would cost

Remark 7.1. Three items were priced and set aside. They are computations outside the formal development and are reported as measurements only.

- *The ball at* $(2, 3)$*, exactly.* Depth-first search over the 26 inverse classes of $B_3 \subset F_2$ with the membership oracle described above visits 2176 nodes in 1.5 s and returns $I_B(2, 3) = 6$, matching Theorem 4.3 and Corollary 4.4. Only the lower bound is proved here. The obstruction to a certified upper bound is the size of the index set: 26 classes and one further mark for the identity, so the enumeration of Lemma 2.4 runs over all $2^{27} \approx 1.3 \times 10^8$ subsets.

The module carrying the same enumeration at $N = 19$ (the ball at $(3, 2)$) elaborates in 188 s, and $N = 27$ is $2^8 = 256$ times as many index sets: an extrapolated cost of order ten core-hours, which we did not spend. This is a limitation of the checking scheme, not of the mathematics: replacing the enumeration of all subsets by a branch-and-bound that visits only circuit-free sets, with its soundness proved once by induction, would bring the cost down to a few thousand nodes.

- *The sphere at* $(2, 4)$. Here there are 54 inverse classes and $\rho(2, 4) = 9$. The plain depth-first search over inverse classes was stopped after 1800 s of a single core with no answer. Symmetry reduction and a colouring bound inside the search are the obvious next moves.
- *Question 5.3 in general.* Nothing here bears on it beyond the four cells.

8. Two observations on the source text

Neither observation below affects the correctness of any statement of [5]. We record them because both cost time when its results were recomputed, and both are stated outside the formal development.

Remark 8.1 (the generator ordering in $Q_r(k)$). The second family of maximal bases in [5] is built from a fixed generator x_1 and an automorphism φ of F_r required to be, verbatim, "an automorphism of F_r that cyclically permutes the set of generators and inverses of generators of F_r": the elements of $Q_r(k)$ are the words $x_1 w\, \varphi^{(m)}(x_1 w)^{-1}$ with $|w| = k-1$, w not beginning with x_1^{-1}, and $1 \le m < 2r$. That requirement determines the cyclic order up to relabelling, which is worth saying explicitly. Let L be the set of $2r$ letters and $\iota\colon L \to L$ the fixed-point-free involution $l \mapsto l^{-1}$. Any automorphism commutes with inversion, so the permutation induced by φ on L commutes with ι; if that permutation is a $2r$-cycle, its centraliser in $\mathrm{Sym}(L)$ is the cyclic group it generates, whose unique involution is its r-th power. Hence $\iota = \varphi^r$ on L, and listing L along the cycle as $l_0, \dots, l_{2r-1}$ gives $l_{i+r} = l_i^{-1}$: after renaming, the cyclic order is $x_1, \dots, x_r, x_1^{-1}, \dots, x_r^{-1}$. Such a φ exists, for instance $\varphi(x_i) = x_{i+1}$ for $i < r$ and $\varphi(x_r) = x_1^{-1}$. The interleaved order $x_1, x_1^{-1}, x_2, x_2^{-1}, \dots$ that one might reach for first does not arise from an automorphism, and the set it produces is not a basis. Take $r = 2$ and implement φ as the induced letter substitution, applied letter by letter and followed by free reduction. Of the six cyclic orders of the four letters, the two the paragraph above allows give the nine distinct length-four words of $Q_2(2)$, a basis of a rank-9 subgroup; each of the other four, the interleaved order among them, gives seven distinct words, of lengths two and four, generating a subgroup of rank 5. That computation, like the centraliser argument, is outside the formal development.

Remark 8.2 (which of the two bounds is smaller). [5] records two upper bounds for the size of an independent subset of B_n: the explicit $\beta(r, n)$ of its §4, quoted in (2), and the refinement $\rho(r, n) + \lfloor \log_2 v(r, n) \rfloor$ stated without proof in its closing commentary. The refinement is not uniformly the smaller of the two. Evaluating both from the displayed formulas: at $(2, 2)$ they are 5 and 4, and at $(3, 2)$ they are 7 and 6, so the explicit bound is strictly better in each case; at $(2, 3)$ they are 8 and 10, and the refinement is better. Both remain valid upper bounds, and the values computed in this paper are strictly below both wherever $n \ge 2$.

9. Formal verification

Every theorem, proposition, corollary and lemma stated above is formalised and machine-checked in Lean 4 [7] against Mathlib [8], with the explicit exception of Remarks 5.4, 5.5, 7.1, 8.1 and 8.2, which are elementary observations, summaries of [5], or measurements outside the development, and are labelled as such where they appear. The development is seven files and 2,556 lines, declaring 96 theorems and 179 definitions — the latter count dominated by the explicit permutations and word lists that the certificates consist of. There is no incomplete proof anywhere in it. The axiom set of each headline statement was printed. Theorem 3.1, the whole of Theorem 4.1 (both bounds, sphere and ball), the lower bound of Theorem 4.3, Corollary 4.4 and Proposition 5.2 rest on the standard triple of propositional extensionality, quotient soundness and choice, and on nothing else: every

finite check inside them is performed by the kernel itself. (What is formalised in Proposition 5.2 is the independence of $\{a^2, a^3\}$ and $\langle a^2, a^3\rangle = \langle a\rangle$; that $\langle a\rangle$ has rank one is read off outside Lean.) Theorems 4.2 and 4.3 additionally invoke compiled evaluation for their two per-cell finite checks — the validity of the circuit identities and the combinatorial bound on circuit-free index sets, the checks described in Section 6 — because the kernel-level enumeration of 2^{15}, 2^{19} and 2^{18} index sets is impractical; Theorem 5.1, which quotes them, inherits that. The development also contains negative controls: that 4 is not the size of any independent subset of S_2 or B_2 in F_2, that 2 is not the largest such size, that neither $\{a, a^{-1}\}$ nor $\{a, a^2\}$ is independent, that the identity lies in every ball, in no sphere of positive radius and in no independent set, and that the combinatorial bound at $(2, 2)$ — for the sphere and for the ball — is tight for the circuit list used, in that replacing 3 by 2 makes it false. The seven files, elaborated one at a time, take a little over seven minutes in total, the longest single file taking 188 s. The searches of Section 6 are separate programs whose correctness is used nowhere. The repository is private; repository reference to be added at submission.

References

[1] I. Kapovich, A. Myasnikov and V. Shpilrain, *Open problems in group theory*, section "Free groups", online problem list, `https://shpilrain.ccny.cuny.edu/gworld/problems/probfree.html`, retrieved 3 September 2026. Problem (F42), attributed there to V. Dotsenko, is the last entry of that section and reads: "Let S_n denote the set of all elements of length n in a free group F_r, and B_n the set of all elements of length $\leq n$. (a) What is the largest number of elements from S_n that freely generate a subgroup of F_r? (b) The same question for elements from B_n." The page carries no status marker and no solution note for (F42). It is the list of the authors of [4], but the wording above is quoted from the page and not from the printed volume, which was not consulted.

[2] M. Hall, Jr., *Coset representations in free groups*, Trans. Amer. Math. Soc. **67** (1949), 421–432. Cited in Section 2 only for the separation property that makes the finite-quotient method of Lemma 2.2 non-restrictive: this is Theorem 5.1, in the section titled "A separation theorem", read there in the publisher's scan on 3 September 2026. No proof in this paper depends on it.

[3] I. Kapovich and A. Myasnikov, *Stallings foldings and subgroups of free groups*, J. Algebra **248** (2002), no. 2, 608–668; DOI `10.1006/jabr.2001.9033` (published title, per Crossref and zbMATH; the preprint arXiv:math/0202285 is titled *… and the subgroup structure of free groups*). Cited for the folding algorithm reimplemented for the searches of Section 6; no theorem depends on it.

[4] I. Kapovich, A. Myasnikov and V. Shpilrain, *Combinatorial group theory: Open problems on infinite groups*, De Gruyter, 2026; DOI `10.1515/9783112208137`, ISBN 978-3-11-914679-1 (print), 978-3-11-220813-7 (electronic); online 22 January 2026, print 16 March 2026, per Crossref, 3 September 2026. The source of Dotsenko's Problem F42, and the reference [5] cites (its Lemma 8.2) for the Stallings rank formula rank $= E - V + 1$. The volume itself was not consulted here, and nothing is quoted from it.

[5] L. Koch-Hyde and É. Olive, *Two problems about subgroups of free groups concerning the lengths of their generators*, arXiv:2609.00382v1 (31 August 2026), 9 pages, 2 figures; primary classification math.GR, cross-listed math.CO. All quotations are from the source of version 1, read in full on 3 September 2026, the only version on the arXiv record then; no journal reference, no DOI, and no citing work on Semantic Scholar. Its Theorem `thm:max rank` is the maximum-rank theorem quoted in (1), its Theorem `thm:independent set limit` the O-bound, the unlabelled theorem of its §4 the explicit bound (2), and its Lemma `lem:odd vertex count` with Corollary `cor:even vertex count` the vertex bound (3). Its environments are auto-numbered, so we quote these internal labels rather than numbers.

[6] E. I. Khukhro and V. D. Mazurov (editors), *Unsolved Problems in Group Theory. The Kourovka Notebook*, No. 21, Novosibirsk, 2026; e-print arXiv:1401.0300. Problem 10.13 is due to R. I. Grigorchuk and dates from the 10th edition (1986), as cited in [5]; Problem 16.45 is due to P. J. Cameron. Both were read from the source of version 46 (1 September 2026), the current one on 3 September 2026; the e-print is updated in place, and neither problem carries a solution note there.

[7] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science, Springer, 2021, pp. 625–635; DOI `10.1007/978-3-030-79876-5_37`.

[8] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381; DOI `10.1145/3372885.3373824`.

[9] W. Magnus, A. Karrass and D. Solitar, *Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations*, Interscience Publishers, New York, 1966. Cited in [5] for the Nielsen-reduction criterion behind its two families of maximal bases; used here only in Remark 8.1.

[10] J. R. Stallings, *Topology of finite graphs*, Invent. Math. **71** (1983), no. 3, 551–565; DOI `10.1007/BF02095993`. Cited for the folding construction used by the searches of Section 6.