A new family of Cayley expanders (?)

Eyal Rozenman *
Hebrew University, Jerusalem, Israel
eyalroz@cs.huji.ac.il

Aner Shalev
Hebrew University, Jerusalem, Israel
shalev@math.huji.ac.il

Avi Wigderson †
Institute for Advanced Study, Princeton
avi@ias.edu

ABSTRACT

We assume that for some *fixed* large enough integer d, the symmetric group S_d can be generated as an expander using $d^{1/30}$ generators. Under this assumption, we explicitly construct an infinite family of groups G_n, and explicit sets of generators $Y_n \subset G_n$, such that all generating sets have bounded size (at most $d^{1/7}$), and the associated Cayley graphs are all expanders.

The groups G_n above are very simple, and completely different from previous known examples of expanding groups. Indeed, G_n is (essentially) all symmetries of the d-regular tree of depth n.

The proof is completely elementary, using only simple combinatorics and linear algebra. The recursive structure of the groups G_n (iterated wreath products of the alternating group A_d) allows for an inductive proof of expansion, using the group theoretic analogue [4] of the zig-zag graph product of [37]. The explicit construction of the generating sets Y_n uses an efficient algorithm for solving certain equations over these groups, which relies on the work of [32] on the commutator width of perfect groups.

We stress that our assumption above on *weak* expansion in the symmetric group is an open problem. We conjecture that it holds for all d. We discuss known results related to its likelihood in the paper.

*Part of this research was performed while visiting the Institute for Advanced Study, Princeton, NJ.
†Partially supported by NSF grants CCR-0324906.

STOC'04, June 13–15, 2004, Chicago, Illinois, USA.

1. INTRODUCTION

1.1 Expander Graphs

Expanders are graphs which are sparse but nevertheless highly connected. Expanders graphs have been used to solve many fundamental problems in computer science, on topics including network design (e.g. [35, 36, 1]), complexity theory ([44, 39, 43]), derandomization ([31, 16, 17]), coding theory ([40, 41]), and cryptography ([14]). Expander graphs have also found some applications in various areas of pure mathematics, such as topology, measure theory, game theory and group theory (e.g. [19, 23, 15, 24]).

Standard probabilistic arguments ([34]) show that almost every constant-degree (≥ 3) graph is an expander. However, most applications demand explicit constructions. Here we take the most stringent definition of explicitness of an infinite family of graphs, requiring that a deterministic polynomial time algorithm can compute the neighbors of any given vertex, from the vertex name and the index of the graph in the family. This challenge of explicit construction led to an exciting and extensive body of research.

Most of this work was guided by the algebraic characterization of expanders, developed in [42, 5, 2]. They showed the intimate relation of (appropriate quantitative versions of) the combinatorial (isoperimetric) notion of expansion above, to the spectral gap in the adjacency matrix (or, almost equivalently, the Laplacian) of the graph. This relationship is tight enough for almost all applications (but there are some exceptions, e.g. see [45, 9]).

Using this connection, an infinite family of regular graphs is defined to be an expander family if for all of them the second largest eigenvalue of the normalized adjacency (i.e. random walk) matrix is bounded above by the same constant that is smaller than 1.

This algebraic definition of expanders by eigenvalues naturally led researchers to consider algebraic constructions, where this eigenvalue can be estimated. The celebrated sequence of papers [27, 13, 5, 3, 18, 25, 28, 30] provided such highly explicit families of constant-degree expanders. All of these constructions are based on groups, and their analysis often appeals to deep results in mathematics.

The algebraic mould was broken recently by [37], where a simple, combinatorial construction of constant-degree expander graphs was presented. The construction is iterative, generating the next graph in the family from two previous ones via a novel graph product, the *zig-zag* product. This product was proved (using simple linear algebra) to simultaneously keep the degree small, and retain expansion. Thus the iteration process need only be provided with an initial, fixed size expander "seed" graph , from which all others are generated. The required parameters of the seed graph are easily shown to hold for a random graph (which would suffice for explicitness, as it is of constant size), but it can also be easily constructed

explicitly.

Our main result in this paper is a similar iterative construction of expanding *Cayley* graphs (which we turn to define next) from one initial "seed" Cayley graph. The major difference is that in our case, the existence of such a seed Cayley graph is still an open question.

Our construction may be seen as another step in exploring this fundamental notion of expansion, and its relations to yet unexplored mathematical structures. It also further explores the power of the zig-zag product in constructing even stronger expanders. It was already shown [9] that it can yield expansion beyond the eigenvalue bound, and is shown here to yield Cayley expanders.

1.2 Expanding Cayley graphs

For a finite group H and a (symmetric) set of elements T in it, the Cayley graph $C(H;T)$ has the elements of H as vertices, and edges connect a pair of vertices g, h if their "ratio" gh^{-1} is in T. We remark that while most applications do not require the expanders to be "Cayley", the recent paper [8] seems to essentially require Cayley expanders to achieve nearly linear-sized locally testable codes (LTCs) and probabilistically checkable proof (PCPs).

Many of the algebraic expander constructions mentioned above are Cayley graphs. In all of these, the groups in question are linear matrix groups over finite fields, and their expansion follows from celebrated results in mathematics, including Kazhdan's work on Property T [20], Selberg's 3/16 theorem [38], and the resolution of the Ramanujan conjecture of Eichler, Deligne and Iguza (starting in [12]). It should be noted that for some of the other algebraic constructions elementary proof of expansion exist, using only a discrete Fourier transform [18].

For other natural families of groups the question was considered both by mathematicians and computer scientists. For example, for Abelian groups it is easy to see that any set of expanding generators has to be at least logarithmic in the size of the group. Thus they cannot provide expanding Cayley graphs of constant degree (a more general result appears in [21]). Lubotzky and Weiss generalized this negative result for all solvable groups of bounded derived length [26].

An interesting open problem, highly relevant to this work, is whether the symmetric group (of all permutations) has a constant number of expanding generators. Much work has been devoted to analyzing the expansion of this group under a variety of generating sets in the context of card shuffling (e.g. see [10, 22]). However in all these papers the generating sets are huge, and do not provide a clue to the status of this problem. The best upper bound known, which applies to *every* finite group, is logarithmic in the group size [6]. For the symmetric group on d letters S_d this bound gives $O(d \log d)$ expanding generators. We conjecture that $d^{1/30}$ generators suffice, and this conjecture will provide the seed Cayley graph to our iterative construction.

The possibility that the zig-zag product and iterative construction may be used for Cayley expanders was first revealed in [4]. They discovered (roughly speaking) that the well known *semi-direct* product on groups may be viewed as a special case of the zig-zag product of graphs. More precisely, the zig-zag product of two Cayley graphs, with certain important restrictions on the structure of their generating sets, is a Cayley graph of the semi-direct product of the associated groups. Thus one can generate larger Cayley expanders of small degree from smaller ones. This observation was used to show that expansion is *not* a group property – in some groups certain constant size sets will expand, while others will not.

However, unlike the case of unstructured graphs, the restrictions on generators alluded to above for applying the zig-zag product on Cayley graphs, make iterations a highly nontrivial (and illuminating) task. In [29] such a construction was given, which falls short of the task at hand on two counts. First, the generating sets (and hence the degrees) of the groups in the family are not of constant size, but rather grow slowly (roughly like $\log^*$ of the group size). Second, these generating sets are shown to exist via a probabilistic argument, hence the resulting family is not explicit. Still, this construction makes no assumptions, as the seed Cayley expander for the iteration is easily seen to exist.

In this paper we fix both problems. Assuming we have the seed Cayley graph from the conjecture above for some fixed (large) d, we give a sequence of groups G_n, and explicit generating sets Y_n for each G_n, such that the Cayley graphs $C(G_n, Y_n)$ are expanding. Moreover, $|Y_n| \leq k = d^{1/7}$ for all n.

1.3 Our construction

Our groups are completely different from groups previously used in this area. Indeed, they are very natural combinatorial objects. Let $T(d, n)$ denote the d-regular tree of depth n. The group of symmetries of this tree allows permuting the children of every internal node arbitrarily. Thus every element of this group may be described by a mapping of the internal nodes to the symmetric group S_d, describing how to permute the children of every such node. Group product of two such elements is simply performing the first set of permutations at every node, and then the next set. Our groups G_n are subgroups of all symmetries, allowing only *even* permutations at every internal node of $T(d, n)$. This natural restriction avoids a huge Abelian quotient that would have rendered expansion (with a constant number of generators) impossible.

There is a very natural inductive definition of the groups G_n. G_1 is the alternating group A_d of all even permutations on d elements (and is essentially the "seed group" of our construction). G_{n+1} can be obtained from d copies of G_n, and one copy of A_d acting on them simply by permuting the copies. Formally, this is called a wreath product, denoted $G_{n+1} = G_n \wr A_d$, and is a special case of a semidirect product, giving equivalently $G_{n+1} = (G_n)^d \rtimes A_d$. Our assumption gives a small expanding set of generators for A_d, and by induction we have such a set for G_n.

How does induction proceed? Naturally, we'd like to use the zig-zag theorem for the semi-direct product [4, 37]. The technical requirement alluded to above is simply that we find an expanding generating set for $(G_n)^d$, which need not be small, but must be an orbit under the action of A_d. A natural candidate for such an orbit is all (even) permutations of the *balanced* d-vector (one which has every one of the k elements of Y_n occurring the same number of times). It is the largest possible orbit, and the projection of a random element of the orbit to any small subset of the coordinates is (almost) a random independent element of Y_n in each coordinate.

We now turn to study the second eigenvalue of the Cayley graph of $(G_n)^d$ under these generators. The associated linear operator acts on the space of real functions on $(G_n)^d$. Luckily, this space of functions is simple to describe - it is the d-fold tensor product of the same space for G_n. What is not so lucky is the dependence between the coordinates of a balanced vector. Indeed, had G_n been Abelian, this orbit would not even be generating (i.e. the graph would not be connected). We have to use our special group structure here. A key fact (proved by Nikolov [32]) is that every element in G_n is a commutator. We construct a new generating set $\tilde{Y}_n$ by adding to Y_n, for each of its elements, the constituents of its representation as a commutator. We use Nikolov's proof to actually give a polynomial time algorithm for finding this representation. We now take the orbit of all balanced vectors over $\tilde{Y}_n$ to be our actual generating

set for $(G_n)^d$.

How can this revision take care of the dependencies? A simpler setting, to which we reduce our analysis, is the following Cayley graph. The group is simply $(G_n)^2$, namely only two copies of G_n. The generators are all pairs (g, g^{-1}) for all $g \in \tilde{Y}_n$. Thus, there is *complete* correlation between the two coordinates. The key point is that, using the special structure of $\tilde{Y}_n$, with positive probability a short word in one of the two components will vanish, while in the second it will give an original generator of Y_n, thereby decoupling the dependence of the two components. So, quite surprisingly, this Cayley graph on two copies is expanding despite the complete correlation (it is a nontrivial exercise to even establish connectivity of this graph – note that it would *not* be connected had G_n been Abelian, or if we took instead the pairs (g, g) for any group G_n). This construction (which we feel is of independent interest) is quite special and mysterious, and naturally the description above hides many essential details. Still, it is the heart of the matter.

1.4 On our assumption

How realistic is our conjecture that the alternating group[1] A_d (with d large enough) has $d^{1/30}$ expanding generators? As mentioned, the best upper bound is $O(d \log d)$, which is logarithmic in the group size – a result that holds for *every* group. What makes A_d special, and indeed leads people to speculate that it even has a constant number of expanding generators independent of d, are the following two results that seem to be in the "right direction". The first is a theorem of Dixon [11], that with probability $1 - o(1)$, two random permutations generate A_d. The second is a result of Babai et al [7] that A_d has seven generators which yield a Cayley graph of logarithmic diameter. Incidentally, both results hold for every non-Abelian finite simple group [**?**].

In another related work, Lubotzky and Pak [24] show that if the automorphism group of the free group F_k on k generators has Kazhdan property (T) then for infinitely many d the group A_d has an expanding generating set of size $O(k^2)$, independent of d.

1.5 Organization of the paper

[Eyal's Note: Fix this section]

In section 2 we define expander graphs and Cayley graphs, and show some useful results. In section 3 we define the sequence of groups we use. In section 4 we describe the expanding generating sets, and prove the main theorem 4 - that they are indeed expanding - by induction. The proof is based on a main lemma (theorem 5). The lemma gives expanding an generating set for the group G^d given an expanding generating set for G (under certain conditions on G). Finally, In section 6 we present an algorithmic version of Nikolov's theorem, that every element in out family of groups has a commutator representation that can be found efficiently.

2. PRELIMINARIES

2.1 Graphs, eigenvalues and adjacency matrices

All graphs discussed in this paper are undirected, regular graphs. We allow multiple edges and self loops, so graphs are best understood as symmetric nonnegative integer matrices with a fixed row-sum, called the *degree*. For a graph X, we let $V(X)$ denote its set of vertices and $E(X)$ its (multiset of) edges.

Let X be a k-regular graph, and $M = M_X$ its normalized adjacency matrix (divide the adjacency matrix by the degree k to make it stochastic). We denote by $\lambda(X)$ the second largest (in absolute value) eigenvalue of M. The *spectral gap* of the graph is $1-\lambda(X)$.

Let W be the vector space of real functions on the set $V(X)$, with its standard L_2 inner product. M_X defines a linear operator on W: For $f \in W$, the value of the function $M_X(f) \in W$ on a vertex x is the average value of f on all the neighbors of x (counted with multiplicities).

Let $W_{||}$ be the one-dimensional subspace consisting of the constant functions, and let $W_\perp$ be the orthogonal complement. Since the constant functions are eigenvectors of M corresponding to the (largest) eigenvalue 1, then

$$\lambda(X) = \max_{w \in W_\perp, \|w\|=1} \|Mw\|$$

where $\|w\|$ is the L_2 norm of w.

DEFINITION 2.1. *An infinite family of graphs X_n is called an* **expander family** *if $\lambda(X_n) \leq \mu$ for some $\mu < 1$ independent of n. The family is said to be* **explicitly described***, if there is a polynomial time algorithm which, on input n and the name of a vertex v in G_n (in binary), outputs the neighbors of v in G_n.*

We will use the following two simple results, which describe how taking the tensor power of a graph, and taking the power of a graph, affect the 2nd eigenvalue λ:

CLAIM 2.2. *Let $X = (V, E)$ be a graph, and let M_X be the normalized adjacency matrix. Let $M_Y = (M_X)^{\otimes d}$, and define Y to be the graph (on the vertex set V^d) with normalized adjacency matrix M_Y. Then $\lambda(Y) = \lambda(X)$.*

OBSERVATION 2.3. *Let $X = (V, E)$ be a graph, M_X the normalized adjacency matrix and $M_Y = (M_X)^t$. Let Y be the graph (on vertex set V) with normalized adjacency matrix M_Y. Then $\lambda(Y) = \lambda(X)^t$.*

We will use the following convexity result later: If the spectral gap $(1 - \lambda(Y))$ of a graph Y is not too small, and Y is a large subgraph of X (on the same vertex set) then the spectral gap of X is also not too small.

CLAIM 2.4. *Let $Y = (V, E_1) \subset X = (V, E_2)$ (i.e. $E_1 \subset E_2$) be s and t regular graphs respectively on the same vertex set V. Then*

$$1 - \lambda(X) \geq \frac{s}{t}(1 - \lambda(Y))$$

2.2 Groups and the wreath product

2.2.1 Cayley graphs

Let G be a finite group. We will represent groups multiplicatively, and 1 will denote the identity of the group. Let Y be a multi-subset of G. We will always use *symmetric* sets Y, namely the number of occurrences of x and x^{-1} in Y is the same for every $x \in G$. $|Y|$ will denote the size of the multiset (counting multiplicities).

The *Cayley graph* $\mathrm{C}(G, Y)$ has vertex set G, and for every vertex $g \in G$ and $x \in Y$ there is an edge (g, gx). The graph $\mathrm{C}(G, Y)$ is undirected (as Y is symmetric) and is $|Y|$-regular. For $x \in G$ let P_x be the permutation matrix corresponding to $g \to gx$ in G. The normalized adjacency matrix of $\mathrm{C}(G, Y)$ is $\sum_{x \in Y} P_x/|Y|$. We will also use the notation $\mathbb{E}_{x \in Y}[P_x]$ to denote this average of operators.

Let $W = W(G)$ be the vector space of functions $G \to \mathbb{R}$ as in the previous section. We will be interested in the expansion properties of Cayley graphs on the group G^d, the Cartesian product of d copies of G. Note that $W(G^d) = W^{\otimes d}$.

[1] Everything in this discussion holds equivalently for the symmetric group S_d up to constants.

OBSERVATION 2.5. *Let $W_{||}, W_{\perp}$ be the constant functions on G and the orthogonal complement as before. Let $\bar{b} = (b_1, \ldots, b_d)$ be a length-d vector where each b_i is in $\{||, \perp\}$, and let $W_{\bar{b}}$ be the vector space $\otimes_{i=1}^{d} W_{b_i}$. The orthogonal decomposition $W = W_{||} + W_{\perp}$ induces an orthogonal decomposition*

$$W^{\otimes d} = \sum_{\bar{b} \in \{||, \perp\}^d} W_{\bar{b}}$$

to 2^d subspaces, by using the distributive law for tensor products. For any $g \in G^d$ the operator P_g preserves the decomposition.

COROLLARY 2.6. *Consider the Cayley graph $C(G^d, Y)$. The normalized adjacency operator $\mathbb{E}_{x \in Y}[P_x]$ preserves the above decomposition, so*

$$\lambda(G^d, Y) = \max_{\bar{b} \neq ||^d} \max_{w \in W_{\bar{b}}} \|\mathbb{E}_{x \in Y}[P_x(w)]\| / \|w\|$$

That is, it suffices to upper bound $\|\mathbb{E}_{x \in Y}[P_x(w)]\|$ for vectors w that are purely in one of these $2^d - 1$ subspaces.

Observation 2.3 translates nicely to the Cayley graph world

OBSERVATION 2.7. *Let G be a group, $Y \subset G$. Define Z to be the set of all words of length k in Y. Then $\lambda(G, Z) = \lambda(G, Y)^k$.*

We end with an observation which simplifies the proof of explicitness for families of Cayley graphs.

OBSERVATION 2.8. *A family of Cayley graphs $C(G_n, Y_n)$ is explicit if there are polynomial time algorithms in $\log |G_n|$ for*

- *performing group multiplication in G_n*
- *computing the set Y_n*

2.2.2 Wreath products and the zigzag product

Let A and B be finite groups. Assume that $B \subset S_d$, that is, it acts by permutations on the set $\{1, \ldots, d\}$. Define the *wreath product* $A \wr B$ of A and B to be the group whose elements are vectors $(a_1, \ldots, a_d, \sigma)$, where $a_i \in A$ for all i, and $\sigma \in B$. The group multiplication rule is

$$(a_1, \ldots, a_d, \sigma) \cdot (\tilde{a}_1, \ldots, \tilde{a}_d, \tau) = (a_{\tau(1)}\tilde{a}_1, \ldots, a_{\tau(d)}\tilde{a}_d, \sigma\tau)$$

One can check that this defines a group structure on $A \wr B$. The wreath product is a special case of a more general construction - the *semi-direct product* of A^d and B, where A^d is the Cartesian product of d copies of A. The groups A^d, B are naturally embedded in $A \wr B$, and we will sometimes refer to elements of A^d and B as elements of $A \wr B$.

Let $\alpha \subset A^d, \beta \subset B$ be sets of generators. Suppose α has a special structure: it is a B-*orbit*. This means that for some arbitrary $\bar{a} \in \alpha$, the set α consists of all vectors obtained from $\bar{a}$ by permuting its coordinates by some permutation $\sigma \in B$. We now define a set γ in $A \wr B$ by $\gamma = \{x\bar{a}y | x, y \in \beta\}$. One can check that γ generates $A \wr B$. The following theorem from [4], following the zigzag theorem of [37], shows that there if α, β are sufficiently good expanding generators then so is γ.

THEOREM 1. [4] *If α is a B-orbit then $\lambda(A \wr B, \gamma) \leq \lambda(A, \alpha) + \lambda(B, \beta)$.*

Note that $|\gamma| = |\beta|^2$ depends only on the size of β, while α could be large (it could be as large as $|B|$). Also, it is easy to compute γ given α and β, as multiplications in $A \wr B$ can be computed efficiently.

2.2.3 The commutator property

Let A be a group. For $g, h \in A$ define the *commutator* $[g, h]$ to be $ghg^{-1}h^{-1}$. A has the *commutator property* if for every element of $a \in A$ there is a solution in the variables x, y to the equation $a = [x, y]$ [2]. Nikolov [32] proves

THEOREM 2. [32] *Let A be a group, and $B \subset S_d$ a group of permutations. If A, B have the commutator property then so does $A \wr B$.*

We shall need an algorithmic version of this theorem. For a group A, a *commutator representation algorithm* gives, for an input $a \in A$, some pair $x, y \in A$ such that $a = [x, y]$.

THEOREM 3. *Let A, B be as in theorem 2. Suppose we are given commutator representaion algorithms for the groups A, B. Then we obtain such an algorithm for $A \wr B$. This algorithm calls the algorithm on B one time, and the algorithm on A at most d times, and uses at most $O(d)$ extra multiplication operations on A, B.*

Since multiplication operations in G_n take time polynomial in $\log |G_n|$ we deduce that

COROLLARY 2.9. *The group G_n has a commutator representation algorithm that runs in time polynomial in $\log |G_n|$.*

We prove the theorem and corollary in section 6.

3. OVERVIEW OF THE CONSTRUCTION

In section 3.1 we will define our sequence of groups G_n. In section 4 we will show how to find generating subsets $Y_n \subset G_n$ that give $\lambda(G_n, Y_n) < 1/1000$ with bounded size $|Y_1|$[4]. This will be based on the assumption that there exists a small enough $Y_1 \subset A_d$ in the alternating group such that $\lambda(A_d, Y_1) < 1/1000$.

3.1 The family of groups

DEFINITION 3.1. *The groups in our construction are defined by $G_1 = A_d$ and, inductively, $G_{n+1} = G_n \wr A_d$.*

Another way to view the group G_n is as a subgroup of the full group of symmetries of the d-regular, depth n tree. Each element in the group of symmetries is uniqely defined by writing a permutation on each internal node of the tree, indicating how the children of this vertex are permuted. In the subgroup G_n all these permutations should be *even*. The representaion of an element of G_n as a list of even permutations is polynomial in $\log |G_n|$. Multiplying two elements and inverting an element can be done in time which is polynomial in the size of this representation

The following important corollary of theorem 3 shows that for our groups G_n there is an efficient algorithm to solve, for any $g \in G_n$, the equation $g = [x, y]$ in the variables x, y.

LEMMA 3.2. *If $d \geq 5$ then the groups G_n have the commutator property of section 2.2.3. Moreover, the commutator representation of an element can be found in time polynomial in $\log |G_n|$.*

PROOF. $G_1 = A_d$, and by [33] it has the commutator property. The result follows by induction using theorem 3. □

[2]Note that this is a stronger property than just the commutator subgroup $[A, A]$ being equal to A.

4. MAIN THEOREM

THEOREM 4. *Suppose that for some d there exists a set of generators $Y_1 \subset A_d$ such that $\lambda(A_d, Y_1) < 1/1000$ and $|Y_1| \leq d^{1/28}/10^{40}$. Then there exist sets $Y_n \subset G_n$ such that $\lambda(G_n, Y_n) < 1/1000$ and $|Y_n| \leq d^{1/7}/10^{40}$. Furthermore, Y_n can be computed in time polynomial in $\log |G_n|$.*

The graphs $C(G_n, Y_n)$ are the required sequence of Cayley graphs. The sets Y_n can be computed efficiently, and we saw in section 3.1 that group operations in G_n can also be computed efficiently, so by observation 2.8 this is an explicit family of Cayley graphs.

We will construct the expanding generators $Y_n \subset G_n$ inductively. The basis of the induction is the (unproved) assumption in the theorem about $G_1 = A_d$.

Let $G = G_n$. We are given $Y \subset G$ such that $\lambda(G, Y) < 1/1000$ and $|Y| \leq d^{1/7}/10^{40}$. We want to find a set $Y' \subset G \wr A_d$ such that $\lambda(G \wr A_d, Y') < 1/1000$ and $|Y'| \leq d^{1/7}/10^{40}$. We will use theorem 1. The theorem requires an expanding generating set for A_d (which we already have), and an expanding generating set $T \subset G^d$ which is exactly one A_d-orbit [3]. Given any element of such T, theorem 1 produces (explicitly) an expanding generating set for $G \wr A_d = G_{n+1}$.

Can we find an expanding, one-orbit generating set for G^d? Here is a simple attempt that fails. Take $T = Y^d$. The set Y^d is expanding, as $\lambda(G^d, Y^d) = \lambda(G, Y)$ by claim 2.2. Unfortunately, Y^d is far from being one orbit - it contains many vectors that are (pairwise) not equal up to permutation. Another natural set to consider in G^d is the set of *balanced vectors*:

DEFINITION 4.1. *Let G be a group, and $Y \subset G$. For $d > |Y|$, define $Y^{(d)}$ to be the vectors in Y^d in which every $u \in Y$ appears exactly $\lfloor d/|Y| \rfloor$ times, and the rest of the elements are $1 \in G$. We call these vectors* **balanced vectors**. *Every two elements in the set $Y^{(d)}$ are equal up to a permutation of the coordinates. Since $d > |Y|$ we may assume that the permutation is even. In other words, the set $Y^{(d)}$ is one A_d-orbit.*

The set $Y^{(d)}$ looks promising, but is it expanding? Not always. If G is Abelian $Y^{(d)}$ does not even *generate* G^d, since every element in $Y^{(d)}$ has product of coordinates equal to 1 (Y is symmetric, and every element of Y appears the same number of times in $Y^{(d)}$). The groups G_n are far from being Abelian. Indeed, every element of G_n has a representation as a commutator. It turns out that this property, along with the existence of a small generating set Y for G (assumed by induction) enables us to find a good generating set for G^d. We will enlarge Y somewhat to a set $X \supset Y$, and see that $X^{(d)}$ is expanding for G^d.

DEFINITION 4.2. *Let G be a group, and let $Y \subset G$. Suppose every element $y \in Y$ can be written as a commutator in G, namely $y = a_y b_y a_y^{-1} b_y^{-1}$ for some $a_y, b_y \in G$. Define*

$$Y^* = \bigcup_{y \in Y} \{a_y, b_y, a_y^{-1}, b_y^{-1}, a_y^{-1} b_y^{-1}, b_y a_y\} \cup \{1\}$$

Y^* is symmetric, and $|Y^*| \leq 7|Y|$.

THEOREM 5. *Let G be a group. Suppose that every element of Y is a commutator in G. Let $c, k \in \mathbb{N}$ be constants (to be chosen later). Define $c \cdot Y \subset G$ to be the multi-subset where every element of Y appears c times. Define $X := (c \cdot Y) \cup Y^*$, and $\lambda = \lambda(G, Y)$. If $d \geq k^2 \cdot |X|^7$ then*

$$\lambda(G^d, X^{(d)}) < 0.01 + \max\left\{(\lambda + 7/c), e^{-k(1-\lambda)c/10^6}\right\}$$

where $X^{(d)}$ is the set of balanced vectors.

The proof is given in section 5. To get a feeling for the constants, note that the larger k, c are, the better inequality we get in the theorem. k is large when X is small. c is large when X is much larger that Y, so k gets smaller when c gets larger. Nevertheless, it is not difficult to make both of them large enough for our purposes.

Theorem 5 is the required result for the inductive step - it remains to show that we can choose c, k properly such that $\lambda(G^d, X^{(d)}$ is small enough for theorem 1.

We proceed with the induction step. We are given a set $Y_n \subset G_n$ of size at most $|Y_1|^4$ such that $\lambda(G_n, Y_n) < 1/1000$. Apply theorem 5 (with $c = 10^3, k = 10^5$). Then the conditions of theorem 5 hold, and we obtain a set $X^{(d)} \subset G^d$ such that $\lambda(G, X^{(d)}) < 1/50$ (just substitute our k, c in the theorem to see this). Apply theorem 1 to obtain a subset $P \subset G_{n+1}$ of size $|Y_1|^2$, and $\lambda(G_{n+1}, P) < 1/1000 + 1/50$. Define Y_{n+1} to be the set of all words of length 2 in P. This is a set of size $|Y_1|^4$ and (by observation 2.7) $\lambda(G_{n+1}, Y_{n+1}) < (1/1000 + 1/50)^2 < 1/1000$. This completes the induction step.

5. PROOF OF THEOREM 5

The theorem appears in section 4. Let G, Y, X, λ be as defined in theorem 5. We will use the notation $W = W(G)$ and $W(G^d), W_{\bar{b}}$ defined in section 2.2.1. We need to prove that for every $w \in W(G^d)_\perp$ such that $\|w\| = 1$, at least one of the following upper bounds holds

$$\|\mathbb{E}_{x \in X^{(d)}}(P_x w)\| \leq 0.01 + \lambda + \frac{7}{c} \quad (1)$$

$$\|\mathbb{E}_{x \in X^{(d)}}(P_x w)\| \leq 0.01 + e^{-kc(1-\lambda)/10^6} \quad (2)$$

We saw in section 2.2.1 that it is enough to prove this for $w \in W_{\bar{b}}$ when $\bar{b} \neq \{||\}^d$. Since $X^{(d)}$ is invariant under permutation of the coordinates it is enough to prove the inequality for every $w \in W_\perp^{\otimes r} \otimes W_{||}^{\otimes(d-r)}$ where $1 \leq r \leq d$ (this is $W_{\bar{b}}$ for $b_i = \perp$ for $1 \leq i \leq r$ and $b_i = ||$ for $r < i \leq d$).

We split the proof to small and large r cases. For small r we will prove inequality (1), and for large r we will prove inequality (2).

Small r case: When $r \leq 0.1\sqrt{d/|X|}$, the first r coordinates of a random element in $X^{(d)}$ are very closely a random element in X^r. As $P_x(w)$ only depends on the first r coordinates of x, it is enough to bound $\|\mathbb{E}_{x \in X^r}(P_x w)\|$ for $w \in W_\perp^{\otimes r}$. By claim 2.2 $\|\mathbb{E}_{x \in X^r}(P_x w)\| \leq \lambda(G, X)^r$. The worst case is when $r = 1$. As $Y \subset X$ we can use claim 2.4 to give an upper bound to $\lambda(G, X)$, and we obtain inequality (1). This part is relatively easy, and we will not give a more detailed proof. Notice however that the argument for small r works for *any* group G, not only for our special sequence of groups, and from the generating set X we only used the Y part - not the Y^* part.

Large r case: When r is large the result is no longer true for any group [4] - we will need the Y^* part of the generating set X (recall that it is only defined when every element of G is a commutator). We will start with the analysis of a different graph - the Cayley graph $C(G \times G, \{(y, y^{-1}) | y \in Y^*\})$. We give a lower bound

[3] Recall that this means that every two elements of T should be equal up to an even permutation of the coordinates. From now on we shall write "one orbit", omitting the A_d

[4] For any abelian group there exists an $f \in W^{\otimes d}$ such that $P_y(f) = f$ for all $y \in Y^{(d)}$

of $(1-\lambda(G,Y))/21|Y^*|^2$ on the spectral gap of this graph in section 5.1. Afterwards, in section 5.2, we will show a reduction giving an upper bound on $\|\mathbb{E}_{x\in X^{(d)}}[P_x(w)]\|$ using this graph on $G\times G$. The reduction is again true for every group G, not only our groups.

Notice that the spectral gap bound we get in the $G\times G$ case is rather weak - much smaller than the spectral gap of the original graph $C(G,Y)$. When r is large enough we are able to apply the $G\times G$ result many times in parallel, amplifying the weaker upper bound in $G\times G$. We will obtain the upper bound (2).

5.1 Expansion of $G\times G$ with correlated generators

DEFINITION 5.1. *Let G be a group. and let $Y\subset G$ be a subset of G. Define*

$$\widetilde{Y}=\{(y,y^{-1})|y\in Y\}$$

THEOREM 6. *Suppose $\lambda(G,Y)<1-\varepsilon$ for some ε, and that every element of Y is a commutator in G. Then*

$$\lambda(G\times G,\widetilde{Y^*})\leq 1-\frac{\varepsilon}{21|Y^*|^2}$$

We find theorem 6 to be quite surprising. In the set $\widetilde{Y^*}$ there is *complete correlation* between the two coordinates, and it would seem that this correlation would prevent the graph from being an expander. For example, if G is Abelian and Y generates G then $\widetilde{Y}$ does not even generate $G\times G$, but only the subgroup $\{(g,g^{-1})|g\in G\}$. Also, for any group G the set $\{(y,y)|y\in Y\}$ only generates the subgroup $\{(g,g)|g\in G\}$. In both cases the correlation in the generating set prevents the graph from being an expander. We manage to decouple this correlation in the case of the special generating set Y^*, whose existence relies on the commutator property of G - that every element $g\in G$ can be represented as $g=xyx^{-1}y^{-1}$.

PROOF. The key observation is that we can represent the element $(y,1)$ for any $y\in Y$ as a word of length 3 in $\widetilde{Y^*}$. We prove this in the following observation.

OBSERVATION 5.2. *Let Z be the set of words of length 3 in the set $\widetilde{Y^*}$. then*

$$C(G\times G,\{(Y,1)\cup(1,Y)\})\subset C(G\times G,Z)$$

PROOF. Recall that for every $y\in Y$ the set Y^* contains the elements $a_y,b_y,a_y^{-1}b_y^{-1}$ where $y=a_yb_ya_y^{-1}b_y^{-1}$. Observe that

$$(a_y,a_y^{-1})\cdot(b_y,b_y^{-1})\cdot((a_y^{-1}b_y^{-1}),(a_y^{-1}b_y^{-1})^{-1})=(y,1)$$

This gives the required representation of $(y,1)$. We can obtain $(1,y)$ similarly. □

It is easy to see that if $C(G,Y)$ has spectral gap ε then the graph $C(G\times G,\{(Y,1)\cup(1,Y)\})$ has spectral gap $\varepsilon/2$. We now have the decoupling we were looking for - the correlated generating set Z contains the uncorrelated one $(Y,1)\cup(1,Y)$. More precisely, apply claim 2.4 to observation 5.2, and deduce that

OBSERVATION 5.3. *$C(G\times G,Z)$ has spectral gap at least $\varepsilon/7|Y^*|^2$*

Recall that Z consists of all words of length 3 in the $\widetilde{Y^*}$. By observation 2.7, the spectral gap of $C(G\times G,\widetilde{Y^*})$ is at most 3 times smaller than the spectral gap of $C(G\times G,Z)$, and the theorem is proved.

5.2 Reduction to $G\times G$

We upper bound the average $\|[\mathbb{E}_{x\in X^{(d)}}(P_xw)]\|$ in terms of $\lambda(G\times G,\widetilde{Y^*})$ from section 5.1.

For $x\in X^d$ write $x=(x_1,x_2,\bar{x})$ where $x_1,x_1\in G$ and $\bar{x}\in G^{d-2}$. By the triangle inequality

CLAIM 5.4. *For every $w\in W^{\otimes d}$*

$$\|\mathbb{E}_{x\in X^{(d)}}P_x(w)\|\leq\mathbb{E}_{x\in X^{(d)}}(\|(P_{x_1,x_2,\bar{x}}+P_{x_2,x_1,\bar{x}})(w)/2\|)$$

Since the value of $\|(P_{x_1,x_2,\bar{x}}+P_{x_2,x_1,\bar{x}})(w)/2\|$ only depends on the first two coordinates of x we group together all x with equal x_1,x_2, replacing $\bar{x}$ by 1^{d-2}, a $(d-2)$-length vector of 1's. It is therfore enough to bound

$$\mathbb{E}_{x\in X^{(d)}}(\|(P_{x_1,x_2,1^{(d-2)}}+P_{x_2,x_1,1^{(d-2)}})(w)/2\|)$$

The number of times each pair x_1,x_2 appears in the average above is proportional to the number of extensions of x_1,x_2 to a vector $(x_1,x_2,\bar{x})\in X^{(d)}$. As d is much larger than 2, the number of such extensions is nearly equal for every pair x_1,x_2, and we get (the 0.01 below pays for the fact that the number of extensions is only nearly equal)

CLAIM 5.5. *If $d\geq 100|X|$ then for every $w\in W^{\otimes d}$*

$$\begin{aligned}&\mathbb{E}_{x\in X^{(d)}}(\|(P_{x_1,x_2,\bar{x}}+P_{x_2,x_1,\bar{x}})(w)/2\|)\\ \leq\ &\mathbb{E}_{y\in X^2}\|(P_{y_1,y_2,1^{(d-2)}}+P_{y_2,y_1,1^{(d-2)}})(w)/2\|+0.01\end{aligned}$$

The following lemma bounds the RHS of claim 5.5

LEMMA 5.6. *If $\lambda(G,Y)<1-\varepsilon$ and $r\geq 2$ then for every $w\in W_\perp^r\otimes W^{\otimes(d-r)}$*

$$\begin{aligned}&\mathbb{E}_{y\in X^2}\|(P_{y_1,y_2,1^{(d-2)}}+P_{y_2,y_1,1^{(d-2)}})(w)/2\|\\ \leq\ &(1-\frac{c\varepsilon}{2\cdot 10^4|X|^3})\|w\|:=\Delta\|w\|\end{aligned}$$

We prove the lemma in section 5.2.1

Lemma 5.6 gives some upper bound $\Delta\|w\|$ on $\|\mathbb{E}_{x\in X^{(d)}}P_x(w)\|$, but Δ is too close to 1. The problem originates from claim 5.4, where we partitioned the set $X^{(d)}$ into pairs based on the value of the first 2 coordinates, and then considered the norm of each pair. This partition turns out to be too coarse. We will use a finer partition of $X^{(d)}$ by looking at the first t pairs of coordinates, for some properly chosen $t\leq r$. This will improve the spectral gap to Δ^t.

We now define this finer partition precisely. Let $H_t<S_d$ be the subgroup (of size 2^t) generated by the transpositions $(2k-1,2k)$ for $1\leq k\leq t$, and group together the elements $\{\sigma(x)|\sigma\in H_t\}$. When $t=1$ we get the grouping into pairs discussed above. The argument leading to claim 5.5 shows

CLAIM 5.7. *If $2t\leq 0.1\sqrt{d/|X|}$ then for every $w\in W^{\otimes d}$*

$$\|\mathbb{E}_{x\in X^{(d)}}P_x(w)\|\leq\mathbb{E}_{y\in X^{2t}}\|\mathbb{E}_{\sigma\in H_t}(P_{\sigma(y,1^{(d-2t)})}(w))\|+0.01$$

The case $t=1$ is claim 5.5. However, the weak upper bound Δ we had for $t=1$ amplifies to Δ^t.

CLAIM 5.8. *Suppose that for every $w\in W_\perp^{\otimes 2}\otimes W^{\otimes d-2}$*

$$\mathbb{E}_{y\in X^2}\|\frac{1}{2}(P_{y_1,y_2,1^{(d-2)}}+P_{y_2,y_1,1^{(d-2)}})(w)\|\leq\Delta\|w\|$$

Then for every $w\in W_\perp^{\otimes 2t}\otimes W^{\otimes d-2t}$

$$\mathbb{E}_{y\in X^{2t}}\|\mathbb{E}_{\sigma\in H_t}P_{\sigma(y,1^{(d-2t)})}(w)\|\leq\Delta^t\|w\|$$

PROOF. The proof is by induction on t. The case $t = 1$ is the assumption of the claim. For general t

$$\begin{aligned} & \mathbb{E}_{y\in X^{2t}} \|\mathbb{E}_{\sigma\in H_t} P_{\sigma(y,1^{(d-2t)})}(w)\| \\ = \; & \mathbb{E}_{z\in X^2, y\in X^{2(t-1)}} \\ \| \; & \mathbb{E}_{\sigma\in H_{t-1}} P_{\sigma(1^2,y,1^{(d-2t)})}[(P_{z_1,z_2,1^{(d-2)}} + P_{z_2,z_1,1^{(d-2)}})(w)]\| \\ \leq \; & \Delta^{t-1}\mathbb{E}_{z\in X^2}\|(P_{z_1,z_2,1^{(d-2)}} + P_{z_2,z_1,1^{(d-2)}})(w)\| \leq \Delta^t\|w\| \end{aligned}$$

Note that in the second line above $\sigma \in H_{t-1}$ acts on the vector y - not on the first $2t-2$ coordinates. The first inequality follows from the induction hypothesis for H_{t-1} The second inequality follows from the induction hypothesis for H_1 □

We can now complete the proof using $\lambda(G, Y) < 1 - \varepsilon$. Pick an integer t satisfying $0.05\sqrt{d/|X|} \leq 2t \leq 0.1\sqrt{d/|X|}$. Then by the claims in this section

$$\begin{aligned} & \|\mathbb{E}_{x\in X^{(d)}} P_x(w)\| \leq 0.01 + (1 - \frac{c\varepsilon}{2\cdot 10^4|X|^3})^t \\ \leq \; & 0.01 + \exp\big(\frac{-ct\varepsilon}{2\cdot 10^4|X|^3}\big) \leq 0.01 + \exp\big(\frac{-kc\varepsilon}{10^6}\big) \end{aligned}$$

We plugged in $2t \geq 0.05\sqrt{d/|X|} \geq 0.05k|X|^3$. This concludes the proof of theorem 5 for large r.

5.2.1 Proof of lemma 5.6

Let $\tau(G \times G, \{(y, y^{-1})|y \in Y^*\})$ be the spectral gap. From theorem 6 we have for every $u \in W_\perp \otimes W$

$$\|\mathbb{E}_{y\in Y^*}\left[P_{y,y^{-1}}(u)\right]\| \leq (1-\tau)\|u\| \tag{3}$$

In lemma 5.6 we want to upper bound

$$\mathbb{E}_{y\in X^2}\|(P_{y_1,y_2,1^{(d-2)}} + P_{y_2,y_1,1^{(d-2)}})(w)/2\| \tag{4}$$

for every $w \in W_\perp^{\otimes r} \otimes W^{d-r}$.

We will start with the case $d = 2$. We will bound (4) in terms of the LHS of (3). In order to do that, we will have to deal with the fact that the norm in (3) appears outside the expectaion, while in (4) it appears inside the expectation (see claim 5.9). Also, the average in (4) is over $y \in X^2$, while in (3) the average is over $y \in Y^*$. (see claim 5.11). After completing the proof in the case $d = 2$, we turn to prove the lemma for general d (claim 5.12).

CLAIM 5.9. *For every $u \in W_\perp \otimes W$*

$$\mathbb{E}_{y\in Y^*}\|\frac{1}{2}(P_{y,y^{-1}} + I)(u)\| \leq (1-\tau/4)\|u\|$$

PROOF. Recall that $\lambda(G\times G, \widetilde{Y^*})$ is the maximal value of $\|\mathbb{E}_{y\in Y^*} P_{y,y^{-1}}(u)\|/\|u\|$ for $u \in W_\perp \otimes W$. In the claim we have a similar expression, but the norm is inside the expectation. It is perhaps surprising at first glance that moving the norm inside the expectation does not change the final value by much, but it is not hard to see: By (3) for every w the average of the vectors $P_{y,y^{-1}}(u)$ has "small" norm. Therefore, it must be that some of the $P_{y,y^{-1}}(u)$ are far away from u, which implies that $(P_{y,y^{-1}} + I)(u)$ has "small" norm for many y, and this proves the claim. The claim below (not proved) makes this argument precise.

CLAIM 5.10. *If for some vectors $w_0, w_1, \ldots, w_L$, all with norm 1,*

$$(1/L)\cdot\|\sum_{i=1}^{L} w_i\| \leq 1 - \varepsilon$$

then

$$(1/L)\cdot\sum_{i=1}^{L}\|w_0 + w_i\|/2 \leq 1 - \varepsilon/4$$

This ends the proof of claim 5.9. □

CLAIM 5.11. *For every $u \in W_\perp \otimes W$*

$$\mathbb{E}_{y\in X^2}\|\frac{1}{2}(P_{y_1,y_2} + P_{y_2,y_1})(u)\| \leq (1 - \frac{\tau}{8c|X|})\|u\|$$

PROOF. By applying the unitary operator $P_{1,y}$ to the y-th summand in claim 5.9 we obtain

$$\mathbb{E}_{y\in Y^*}\|\frac{1}{2}(P_{y,1} + P_{1,y})(u)\| \leq (1-\tau/4)\|u\|$$

Consider $\mathbb{E}_{y\in X^2}\|\frac{1}{2}(P_{y_1,y_2} + P_{y_2,y_1})(u)\|$. Let p be the probability that for a random $y \in X^2$ we have $y_1 \in Y^*$ and $y_2 = 1$. Then $p \geq (1/2c)\cdot 1/|X|$ (as $X = c\cdot Y \cup Y^*$ and Y^* is larger than Y). Using a convexity argument similar to claim 2.4 we see that

$$\begin{aligned} & \mathbb{E}_{y\in X^2}\|\frac{1}{2}(P_{y_1,y_2} + P_{y_2,y_1})(u)\| \\ \leq \; & p\cdot\mathbb{E}_{y\in Y^*}\|\frac{1}{2}(P_{y,1} + P_{1,y})(u) \; + (1-p)\cdot\|u\| \\ \leq \; & p\cdot(1-\tau/4)\|u\| + (1-p)\|u\| \\ \leq \; & (1-p\tau)\|u\| \leq (1 - \frac{\tau}{8c|X|})\|u\| \end{aligned}$$

which proves claim 5.11 □

We have shown that for every $u \in W_\perp \otimes W$

$$\begin{aligned} & \mathbb{E}_{y\in X^2}\|\frac{1}{2}(P_{y_1,y_2} + P_{y_2,y_1})(u)\| \leq (1 - \frac{\tau}{8c|X|})\|u\| \\ = (1 - \frac{\varepsilon}{21\cdot 8|Y^*|^2\cdot|X|})\|u\| \leq (1 - \frac{c\varepsilon}{2\cdot 10^4|X|^3})\|u\| \end{aligned}$$

The last step follows from $Y^* \leq 10|X|/c$ (which is true since $X = cY \cup Y^*$ and $Y^* \leq 10|Y|$).

We have almost completed proving the lemma. We have the right upper bound, but for $u \in W^{\otimes 2}$ instead of in $W^{\otimes d}$.

CLAIM 5.12. *If there is a $\lambda > 0$ such that for every $u \in W_\perp^{\otimes 2}$*

$$\mathbb{E}_{y\in X^2}\|[\frac{1}{2}(P_{y_1,y_2} + P_{y_2,y_1})(u)]\| \leq \lambda\|u\|$$

then for every $w \in W_\perp^{\otimes r} \otimes W^{\otimes(d-r)}$

$$\mathbb{E}_{y\in X^2}\|[\frac{1}{2}(P_{y_1,y_2,1^{(d-2)}} + P_{y_2,y_1,1^{(d-2)}})(w)]\| \leq \lambda\|w\|$$

PROOF. Write $w \in W_\perp^{\otimes r}\otimes W^{\otimes(d-r)}$ as $w = \sum u_i \otimes v_i$ where $u_i \in W_\perp^{\otimes 2}$ and $v_i \in W^{\otimes(d-2)}$, such that the v_i are orthogonal and $\|v_i\| = 1$. We have

$$\begin{aligned} & \mathbb{E}_y\|[\frac{1}{2}(P_{y_1,y_2,1^{(d-2)}} + P_{y_2,y_1,1^{(d-2)}})(w)]\|^2 \\ = \; & \mathbb{E}_y\sum_i\|[\frac{1}{2}(P_{y_1,y_2} + P_{y_2,y_1})(u_i)]\|^2 \leq \lambda^2\|w\|^2 \end{aligned}$$

And the result follows since $\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)$ for any random variable X. □

6. PROOF OF THEOREM 3

The theorem appears in section 2.2.3.

REMARK 6.1. *This section contains equations in groups. Constants in the equations will be written in Greek letters. Variables will be written in small Latin letters.*

Let $C = A \wr B$, where A is any group and $B \subset S_d$. Given an element $\gamma \in C$ we look for a "commutator representation algorithm" that solves the equation $\gamma = [c_1, c_2] := c_1 c_2 {c_1}^{-1} {c_2}^{-1}$. By assumption we have such an algorithm for A and B. The proof below extends Nikolov's proof in [32].

Any element $\gamma \in A \wr B$ has a unique representation $c = \beta \cdot \underline{\alpha}$ with $\beta \in B$, $\underline{\alpha} \in A^d$, so it is enough to solve, for every pair $(\beta \in B, \underline{\alpha} \in A^d)$, the equation $\beta\underline{\alpha} = [b_1\underline{x}, b_2\underline{y}]$ Now

$$[b_1\underline{x}, b_2\underline{y}] = [b_1, b_2] \cdot \underline{x}^{[b_1,b_2]} \underline{y}^{b_2 b_1^{-1} b_2^{-1}} \underline{x}^{-b_2^{-1}} \underline{y}^{-1}$$

where $\underline{x}^b = b\underline{x}b^{-1}$. In our case $\underline{x}^b$ is simply a permutation of the coordinates of $\underline{x}$ by $b \in B \subset S_d$.

We obtain a pair of equations

$$\beta = [b_1, b_2]$$
$$\underline{\alpha} = \underline{x}^{[b_1,b_1]} \underline{y}^{b_2^{-1} b_1 b_2} \underline{x}^{-b_2} \underline{y}^{-1}$$

By assumption there is an algorithm that solves $\beta = [b_1, b_2]$. Fix some solution $b_1 = \beta_1, b_2 = \beta_2$. It remains to solve

$$\underline{\alpha} = \underline{x}^{-[\beta_1,\beta_2]} \underline{y}^{-\beta_2^{-1}\beta_1\beta_2} \underline{x}^{b_2} \underline{y}$$

Since $\underline{x}^\beta$ is a permutation (depending on β) of the coordinates of $\underline{x}$, the following lemma solves a more general system of equations.

LEMMA 6.2. *For any four permutations $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in S_d$ and any $\underline{\alpha} = \alpha_1, \ldots, \alpha_d \in A^d$, the following system of d equations, one for each $1 \le i \le d$:*

$$\alpha_i = x_{\sigma_1(i)} y_{\sigma_2(i)} x^{-1}_{\sigma_3(i)} y^{-1}_{\sigma_4(i)}$$

has a solution algorithm that calls the commutator representation algorithm on A at most d times, and does at most $O(d)$ operations in the group A.

The rest of this section is dedicated to the proof of this lemma.

DEFINITION 6.3. *We shall refer to the α_i as* **constants** *and to the $x_i, y_i, x_i^{-1}, y_i^{-1}$ as* **literals**

There are d constants and $4d$ literals i our system. An important fact is that each literal appears *exactly once* in the system.

Let as solve first in the case that all four σ_i are the identity permutation. The system in this case is:

$$\alpha_1 = [x_1, y_1]$$
$$\alpha_2 = [x_2, y_2]$$
$$\ldots$$
$$\alpha_d = [x_d, y_d]$$

In this case the equations are independent (no variable appears in more than one equation). Each equation asks for a commutator representation for $\alpha_i \in A$. We solve the system of equations by calling the commutator representation algorithm for A for each equation separately.

The solution for general σ_i is by reduction to a system similar to the one we obtained for the $\sigma_i = 1$ case. As long as there are variables that appear in more than one equation, we will remove equations by "Gaussian elimination", until we obtain a system of independent equations. We will then translate each equation to a commutator representation equation like the ones above.

As mentioned, each literal appears exactly once in the system. If x_i, x_i^{-1} do not both appear in the same equation, then we can eliminate x_i, x_i^{-1} from the system by substitution (paying $O(1)$ multiplications in A). This reduces the number of equations in the system by 1. Repeat the substitution operation until it is no longer possible. Notice that the property that each literal appears exactly once is preserved along the way.

CLAIM 6.4. *The substitution process ends with $L \le d$ equations*

$$v_l = W_l \quad \forall l \in \{1, \ldots, L\}$$

where v_l is a constant, W_l is some word in literals and constants. The equations are now independent - every literal appears in the same equation as its inverse, ot they both do not appear in the system. □

We will now reduce this system to L commutator representation problems in the group A. The following lemma finds a "hidden commutator" in each of the words W_l:

LEMMA 6.5. *[32] In every W_l there exist $g, h \in \{1, 2, \ldots, d\}$ depending on l, such that*

$$W_l = Z_1 x_g Z_2 y_h Z_3 x_g^{-1} Z_4 y_h^{-1} Z_5$$

where the Z_i are words in literals and constants from the word W_l (they do not contain $x_g^{\pm 1}, x_h^{\pm 1}$ since each literal appears at most once in the system of equations).

The proof is in [32]. Given that such a hidden commutator exists, it is easy to find one in time polynomial in d by looking at all the literals appearing in W_l (there are at most $2d$ of those). Substitute every variable appearing in the Z_i by 1. This does not affect any other equation - the equations are independent at this point. We obtain a new equation

$$v_l = \tilde{W}_l = \zeta_1 x_g \zeta_2 y_h \zeta_3 x_g^{-1} \zeta_4 y_h^{-1} \zeta_5$$

This is now an equation in two variables x_g, x_h - all the other words are constants. This is almost a "commutator representation" equation. Indeed, if the five ζ_i are all equal 1, we obtain the equation

$$v_l = [x_g, y_h]$$

which is solved by calling the commutator algorithm on A. For general ζ_i we transform the "hidden" commutator to a "real" commutator by changing variables. Define $\tilde{x}_g = \zeta_3 x_g \zeta_4$ and $\zeta_h = y_h \zeta_2^{-1} \zeta_3^{-1}$. Observe that

$$v_l = \zeta_1 \zeta_4 [\tilde{x}_g, \tilde{y}_h] \zeta_3 \zeta_2 \zeta_5$$

Rewrite this equation as

$$(\zeta_1 \zeta_4)^{-1} v_l (\zeta_3 \zeta_2 \zeta_5)^{-1} = [\tilde{x}_g, \tilde{y}_h]$$

The LHS is some constant element in A, and the equation requests a representation of this element as a commutator. We can find a solution by calling the commutator representation algorithm on A. The solution is in the variables $\tilde{x}_g, \tilde{y}_h$, but this is easily translated to a solution in our original variables x_g, y_h.

How many operations did we use? We called the commutator representation algorithm in A At most d times (one call for each final equation $v_l = W_l$). We called the commutator representation algorithm on B one time. We used $O(1)$ multiplications in B, and $O(d)$ multiplications in A (there were $O(1)$ per either removing an equation or solving a final equation).

We can now deduce corollary 2.9. Define $m(n)$ to be the cost (in bit operations) of multiplication in G_n, and define $c(n)$ to be the cost of computing the commutator representation of an element in G_n. As $m(n+1) < (d+1)m(n)$ and $m(1) = O(d^2)$ we deduce that $m(n) < (d+1)^{n+2} \cdot O(1)$. From the discussion above we see that $c(n+1) < (d+1)c(n) + m(n) \cdot O(d) < (d+1)c(n) + d^{n+3} \cdot O(1)$. This implies that $c(n) < d^{4n} \cdot O(1)$ for large enough d. It remains to verify that $c(n)$ is polynomial in $\log |G_n|$. Denote $s(n) = \log |G_n|$. Clearly $s(n+1) > d \cdot s(n)$, as G_{n+1} consists of all the vectors of length $d+1$ whose first element is in A_d and the rest are in G_n. Therefore $s(n) > d^n$ and the corollary follows.

Acknowledgments

We are grateful to Alex Lubotzky for many insightful conversations, in part supplying the group theoretic tools we ended up needing for the proof.

7. REFERENCES

[1] M. Ajtai, J. Komlós, and E. Szemerédi. Sorting in $c \log n$ parallel steps. *Combinatorica*, 3(1):1–19, 1983.

[2] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[3] N. Alon, Z. Galil, and V. D. Milman. Better expanders and superconcentrators. *Journal of Algorithms*, 8(3):337–347, 1987.

[4] N. Alon, A. Lubotzky, and A. Wigderson. Semi-direct product in groups and zig-zag product in graphs: connections and applications (extended abstract). In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 630–637. IEEE Computer Soc., Los Alamitos, CA, 2001.

[5] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory. Series B*, 38(1):73–88, 1985.

[6] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994.

[7] L. Babai, G. Hetyei, W. M. Kantor, A. Lubotzky, and Á. Seress. On the diameter of finite groups. In *31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990)*, pages 857–865. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990.

[8] E. Ben-Sasson, M. Sudan, S. Vadhan, and A. Wigderson. Randomness-efficient low degree tests and short pcps via epsilon-biased sets. In *Proceedings of the thirty-fifth ACM symposium on Theory of computing*, pages 612–621. ACM Press, 2003.

[9] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 659–668. ACM Press, 2002.

[10] P. Diaconis and M. Shahshahani. On the eigenvalues of random matrices. *J. Appl. Probab.*, 31A:49–62, 1994. Studies in applied probability.

[11] J. D. Dixon. The probability of generating the symmetric group. *Math. Z.*, 110:199–205, 1969.

[12] M. Eichler. Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion. *Arch. Math.*, 5:355–366, 1954.

[13] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. *J. Comput. Syst. Sci.*, 22(3):407–420, June 1981.

[14] O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, and D. Zuckerman. Security preserving amplification of hardness. In *31st Annual Symposium on Foundations of Computer Science*, volume I, pages 318–326, St. Louis, Missouri, 22–24 Oct. 1990. IEEE.

[15] M. Gromov. Spaces and questions. *Geometric and Functional Analysis*, pages 118–161, 2000. Part I of Special Volume on GAFA 2000 (Tel Aviv, 1999).

[16] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on the Theory of Computing*, pages 356–364, Montréal, Québec, Canada, 23–25 May 1994.

[17] R. Impagliazzo and A. Wigderson. $P = BPP$ if E requires exponential circuits: Derandomizing the XOR lemma. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 220–229, El Paso, Texas, 4–6 May 1997.

[18] S. Jimbo and A. Maruoka. Expanders obtained from affine transformations. *Combinatorica*, 7(4):343–355, 1987.

[19] N. J. Kalton and J. W. Roberts. Uniformly exhaustive submeasures and nearly additive set functions. *Transactions of the American Mathematical Society*, 278(2):803–816, 1983.

[20] D. Kazhdan. On the connection of the dual space of a group with the structure of its closed subgroups (russian). *Funkcional. Anal. i Prilozh.*, 1:71–74, 1967.

[21] M. Klawe. Limitations on explicit constructions of expanding graphs. *SIAM J. Comput.*, 13(1):156–166, 1984.

[22] L. Lovász and P. Winkler. Mixing times. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 85–133. Amer. Math. Soc., Providence, RI, 1998.

[23] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures*. Birkhäuser Verlag, Basel, 1994.

[24] A. Lubotzky and I. Pak. The product replacement algorithm and Kazhdan's property (T). *Journal of the American Mathematical Society*, 14(2):347–363 (electronic), 2001.

[25] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[26] A. Lubotzky and B. Weiss. Groups and expanders. In *Expanding graphs (Princeton, NJ, 1992)*, volume 10 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 95–109. Amer. Math. Soc., Providence, RI, 1993.

[27] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredači Informacii*, 9(4):71–80, 1973.

[28] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988.

[29] R. Meshulam and A. Wigderson. Expanders from symmetric codes. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 669–677. ACM Press, 2002.

[30] M. Morgenstern. Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q. *Journal of Combinatorial Theory. Series B*, 62(1):44–62, 1994.

[31] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, Aug. 1993.

[32] N. Nikolov. On the commutator width of perferct groups. *to appear*.

[33] O. Ore. Some remarks on commutators. *Proc. Amer. Math. Soc.*, 2:307–314, 1951.

[34] M. S. Pinsker. On the complexity of a concentrator. In *7th Annual Teletraffic Conference*, pages 318/1–318/4, Stockholm, 1973.

[35] N. Pippenger. Sorting and selecting in rounds. *SIAM J. Comput.*, 16(6):1032–1038, Dec. 1987.

[36] N. Pippenger and A. C. Yao. Rearrangeable networks with limited depth. *SIAM Journal on Algebraic and Discrete Methods*, 3:411–417, 1982.

[37] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Ann. of Math. (2)*, 155(1):157–187, 2002.

[38] A. Selberg. On the estimation of Fourier coefficients of modular forms. In *Proc. Sympos. Pure Math., Vol. VIII*, pages 1–15. Amer. Math. Soc., Providence, R.I., 1965.

[39] M. Sipser. Expanders, randomness, or time versus space. *J. Comput. Syst. Sci.*, 36(3):379–383, June 1988.

[40] M. Sipser and D. A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6, part 1):1710–1722, 1996.

[41] D. A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 42(6, part 1):1723–1731, 1996.

[42] M. R. Tanner. Explicit concentrators from generalized n-gons. *SIAM Journal on Algebraic Discrete Methods*, 5(3):287–293, 1984.

[43] A. Urquhart. Hard examples for resolution. *Journal of the Association for Computing Machinery*, 34(1):209–219, 1987.

[44] L. G. Valiant. Graph-theoretic arguments in low-level complexity. In *Mathematical foundations of computer science (Proc. Sixth Sympos., Tatranská Lomnica, 1977)*, pages 162–176. Lecture Notes in Comput. Sci., Vol. 53. Springer, Berlin, 1977.

[45] A. Wigderson and D. Zuckerman. Expanders that beat the eigenvalue bound: explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999.