# Expansion in Soluble Groups

**Mika Göös**

Keble College

University of Oxford

A thesis submitted in partial fulfillment of the MSc in

*Mathematics and the Foundations of Computer Science*

September 2011

# EXPANSION IN SOLUBLE GROUPS

MIKA GÖÖS

## Contents

## 1. Introduction

The purpose of this dissertation is to explicitly construct *families of **expander graphs*** out of *families of **finite soluble groups***. In short, we study *expansion in soluble groups*. This line of research is relatively new: most of the major results in this area are only a decade old. We present a self-contained exposition of some the known results and offer simple alternative proofs and observations in the process. Our focus is on communicating the flavour of the field; we do not aim to present the most general theorems available.

The study of expander graphs *in general* is motivated by the diverse applications that they have in computer science and elsewhere. However, for us, the main motivation to study expansion *in soluble groups* is in the interplay between the different areas of mathematics that are involved: algebraic graph theory, finite groups and their linear representations, probabilistic combinatorics, and randomized algorithms. We assume the reader has some familiarity with these topics.

We proceed to give a brief overview of our subject matter. We keep the discussion relatively informal at this stage; the precise definitions are given in due course.

**Expander graphs and their importance.** *Expander graphs* (or *expanders*, for short) are graphs that have two seemingly contradictory properties: they are *sparse* (i.e., they have very few edges), yet *well-connected.* Here, the property of well-connectedness can be made precise in a variety of ways. In fact, the wide applicability of expander graphs stems from there being many equivalent definitions. We list three, informally:

(1) **Combinatorial.** Every not-too-large set $T \subseteq V(\mathcal{G})$ of vertices in an expander graph $\mathcal{G}$ has relatively many edges leaving $T$.
(2) **Probabilistic.** A random walk on an expander graph converges to the uniform distribution very rapidly.
(3) **Algebraic.** The second largest eigenvalue of the adjacency matrix of an expander graph is small.

Expander graphs are of major importance in theoretical computer science. Most often, expanders are utilized as *pseudorandom* objects, since they share properties with "typical" random graphs. For example, whenever $T, S \subseteq V(\mathcal{G})$ are sets of vertices in an expander graph, the number of edges between $T$ and $S$ is close to what one expects in a suitably defined random graph of the same edge density (this is the Expander Mixing Lemma [20, p. 454])—loosely speaking, expander graphs *look* random. In fact, one concrete application of expanders is in reducing the amount of randomness needed in an algorithm: one might be able to replace a probabilistically generated random graph with a deterministically generated expander graph. These kinds of applications call for *explicit* constructions of expanders graphs, i.e., we want the expander graphs to have efficiently computable descriptions.

Some further applications of expander graphs are found in the design of computer networks, in constructing error-correcting codes, and in computational complexity (most notably, Reingold's log-space algorithm for undirected connectivity [36]). The award-winning survey of Hoory, Linial & Wigderson [20] covers a broad range of applications of expanders in theoretical computer science. Very recently, expanders have also found applications in pure mathematics; see the survey of Lubotzky [28].

Because of this wealth of material, we do not discuss the applications of expander graphs in this dissertation. Instead, we study expanders in their own right: we concentrate on constructing expanders, and in the process, we ask questions about how to optimize the parameters that appear in these constructions.

**Using finite groups to construct expanders.** Given a finite group $G$ and a generating set $S \subseteq G$ one can construct the *Cayley graph* $\mathcal{C}(G; S)$ of $G$ with respect to $S$. This is the graph that has the elements of $G$ as vertices and two elements $g, h \in G$ are joined by an edge if $gs = h$ for some $s \in S$.

Many explicit constructions of expander graphs are obtained as Cayley graphs of a family of finite groups. Most notably, this is the case with the *Ramanujan graphs* found in 1988 by Lubotzky, Phillips & Sarnak and, independently, by Margulis (Section 2.3). Their construction involves a family of finite *simple groups*. Subsequently, expansion in simple groups has been extensively studied: this research has culminated in the recent result that all non-abelian finite simple groups can be made uniformly into *constant-degree* expander graphs by considering their Cayley graphs (Section 3.4).

By contrast, it has been long known that many natural families of *soluble groups* can not be made into constant-degree expanders. This is a huge drawback from the point of view of applications, and consequently, very little research has been done on expansion in soluble groups—this is the research we will explore.

**Our focus.** Our main goal in this work is to quantify the extent to which specific soluble groups fail to be made into constant-degree expanders: We seek *upper bounds* on the number of generators that are needed to make specific groups expanding. Also, we want to find *matching lower bounds* in order to prove that our constructions are asymptotically optimal.

In finding generating sets, we keep the issue of *explicitness* in mind by accompanying our existence proofs with efficient algorithms that construct the relevant generating sets. To do so, we introduce methods to *derandomize* our probabilistic existence proofs.

**Our contributions.** In order to ensure a unified exposition we have adapted many proofs to our special case of Cayley graphs. We always give our proofs in full detail, even though the original articles might not have done so. In some instances we offer alternative proofs to those found in the literature.

In particular, we believe the following small contributions are novel.

(1) In Section 6.4 we give a representation theoretic proof of a theorem of Alon, Lubotzky & Wigderson [2].
(2) In Section 7.1 we (weakly) derandomize a variant of a theorem in [2].
(3) In Section 8 we make the observation that the method of *derandomized squaring* can be used to improve some results in [19, 30].

**Organization of the thesis.** In Section 2 we will formally introduce expander graphs and survey their general properties. From Section 3 onwards we restrict our attention to Cayley expander graphs: we explain how their algebraic origins allow their expansion properties to be analyzed. In Section 4 we study expansion in abelian groups—the building blocks of soluble groups—where a theorem of Alon & Roichman will be of central importance. In Section 5 we discuss how the expansion properties of a group are related to the expansion properties of its quotients and subgroups. Here, the main applications will be in proving lower bounds on the number of generators that are required to achieve expansion.

In Sections 6 and 7 we describe the celebrated zig-zag graph product, due to Reingold, Vadhan & Wigderson, and how it is related to semi-direct products of groups. This will be the key ingredient in constructing sparse expanding families from non-abelian soluble groups. In Section 8 we discuss the method of *derandomized squaring*, due to Rozenman & Vadhan. Finally, in Section 9, we mention some further directions.

**Acknowledgements.** We thank our supervisors M. J. Collins and O. Riordan for taking the time to offer suggestions and improvements on our early drafts.

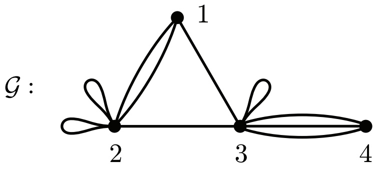
## 2. Expander Graphs

Expander graphs are highly-connected sparse graphs. Their precise definition will be given in Section 2.2 after we first motivate their study by reviewing some concepts from random walks on graphs. Even though this section is concerned with *general* expanders (not necessarily derived from finite groups) we prove only results that will be used later in this work. However, we mention without proof some important results in Section 2.3 in order to give a taste of the vast body of research that is being done on expander graphs.

2.1. **Graphs and random walks.** The material in this subsection is standard. Our main references are the text book of Godsil & Royle [18] and the survey of Hoory et al. [20].

2.1.1. *Multigraphs.* Our graphs $\mathcal{G} = (V, E)$ are always undirected. We allow our graphs to have self-loops (i.e., $\{v\} \in E$ for some $v \in V$) and parallel edges (i.e., the multiplicity of an edge in the *multiset* $E$ can exceed 1). In fact, a graph $\mathcal{G}$ is best understood via its *adjacency matrix* $\mathbf{M} = \mathbf{M}(\mathcal{G}) = (m_{uv})$, $(u, v) \in V \times V$, where $m_{uv}$ equals the number of edges from $v$ to $u$. In particular, the degree of a vertex $v$ is given by $\deg(v) = \sum_{u \in V} m_{uv}$, and $m_{vv}$ counts self-loops on $v$. More generally, the $uv^{\text{th}}$ entry of $\mathbf{M}^k$ gives the number of walks of length $k$ starting at $v$ and ending in $u$. We define the *$k^{th}$ power* $\mathcal{G}^k$ of $\mathcal{G}$ to be the graph with adjacency matrix $\mathbf{M}^k$, i.e., for each $k$-walk on $\mathcal{G}$ there corresponds an edge in $\mathcal{G}^k$. Note that $\mathbf{M}^T = \mathbf{M}$ as our graphs are undirected.

The following is an example on the vertex set $V = \{1, 2, 3, 4\}$:



$$\mathbf{M}(\mathcal{G}) = \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 2 & 1 & 0 \\ 1 & 1 & 1 & 3 \\ 0 & 0 & 3 & 0 \end{bmatrix} .$$

2.1.2. *Random walks on graphs.* A *simple random walk* on a graph $\mathcal{G}$ is a sequence of $V(\mathcal{G})$-valued random variables $X_0, X_1, X_2, \ldots$ such that when conditioned on the event $X_t = v$, the distribution of $X_{t+1}$ is given by randomly walking along one of the $\deg(v)$ edges leaving $v$, each edge with probability $1/\deg(v)$. Formally,

$$\mathbf{Pr}[X_{t+1} = u \mid X_t = v, \ldots, X_0 = v_0] = \mathbf{Pr}[X_{t+1} = u \mid X_t = v] = m_{uv}/\deg(v) \ .$$

Here, the first equality is known as the *Markov property.* Consequently, the variables $\{X_t\}_{t \in \mathbb{N}}$ form a *Markov chain.*

Define the *random walk matrix* $\mathbf{A} = \mathbf{A}(\mathcal{G}) = (a_{uv})$ to be the normalized adjacency matrix of $\mathcal{G}$ so that $a_{uv} = m_{uv}/\deg(v)$. The column sums of $\mathbf{A}$ satisfy $\sum_u a_{uv} = 1$, so $\mathbf{A}$ is a *left stochastic matrix.* Now, if $X_t$ has distribution $\mathbf{p} \in \mathbb{R}^V$ (i.e., $\mathbf{Pr}[X_t = v] = p_v$) then the distribution of $X_{t+1}$ is given by $\mathbf{Ap}$, since

$$\mathbf{Pr}[X_{t+1} = u] = \sum_v \mathbf{Pr}[X_{t+1} = u \mid X_t = v] \cdot \mathbf{Pr}[X_t = v] = \sum_v a_{uv} p_v = (\mathbf{Ap})_u \ .$$

More generally, the distribution of $X_{t+s}$ is given by $\mathbf{A}^s \mathbf{p}$. Indeed, the Markov chain $\{X_t\}_{t \in \mathbb{N}}$ is completely determined by the distribution of $X_0$ and the transition probabilities that are recorded in $\mathbf{A}$.

Aspects of the long-term behaviour of the random walk determined by a non-negative matrix $\mathbf{A}$ can be analyzed using the Perron-Frobenius Theorem [18, p. 178]. Instead of pursuing random walk analysis in full generality we restrict our considerations to *regular* graphs that are particularly well-behaved. On a $d$-regular graph $\mathcal{G}$ the random walk matrix is simply given by

$$\mathbf{A}(\mathcal{G}) = \frac{1}{d}\mathbf{M}(\mathcal{G}) \ . \tag{1}$$

2.1.3. *Eigenvalues of graphs.* In what follows we will always assume our graphs are regular. In this case $\mathbf{A}$ is symmetric and thus *doubly stochastic* since both the columns and rows sum up to 1. As a real symmetric matrix, $\mathbf{A}$ has a full set of eigenvectors that form an *orthogonal* basis of $\mathbb{R}^V$ (w.r.t. the natural inner product on $\mathbb{R}^V$) [18, p. 170]. Enumerate the eigenvalues of $\mathbf{A}$ as

$$\lambda_1(\mathcal{G}) \geq \lambda_2(\mathcal{G}) \geq \cdots \geq \lambda_n(\mathcal{G}) \ ,$$

where $n = |\mathcal{G}|$ is the order of $\mathcal{G}$. The (multi-)set of all the eigenvalues $\lambda_i$ is called the *spectrum* of $\mathbf{A}$ (or of $\mathcal{G}$) and it is denoted by $\operatorname{spec}(\mathbf{A})$ (or $\operatorname{spec}(\mathcal{G})$).

We collect some well-known properties of the spectrum in the following.

**Theorem 2.1** (e.g., [18]). *Let $\mathbf{A}$ be the random walk matrix of a d-regular graph $\mathcal{G}$. Then,*

*(i)* $\mathbf{A}(\mathcal{G}^k) = \mathbf{A}(\mathcal{G})^k$ *and* $\operatorname{spec}(\mathcal{G}^k) = \{\lambda^k : \lambda \in \operatorname{spec}(\mathcal{G})\}$.

*(ii)* $\lambda_1 = 1$ *and* $|\lambda| \leq 1$ *for all* $\lambda \in \operatorname{spec}(\mathbf{A})$.

*(iii)* $\lambda_1 > \lambda_2$ *iff* $\mathcal{G}$ *is connected.*

*(iv) Suppose* $\mathcal{G}$ *is connected. Then* $\lambda_n = -1$ *iff* $\mathcal{G}$ *is bipartite.*

*Proof. (i)* Immediate from equation (1).

*(ii)* Let $\mathbf{x}$ be an eigenvector with eigenvalue $\lambda \in \operatorname{spec}(\mathbf{A})$. Fix $v \in V$ so that $|x_v| \geq |x_u|$ for all $u \in V$. Then $|(\mathbf{A}\mathbf{x})_v| = |\sum_u a_{vu}x_u| \leq (\sum_u a_{vu})|x_v| = |x_v|$. On the other hand, $|(\mathbf{A}\mathbf{x})_v| = |\lambda| \cdot |x_v|$ so that $|\lambda| \leq 1$. Furthermore, the uniform all-ones vector $\mathbf{1} = [1 \ 1 \ \cdots \ 1]^T$ has eigenvalue 1.

*(iii)* Let $\mathcal{G}$ be connected and $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{x} = \mathbf{x}$. Denote by $U \subseteq V$ the set of vertices $u$ with $x_u = \max_v x_v$. Assume $U \neq V$. Using the fact that $\mathcal{G}$ is connected we can find a vertex $u \in U$ that has a neighbour $v \notin U$. Now $x_u = (\mathbf{A}\mathbf{x})_u = \sum_w a_{uw}x_w \leq (1/d)x_v + (1 - 1/d)x_u < x_u$, a contradiction. Hence, $U = V$ and we have that $\mathbf{x}$ is a constant vector, i.e., a multiple of $\mathbf{1}$. This shows that the eigenspace corresponding to eigenvalue 1 is one-dimensional and hence $\lambda_1 > \lambda_2$.

Conversely, suppose $\mathcal{G}$ is not connected and $U \subsetneq V$ is the vertex set of some connected component. Now, the 01-vector $\mathbf{1}_U$ that has $v^{\text{th}}$ entry 1 iff $v \in U$ is not a multiple of $\mathbf{1}$ but it is still an eigenvector with eigenvalue 1. Thus, $\lambda_1 = \lambda_2 = 1$.

*(iv)* Let $\mathcal{G}$ be bipartite with bipartition $V = X \cup Y$. Noting that $|X| = |Y|$ (as $\mathcal{G}$ is regular) it is easy to see that $\mathbf{1}_X - \mathbf{1}_Y$ is an eigenvector with eigenvalue $-1$.

Conversely, suppose $\mathcal{G}$ is connected and $\lambda_n = -1$. Then the graph $\mathcal{G}^2$ has spectrum $\lambda_1^2, \ldots, \lambda_n^2$. In particular, the 1-eigenspace has dimension at least 2, which implies by part *(iii)* that $\mathcal{G}^2$ is disconnected. This is possible only if $\mathcal{G}$ is bipartite. □

By the above discussion the *uniform distribution* $\mathbf{u} \coloneqq \mathbf{1}/n$, $\|\mathbf{u}\|_1 = 1$, is an eigenvector with eigenvalue $\lambda_1 = 1$. Thus, it is always a stationary distribution of the random walk on $\mathcal{G}$. It is natural to ask whether a random walk $\{X_t\}_{t\in\mathbb{N}}$ always converges to this distribution, i.e., whether $\|\mathbf{A}^t\mathbf{p} - \mathbf{u}\|_1 \to 0$ as $t \to \infty$, where $\mathbf{p}$ is the distribution of $X_0$. To answer this, we introduce an algebraic measure of connectivity that controls the speed of convergence.

**Definition 2.1.** The *second largest eigenvalue (in absolute value)* of $\mathcal{G}$ is defined to be $\lambda(\mathcal{G}) = \lambda(\mathbf{A}) \coloneqq \max\{|\lambda_2|, |\lambda_3|, \ldots, |\lambda_n|\} = \max\{|\lambda_2|, |\lambda_n|\}$.

The parameter $\lambda = \lambda(\mathcal{G})$ measures the amount by which the orthogonal complement of $\mathbf{1}$ gets compressed under the action of $\mathbf{A}$: let $\mathbf{e}^1, \dots, \mathbf{e}^n \in \mathbb{R}^V$, $\|\mathbf{e}^i\|_2 = 1$, $\mathbf{A}\mathbf{e}^i = \lambda_i \mathbf{e}^i$, be an orthonormal basis of eigenvectors with the convention that $\mathbf{e}^1$ and $\mathbf{u} = \mathbf{1}/n$ are parallel. Then every vector $\mathbf{x} \perp \mathbf{1}$ can be expanded in the eigenvector basis as $\mathbf{x} = \sum_{i=2}^n \alpha_i \mathbf{e}^i$ for some $\alpha_i \in \mathbb{R}$. Calculating

$$\|\mathbf{A}\mathbf{x}\|_2^2 = \|\sum_{i=2}^n \alpha_i \mathbf{A}\mathbf{e}^i\|_2^2 = \|\sum_{i=2}^n \alpha_i \lambda_i \mathbf{e}^i\|_2^2 = \sum_{i=2}^n \alpha_i^2 \lambda_i^2 \le \lambda^2 (\sum_{i=2}^n \alpha_i^2) = \lambda^2 \|\mathbf{x}\|_2^2 \ ,$$

we have $\|\mathbf{A}\mathbf{x}\|_2 \le \lambda \|\mathbf{x}\|_2$. Indeed, the parameter $\lambda$ can be equivalently defined as

$$\lambda(\mathcal{G}) = \max_{\mathbf{x} \perp \mathbf{1}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \in [0, 1] \ .$$

Using the matrix norm $\|\mathbf{A}\|_2 := \max_{\mathbf{x}} \|\mathbf{A}\mathbf{x}\|_2 / \|\mathbf{x}\|_2$ the above can be written concisely as $\lambda(\mathcal{G}) = \|\mathbf{A} \restriction \mathbf{1}^\perp\|_2$, where $\mathbf{A} \restriction \mathbf{1}^\perp$ is the restriction of the linear map $\mathbf{A}$ to the subspace of all vectors orthogonal to $\mathbf{1}$.

Now, let $\mathbf{p} \in \mathbb{R}^V$ be an arbitrary probability distribution on the vertices of $\mathcal{G}$. Then $\mathbf{p} - \mathbf{u} \perp \mathbf{u}$, so the above considerations give us

$$\|\mathbf{A}^t \mathbf{p} - \mathbf{u}\|_2 = \|\mathbf{A}^t (\mathbf{p} - \mathbf{u})\|_2 \le \lambda^t \|\mathbf{p} - \mathbf{u}\|_2 \le \lambda^t \ ,$$

where we simplified $\|\mathbf{p} - \mathbf{u}\|_2^2 = \|\mathbf{p}\|_2^2 - 2\langle \mathbf{p}, \mathbf{u} \rangle + \|\mathbf{u}\|_2^2 \le 1 - 2/n + 1/n \le 1$. In $\ell_1$-norm the analogous bound reads (using $\|\mathbf{x}\|_1 \le \sqrt{n} \|\mathbf{x}\|_2$)

$$\|\mathbf{A}^t \mathbf{p} - \mathbf{u}\|_1 \le \sqrt{n} \lambda^t \ . \tag{2}$$

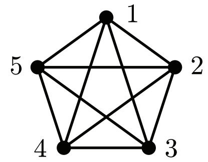
Thus, provided that $\lambda < 1$ (or equivalently that $\mathcal{G}$ is connected and non-bipartite), every random walk on $\mathcal{G}$ converges to $\mathbf{u}$ in distribution regardless of what the initial distribution of $X_0$ is.

In order to ensure fast convergence we want $\lambda(\mathcal{G})$ to be small. Indeed, a more quantitative version of part *(iii)* of Theorem 2.1 is often used in applications. Namely, it can be shown that on any connected non-bipartite graph $\mathcal{G}$ the second largest eigenvalue is not too large: $\lambda(\mathcal{G}) = 1 - \Omega(n^{-2})$ [5, p. 424]. Instead of studying regular graphs in general, the purpose of this dissertation is to construct special families of highly connected graphs where the convergence to uniform is fast.

2.1.4. *Eigenvalues of the complete graph.* As an example, we find the spectrum of the complete graph $\mathcal{K}_n$ on $n$ vertices. This is relatively easy to do by hand.

On $\mathcal{K}_n$, the random walk matrix is given by $\mathbf{A}(\mathcal{K}_n) = \frac{1}{n-1}(\mathbf{J} - \mathbf{I})$, where $\mathbf{J}$ is the all-ones matrix and $\mathbf{I} = \mathbf{I}_n$ is the identity matrix. For instance, when $n = 5$:

$\mathcal{K}_5$ :



$$\mathbf{M}(\mathcal{K}_5) = \mathbf{J} - \mathbf{I} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \ .$$

We have that $\operatorname{spec}(\mathbf{J}) = \{n, 0^{(n-1)}\}$, where we write multiplicities of eigenvalues in the superscript. Writing $\mathbf{A} = p(\mathbf{J})$ for $p(x) = \frac{1}{n-1}(x-1)$ we obtain $\operatorname{spec}(\mathbf{A}) = \{p(\lambda) : \lambda \in \operatorname{spec}(\mathbf{J})\} = \{1, (-1/(n-1))^{(n-1)}\}$. In particular, $\lambda(\mathcal{K}_n) = 1/(n-1) = o(1)$.

2.2. **Expander graphs and their properties.** A random walk on an *expander graph* is guaranteed to converge to the uniform distribution rapidly. We adopt the following algebraic definition of expansion as fundamental.

**Definition 2.2 (Algebraic definition of expanders).** Let $\mathscr{F} = \{\mathcal{G}_i\}_{i \in \mathbb{N}}$ be a family of graphs where $\mathcal{G}_i$ is $d_i$-regular and $|\mathcal{G}_i| \to \infty$. The family $\mathscr{F}$ is called a *family of expander graphs (of degree $d_i$)* if there exists a constant $\lambda < 1$ such that $\lambda(\mathcal{G}_i) \le \lambda$ for all $i \in \mathbb{N}$. A member $\mathcal{G}_i$ of $\mathscr{F}$ is called a $(|\mathcal{G}_i|, d_i, \lambda)$*-graph* for short.

Under the current definition, the complete graphs $\mathcal{K}_n$ (for $n \ge 3$) form an $(n-1)$-degree expander family. However, the definition of expander families is traditionally given with the requirement that the family $\mathscr{F}$ has *fixed degree* $d_i = d \in \mathbb{N}$. We relax this constraint by allowing $d_i$ to grow with $n = |\mathcal{G}_i|$—preferably slowly. In Section 4, we argue that this relaxation is necessary in order to make the Cayley graphs of certain families of finite groups expanding.

*Remark* 2.1. We stress that a single graph is an expander only in the context of an expander family. Nevertheless, we might abuse terminology and say, e.g., that a complete graph is an expander with the understanding that we are implicitly talking about the *infinite family* of complete graphs.

*Remark* 2.2 (Lazy random walk). Every bipartite family $\mathscr{F} = \{\mathcal{G}_i\}$ of degree $d_i$ has trivially $\lambda(\mathcal{G}_i) = 1$. If it is given that $\lambda_2(\mathcal{G}_i) \le \mu < 1$, we can still turn $\mathscr{F}$ into an expander family of degree $2d_i$: add $d_i$ self-loops to every vertex of $\mathcal{G}_i$. The random walk matrices get transformed as $\mathbf{A} \mapsto \frac{1}{2}(\mathbf{A} + \mathbf{I})$ and so the spectrum of $\mathcal{G}_i$ gets transformed as $x \mapsto \frac{1}{2}(x+1) : [-1,1] \to [0,1]$. This gives $\lambda(\mathcal{G}_i) = \frac{1}{2}(\lambda_2(\mathcal{G}_i) + 1) \le \frac{1}{2}(\mu + 1) < 1$. [20, 27]

Next, we present some useful general properties of expander graphs that will be used extensively later.

2.2.1. *Expanders as almost-complete graphs.* The random walk matrix of the complete graph was introduced as $\frac{1}{n-1}(\mathbf{J} - \mathbf{I})$ in Section 2.1.4 above. This matrix can be written (somewhat unnaturally) as a convex combination:

$$(1 - \tfrac{1}{n-1})\tilde{\mathbf{J}} + \tfrac{1}{n-1}(\tfrac{2}{n}\mathbf{J} - \mathbf{I}) \ ,$$

where $\tilde{\mathbf{J}} = \frac{1}{n}\mathbf{J}$ is the random walk matrix of the complete graph *with self-loops* $\tilde{\mathcal{K}}_d$ (these graphs are the best possible expanders with $\lambda(\tilde{\mathcal{K}}_d) = 0$) and $\|\frac{2}{n}\mathbf{J} - \mathbf{I}\|_2 \le 1$, because $(\frac{2}{n}\mathbf{J} - \mathbf{I})\mathbf{1} = \mathbf{1}$ and for $\mathbf{x} \perp \mathbf{1}$ we have $(\frac{2}{n}\mathbf{J} - \mathbf{I})\mathbf{x} = -\mathbf{x}$.

More generally, the random walk matrices of expander graphs have analogous decompositions. The following simple proposition is both conceptually and technically useful.

**Proposition 2.2** ([44, p. 63])**.** *Let* $\mathbf{A}$ *be a doubly stochastic symmetric matrix with* $\lambda(\mathbf{A}) \le \lambda$. *Then there exists a symmetric matrix* $\mathbf{E}$ *with* $\|\mathbf{E}\|_2 \le 1$ *such that*

$$\mathbf{A} = (1 - \lambda)\tilde{\mathbf{J}} + \lambda\mathbf{E} \ . \tag{3}$$

*Proof.* Set $\mathbf{E} = \frac{1}{\lambda}(\mathbf{A} - (1-\lambda)\tilde{\mathbf{J}})$. To see that $\|\mathbf{E}\|_2 \le 1$ note that $\mathbf{E}\mathbf{1} = \frac{1}{\lambda}(\mathbf{1} - (1-\lambda)\mathbf{1}) = \mathbf{1}$ and if $\mathbf{x} \perp \mathbf{1}$ then $\mathbf{E}\mathbf{x} = \frac{1}{\lambda}(\mathbf{A}\mathbf{x} - (1-\lambda)\mathbf{0}) = \frac{1}{\lambda}\mathbf{A}\mathbf{x}$ so that $\|\mathbf{E}\mathbf{x}\|_2 \le \|\mathbf{x}\|_2$. □

2.2.2. *Diameter of expanders.* As expander families have their second largest eigenvalue uniformly bounded away from 1, the next well-known proposition implies that expanders have logarithmic diameter. We write $f(n) = O_\lambda(g(n))$ when $f(n) = O(g(n))$ and the implied constant depends on the parameter $\lambda$.

**Proposition 2.3.** *Let* $\lambda < 1$. *An* $(n, d, \lambda)$*-graph* $\mathcal{G}$ *has diameter* $O_\lambda(\log n)$.

*Proof.* Let $v, w \in V(\mathcal{G})$ and consider the random walk on $\mathcal{G}$ starting from the distribution $\mathbf{1}_v$ that has $v^{\text{th}}$ entry 1. Then, by (2) we have that $\|\mathbf{A}^t\mathbf{1}_v - \mathbf{u}\|_1 \le 1/(2n)$ for $t = -\log(2n^{3/2})/\log\lambda$. This means the distribution $\mathbf{A}^t\mathbf{1}_v$ has $w^{\text{th}}$ component non-zero and so there is a way of reaching $w$ from $v$ within $t = O_\lambda(\log n)$ steps. Since $v$ and $w$ were arbitrary this implies that $\mathcal{G}$ has diameter at most $t$. □

We will see in Section 3.3.2 that this property does not characterize expanders, i.e., small (logarithmic) diameter does not imply small $\lambda(\mathcal{G})$.

2.3. **Constant-degree expanders.** It is a surprising fact that there exist *constant-degree* expander families. We conclude this section by giving a brief overview of some of the important results in this well-developed area of research; for comprehensive surveys, see [20, 27, 28].

The discovery of constant-degree expanders is often attributed to Pinsker [34] in 1973 even though Lubotzky [28] cites an earlier example. These early articles use an alternative combinatorial characterization of expanders that we introduce next.

Define the *edge expansion ratio* $h(\mathcal{G})$ of a $d$-regular graph $\mathcal{G}$ as

$$h(\mathcal{G}) = \min_{S\subseteq V:|S|\leq n/2} \frac{e(S,\overline{S})}{d|S|}, \tag{4}$$

where $e(S,\overline{S})$ is the number of edges from $S \subseteq V$ to its complement. The edge expansion ratio is a *combinatorial* measure for the connectivity of a graph: $h(\mathcal{G})$ is large if every not-too-large set of vertices has relatively many outgoing edges. This parameter is related to the second largest eigenvalue $\lambda(\mathcal{G})$ (an *algebraic* measure of connectivity) via the following theorem.

**Theorem 2.4** (Cheeger inequality, e.g. [12, 13]). *If $\mathcal{G}$ is a regular graph, then*

$$\frac{1-\lambda(\mathcal{G})}{2} \leq h(\mathcal{G}) \leq \sqrt{2(1-\lambda(\mathcal{G}))}\ . \tag{5}$$

In particular, $\{\mathcal{G}_i\}_{i\in\mathbb{N}}$ is a family of expander graphs iff $h(\mathcal{G}_i) \geq \epsilon$ for some $\epsilon > 0$ and all $i \in \mathbb{N}$.

The computational problem of deciding whether $h(\mathcal{G}) \geq q$ for given $\mathcal{G}$ and $q \in \mathbb{Q}$ has been shown coNP-complete [8]. Thus, Theorem 2.4 is useful in providing an approximation algorithm for estimating $h(\mathcal{G})$, since the algebraic parameter $\lambda(\mathcal{G})$ can be computed efficiently using Gaussian elimination. Despite the complexity of determining $h(\mathcal{G})$ it is the most convenient parameter when proving the existence of constant-degree expanders using probabilistic methods (e.g. [27, p. 5]).

The probabilistic existence proofs are often considered relatively easy, while optimizing the parameters in such results requires hard work. As an example of the current state-of-the-art in the study of expansion in random $d$-regular graphs we quote a representative result from the recent monograph of Friedman [17]: For $d$ an even number let $\mathcal{G}_{n,d}$ denote the random $d$-regular graph on the vertex set $[n]$ obtained by choosing $d/2$ random permutations $\pi_i : [n] \to [n]$, $i = 1, \dots, d/2$, and having an edge $uv$ present whenever $\pi_i(u) = v$ (counting multiplicities) with the convention that every fixed point $v = \pi_i(v)$ contributes two self-loops to $v$. We have

**Theorem 2.5** (Friedman [17]). *Fix $d$ and $\epsilon > 0$. Then,*

$$\lambda(\mathcal{G}_{n,d}) \leq 2\sqrt{d-1}/d + \epsilon \tag{6}$$

*with probability tending to 1 as $n \to \infty$.*

This result means that (suitably defined) random regular graphs are nearly optimal expanders as (6) almost matches the following classical lower bound of Alon and Boppana.

**Theorem 2.6** (Alon–Boppana lower bound, e.g. [20, 5.2]). *Let $\{\mathcal{G}_i\}_{i\in\mathbb{N}}$ be a family of expander graphs of constant degree $d$. Then*

$$\liminf_{i\to\infty} \lambda(\mathcal{G}_i) \geq 2\sqrt{d-1}/d\ . \tag{7}$$

Graphs having $\lambda(\mathcal{G}) \leq 2\sqrt{d-1}/d$ are called *Ramanujan graphs.* Remarkably, explicit constructions for such graphs exist: Davidoff et al. [13] give a book-length treatment of a class of constant-degree Ramanujan families that are obtained as Cayley graphs of families of finite (simple) groups. This is the group theoretic setting we begin studying in the next section.

## 3. Cayley Expanders

Virtually all explicit constructions of expander graphs that have been proposed have connections to group theory. In fact, many constructions use Cayley graphs of finite groups. These will be the fundamental objects of our study.

In this section we formalize our research questions and develop the tools and techniques by which we can begin answering these questions.

### 3.1. **Cayley graphs.**

Our purpose in this dissertation is to study which groups $G$ can be turned into low-degree expanders by choosing a generating set $S \subseteq G$ and constructing the associated *Cayley graph*:

**Definition 3.1.** The *Cayley graph* $\mathcal{C}(G;S)$ of $G$ with respect to a multiset $S \subseteq G$ is a graph on $G$ where two elements $g, h \in G$ are joined by an edge iff $gs = h$ for some $s \in S$ (counting multiplicities).

In accordance with our previous convention we require $S$ to be *symmetric*, $S = S^{-1}$ (i.e., $s$ and $s^{-1}$ have the same multiplicities in $S$), in order that $\mathcal{C}(G;S)$ can be considered undirected.

We note that the graph $\mathcal{C}(G;S)$ is $d$-regular for $d = |S|$, and it is connected if and only if $S$ generates $G$. Every occurrence of the identity in $S$ contributes a self-loop to each vertex. Moreover, the $k^{\text{th}}$ power of a Cayley graph is a Cayley graph as well. Namely, $\mathcal{C}(G;S)^k = \mathcal{C}(G;S^{(k)})$, where $S^{(k)} = \{s_1 s_2 \cdots s_k \in G : s_i \in S\}$ is the set of all words of length $k$ in the elements of $S$.

Some simple examples of Cayley graphs are given in Figure 1.

Cayley graphs form a very special class of graphs. For instance, if $g, h \in G$ are two vertices of $\mathcal{C}(G;S)$ the mapping $x \mapsto hg^{-1}x$ is an automorphism of $\mathcal{C}(G;S)$ that maps $g$ to $h$, i.e., $\mathcal{C}(G;S)$ is a *vertex-transitive* graph. Furthermore, the fact that Cayley graphs are constructed out of algebraic objects makes, e.g., their spectral analysis tractable (Section 3.2).

Write $\lambda(G;S) = \lambda(\mathcal{C}(G;S))$ for short. We say that $S \subseteq G$ is $\lambda$-*expanding* if $\lambda(G;S) \leq \lambda$. The fundamental question is now the following.

**Question 1.** *Given a group $G$ and $\lambda < 1$ how large a generating set $S \subseteq G$ is needed to achieve $\lambda(G;S) \leq \lambda$?*

Or, alternatively:

**Question 2.** *Given a family of groups $\{G_i\}_{i\in\mathbb{N}}$ does there exist* small *generating sets $S_i \subseteq G_i$ so that $\{\mathcal{C}(G_i;S_i)\}_{i\in\mathbb{N}}$ is a family of expander graphs?*

Our focus is on finding *sparse* families of Cayley expanders; each family $\{G_i\}_{i\in\mathbb{N}}$ can be made trivially into *dense* expanders by choosing $S_i = G_i \smallsetminus \{\text{id}\}$ which turns $\mathcal{C}(G_i;S_i)$ into a complete graph on $|G_i|$ vertices (Section 2.1.4).

### 3.2. **Eigenvalues of Cayley graphs.**

In laying out the basic framework for the eigenvalue analysis we assume familiarity with the basic notions from (ordinary) representation theory of finite groups (e.g., [22] is a highly readable introduction). We follow Diaconis & Shahshahani [14, 15], who first applied this framework to study card shuffling, or equivalently, to study random walks on (Cayley graphs of) symmetric groups $\mathfrak{S}_n$.

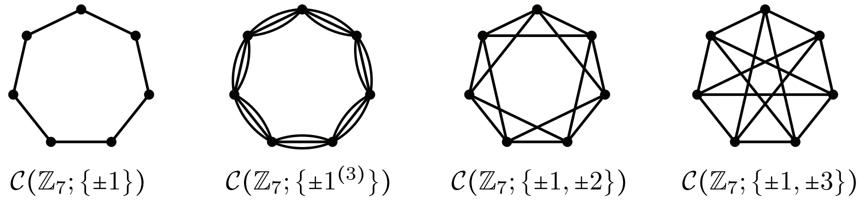

$\mathcal{C}(\mathbb{Z}_7;\{\pm 1\})$
$\mathcal{C}(\mathbb{Z}_7;\{\pm 1^{(3)}\})$
$\mathcal{C}(\mathbb{Z}_7;\{\pm 1,\pm 2\})$
$\mathcal{C}(\mathbb{Z}_7;\{\pm 1,\pm 3\})$

FIGURE 1. Several Cayley graphs for the group $G = \mathbb{Z}_7$.

3.2.1. *Regular representation and the adjacency matrix.* Let $G$ be a finite group and let $\{\mathbf{g}\}_{g\in G}$, $\mathbf{g} = \mathbf{1}_g$, be a basis for the vector space $\mathbb{C}^G$. This space can be equipped with a ring multiplication by letting $\mathbf{g}\cdot\mathbf{h} := \mathbf{gh}$ on the basis elements and extending this linearly to all of $\mathbb{C}^G$. The resulting structure is called the *group algebra* of $G$ and it is usually denoted by $\mathbb{C}[G]$.

The group algebra $\mathbb{C}[G]$ can be considered a right $G$-module in the natural way. The *right regular representation* of $G$ is a homomorphism $\rho_{\mathrm{reg}} : G \to \mathrm{GL}_n(\mathbb{C})$, $n = |G|$, corresponding to this module action, i.e., $\rho_{\mathrm{reg}}$ associates with $h \in G$ the permutation matrix (w.r.t. the basis $\{\mathbf{g}\}_{g\in G}$) corresponding to the action of $h$ on $\mathbb{C}[G]$: $\rho_{\mathrm{reg}}(h)$ maps a basis vector $\mathbf{g}$ to $\mathbf{g}\rho_{\mathrm{reg}}(h) = \mathbf{gh}$. Now, with this notation, the key observation is that the adjacency matrix of a Cayley graph $\mathcal{C}(G;S)$ is given by (the transpose of) $\sum_{s\in S}\rho_{\mathrm{reg}}(s)$. Normalizing by the degree of regularity we get the random walk matrix:

$$\mathbf{A} = \frac{1}{|S|}\sum_{s\in S}\rho_{\mathrm{reg}}(s) \ .$$

This invites the use of representation theory to analyze the eigenvalues of $\mathbf{A}$ via decomposing the regular representation.

3.2.2. *Fourier transform.* Let $\rho_i : G \to \mathrm{GL}_{d_i}(\mathbb{C})$, $i = 1,\dots,r$, be the irreducible representations of $G$ with the convention that $\rho_1$ is the trivial representation ($\rho_1(g) = \mathbf{I}_1 = 1 \in \mathbb{C}$, for all $g \in G$) of dimension $d_1 = 1$. Recall that the regular representation $\rho_{\mathrm{reg}}$ decomposes as

$$\rho_{\mathrm{reg}} \simeq \bigoplus_{i=1}^{r} d_i\rho_i = \bigoplus_{i=1}^{r}\bigoplus_{j=1}^{d_i}\rho_i \ .$$

From a computational perspective this means that there exists a *change of basis* $\mathfrak{F} \in \mathrm{GL}_n(\mathbb{C})$—sometimes called the *Fourier transform*—such that $\mathfrak{F}\rho_{\mathrm{reg}}\mathfrak{F}^{-1}$ has the following *block diagonal* form

$$\mathfrak{F}\rho_{\mathrm{reg}}(g)\mathfrak{F}^{-1} = \begin{bmatrix} \rho_1(g) & & \\ & \ddots & \\ & & \rho_r(g) \end{bmatrix} \qquad \text{for all } g \in G \ , \tag{8}$$

where each irreducible $\rho_i$ appears $d_i$ times on the diagonal. For notational convenience we extend linearly all homomorphisms $G \to \mathrm{GL}_d(\mathbb{C})$ to $\mathbb{C}$-algebra homomorphisms $\mathbb{C}[G] \to \mathrm{Mat}_d(\mathbb{C})$, where $\mathrm{Mat}_d(\mathbb{C})$ is the $\mathbb{C}$-algebra of all $d\times d$ matrices. This allows us to write, for $\boldsymbol{\delta} = 1/|S|\cdot\sum_{s\in S}\mathbf{s} \in \mathbb{C}[G]$,

$$\mathbf{A} = \rho_{\mathrm{reg}}(\boldsymbol{\delta}) \qquad \text{and} \qquad \mathfrak{F}\mathbf{A}\mathfrak{F}^{-1} = \bigoplus_{i=1}^{r} d_i\rho_i(\boldsymbol{\delta}) \ .$$

Since similar matrices have the same eigenvalues, we can determine the eigenvalues of $\mathbf{A}$ from $\mathfrak{F}\mathbf{A}\mathfrak{F}^{-1}$ that is more close to being diagonal. Indeed, by the block diagonal form (8) the eigenvalues are precisely those of the blocks:

$$\mathrm{spec}(\mathcal{C}(G;S)) = \mathrm{spec}(\mathbf{A}) = \mathrm{spec}(\mathfrak{F}\mathbf{A}\mathfrak{F}^{-1}) = \bigcup_{i=1}^{r}\mathrm{spec}(\rho_i(\boldsymbol{\delta})) \ . \tag{9}$$

In particular, the block of the trivial representation at $\boldsymbol{\delta}$, $\rho_1(\boldsymbol{\delta}) = \mathbf{I}_1 = 1$, always corresponds to the trivial eigenvalue $\lambda_1 = 1$ and so

$$\lambda(G;S) = \max_{i \neq 1} \|\rho_i(\boldsymbol{\delta})\|_2 \ .$$

Furthermore, if we denote by $m(\lambda, \rho)$ the multiplicity of an eigenvalue $\lambda$ in $\rho(\boldsymbol{\delta})$, then $m(\lambda, \rho_{\text{reg}}) = \sum_{i=1}^{r} d_i m(\lambda, \rho_i)$ (again, from (8)).

Hence, the task of finding the spectrum of $\mathcal{C}(G;S)$ reduces to understanding the eigenvalues of the irreducible representations of $G$.

3.2.3. *Conjugacy classes.* An interesting special case to consider is when the generating set is a union of conjugacy classes: $S = \bigcup_{j=1}^{t} \mathscr{C}_j$. Here, $\boldsymbol{\delta} = 1/|S| \sum_j \mathbf{c}_j$, where $\mathbf{c}_j = \sum_{g \in \mathscr{C}_j} \mathbf{g}$ are *class sums.* We recall that the $\mathbf{c}_j$ are central in $\mathbb{C}[G]$ (indeed, $Z(\mathbb{C}[G])$ is spanned by the class sums) and so $\boldsymbol{\delta} \in Z(\mathbb{C}[G])$. Using this, for $\mathbf{A}_i = \rho_i(\boldsymbol{\delta})$, we have

$$\mathbf{A}_i \rho_i(g) = \rho_i(g) \mathbf{A}_i \quad \text{for all } g \in G \ .$$

This condition gives, by Schur's lemma, that $\mathbf{A}_i = \rho_i(\boldsymbol{\delta})$ is a multiple of the identity: $\lambda \mathbf{I}_{d_i}$. That is, $\rho_i(\boldsymbol{\delta})$ has only a single eigenvalue $\lambda$ (with multiplicity $d_i$), $\text{spec}(\rho(\boldsymbol{\delta})) = \{\lambda^{(d_i)}\}$, and $\mathbf{A}$ is fully diagonalized by $\mathfrak{F}$ in (8). Expressing $\lambda$ in terms of the character $\chi_i$ of $\rho_i$ we have

$$\lambda = \frac{1}{d_i} \text{Trace}(\rho_i(\boldsymbol{\delta})) = \frac{1}{|S| d_i} \sum_{j=1}^{t} |\mathscr{C}_j| \chi_i(x_j) \ , \tag{10}$$

where $x_j$ is some representative from $\mathscr{C}_j$. In the light of (9) and (10), we conclude that in case $S$ is a union of conjugacy classes the eigenvalues of $\mathcal{C}(G;S)$ are given by the information in the character table of $G$.

Finally, we note that in case $G$ is abelian, all the $n = |G|$ irreducible representations are one-dimensional, and the scalars $\rho_i(\boldsymbol{\delta})$ *are* the eigenvalues.

3.3. **Examples.** To illustrate the use of the above tool set in computing eigenvalues we give some easy examples on the groups $\mathbb{Z}_n$, $\mathbb{Z}_2^d$ and $\mathfrak{S}_n$.

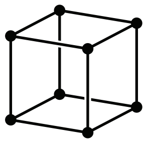
3.3.1. *Cycles.* We determine the eigenvalues of the $n$-cycle $\mathcal{C}_n = \mathcal{C}(\mathbb{Z}_n, \{\pm 1\})$. The characters of $\mathbb{Z}_n = \{0, \ldots, n-1\}$ are indexed by elements $x$ of $\mathbb{Z}_n$ in a natural way: $\chi_x(y) = \rho_x(y) = \omega^{xy}$, for $\omega = e^{2\pi i/n}$ a primitive $n^{\text{th}}$ root of unity. For the generating set $S = \{\pm 1\}$ we get as eigenvalues

$$\lambda_x = \rho_x(\boldsymbol{\delta}) = 1/2(\rho_x(1) + \rho_x(-1)) = 1/2(\omega^x + \omega^{-x}) = \cos(2\pi x/n) \ .$$

If $n$ is even, the cycle $\mathcal{C}_n$ is bipartite, and sure enough, $\lambda_{n/2} = \cos(2\pi(n/2)/n) = -1$ is an eigenvalue of $\mathcal{C}_n$. If $n$ is odd, the second largest eigenvalue is given at $x = \lfloor n/2 \rfloor$ with $\lambda(\mathcal{C}_n) = |\cos(2\pi x/n)| = \cos(\pi/n) = 1 - \Theta(n^{-2})$. The odd cycles do not form an expander family.

3.3.2. *Hypercubes.* Denote by $\mathcal{Q}_d = \mathcal{C}(\mathbb{Z}_2^d, \{\mathbf{e}^1, \ldots, \mathbf{e}^d\})$, the $d$-regular hypercube, where $\mathbf{e}^i$ is the standard basis vector having 1 in $i^{\text{th}}$ coordinate.

$\mathcal{Q}_3$ :



$$\mathbf{M}(\mathcal{Q}_3) = \begin{bmatrix} \mathbf{M}(\mathcal{Q}_2) & \mathbf{I}_4 \\ \mathbf{I}_4 & \mathbf{M}(\mathcal{Q}_2) \end{bmatrix} = \begin{bmatrix} & 1 & 1 & & 1 & & & \\ 1 & & & 1 & & 1 & & \\ 1 & & & 1 & & & 1 & \\ & 1 & 1 & & & & & 1 \\ 1 & & & & & 1 & 1 & \\ & 1 & & & 1 & & & 1 \\ & & 1 & & 1 & & & 1 \\ & & & 1 & & 1 & 1 & \end{bmatrix} .$$

The characters of $\mathbb{Z}_2^d$ are given by $\chi_{\mathbf{x}}(\mathbf{y}) = (-1)^{\mathbf{x}\cdot\mathbf{y}}$, $\mathbf{x},\mathbf{y} \in \mathbb{Z}_2^d$, where $\mathbf{x}\cdot\mathbf{y} = \sum_{i=1}^d x_i y_i$. The eigenvalues for $S = \{\mathbf{e}^1,\dots,\mathbf{e}^d\}$ are then $\lambda_{\mathbf{x}} = \frac{1}{d}\sum_{i=1}^d \chi_{\mathbf{x}}(\mathbf{e}^i) = \frac{1}{d}\sum_{i=1}^d (-1)^{x_i}$, from which we get $\mathrm{spec}(\mathcal{Q}_d) = \{k/d : k \in \{-d,-d+2,\dots,d-2,d\}\}$. In particular, $\lambda_2 = 1 - 2/d$, so even the lazy walk (Remark 2.2) version of $\mathcal{Q}_d$ is not an expander.

An alternative way to see that $\mathcal{Q}_d$ is not an expander is to calculate combinatorial edge expansion (4): the set $T \subseteq V(\mathcal{Q}_d)$ of vectors $\mathbf{x}$ with $x_1 = 0$ has size $|T| = |\mathcal{Q}_d|/2$ but there are only $|T|$ edges going out of $T$, i.e., $h(\mathcal{Q}_d) \le e(T,\overline{T})/(d|T|) \to 0$ as $d \to \infty$.

Note that the diameter $d = \log_2 n$ of $\mathcal{Q}_d$ is still logarithmic as is necessary in expander families (Proposition 2.3).

3.3.3. *Random transpositions.* In their classic article Diaconis & Shahshahani [15] analyze the speed of convergence of a random walk on $\mathcal{C}(\mathfrak{S}_n;T)$, where $T \subseteq \mathfrak{S}_n$ is the conjugacy class of all transpositions in the symmetric group on $n$ elements. They bound the distribution of *all* eigenvalues of $\mathcal{C}(\mathfrak{S}_n;T)$ by making use of the well-developed theory of symmetric groups and their representations. Here, we are content with calculating the second largest eigenvalue only. No familiarity with the representation theory of symmetric groups is assumed.

The graph $\mathcal{C}(\mathfrak{S}_n;T)$ is bipartite as the division of $\mathfrak{S}_n$ into even and odd permutations forms a bipartition of $\mathcal{C}(\mathfrak{S}_n;T)$. By virtue of Remark 2.2 we focus on bounding the eigenvalue $\lambda_2$.

We consider the natural action of $\mathfrak{S}_n$ on $[n]$. Let $\rho : \mathfrak{S}_n \to \mathrm{GL}_n(\mathbb{C})$ be the corresponding linear representation and $\chi$ its character. Here, $\chi$ is given by $\chi(\pi) = \mathrm{Fix}_{[n]}(\pi)$, where $\mathrm{Fix}_X(\pi)$ counts the number of points in a $\mathfrak{S}_n$-set $X$ fixed by $\pi$. We present the standard character theoretic analysis of $\rho$ (e.g. [10, p. 42]).

The set $[n]^2$ of ordered pairs is split into two orbits under the diagonal action of $\mathfrak{S}_n$: $\{(i,i) \in [n]^2 : i \in [n]\}$ and $\{(i,j) \in [n]^2 : i \ne j\}$. This number of $\mathfrak{S}_n$-orbits can be calculated alternatively by employing (not-)Burnside's lemma:

$$2 = \frac{1}{|\mathfrak{S}_n|}\sum_{\pi\in\mathfrak{S}_n} \mathrm{Fix}_{[n]^2}(\pi) \ .$$

But now, $\mathrm{Fix}_{[n]^2}(\pi) = (\mathrm{Fix}_{[n]}(\pi))^2 = (\chi(\pi))^2$, so that

$$\langle \chi,\chi\rangle_{\mathfrak{S}_n} = \frac{1}{|\mathfrak{S}_n|}\sum_{\pi\in\mathfrak{S}_n} \chi(\pi)\cdot\chi(\pi) = \frac{1}{|\mathfrak{S}_n|}\sum_{\pi\in\mathfrak{S}_n} \mathrm{Fix}_{[n]^2}(\pi) = 2 \ .$$

This means that $\chi$ has exactly 2 irreducible constituents. One of these is the trivial character corresponding to the $\mathfrak{S}_n$-invariant subspace $\mathbb{C}\mathbf{1} \subseteq \mathbb{C}^n$. This leaves a non-trivial $(n-1)$-dimensional character $\chi^\star = \chi - 1$ that gives us a non-trivial eigenvalue of $\mathcal{C}(\mathfrak{S}_n;T)$ via equation (10):

$$\lambda^\star = \frac{1}{n-1}\chi^\star((1,2)) = \frac{1}{n-1}(\mathrm{Fix}_{[n]}((1,2)) - 1) = \frac{n-3}{n-1} \ .$$

Since $\lambda^\star \le \lambda_2$ we have that $\lambda_2 \to 1$ as $n \to \infty$. (In fact, it can be shown that $\lambda^\star = \lambda_2$ [11, p. 345].) Thus, even the lazy walk version of $\mathcal{C}(\mathfrak{S}_n;T)$ is not an expander.

3.4. **Simple groups.** The above examples of Cayley graphs all turned out not to be expanders. It is natural to ask whether some *different* generating sets would turn these groups into low-degree expanders.

Deciding whether symmetric groups (or, equivalently, alternating groups) have constant-size expanding generating sets was an outstanding question for a long while until it was answered in the affirmative by Kassabov [24]. This result is part of a recently completed project by several mathematicians to investigate expansion in simple groups. We mention without proof their major achievement.

**Theorem 3.1** ([9, 25] etc.)**.** *The family of non-abelian finite simple groups can be made into constant-degree Cayley expanders.*

Our goal in the rest of this dissertation is somewhat orthogonal to this result: we try to quantify the extent to which Theorem 3.1 *fails* for specific families of soluble groups.

## 4. Abelian Groups

In order to understand expansion properties of soluble groups it is fundamental to thoroughly understand expansion in abelian groups first. In this section we argue (following [3, 41, 46]) that to generate an abelian group as an expander it is necessary and sufficient to have $\Theta(\log n)$ generators. In addition, we develop a deterministic algorithm for finding expanding generating sets of this size.

4.1. **A lower bound.** To generate the abelian group $\mathbb{Z}_2^d$ one needs $d = \log_2 n$ generators; there is no hope of generating the groups $\{\mathbb{Z}_2^d\}_{d\in\mathbb{N}}$ as constant-degree expanders. More generally, the following result shows that *every* abelian group needs at least this many generators for the second largest eigenvalue to be bounded away from 1.

**Theorem 4.1** ([3, 29])**.** *Let $G$ be an abelian group and $S \subseteq G$ a $\lambda$-expanding generating set for $\lambda < 1$. Then $|S| = \Omega_\lambda(\log n)$.*

*Proof.* We fill in the details of the combinatorial proof given in [3]. Alternatively, Lubotzky & Weiss [29] give a proof using characters of abelian groups.

Write $S = S^{-1} = \{s_1, \dots, s_d\}$. Proposition 2.3 tells us that $\mathcal{C}(G;S)$ has logarithmic diameter, say $D \leq C_\lambda \log n$. Since $G$ is abelian, every element in $G$ can be written as

$$s_1^{\alpha_1} \cdot s_2^{\alpha_2} \cdots s_d^{\alpha_d} \quad , \quad \text{where } \alpha_i \in \mathbb{N} \text{ and } \sum_{i=1}^d \alpha_i \leq D \quad .$$

Or, equivalently, as $\mathrm{id}^{\alpha_0}\, s_1^{\alpha_1} \cdots s_d^{\alpha_d}$ with $\sum_{i=0}^d \alpha_i = D$. There are $\binom{D+d}{d}$ ways of choosing the numbers $\alpha_i$. This gives, using $\binom{a}{b} \leq (ea/b)^b$,

$$n = |G| \leq \binom{D+d}{d} \leq (e(D+d)/d)^d \quad .$$

Taking logarithms:

$$\log n \leq d \log(e(D+d)/d) \leq d \log(e(C_\lambda \log n + d)/d) \quad .$$

Writing $\omega(n) = \log n / d$ this becomes

$$\omega(n) \leq \log(eC_\lambda \omega(n) + e) \quad ,$$

so we must have that $\omega(n)$ is bounded, i.e., $|S| = d = \Omega_\lambda(\log n)$. □

4.2. **A randomized upper bound.** The most fundamental existence proof for expanding generating sets for general groups was given by Alon & Roichman in 1997. We measure any attempt to find expanding generating sets for general groups against this theorem.

**Theorem 4.2** (Alon & Roichman [3])**.** *Fix $\lambda > 0$. Every finite group $G$ has a generating set $S \subseteq G$ of size $|S| = O_\lambda(\log n)$ such that $\lambda(G;S) \leq \lambda$. Moreover, a random subset of this size has the $\lambda$-expanding property with high probability.*

Alternative proofs and extensions for this theorem have been subsequently given by several authors [6, 26, 33, 41, 46].

Here, we give a proof of Theorem 4.2 only for abelian groups. This special case admits a relatively simple proof that only uses elementary probabilistic methods. We assume some familiarity with these basic methods; the text books [4, 31, 32] give nice (algorithmic) introductions.

We will make use of the following Chernoff-type estimate that bounds the tail distribution of a sum of independent random variables. Even though the proof of this lemma is standard, we include a sketch of it here because we will need some of the details later in Section 4.4.2.

**Lemma 4.3** ([4, p. 313])**.** *Let $X_i$, $i = 1,\dots,d$, be independent random variables with $X_i \in [-1,1] \subseteq \mathbb{R}$ and $\mathbf{E}[X_i] = 0$. Then, for $a > 0$,*

$$\mathbf{Pr}[\sum_{i=1}^{d} X_i > a] \leq e^{-a^2/(2d)} \ . \tag{11}$$

*Proof.* Let $\theta > 0$ be arbitrary. Then,

$$\mathbf{Pr}[\sum_{i=1}^{d} X_i > a] = \mathbf{Pr}[e^{\theta \sum_i X_i} > e^{\theta a}] \leq \frac{\mathbf{E}[e^{\theta \sum_i X_i}]}{e^{\theta a}} = e^{-\theta a} \prod_{i=1}^{d} \mathbf{E}[e^{\theta X_i}] \ , \tag{12}$$

where we first use Markov's inequality and then the independence of the $X_i$. The claim now follows from the estimate

$$\mathbf{E}[e^{\theta X_i}] \leq e^{\theta^2/2} \tag{13}$$

and setting $\theta = a/d$. □

The estimate (13) will be used extensively later.

In [41, 46] Theorem 4.2 is proved by using a more general Chernoff bound for *matrix-valued* random variables due to Ahlswede & Winter [1]. For our purposes Lemma 4.3 suffices; our proof is essentially a simplified version of [41, 46] with this modification.

*Proof of Theorem 4.2 for abelian groups.* Let $G$ be an abelian group and fix $\lambda > 0$. We pick a (multi-)set $S \subseteq G$ of size $|S| = 2d$ by first choosing $d$ elements $s_1,\dots,s_d$ uniformly and independently at random from $G$ and then setting $S = \{s_1, s_1^{-1}, \dots, s_d, s_d^{-1}\}$ (counting multiplicities). We proceed to show that for a suitable $d = O_\lambda(\log n)$ we have $\lambda(G;S) \leq \lambda$ with high probability.

The group $G$, being abelian, has $n-1$ non-trivial irreducible representations—or characters; fix one such character $\chi = \rho$. This determines the non-trivial eigenvalue

$$\lambda_\chi = \frac{1}{|S|} \sum_{s \in S} \chi(s) = \frac{1}{d} \sum_{i=1}^{d} \frac{\chi(s_i) + \chi(s_i^{-1})}{2} = \frac{1}{d} \sum_{i=1}^{d} X_i \ , \tag{14}$$

where we set $X_i = (\chi(s_i) + \chi(s_i^{-1}))/2 \in [-1,1]$, $i = 1,\dots,d$. We argue that the $X_i$ satisfy the conditions of Lemma 4.3: Firstly, $\rho_{\text{reg}}(\sum_{g \in G} \mathbf{g}) = \sum_{g \in G} \rho_{\text{reg}}(\mathbf{g}) = \mathbf{J}$ is the all-ones matrix (from Section 2.1.4) and has spectrum $\{n, 0^{(n-1)}\}$. The 0-eigenvalues correspond to Fourier transforms at non-trivial irreducible representations and hence $0 = \rho(\sum_g \mathbf{g}) = \sum_g \chi(g)$ yielding

$$\mathbf{E}[\chi(s_i)] = \sum_{g \in G} \chi(g) \mathbf{Pr}[s_i = g] = \frac{1}{n} \sum_{g \in G} \chi(g) = 0 \ .$$

This gives $\mathbf{E}[X_i] = 0$ by linearity of expectation. Also, the $X_i$ are independent as the $s_i$ are.

We are now ready to bound the failure probability that the eigenvalue $\lambda_\chi$ exceeds (in absolute value) the threshold $\lambda$. Applying Lemma 4.3:

$$\mathbf{Pr}[\lambda_\chi > \lambda] = \mathbf{Pr}[\frac{1}{d}\sum_{i=1}^{d} X_i > \lambda] = \mathbf{Pr}[\sum_{i=1}^{d} X_i > d\lambda] \leq e^{-(d\lambda)^2/(2d)} = e^{-d\lambda^2/2} \quad . \tag{15}$$

Similarly, $\mathbf{Pr}[\lambda_\chi < -\lambda] \leq e^{-d\lambda^2/2}$, so by the union bound

$$\mathbf{Pr}[|\lambda_\chi| > \lambda] \leq 2e^{-d\lambda^2/2} \quad .$$

Choosing $d = \lceil 2\log(2n^2)/\lambda^2 \rceil = O_\lambda(\log n)$ this probability is at most $1/n^2$. To conclude the proof, we apply the union bound over all $n-1$ non-trivial characters to see that the probability that $S$ fails to achieve $\lambda(G;S) \leq \lambda$ is at most $(n-1)/n^2 \leq 1/n = o(1)$. That is, we succeed with high probability. □

*Remark* 4.1. The generating set $S$ in Theorem 4.2 can be taken to have only elements with multiplicity 1, since a random sample of $O(\log n)$ elements has no collisions with high probability. Indeed, the expected number of collisions is $O((\log n)^2/n) = o(1)$.

4.3. **A randomized algorithm.** Theorem 4.2 immediately gives an efficient randomized algorithm for finding $O_\lambda(\log n)$-sized $\lambda$-expanding generating sets for abelian groups that runs in time polynomial in $n = |G|$: Given (the multiplication table of) $G$ and $\lambda < 1$ we can simply choose $d$ elements $s_1, \ldots, s_d$ at random from $G$ and verify that $\lambda(G;S) \leq \lambda$ using the formulas (14). This algorithm fails with probability at most $1/n$ when $d$ is chosen as in the proof. As is true of all randomized algorithms, the failure probability can be reduced even further by simply repeating the algorithm $\mathsf{poly}(n)$ times.

Even though this straightforward algorithm is good enough for most practical purposes, there still remains the theoretical question: Can the randomness be eliminated? Does there exist an efficient *deterministic* algorithm for finding small expanding generating sets? Notice that going through all the $n^d = n^{\Omega(\log n)}$ different candidates for $S$ does not yield a polynomial time algorithm.

In fact, it is widely believed that all efficient (polynomial-time) randomized algorithms can be *derandomized*, i.e., replaced with an error-free deterministic algorithm solving the same problem—this is the $\mathsf{P} = \mathsf{BPP}$ conjecture [5, Ch. 19 & 20].

4.4. **Derandomizing the upper bound proof.** A full derandomization of Theorem 4.2 was given by Wigderson & Xiao [46], who gave *pessimistic estimators* for the Ahlswede–Winter bound [1].

In the present exposition, we will carry out the derandomization in our special case of abelian groups by specializing the proof of Wigderson & Xiao. Again, the resulting treatment is vastly less technical: we only have to derandomize the Chernoff bound of Lemma 4.3 instead of the more general Ahlswede–Winter bound.

We begin by describing a well-known technique called the method of conditional probabilities that underlies many derandomization constructions. We use [32, 35, 46] as reference.

4.4.1. *Method of conditional probabilities.* Our randomized algorithm of Section 4.3 starts by choosing $d$ random elements $s_1, \ldots, s_d$ each $s_i$ assuming one of $n$ different values in $G$. After these initial choices the algorithm proceeds deterministically. The random choices in this computation can be visualized as a random walk down a rooted $n$-ary tree of height $d$ where the element $s_i$ is chosen at the $i^{\text{th}}$ step. That is, each node at the $i^{\text{th}}$ level corresponds to a partial assignment $s_1 = g_1, s_2 = g_2, \ldots, s_i = g_i$ for some $g_1, \ldots, g_i \in G$. Indeed, one can think of the nodes at the $i^{\text{th}}$ level as being labelled with $i$-tuples $(g_1, g_2, \ldots, g_i) \in G^{\times i}$, i.e., formally we have

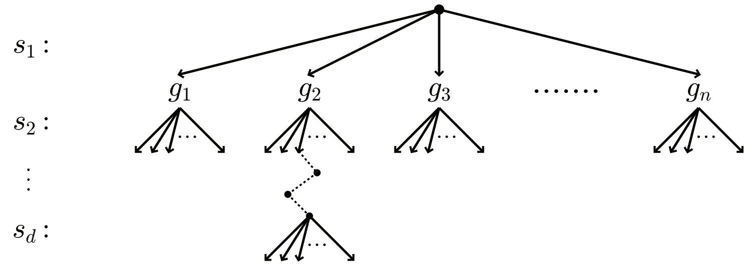


FIGURE 2. The random choice tree $\mathcal{T}$ corresponding to choosing the values $s_1, \ldots, s_d$ from $G = \{g_1, \ldots, g_n\}$.

a rooted tree $\mathcal{T}$ on $G^{\le d} = \bigcup_{i=0}^{d} G^{\times i}$ (see Figure 2). In particular, every leaf node corresponds to a total assignment of all the variables $s_1, \ldots, s_d$.

Denote by $P_i(g_1, \ldots, g_i)$ the conditional probability of failure given that $s_1 = g_1, \ldots, s_i = g_i$, i.e.,

$$P_i(g_1, \ldots, g_i) = \mathbf{Pr}[\lambda(G; S) > \lambda \mid s_1 = g_1, \ldots, s_i = g_i] \ .$$

Then, in particular, $P_0 \le 1/n$ is the original failure probability in the proof of Theorem 4.2 and the leaves have either $P_d(g_1, \ldots, g_d) = 0$ or 1 depending on whether the $g_i$ achieve the required $\lambda$-expansion. Now, the crucial observation is that regardless of the choices $g_1, \ldots, g_i$ made during the first $i$ steps, there is always a way of choosing $g_{i+1} \in G$ at the $(i+1)^{\text{st}}$ step so that $P_i(g_1, \ldots, g_i) \ge P_{i+1}(g_1, \ldots, g_i, g_{i+1})$. This follows from the fact that

$$P_i(g_1, \ldots, g_i) = \sum_{g \in G} P_{i+1}(g_1, \ldots, g_i, g)\mathbf{Pr}[s_{i+1} = g]$$

is a convex combination of the conditional probabilities of the children of $(g_1, \ldots, g_i)$ in $\mathcal{T}$.

The idea for an efficient deterministic algorithm is now the following: Start a deterministic walk down the tree $\mathcal{T}$. After having chosen $s_1 = g_1, \ldots, s_i = g_i$ during the first $i$ steps compute the probabilities $P_{i+1}(g_1, \ldots, g_i, g)$, $g \in G$, of the children and greedily move to the child that minimized this probability. This yields a sequence of conditional probabilities

$$1 > P_0 \ge P_1(g_1) \ge P_2(g_1, g_2) \ge \cdots \ge P_d(g_1, \ldots, g_d) \ , \tag{16}$$

where we must have that $P_d(g_1, \ldots, g_d) = 0$, i.e., the assignment $s_1 = g_1, \ldots, s_d = g_d$ achieves $\lambda$-expansion. In this way we have found a suitable generating set, and done only polynomial amount of work *in case the $P_i$ can be computed efficiently.*

Unfortunately, we do not know of a way to compute the probabilities $P_i$ using only $\mathsf{poly}(n)$ operations: the naive approach to calculate $P_i$ would involve iterating through all the $n^{d-i}$ possible values for the remaining $d-i$ variables that are yet to be fixed. However, if we can find efficiently computable estimates for the $P_i$, we might hope to use them as guides in the deterministic walk. This is made precise by the concept of *pessimistic estimators* initially proposed by Raghavan [35].

4.4.2. *Pessimistic estimators.* Suppose that we could compute upper bounds $U_i$ for the $P_i$ at each node of the tree $\mathcal{T}$. We call the functions $U_i$ *(efficient) pessimistic estimators* for the $P_i$ (or for the event $\lambda(G; S) > \lambda$) if the following hold

(E1) The $U_i$ can be computed in time polynomial in $n$.
(E2) $U_0 < 1$ and $U_i(g_1, \ldots, g_i) \ge P_i(g_1, \ldots, g_i)$ for all $g_1, \ldots, g_i \in G$.
(E3) $U_i(g_1, \ldots, g_i) \ge \mathbf{E}_{s \in_{\mathsf{R}} G}[U_{i+1}(g_1, \ldots, g_i, s)]$, where $s \in_{\mathsf{R}} G$ means that $s$ is chosen uniformly at random from $G$.

Using the above properties it is easy to see that performing the greedy walk on $\mathcal{T}$ with the $U_i$ in place of the $P_i$ we obtain the following sequence of estimates

$$1 > U_0 \geq U_1(g_1) \geq U_2(g_1, g_2) \geq \cdots \geq U_d(g_1, \ldots, g_d) \ ,$$

that dominates the sequence (16). In particular, $P_d(g_1, \ldots, g_d) = 0$, because

$$1 > U_d(g_1, \ldots, g_d) \geq P_d(g_1, \ldots, g_d) \in \{0, 1\} \ .$$

Thus, our probabilistic algorithm can be derandomized by finding efficiently computable pessimistic estimators for the $P_i$. We extract such $U_i$ from the proof of Lemma 4.3.

*Remark* 4.2. For simplicity, our discussion assumes a model of computation with infinite precision arithmetic. We measure the complexity of our algorithms by the number of elementary arithmetic operations and function evaluations that are needed. In an actual implementation these calculations have to be carried out approximatively using polynomially many bits of precision.

The failure event $\lambda(G; S) > \lambda$ can be expressed as $\bigcup_{\mathcal{E} \in \mathscr{E}} \mathcal{E}$ where $\mathscr{E}$ consists of events of type $\mathcal{E}_\chi^+ = \{\lambda_\chi > \lambda\}$ and $\mathcal{E}_\chi^- = \{\lambda_\chi < -\lambda\}$ for $\chi$ non-trivial. We focus on an event $\mathcal{E} = \mathcal{E}_\chi^+$ first. We define

$$U_i^{\mathcal{E}}(g_1, \ldots, g_i) = e^{-d\lambda^2 + (d-i)\lambda^2/2} \prod_{j=1}^{i} e^{\lambda \tilde{\chi}(g_j)} \ , \tag{17}$$

where $\tilde{\chi}(g) = (\chi(g) + \chi(g^{-1}))/2$. The functions $U_i^{\mathcal{E}}$ are efficiently computable, i.e, they satisfy the property (E1). Also, we can repeat calculation (12) conditional on $s_1 = g_1, \ldots, s_i = g_i$ to get

$$\begin{aligned} \mathbf{Pr}[\mathcal{E} \mid s_1 = g_1, \ldots, s_i = g_i] &\leq e^{-d\lambda^2} \prod_{j=1}^{d} \mathbf{E}[e^{\lambda \tilde{\chi}(s_j)} \mid s_1 = g_1, \ldots, s_i = g_i] \qquad (18) \\ &\leq e^{-d\lambda^2} \Big( \prod_{j=1}^{i} e^{\lambda \tilde{\chi}(g_i)} \Big) \Big( \prod_{j=i+1}^{d} e^{\lambda^2/2} \Big) = U_i^{\mathcal{E}}(g_1, \ldots, g_i) \end{aligned}$$

establishing the property (E2). Finally, we have the property (E3):

$$\begin{aligned} \mathbf{E}_{s \in_{\mathrm{R}} G}[U_{i+1}^{\mathcal{E}}(g_1, \ldots, g_i, s)] &= e^{-d\lambda^2 + (d-(i+1))\lambda^2/2} \Big( \prod_{j=1}^{i} e^{\lambda \tilde{\chi}(g_i)} \Big) \underbrace{\mathbf{E}_{s \in_{\mathrm{R}} G}[e^{\lambda \tilde{\chi}(s)}]}_{\leq e^{\lambda^2/2}} \\ &\leq U_i^{\mathcal{E}}(g_1, \ldots, g_i) \ . \end{aligned} \tag{19}$$

Hence, we have checked that $U_i^{\mathcal{E}}$ is a pessimistic estimator for the event $\mathcal{E} = \mathcal{E}_\chi^+$. Similarly, pessimistic estimators $U_i^{\mathcal{E}}$ for $\mathcal{E} = \mathcal{E}_\chi^-$ can be defined by choosing suitable signs in (17).

Our final step is to set

$$U_i(g_1, \ldots, g_i) = \sum_{\mathcal{E} \in \mathscr{E}} U_i^{\mathcal{E}}(g_1, \ldots, g_i) \ .$$

This is a pessimistic estimator for the failure event $\lambda(G; S) > \lambda$: (E1): The function $U_i$ is efficiently computable as each of the $|\mathscr{E}| = 2(n-1)$ many functions $U_i^{\mathcal{E}}$ are; (E2): We have that $P_i(g_1, \ldots, g_i) \leq U_i(g_1, \ldots, g_i)$ by (18) and the union bound. Also, our proof of Theorem 4.2 shows, in fact, that $1 > 1/n \geq U_0$; (E3): This follows from the linearity of expectation and (19).

We have proved the following theorem.

**Theorem 4.4.** *Given an abelian group $G$ and $\lambda > 0$ a generating set $S \subseteq G$ of size $|S| = O_\lambda(\log |G|)$ satisfying $\lambda(G; S) \leq \lambda$ can be found in polynomial time.*

TABLE 1. Values of some of the parameters during the search for a 1/2-expanding set of generators for $\mathbb{Z}_{128}$.

| Step $i$ | $s_i$ | $U_i$ | $\lambda(\mathbb{Z}_{128}; S_i)$ |
|---|---|---|---|
| 0 | | 0.924063846074 | 1.0 |
| 1 | 71 | 0.923206260462 | 1.0 |
| 2 | 106 | 0.913637256796 | 0.977867895969 |
| 3 | 24 | 0.909093475502 | 0.876995437899 |
| 4 | 123 | 0.896481823197 | 0.753417436516 |
| 5 | 29 | 0.885628948485 | 0.592231589770 |
| 6 | 16 | 0.866182998902 | 0.514882038890 |
| 7 | 34 | 0.850314949958 | 0.482518574902 |

4.5. **Computed examples.** We made no effort to optimize any of the parameters in the preceding discussion. A few easy computational optimizations suggest themselves:

(1) In the definition of $U_i^{\mathcal{E}}$ replace the terms $e^{\lambda^2/2}$ with the better estimate $\mathbf{E}_{s\in_{\mathsf{R}}G}[e^{\lambda\tilde{\chi}(s)}]$ that is computable in $\mathsf{poly}(n)$ operations.

(2) Instead of using the $d$ specified in the proof, one can compute the smallest $d$ for which the associated pessimistic estimator satisfies $U_0 < 1$.

(3) The search can be terminated early if some $S_i = \{s_1^{\pm1}, \dots, s_i^{\pm1}\}$ with $i < d$ already achieves the required expansion.

We implemented the derandomized algorithm with these optimizations. Some examples of Cayley graphs for small $n$ are given in Figure 3. A single example run of the algorithm in Table 1 illustrates how the parameters $s_i$, $U_i$ and $\lambda(G; S_i)$ evolve during computation.

## 5. QUOTIENTS AND SUBGROUPS

Next, we discuss how an expanding generating set $S \subseteq G$ can be used to construct expanding generating sets for quotients and subgroups of $G$. Main applications of these considerations will be in proving lower bounds for the number of generators needed to make specific families of groups expanding. These results extend the theorem proved in the previous section that expanding generating sets for abelian groups are necessarily logarithmic in the size of the group.

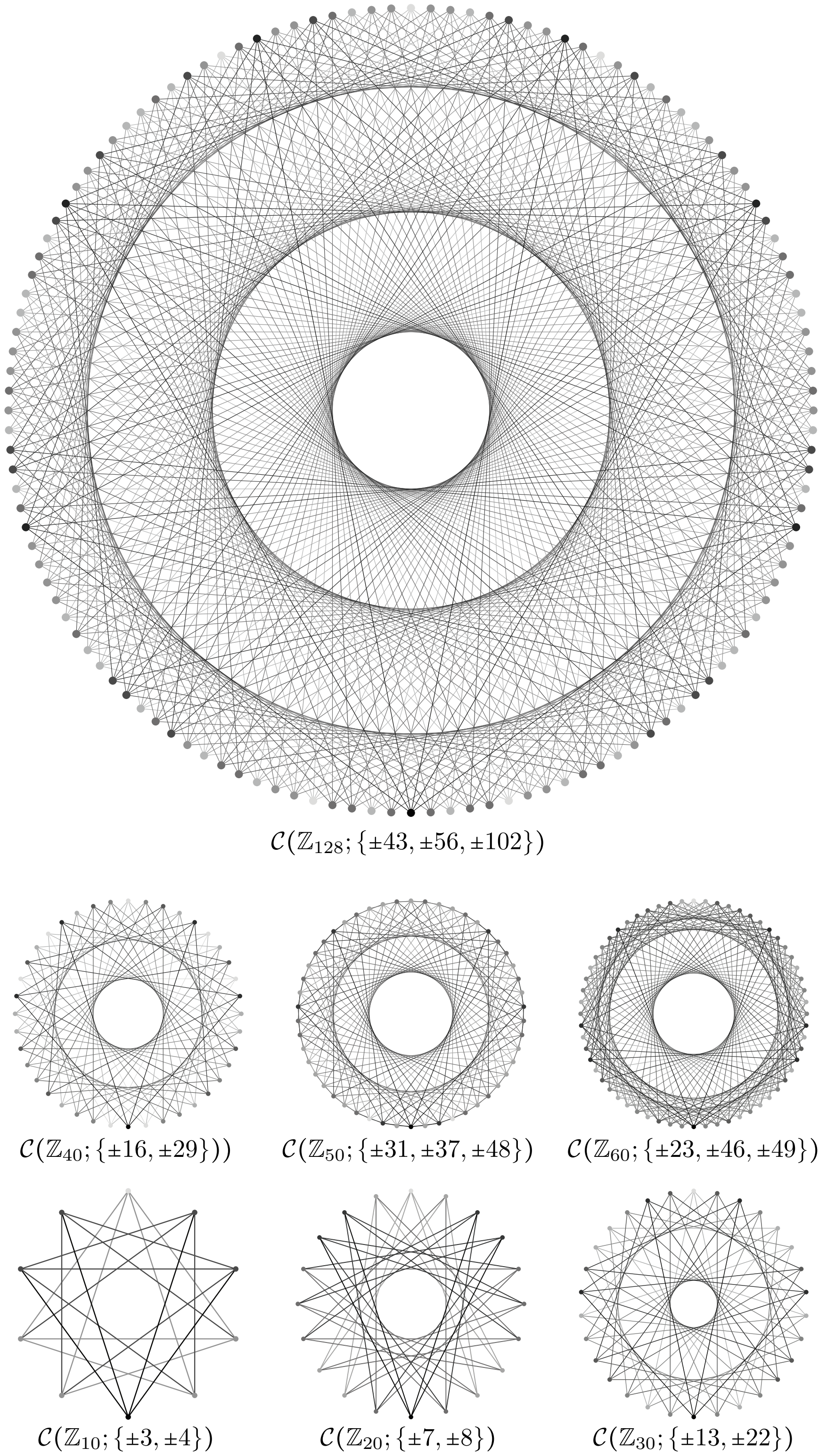
5.1. **Quotients.** If $N \lhd G$ is a normal subgroup and $S \subseteq G$, we think of $S/N = \{sN : s \in S\}$ as a multiset with $|S|$ elements. Expanding generating sets behave well with respect to taking quotients:

**Proposition 5.1** (e.g. [27]). *Let $N \lhd G$ and $S \subseteq G$. Then $\lambda(G/N; S/N) \le \lambda(G; S)$.*

*Proof.* Every irreducible representation $\rho$ of $G/N$ can be lifted to an irreducible representation $\rho \circ \pi$ of $G$, where $\pi : G \to G/N$ is the canonical surjection. Thus,

$$\begin{aligned}\lambda(G/N; S/N) &= \max_{\rho\neq 1} \Big\| \frac{1}{|S/N|} \sum_{sN \in S/N} \rho(sN) \Big\|_2 = \max_{\rho \neq 1} \Big\| \frac{1}{|S|} \sum_{s\in S} (\rho \circ \pi)(s) \Big\|_2 \\ &\le \lambda(G; S) \ . \end{aligned}$$ □

The above observation is widely exploited in constructing constant-degree expanders [13, 20, 27–29]: If an *infinite* group $G$ has a finite generating set $S$ that satisfies a certain *Kazhdan's property (T)*, then $\{\mathcal{C}(G/N; S/N)\}_N$, $[G : N] < \infty$, is a family of constant-degree expanders. The mathematics involved here are outside the scope of the present work. We are mostly interested in constructing expanding



$\mathcal{C}(\mathbb{Z}_{128}; \{\pm 43, \pm 56, \pm 102\})$

$\mathcal{C}(\mathbb{Z}_{40}; \{\pm 16, \pm 29\}))$

$\mathcal{C}(\mathbb{Z}_{50}; \{\pm 31, \pm 37, \pm 48\})$

$\mathcal{C}(\mathbb{Z}_{60}; \{\pm 23, \pm 46, \pm 49\})$

$\mathcal{C}(\mathbb{Z}_{10}; \{\pm 3, \pm 4\})$

$\mathcal{C}(\mathbb{Z}_{20}; \{\pm 7, \pm 8\})$

$\mathcal{C}(\mathbb{Z}_{30}; \{\pm 13, \pm 22\})$

FIGURE 3. Cayley graphs with $\lambda \le 0.9$ produced by the derandomized algorithm. Vertices are shaded based on their distance from the bottom-most vertex.

generating sets in a *bottom-up* way, whereas the quotient map constructs them in a *top-down* fashion.

5.1.1. *Example: Affine groups over $\mathbb{F}_q$.* Let $\mathbb{F} = \mathbb{F}_q$ be the finite field of prime power order $q$ and let $\mathbb{F}^\times$ denote the multiplicative group of $\mathbb{F}$. The *affine group of degree one* over $\mathbb{F}$, denoted by $\mathrm{Aff}(\mathbb{F})$, is the set of all invertible affine transformations

$$x \longmapsto \alpha x + \beta \ : \ \mathbb{F} \to \mathbb{F} \ , \qquad \alpha \in \mathbb{F}^\times, \ \beta \in \mathbb{F} \ ,$$

with the group operation given by function composition. We will show that the group $\mathrm{Aff}(\mathbb{F})$ needs $\Omega_\lambda(\log n)$ generators to achieve $\lambda$-expansion.

The group $\mathrm{Aff}(\mathbb{F})$ is not abelian so the lower bound of Theorem 4.1 does not directly apply. However, $\mathrm{Aff}(\mathbb{F})$ has a *large abelian quotient*: the translations $N = \{x \mapsto x + \alpha : \alpha \in \mathbb{F}\}$ form a normal subgroup of $\mathrm{Aff}(\mathbb{F})$ with $\mathrm{Aff}(\mathbb{F})/N \simeq \mathbb{F}^\times$ abelian. Now, every $\lambda$-expanding generating set $S \subseteq \mathrm{Aff}(\mathbb{F})$ implies (by Proposition 5.1) a $\lambda$-expanding generating set $S/N$ for the abelian group $\mathbb{F}^\times$. By Theorem 4.1 we must have that $|S| = |S/N| = \Omega_\lambda(\log|\mathbb{F}^\times|) = \Omega_\lambda(\log(q-1)) = \Omega_\lambda(\log(q(q-1))) = \Omega_\lambda(\log|\mathrm{Aff}(\mathbb{F})|)$, as required.

Thus, even though $\mathrm{Aff}(\mathbb{F})$ is not abelian, the family $\{\mathrm{Aff}(\mathbb{F}_q)\}_q$ cannot be made into $o(\log n)$-degree expanders—the Alon–Roichman bound (Theorem 4.2) is optimal for $\mathrm{Aff}(\mathbb{F})$.

5.2. **Subgroups.** We start by describing the standard procedure for obtaining a generating set of a subgroup of $G$ from a generating set of $G$.

**Lemma 5.2** (Schreier's subgroup lemma, e.g. [23, Ch. 2])**.** *Let $H \le G$ and $S \subseteq G$ a generating set. Fix a right transversal $R \subseteq G$ of $H$ in $G$ and denote by $\overline{g}$ the representative of $Hg$, i.e., the unique element in $Hg \cap R$. Then $H$ is generated by*

$$S_H \coloneqq \{rs(\overline{rs})^{-1} : s \in S, r \in R\} \subseteq H \ , \qquad |S_H| = [G : H] \cdot |S| \ .$$

*Proof.* Assume for convenience that $1 \in R$ is the representative of $\overline{H}$. Let $h \in H$ and write $h = s_1 s_2 \cdots s_n$ with $s_i \in S$. Making use of the property $\overline{\overline{g}h} = \overline{gh}$ we see that $h$ can be expressed as a product of elements in $S_H$:

$$\begin{aligned} & 1s_1(\overline{\overline{1}s_1})^{-1} \cdot \overline{s_1}s_2(\overline{\overline{s_1}s_2})^{-1} \cdot \overline{s_1s_2}s_3(\overline{\overline{s_1s_2}s_3})^{-1} \cdot \ldots \cdot \overline{s_1\cdots s_{n-1}}s_n(\overline{\overline{s_1\cdots s_{n-1}}s_n})^{-1} \\ = \ & s_1(\overline{s_1})^{-1} \quad \cdot \overline{s_1}s_2(\overline{s_1s_2})^{-1} \cdot \overline{s_1s_2}s_3(\overline{s_1s_2s_3})^{-1} \cdot \ldots \cdot \overline{s_1\cdots s_{n-1}}s_n(\overline{s_1\cdots s_n})^{-1} \\ = \ & s_1\cdots s_n(\overline{s_1\cdots s_n})^{-1} = h(\overline{h})^{-1} = h1^{-1} = h \ . \end{aligned} \qquad \square$$

Schreier's generators $S_H \subseteq H$ preserve expansion in the following sense.

**Theorem 5.3** (Lubotzky & Weiss [29])**.** *Let $H \le G$ and construct $S_H \subseteq H$ from $S \subseteq G$ as in Lemma 5.2. Then $\lambda(H; S_H) \le \lambda(G; S)$.*

Instead of the second largest eigenvalue Lubotzky & Weiss use a different measure of expansion ("Kazhdan's constant") that is not very well suited to our setting where generating sets may not be of bounded size. Nevertheless, their method of proof can be used to establish the above claim, which is—a priori—incomparable in strength to what they state in [29]. In fact, we claim our calculations for bounding the second largest eigenvalue are a bit simpler.

*Proof of Theorem 5.3.* Let $\rho$ be a non-trivial irreducible representation of $H$; we need to show that $\|\rho(\boldsymbol{\delta})\|_2 \le \lambda(G; S)$ for $\boldsymbol{\delta} = \frac{1}{|S||R|}\sum_{r\in R}\sum_{s\in S}\mathbf{rs}(\overline{\mathbf{rs}})^{\mathbf{-1}}$.

In order to relate $\rho$ to representations of $G$ we consider the induced representation $\rho^G = \mathrm{Ind}_H^G\,\rho$. Here, $\rho^G$ does not contain the trivial representation by Frobenius reciprocity and thus

$$\|\rho^G\big(\frac{1}{|S|}\sum_{s\in S}\mathbf{s}\big)\|_2 \le \lambda(G; S) \ .$$

Since $\boldsymbol{\delta} \in \mathbb{C}[H]$, the matrix $\rho^G(\boldsymbol{\delta})$ has a block diagonal form with $|R|$ copies of $\rho(\boldsymbol{\delta})$ on the diagonal. Hence, $\|\rho^G(\boldsymbol{\delta})\|_2 = \|\rho(\boldsymbol{\delta})\|_2$. We calculate

$$\begin{aligned}
\|\rho(\boldsymbol{\delta})\|_2 = \|\rho^G(\boldsymbol{\delta})\|_2 &= \|\rho^G(\tfrac{1}{|S||R|}\sum_{r,s}\mathbf{rs}(\overline{\mathbf{rs}})^{-1})\|_2 \\
&= \|\tfrac{1}{|R|}\sum_r \rho^G(\mathbf{r})\rho^G(\tfrac{1}{|S|}\sum_s \mathbf{s})\rho^G((\overline{\mathbf{rs}})^{-1})\|_2 \\
&\le \tfrac{1}{|R|}\sum_r \|\rho^G(\mathbf{r})\|_2 \cdot \|\rho^G(\tfrac{1}{|S|}\sum_s \mathbf{s})\|_2 \cdot \|\rho^G((\overline{\mathbf{rs}})^{-1})\|_2 \\
&\le \tfrac{1}{|R|}\sum_r 1 \cdot \lambda(G;S) \cdot 1 = \lambda(G;S) \ . \qquad \square
\end{aligned}$$

We note that the set $S_H$ above may not be symmetric. Still, one can always replace $S_H$ with $S_H \cup S_H^{-1}$ (disjoint union of multisets) and only pay a factor of 2 in generating set size. It is easy to see that $S_H \cup S_H^{-1}$ satisfies Theorem 5.3 in place of $S_H$.

5.2.1. *Example: Dihedral groups.* The dihedral group of order $2n$ has presentation

$$D_{2n} = \langle a, b \mid a^n = b^2 = 1, \ \ bab = a^{-1} \rangle \ .$$

Even though dihedral groups are not abelian, they require logarithmically many generators in order to expand: If $S \subseteq D_{2n}$ is an $\lambda$-expanding generating set, Theorem 5.3 implies that the abelian subgroup $H := \langle a \rangle$ of index 2 has a $\lambda$-expanding symmetric generating set $S_H \cup S_H^{-1} \subseteq H$ of size $4|S|$. But, Theorem 4.1 requires $S_H \cup S_H^{-1}$ to be large: $|S_H \cup S_H^{-1}| = \Omega_\lambda(\log |H|)$. Thus, $|S| = \Omega_\lambda(\log |D_{2n}|)$ matching the Alon–Roichman bound (Theorem 4.2).

5.3. **A lower bound for soluble groups.** Denote by $G' = [G, G]$ the derived subgroup of $G$. Recall that $G' \lhd G$, $G/G'$ is abelian, and if $G/H$ is abelian then $G' \le H$. Letting $G^{(0)} = G$ and $G^{(i+1)} = (G^{(i)})'$ the derived series of $G$ is

$$\cdots \lhd G^{(2)} \lhd G^{(1)} \lhd G^{(0)} = G \ ,$$

which terminates at $\{1\}$ precisely when $G$ is soluble. The *derived length* of $G$ is the smallest $l$ such that $G^{(l)} = \{1\}$. In particular, abelian groups can be characterized as groups having derived length $\le 1$.

Meshulam & Wigderson [30] observed that the subgroup and quotient constructions of the previous sections imply a lower bound for the size of an expanding generating set of a soluble group. We let $\log^{(l)}$ denote the $l$ times iterated base-$e$ logarithm.

**Theorem 5.4** ([29, 30])**.** *Let $S \subseteq G$ be $\lambda$-expanding and suppose $G$ has derived length $l$. Then $|S| = \Omega_\lambda(\log^{(l)} |G|)$.*

*Proof.* Meshulam & Wigderson [30] give only a three-line proof sketch. We present the computational details.

To avoid some technical inconveniences we define a function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) := \begin{cases} \log x & \text{if } x \ge 2 \\ \frac{\log 2}{2} \cdot x & \text{if } x \le 2 \end{cases} ,$$

and we prove the theorem for $f$ in place of log. Note that $f$ is Lipschitz continuous with Lipschitz constant 1/2 so that for $y \ge 0$, $f(x+y) \le f(x) + y/2$ and more generally $f^{(i)}(x+y) \le f^{(i)}(x) + y/2^i$.

The subgroup $G^{(i)}$, $0 \le i \le l$, has a $\lambda$-expanding generating set of size $[G : G^{(i)}] \cdot |S|$ (Theorem 5.3) and this generating set can be mapped into the factor group $G^{(i)}/G^{(i+1)}$ while preserving expansion (Proposition 5.1). But this factor group is abelian, so Theorem 4.1 requires that for some constant $C = C_\lambda \geq 1$ we have

$$f([G^{(i)} : G^{(i+1)}]) \leq C[G : G^{(i)}] \cdot |S| \ . \tag{20}$$

To prove the theorem we verify by induction on $i$ the claim that

$$f^{(i)}([G : G^{(i)}]) \leq (2 - 2^{-i})4C|S| \ . \qquad (i^{\text{th}} \text{ claim})$$

The claim holds for $i = 1$ by (20). Suppose it holds for $i \geq 1$; we show it holds for $i + 1$. We calculate

$$\begin{aligned} f([G : G^{(i+1)}]) &= f([G : G^{(i)}]) + f([G^{(i)} : G^{(i+1)}]) \\ &\leq f([G : G^{(i)}]) + C[G : G^{(i)}] \cdot |S| && \text{(using (20))} \\ &\leq [G : G^{(i)}] \cdot 2C|S| \ . \end{aligned}$$

Applying $i$ times the non-decreasing function $f$ we obtain the $(i+1)^{\text{st}}$ claim:

$$\begin{aligned} f^{(i+1)}([G : G^{(i+1)}]) &\leq f^{(i)}([G : G^{(i)}] \cdot 2C|S|) \\ &= f^{(i-1)}(f([G : G^{(i)}]) + f(2C|S|)) \\ &\leq f^{(i)}([G : G^{(i)}]) + 2C|S|/2^i \\ &\leq (2 - 2^{-i})4C|S| + 4C|S|/2^{i+1} = (2 - 2^{-(i+1)})4C|S| \ . \end{aligned}$$ □

**Corollary 5.5.** *If $\mathscr{F}$ is a family of soluble groups of bounded derived length, then $\mathscr{F}$ cannot be made into constant-degree Cayley expanders.*

In their article Meshulam & Wigderson [30] construct expanding generating sets for a family $\{G_i\}_{i\in\mathbb{N}}$ of soluble groups where $G_i$ has derived length $i$. The resulting degree for the family is $O(\log^{(i-\log^* i)}|G_i|)$, where $\log^*$ is the iterated logarithm. This almost matches the above lower bound (see also Remark 8.1). Their construction uses the celebrated zig-zag graph product as a starting point; we describe this graph product in the next section.

## 6. Semi-direct Products

The zig-zag graph product of Reingold, Vadhan & Wigderson [38] was the first purely combinatorial tool to explicitly construct expander graphs. Somewhat ironically Alon, Lubotzky & Wigderson [2] pointed out in later work that the zig-zag product has a natural counterpart in constructing expanding Cayley graphs. Namely, for suitable generating sets, the Cayley graph of a *semi-direct product* of groups admits—roughly—the structure of the zig-zag product of the Cayley graphs of the factors.

The zig-zag construction gives a black-box way of finding small expanding generating sets for semi-direct products provided we have good expanding generating sets for the factors.

Originally Alon et al. [2] reduced the analysis of the semi-direct product to that of the zig-zag product. In this section we give a *direct* proof of the theorem of Alon et al. by means of the representation theory of semi-direct products. We build some intuition by reviewing the original combinatorial construction first, even though our eventual proof makes no use of this.

### 6.1. Replacement and zig-zag products.

Let $\mathcal{G}$ be an $(N, D, \lambda)$-graph on the vertex set $[N]$ and $\mathcal{H}$ an $(D, d, \mu)$-graph on the vertex set $[D]$. We start by describing the replacement product of $\mathcal{G}$ and $\mathcal{H}$, denoted by $\mathcal{G} \circledR \mathcal{H}$.

The replacement product is defined with respect to an *two-way* edge labelling of $\mathcal{G}$ that orders the $D$ outgoing edges on each vertex: each edge carries two numbers, one for each endpoint. Formally, such a labelling is given by a *rotation map* $\mathrm{Rot}_{\mathcal{G}}$ :

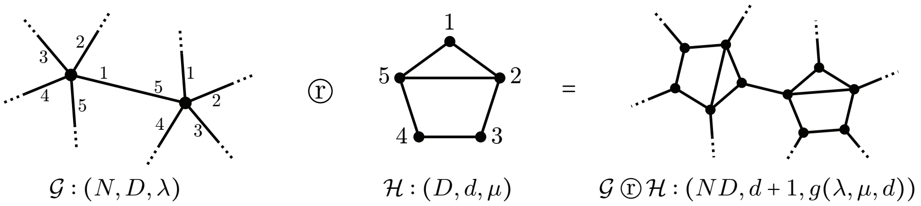

1
2
3
4
5
1
2
3
4
5
ⓡ
1
2
3
4
5
=
$\mathcal{G} : (N, D, \lambda)$
$\mathcal{H} : (D, d, \mu)$
$\mathcal{G} \textcircled{r} \mathcal{H} : (ND, d+1, g(\lambda, \mu, d))$

FIGURE 4. The replacement product of graphs $\mathcal{G}$ and $\mathcal{H}$. Each vertex in $\mathcal{G}$ is blown up into a cloud isomorphic to $\mathcal{H}$.

$[N] \times [D] \to [N] \times [D]$ that maps $(v, i) \mapsto (u, j)$ iff the $i^{\text{th}}$ edge incident to $v$ leads to $u$ and this edge is the $j^{\text{th}}$ edge incident to $u$. Note that $\mathrm{Rot}_{\mathcal{G}}$ is an involution (i.e., $\mathrm{Rot}_{\mathcal{G}}^2 = \mathrm{id}$). We don't usually care which labelling is chosen since the analysis of the replacement product does not depend on the choice.

Now, $\mathcal{G} \textcircled{r} \mathcal{H}$ is the graph on the vertex set $[N] \times [D]$ defined by having $N$ disjoint copies of $\mathcal{H}$ (called the *clouds*), one for each vertex set $\{v\} \times [D]$, $v = 1, \dots, N$, in the natural way; furthermore, the clouds are linked together by having an edge between $(v, i)$ and $(u, j)$ if $\mathrm{Rot}_{\mathcal{G}}(v, i) = (u, j)$. The resulting graph has degree $d + 1$. See Figure 4.

The zig-zag product $\mathcal{G} \textcircled{z} \mathcal{H}$ on $[N] \times [D]$ is now an easily defined variant of $\mathcal{G} \textcircled{r} \mathcal{H}$. There is an edge between $(v, i)$ and $(u, j)$ in $\mathcal{G} \textcircled{z} \mathcal{H}$ iff there are vertices $i', j' \in [D]$ such that $\{i, i'\}, \{j, j'\} \in E(\mathcal{H})$ and $\mathrm{Rot}_G(v, i') = (u, j')$. That is, every edge in $\mathcal{G} \textcircled{z} \mathcal{H}$ corresponds to a *zig-zag* 3-step walk on $\mathcal{G} \textcircled{r} \mathcal{H}$:

(1) Starting at $(v, i)$ take a single step inside the $v^{\text{th}}$ cloud to arrive at $(v, i')$.
(2) Take the *unique* inter-cloud edge at $(v, i')$ arriving at $(u, j') = \mathrm{Rot}_{\mathcal{G}}(v, i')$.
(3) Take a single step inside the $u^{\text{th}}$ cloud to arrive at some $(u, j)$.

Counting the number of such walks we have that $\mathcal{G} \textcircled{z} \mathcal{H}$ is of degree $d \cdot 1 \cdot d = d^2$.

The zig-zag product preserves expansion while reducing the degree:

**Theorem 6.1** (Zig-zag theorem, Reingold et al. [38])**.** *Let $\mathcal{G}$ be an $(N, D, \lambda)$-graph and $\mathcal{H}$ a $(D, d, \mu)$-graph. Then $\mathcal{G} \textcircled{z} \mathcal{H}$ is an $(ND, d^2, f(\lambda, \mu))$-graph where the function $f$ satisfies*

*(1) If $\lambda, \mu < 1$, then $f(\lambda, \mu) < 1$.*
*(2) $f(\lambda, \mu) \leq \lambda + \mu$.*

Usually, we think of a large $D$ and a small $d$. For example, if $\{\mathcal{G}_i\}_{i \in \mathbb{N}}$ is a $D$-regular expander family, then $\{\mathcal{G}_i \textcircled{z} \mathcal{H}\}_{i \in \mathbb{N}}$ is a $d^2$-regular expander family, where possibly $D \gg d^2$.

To illustrate the power of this theorem Reingold et al. give the following relatively simple combinatorial construction of a constant-degree expander family.

*Example* 1. Let $\mathcal{H}$ be a fixed $(d^4, d, 1/5)$-graph; graphs with these parameters abound, e.g., our abelian expanders from Section 4 suffice. Consider the family $\{\mathcal{G}_i\}_{i \in \mathbb{N}}$ defined by $\mathcal{G}_0 = \mathcal{H}^2$ and $\mathcal{G}_{i+1} = \mathcal{G}_i^2 \textcircled{z} \mathcal{H}$. Using Theorem 6.1 together with the fact that $\lambda(\mathcal{G}^2) = \lambda(\mathcal{G})^2$ (Theorem 2.1) it is easy to verify that this is a $d^2$-regular expander family with $\lambda(\mathcal{G}_i) \leq 2/5$ for all $i \in \mathbb{N}$. □

The problem with replicating the above construction in the setting of Cayley graphs is that Cayley graphs in general do not admit the zig-zag structure: if $\mathcal{G} = \mathcal{C}(G; S_G)$ and $\mathcal{H} = \mathcal{C}(H; S_H)$ are Cayley graphs, the graph $\mathcal{G} \textcircled{z} \mathcal{H}$ might not be a Cayley graph of any finite group. However, for suitable generating sets $S_G$ and $S_H$ we can come close.

6.2. **Semi-direct products.** Let $A$ and $H$ be groups with $H$ acting on $A$ on the left (written ${}^h a$, $h \in H$, $a \in A$). To fix notation we define the semi-direct product

$A \rtimes H$ as the set $A \times H$ with the law of composition given by

$$(a,h) \cdot (a',h') = (a({}^{h}a'), hh') \ .$$

We identify the groups $A$ and $H$ with their natural embeddings in $A \rtimes H$. Thus, $A \lhd A \rtimes H$ and ${}^{h}a = hah^{-1} \in A$ for $h \in H$ and $a \in A$. The $H$-orbit of $a \in A$ is denoted by ${}^{H}a$; for convenience we think of ${}^{H}a$ as a multiset of size $|H|$.

To build the connection between the semi-direct product and the zig-zag product we start with the following lemma.

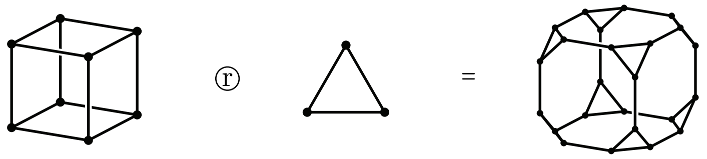
**Lemma 6.2** (Alon et al. [2])**.** *Let $S \subseteq H$ be a generating set and suppose that $A$ is generated by some $H$-orbit ${}^{H}b$ of involutions (i.e., $b^2 = 1$). Then*

$$\mathcal{C}(A \rtimes H; \{b\} \cup S) \simeq \mathcal{C}(A; {}^{H}b) \mathbin{\textcircled{r}} \mathcal{C}(H; S) \ .$$

*Proof.* First, note that the graph $\mathcal{C}(A \rtimes H; \{b\} \cup S)$ has $|A|$ many clouds isomorphic to $\mathcal{C}(H;S)$, one for each coset $aH$, $a \in A$. To establish the claim it remains to argue that the element $b$ generates the inter-cloud edges.

The edges of the $|H|$-regular graph $\mathcal{A} = \mathcal{C}(A; {}^{H}b)$ are naturally labelled by the elements of $H$: both endpoints of the edge $\{a, a({}^{h}b)\}$ have the label $h$. The rotation map for this labelling is $\mathrm{Rot}_{\mathcal{A}} : A \times H \to A \times H$, $\mathrm{Rot}_{\mathcal{A}}(a,h) = (a({}^{h}b), h)$ (indeed, this is a rotation map since $({}^{h}b)^2 = 1$ by assumption). Now, an edge in $\mathcal{C}(A \rtimes H; \{b\} \cup S)$ corresponding to multiplication by $b$ on the right joins a vertex $(a,h)$ to $(a,h)b = (a({}^{h}b), h) = \mathrm{Rot}_{\mathcal{A}}(a,h)$ as required. □

*Example* 2. Consider the vector space $\mathbb{Z}_2^d$ with the standard basis $S = \{\mathbf{e}^1, \dots, \mathbf{e}^d\}$. The cyclic group $\mathbb{Z}_d$ acts on this basis by cyclically permuting the basis elements: ${}^{j}\mathbf{e}^i = \mathbf{e}^{i+j}$ for $i, j \in \mathbb{Z}_d$. This action can be linearly extended to all of $\mathbb{Z}_2^d$. Now, $S = {}^{\mathbb{Z}_d}\mathbf{e}^1$, so the Cayley graph of the semi-direct product $\mathbb{Z}_2^d \rtimes \mathbb{Z}_d$ admits the structure of a replacement product of the graphs $\mathcal{C}(\mathbb{Z}_2^d; S)$ and $\mathcal{C}(\mathbb{Z}_d; \{\pm 1\})$ by choosing $\{\mathbf{e}^1, \pm 1\}$ as the generating set. For instance, when $d = 3$:


ⓡ
=
$\mathcal{C}(\mathbb{Z}_2^3; \{\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3\})$
$\mathcal{C}(\mathbb{Z}_3; \{\pm 1\})$
$\mathcal{C}(\mathbb{Z}_2^3 \rtimes \mathbb{Z}_3; \{\mathbf{e}^1, \pm 1\})$

□

Alternatively, in the setting of Lemma 6.2 we can form the set $SbS \subseteq A \rtimes H$ whose elements correspond precisely to the zig-zag walks on $\mathcal{C}(A; {}^{H}b) \mathbin{\textcircled{r}} \mathcal{C}(H;S)$:

**Corollary 6.3.** *With the assumptions of Lemma 6.2, we have that*

$$\mathcal{C}(A \rtimes H; SbS) \simeq \mathcal{C}(A; {}^{H}b) \mathbin{\textcircled{z}} \mathcal{C}(H; S) \ .$$

More generally, we can choose generating sets for $A$ that are *unions* of $H$-orbits: $\bigcup_{i=1}^{t} {}^{H}a_i$ with $\{a_1, \dots, a_t\}$ symmetric. In this setting, consider the following generating sets for $A \rtimes H$ (the following "wide"-terminology is from [47]) :

(1) $\{a_1, \dots, a_t\} \cup S$ (the "wide" replacement product)
(2) $S\{a_1, \dots, a_t\}S$ (the "wide" zig-zag product)

The first yields a Cayley graph that has $t$ inter-cloud edges incident to every vertex; the second yields the Cayley graph that is the zig-zag version of the first. The original analysis of the zig-zag product in [38] makes it evident that the "wide" zig-zag product also satisfies the eigenvalue bounds of Theorem 6.1.

Next, we proceed to give our representation theoretic proof of the combined content of Theorem 6.1 and the preceding discussion. That is, in short, the above choices of small generating sets preserve expansion.

6.3. **Representation theory of semi-direct products.** Let $H$ act on $A$ as in the previous section and set $G = A \rtimes H$. We will restrict our considerations to the case where $A$ is abelian. Even though this loses some generality, the abelian case is what is usually considered in the literature [2, 19, 30] (with the exception of [39]).

The representation theory of semi-direct products by an abelian group is well-known; we follow the accounts of Serre [42, p. 62] and Etingof et al. [16, p. 76].

The action of $H$ on $A$ induces an action of $H$ on the characters $\chi$ of $A$ (which are 1-dimensional) by setting $({}^{h}\chi)(a) = \chi(h^{-1}ah)$. This action partitions the characters of $A$ into some $s$ orbits. Choose representatives $\chi_i$, $i = 1, \ldots, s$, from each orbit. Fix some $1 \leq i \leq s$. Denote by $H_i \leq H$ the stabilizer of $\chi_i$, i.e., the set of elements $h \in H$ with ${}^{h}\chi_i = \chi_i$. We can extend $\chi_i$ to the subgroup $G_i = AH_i$ by defining $\chi_i(ah) = \chi_i(a)$. This makes $\chi_i$ a 1-dimensional representation of $G_i$, as is checked by the calculation

$$\begin{aligned}\chi_i((ah)(a'h')) &= \chi_i(a({}^{h}a')hh') = \chi_i(a({}^{h}a')) = \chi_i(a)\chi_i({}^{h}a') = \chi_i(a)\chi_i(a') \\ &= \chi_i(ah)\chi_i(a'h') \ .\end{aligned}$$

Let $\rho$ be an irreducible representation of $H_i$. This can be lifted to an irreducible representation of $G_i$; we call it still $\rho$. Finally, we define an induced representation

$$\theta_{i,\rho} = \mathrm{Ind}_{G_i}^{G}(\chi_i \otimes \rho) \ ,$$

where $(\chi_i \otimes \rho)(g)$ is simply $\chi_i(g) \cdot \rho(g)$ as $\chi_i(g)$ is only a scalar.

**Theorem 6.4** ([16, 42])**.** *Let $\chi_i$, $i = 1, \ldots, s$, and $\theta_{i,\rho}$ be as above. Then*

*(i) Each $\theta_{i,\rho}$ is irreducible.*
*(ii) The $\theta_{i,\rho}$ are pairwise inequivalent.*
*(iii) Every irreducible representation of $A \rtimes H$ is equivalent to one of the $\theta_{i,\rho}$.*

*Proof.* Serre's [42] proofs involve the use of Mackey's criterion; we'll avoid it's use by giving a direct argument similar to that in [16].

We'll use the language of modules. Let $M$ be a simple $G_i$-module affording the irreducible representation $\chi_i \otimes \rho$. The induced module affording $\theta_{i,\rho}$ is then

$$M^G = \mathrm{Ind}_{G_i}^{G} M = M \otimes_{\mathbb{C}[G_i]} \mathbb{C}[G] \ .$$

*(i)* Let $h_j \in H$, $j = 1, \ldots, t$ be a transversal for the right cosets of $G_i$ in $G$. As an $A$-module $M^G$ can be *uniquely* decomposed into *homogeneous components*:

$$\mathrm{Res}_{A}^{G} M^G = \bigoplus_{j=1}^{t} M \otimes h_j \ .$$

Here, the action of $a \in A$ on $M \otimes h_j$ is multiplication by the scalar ${}^{h_j}\chi_i(a)$. Hence, for $j_1 \neq j_2$ the components $M \otimes h_{j_1}$ and $M \otimes h_{j_2}$ share no simple $A$-modules.

Let $V \leq M^G$ be a $G$-submodule. The above homogeneous decomposition can be carried out for $\mathrm{Res}_{A}^{G} V$ so that we can write $\mathrm{Res}_{A}^{G} V = \bigoplus_j V_j$ with $V_j \leq M \otimes h_j$. Now $M \otimes h_j$ is a simple $h_j^{-1}G_ih_j$-module (since $M$ is a simple $G_i$-module) so $V_j = 0$ or $V_j = M \otimes h_j$ for all $j$. But $G$ acts transitively on the homogeneous components, so that $V_j = 0$ for all $j$ or $V_j = M \otimes h_j$ for all $j$. That is, $V = 0$ or $V = M^G$.

*(ii)* Suppose we are given an unknown $M^G$; we show how to recover $\chi_i$ and $\rho$. Firstly, the orbit of $\chi_i$ (and thus the representative $\chi_i$ of our choice) is determined by the decomposition of $\mathrm{Res}_{A}^{G} M^G$ into homogeneous components. Secondly, picking out $V = \{v \in M^G : va = v\chi_i(a),\ a \in A\}$ we have $V = M \otimes \mathrm{id} \simeq M$ as $H_i$-modules. This determines $\rho$.

*(iii)* All the irreducible representations are accounted for:

$$\begin{aligned}\sum_{i,\rho}(\dim\theta_{i,\rho})^2 &= \sum_{i,\rho}[H:H_i]^2(\dim\rho)^2 = \sum_i[H:H_i]^2\sum_\rho(\dim\rho)^2\\ &= \sum_i[H:H_i]^2|H_i| = |H|\sum_i[H:H_i] = |H|\cdot|A| = |G| \ . \qquad \square\end{aligned}$$

6.4. **Expanding generating sets for semi-direct products.** We are ready to prove the main theorem of this section, a group theoretic analogue of Theorem 6.1.

The proof is carried out by a direct analysis of the irreducible representations. The combinatorial connection to the zig-zag product is not explicitly used, e.g., we do not make use of rotation maps etc.

**Theorem 6.5** (Wide zig-zag product)**.** *Let $H$ act on an abelian group $A$. Suppose that we have generating sets $S_A \subseteq A$ and $S_H \subseteq H$ with $S_A = \bigcup_{i=1}^t {}^H a_i$. Then*

$$\lambda(A \rtimes H; S_H\{a_1,\dots,a_t\}S_H) \le f(\lambda_A,\lambda_H) \ ,$$

*where $\lambda_A = \lambda(A;S_A)$, $\lambda_H = \lambda(H;S_H)$ and $f$ satisfies*

*(F1) If $\lambda,\mu < 1$, then $f(\lambda,\mu) < 1$.*

*(F2) $f(\lambda,\mu) \le \lambda + 2\mu$.*

*Proof.* Write $S = S_H\{a_1,\dots,a_t\}S_H$ and let $\|\cdot\| = \|\cdot\|_2$ be the $\ell_2$-norm. The Fourier transform of $\boldsymbol{\delta} = 1/|S|\sum_{s\in S}\mathbf{s}$ at an irreducible representation $\theta = \theta_{i,\rho}$ is

$$\theta(\boldsymbol{\delta}) = \Big(\frac{1}{|S_H|}\sum_{h\in S_H}\theta(\mathbf{h})\Big)\Big(\frac{1}{t}\sum_{j=1}^t\theta(\mathbf{a}_j)\Big)\Big(\frac{1}{|S_H|}\sum_{h\in S_H}\theta(\mathbf{h})\Big) =: \mathbf{PDP} \ ,$$

where $\mathbf{P}$ and $\mathbf{D}$ are symmetric matrices with $\|\mathbf{P}\| \le 1$, $\|\mathbf{D}\| \le 1$. We need to show that $\|\theta(\boldsymbol{\delta})\|$ is small when $\theta \ne 1$.

**Case 1:** *Suppose $\rho \ne 1$.* Using Frobenius reciprocity we get that the representation $\rho^H = \mathrm{Ind}_{H_i}^H\rho$ does not contain the trivial representation. But $\mathrm{Res}_H^G\theta = \rho^H$, so that we have

$$\|\mathbf{P}\| = \|\frac{1}{|S_H|}\sum_{h\in S_H}\theta(\mathbf{h})\| = \|\frac{1}{|S_H|}\sum_{h\in S_H}\rho^H(\mathbf{h})\| \le \lambda_H$$

by the definition of $\lambda_H$. Hence $\|\theta(\boldsymbol{\delta})\| \le \lambda_H^2$ and we are done in this case.

**Case 2:** *Suppose $\rho = 1$ and $\chi_i \ne 1$.* Here, $\mathrm{Res}_H^G\theta = \rho^H$ is the $[H:H_i]$-dimensional permutation representation on the cosets of $H_i$. From Frobenius reciprocity (or directly from the transitivity of the action) we see that $\rho^H$ contains the trivial representation only once: the trivial representation corresponds to the subspace spanned by the all-ones vector $\mathbf{1}$. Consequently, $\mathbf{P1} = \mathbf{1}$ and $\lambda(\mathbf{P}) \le \lambda_H$.

On the other hand, $\mathbf{D}$ is a linear combination of images of $\mathrm{Res}_A^G\theta$, which are diagonal matrices having the orbit ${}^{h_1}\chi_i,\dots,{}^{h_r}\chi_i$ of $\chi_i$ on the diagonal. We have

$$\begin{aligned}|\langle\mathbf{1},\mathbf{D1}\rangle| &= \left|\frac{1}{t}\sum_{j=1}^t\sum_{k=1}^r {}^{h_k}\chi_i(a_j)\right| = \left|\frac{1}{t}\sum_{j=1}^t\frac{1}{|H_i|}\sum_{h\in H}{}^h\chi_i(a_j)\right|\\ &= \frac{1}{t|H_i|}\left|\sum_{j=1}^t\sum_{h\in H}\chi_i(h^{-1}a_jh)\right| = \frac{1}{t|H_i|}\left|\sum_{s\in S_A}\chi_i(s)\right|\\ &\le \frac{1}{t|H_i|}\cdot\lambda_A|S_A| = \lambda_A[H:H_i] = \lambda_A\langle\mathbf{1},\mathbf{1}\rangle \ . \qquad (21)\end{aligned}$$

This implies $\|\tilde{\mathbf{J}}\mathbf{D}\tilde{\mathbf{J}}\| \le \lambda_A$ for the normalized all-ones matrix $\tilde{\mathbf{J}}$.

The proof could be now concluded by following the original analysis of Reingold et al. [38] which would give the bound $\|\theta(\boldsymbol{\delta})\| \le f(\lambda_A,\lambda_H)$ for $f$ as in Theorem 6.1. Instead, we adapt the cleaner analysis of [37, 44] that yields slightly worse bounds.

For convenience, set $\epsilon_H = 1 - \lambda_H$ and $\epsilon_A = 1 - \lambda_A$. Using Proposition 2.2 write $\mathbf{P} = \epsilon_H \tilde{\mathbf{J}} + \lambda_H \mathbf{E}$ for some $\mathbf{E}$, $\|\mathbf{E}\| \le 1$. Then

$$\begin{aligned}\theta(\boldsymbol{\delta}) = \mathbf{PDP} &= \left(\epsilon_H \tilde{\mathbf{J}} + \lambda_H \mathbf{E}\right) \mathbf{D} \left(\epsilon_H \tilde{\mathbf{J}} + \lambda_H \mathbf{E}\right) \\ &= \epsilon_H^2 \tilde{\mathbf{J}} \mathbf{D} \tilde{\mathbf{J}} + (1 - \epsilon_H^2) \mathbf{F} \qquad \text{(for some } \mathbf{F},\ \|\mathbf{F}\| \le 1)\end{aligned}$$

Finally,

$$\begin{aligned}\|\theta(\boldsymbol{\delta})\| &\le \epsilon_H^2 \|\tilde{\mathbf{J}} \mathbf{D} \tilde{\mathbf{J}}\| + (1 - \epsilon_H^2) \|\mathbf{F}\| \\ &\le \epsilon_H^2 \lambda_A + (1 - \epsilon_H^2) \\ &= 1 - \epsilon_H^2 (1 - \lambda_A) = 1 - \epsilon_H^2 \epsilon_A \ .\end{aligned}$$

We can choose the final expression as our $f(\lambda_A, \lambda_H)$ satisfying *(F1)* and *(F2)*. □

**Corollary 6.6** (Wide replacement product)**.** *In the notation of Theorem 6.5,*

$$\lambda(A \rtimes H; \{a_1, \dots, a_t\} \cup S_H) \le g(\lambda_A, \lambda_H, \alpha) \ ,$$

*where* $\alpha = t/(t + |S_H|)$ *and* $g(\lambda_A, \lambda_H, \alpha) < 1$ *whenever* $\lambda_A, \lambda_H < 1$.

*Proof.* The proof is by a reduction to the above zig-zag analysis as in [38].

This time, we need to bound the $\ell_2$-norm of the matrix $\theta(\boldsymbol{\delta}) = \alpha \mathbf{D} + (1 - \alpha) \mathbf{P}$ for $\theta$ non-trivial. Note that

$$(\theta(\boldsymbol{\delta}))^3 = (\alpha \mathbf{D} + (1 - \alpha) \mathbf{P})^3 = \alpha (1 - \alpha)^2 \mathbf{PDP} + (1 - \alpha (1 - \alpha)^2) \mathbf{E} \ ,$$

for some $\mathbf{E}$ with $\|\mathbf{E}\| \le 1$. Using the symmetricity of $\theta(\boldsymbol{\delta})$ we have $\|(\theta(\boldsymbol{\delta}))^3\| = \|\theta(\boldsymbol{\delta})\|^3$ so that

$$\|\theta(\boldsymbol{\delta})\| \le \left[\alpha (1 - \alpha)^2 f(\lambda_A, \lambda_H) + (1 - \alpha (1 - \alpha)^2)\right]^{1/3} =: g(\lambda_A, \lambda_H, \alpha) \ . \qquad \square$$

The wide zig-zag generating set $S = S_H \{a_1, \dots, a_t\} S_H$ for $A \rtimes H$ preserves expansion in a very strong sense, in particular,

$$\lambda(A \rtimes H; S) \longrightarrow 0 \qquad \text{whenever} \quad \lambda_A, \lambda_H \longrightarrow 0 \ . \tag{22}$$

The alternative choice $S = \{a_1, \dots, a_t\} \cup S_H$ gives a somewhat smaller generating set, but the limiting behaviour (22) is not guaranteed by Corollary 6.6. The next example shows that this is not an artefact of sloppy analysis.

*Example* 3. Let $S_A \subseteq A$ be a generating set of an abelian group $A$. Denote by $\chi_1, \dots, \chi_n$, $\chi_1 = 1$, the characters of $A$. Consider the abelian group $A \times A$ whose characters are given by $\chi_{i,j}(a, b) = \chi_i(a) \chi_j(b)$, $1 \le i, j \le n$. Choosing $S = S_A \times \{0\} \cup \{0\} \times S_A \subseteq A \times A$ as in Corollary 6.6 we get a non-trivial eigenvalue

$$\lambda_{1,2} = \frac{1}{|S|} \sum_{(a,b) \in S} \chi_{1,2}(a, b) = \frac{1}{|S|} \sum_{a \in S_A} \underbrace{\chi_1(a) \chi_2(0)}_{=1} + \frac{1}{|S|} \sum_{b \in S_A} \chi_1(0) \chi_2(b) \ge \frac{1}{2} \ .$$

Thus, $\lambda(A \times A; S) \ge 1/2$ regardless of how small $\lambda(A; S_A)$ is. □

## 7. Expanding Orbits

In the previous section we saw that the zig-zag and replacement product constructions allow us to generate a semi-direct product $A \rtimes H$ as an expander by picking a single representative $a_i$ from each orbit ${}^H a_i$ which together generate $A$ as an expander. Hence, the problem of finding good generating sets is reduced to finding *expanding orbits* in $A$. The following theorem supplies such orbits in a very general setting.

**Theorem 7.1** ([2, 30])**.** *Let* $\lambda > 0$ *and let* $\mathbb{F} = \mathbb{F}_p$ *be the field of prime order* $p$. *Let* $H$ *act on* $\mathbb{F}^d$ *as an irreducible representation* $\rho : H \to \mathrm{GL}_d(\mathbb{F})$. *Then there exist* $t = t(\lambda, p)$ *vectors* $\mathbf{a}^1, \dots, \mathbf{a}^t \in \mathbb{F}^d$ *such that* $\lambda(\mathbb{F}^d; \bigcup_{i=1}^t {}^H \pm \mathbf{a}^i) \le \lambda$.

Alon et al. [2, Thm 3.1] state this theorem in the form that the union of *two* random orbits forms a $\lambda_p$-expanding generating set with high probability for some $\lambda_p < 1$ depending on $p$ (they only sketch the proof in the case $p = 2$). Meshulam & Wigderson [30, Thm 1.2] prove a generalization of Theorem 7.1 where $\rho$ is the regular representation and $\lambda > 1/2$. The following proof assuming only $\lambda > 0$ is an easy modification of the proofs in [2, 30].

*Proof of Theorem 7.1.* Choose $\mathbf{a}^1, \dots, \mathbf{a}^t \in \mathbb{F}^d$ (for $t$ to be determined later) uniformly and independently at random. The characters of the abelian group $\mathbb{F}^d$ are given by $\chi_{\mathbf{x}}(\mathbf{y}) = \omega^{\mathbf{x}\cdot\mathbf{y}}$, $\mathbf{x}, \mathbf{y} \in \mathbb{F}^d$, where $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^d x_i y_i$ and $\omega = e^{2\pi i/p}$. Fix $\mathbf{x} \in \mathbb{F}^d \smallsetminus \{\mathbf{0}\}$; we want to bound the associated non-trivial eigenvalue

$$\lambda_{\mathbf{x}} = \frac{1}{t|H|} \sum_{i=1}^{t} \sum_{h \in H} \tilde{\chi}_{\mathbf{x}}({}^{h}\mathbf{a}^i) \ , \tag{23}$$

where we write $\tilde{\chi}_{\mathbf{x}}(\mathbf{y}) = (\chi_{\mathbf{x}}(\mathbf{y}) + \chi_{\mathbf{x}}(-\mathbf{y}))/2$ as before.

Because the action $\rho$ is irreducible, we can find a set $R \subseteq H$ of size $|R| = d$ with the property that the $d$ vectors $\{\rho(r)^T \mathbf{x}\}_{r \in R}$ are linearly independent. Note that every translate $hR$ of $R$ also has this property. Since the translates $\{hR\}_{h \in H}$ cover every element of $H$ exactly $d$ times we can write (23) as

$$\lambda_{\mathbf{x}} = \frac{1}{|H|} \sum_{h \in H} \left[ \frac{1}{td} \sum_{i=1}^{t} \sum_{r \in hR} \tilde{\chi}_{\mathbf{x}}({}^{r}\mathbf{a}^i) \right] =: \frac{1}{|H|} \sum_{h \in H} A_h \ .$$

Fix $h \in H$; we show that $|A_h|$ is small with high probability. Note that for any transformation $\mathbf{T} \in \mathrm{GL}_d(\mathbb{F})$ the vector $\mathbf{T}\mathbf{a}^i$ is uniformly distributed in $\mathbb{F}^d$ and moreover the coordinate entries $(\mathbf{T}\mathbf{a}^i)_j$, $j = 1, \dots, d$, are mutually independent random variables. In particular, setting $\mathbf{T} = [\rho(r)^T \mathbf{x}]_{r \in R}$ we have that

$$\chi_{\mathbf{x}}({}^{r}\mathbf{a}^i) = \omega^{\mathbf{x}\cdot(\rho(r)\mathbf{a}^i)} = \omega^{(\mathbf{T}\mathbf{a}^i)_r} \ ,$$

and consequently the random variables defined as $X_{i,r} := \tilde{\chi}_{\mathbf{x}}({}^{r}\mathbf{a}^i)$, $1 \le i \le t$, $r \in hR$, are independent. Now, using the Chernoff bound of Lemma 4.3 for the $X_{i,r}$ we conclude that

$$\mathbf{Pr}[|A_h| > \lambda/2] \le 2e^{-dt\lambda^2/8} =: \epsilon \qquad \text{for all } h \in H \ . \tag{24}$$

Denote by $\mathcal{E}_h = \{|A_h| > \lambda/2\}$ the above bad event and let $Y = \sum_{h \in H} \mathbf{1}_{\mathcal{E}_h}$ count the number of bad events holding. The bound (24) implies that $\mathbf{E}[Y] \le \epsilon|H|$. If $Y \le \frac{\lambda}{2}|H|$ (i.e., $|A_h| > \lambda/2$ holds for at most a $\lambda/2$ fraction of $h \in H$), then

$$|\lambda_{\mathbf{x}}| \le \frac{1}{|H|} \sum_{h \in H} |A_h| \le \lambda/2 \cdot 1 + (1 - \lambda/2) \cdot \lambda/2 \le \lambda \ .$$

But Markov's inequality gives $\mathbf{Pr}[Y > \frac{\lambda}{2}|H|] \le 2\epsilon/\lambda$ so that choosing $t = t(\lambda, p) = \lceil 8/\lambda^2(\log p + \log(4/\lambda)) \rceil$ we get

$$\mathbf{Pr}[|\lambda_{\mathbf{x}}| > \lambda] \le \mathbf{Pr}[Y > \frac{\lambda}{2}|H|] \le p^{-d} \ . \tag{25}$$

Finally, taking the union bound over all $p^d - 1$ many $\mathbf{x} \neq \mathbf{0}$ we have that $\mathbf{Pr}[\exists \mathbf{x} \neq \mathbf{0} : |\lambda_{\mathbf{x}}| > \lambda] < 1$ and we are done. □

7.1. **Derandomizing the search for expanding orbits.** Next, we address the question of whether there exists an efficient deterministic algorithm for finding the $t$ orbit representatives $\mathbf{a}^1, \dots, \mathbf{a}^t$ in Theorem 7.1. Meshulam & Wigderson [30] leave this question open for their generalization of this theorem. In our case of an irreducible $\rho$ we will show the following.

**Theorem 7.2.** *Given $\lambda > 0$ and an irreducible $\rho : H \to \mathrm{GL}_d(\mathbb{F})$, $\mathbb{F} = \mathbb{F}_p$, we can find $\mathbf{a}^1, \dots, \mathbf{a}^t \in \mathbb{F}^d$, $t = t(\lambda, p)$, satisfying $\lambda(\mathbb{F}^d; \bigcup_{i=1}^{t} {}^{H}\pm\mathbf{a}^i) \le \lambda$ in time $p^d \cdot \mathsf{poly}(|H|, p, \lambda^{-1})$.*

Denote by $\mathcal{G} = \mathcal{C}(\mathbb{F}^d; \bigcup_{i=1}^{t} {}^{H}\pm\mathbf{a}^i)$, $|\mathcal{G}| = p^d$, the resulting expander graph.

To better understand the time bound $p^d \cdot \mathsf{poly}(|H|, p, \lambda^{-1})$ and the strength of our claim we offer some observations.

(1) The running time $p^d \cdot \mathsf{poly}(|H|, p, \lambda^{-1})$ is that of the immediate randomized algorithm implied by Theorem 7.1 (recall Section 4.3), and thus, it is hard to improve on it.

(2) In applications, it is usually the case that $d = \Theta(|H|^\epsilon)$ so that $p^d \gg |H|$. That is, the factor $p^d = |\mathcal{G}|$ might be exponential in the size of the input.

(3) Even if we make $t = t(p, \lambda)$ constant by fixing $p$ and $\lambda$, the naive algorithm that finds $\mathbf{a}^1, \dots, \mathbf{a}^t$ by considering all possible $t$-subsets of $\mathbb{F}^d$ runs in time $\Omega(p^{td})$, which is far worse than the time bound we claim.

It would be ideal if the above running time could be reduced to $\mathsf{poly}(|H|, p, \lambda^{-1})$, because this would imply that $\mathcal{G}$ could be described *implicitly* in the following sense: Given two vertices $u, v \in \mathcal{G}$ (represented as binary strings of length $O(\log|\mathcal{G}|)$) we could decide if $u$ and $v$ are adjacent by finding the generators $\mathbf{a}^1, \dots, \mathbf{a}^t$ without the need to consider every vertex of $\mathcal{G}$ in the computation.

Instead, we end up having to pay a factor $p^d = |\mathcal{G}|$ in running time since we check separately that each $\lambda_{\mathbf{x}}$, $\mathbf{x} \in \mathbb{F}^d \smallsetminus \{\mathbf{0}\}$, satisfies $|\lambda_{\mathbf{x}}| \le \lambda$. More precisely, using the framework of conditional probabilities from Section 4 our goal is to argue that there are good pessimistic estimators for the events $\mathcal{E}_{\mathbf{x}} = \{|\lambda_{\mathbf{x}}| > \lambda\}$ computable in time $\mathsf{poly}(|H|, p, \lambda^{-1})$.

7.1.1. *Derandomized algorithm.* Our algorithm proceeds by picking the vectors $\mathbf{a}^1, \dots, \mathbf{a}^t$ (where $\mathbf{a}^i = [a_1^i \; a_2^i \; \dots \; a_d^i]^T$) one coordinate $a_j^i \in \mathbb{F}$ at a time. This is to say that our random choice tree $\mathcal{T}$ has depth $dt$ and arity $|\mathbb{F}| = p$. (Note that picking a $d$-dimensional vector $\mathbf{a}^i \in \mathbb{F}^d$ at a single step would not result in the running time we are aiming for.) For a fixed $\mathbf{x} \ne \mathbf{0}$ we begin re-examining the proof of Theorem 7.1.

To shorten notation we denote by $\mathcal{F}_{i,j}$ an event of the form $\{a_1^i = \alpha_1^i, \dots, a_j^i = \alpha_j^i\}$ where $\alpha_k^i \in \mathbb{F}$, $1 \le k \le j$. In the proof we analyzed

$$A_h = \frac{1}{td} \sum_{i=1}^{t} \sum_{r \in hR} X_{i,r}$$

using the Chernoff bound for the independent $X_{i,r}$, $|X_{i,r}| \le 1$. Accordingly, we'll develop Chernoff-type pessimistic estimators for $\mathcal{E}_h = \{|A_h| > \lambda/2\}$. The key observation here is

**Proposition 7.3.** *Either (i) $X_{i,r}$ is constant on $\mathcal{F}_{i,j}$; or (ii) $\mathbf{E}[X_{i,r} \mid \mathcal{F}_{i,j}] = 0$.*

*Proof.* Recall that $X_{i,r} = \tilde{\chi}_{\mathbf{x}}({}^{r}\mathbf{a}^i) = \frac{1}{2}(\chi({}^{r}\mathbf{a}^i) + \chi({}^{r}(-\mathbf{a}^i)))$ and $\chi({}^{r}\mathbf{a}) = \omega^{(\mathbf{Ta})_r} = \omega^{t_{r,1}a_1 + t_{r,2}a_2 + \dots + t_{r,d}a_d}$ where $\mathbf{T} = (t_{r,k})$. The case *(i)* occurs when $t_{r,k} = 0$ for $j+1 \le k \le d$ and the case *(ii)* otherwise. □

This proposition allows us to (efficiently) estimate

$$\mathbf{E}[e^{\frac{\lambda}{2}X_{i,r}} \mid \mathcal{F}_{i,j}] \le \begin{cases} \text{the exact value} & \text{in case } \textit{(i)} \\ e^{(\frac{\lambda}{2})^2/2} & \text{in case } \textit{(ii)} \end{cases}.$$

Analogously to (17) we can use the above to define Chernoff-type pessimistic estimators

$$U_{i,j}^{\mathcal{E}_h}(\alpha_1^i, \dots, \alpha_j^i) \ge \mathbf{Pr}[\mathcal{E}_h \mid \mathcal{F}_{i,j}]$$

satisfying $U_{0,0}^{\mathcal{E}_h} \le \epsilon$ as in (24).

Now, similarly to (25), we calculate:

$$\begin{aligned}\mathbf{Pr}[|\lambda_{\mathbf{x}}| > \lambda \mid \mathcal{F}_{i,j}] &\le \mathbf{Pr}[Y > \frac{\lambda}{2}|H| \mid \mathcal{F}_{i,j}] \\ &\le \frac{\mathbf{E}[Y \mid \mathcal{F}_{i,j}]}{\lambda|H|/2} = \frac{\sum_{h\in H}\mathbf{Pr}[\mathcal{E}_h \mid \mathcal{F}_{i,j}]}{\lambda|H|/2} \\ &\le \frac{2}{\lambda|H|}\sum_{h\in H} U_{i,j}^{\mathcal{E}_h}(\alpha_1^1,\dots,\alpha_j^i) \ . \qquad (26)\end{aligned}$$

This shows that setting $U_{i,j}^{\mathcal{E}_{\mathbf{x}}}(\alpha_1^1,\dots,\alpha_j^i)$ equal to (26) gives pessimistic estimators for the event $\mathcal{E}_{\mathbf{x}} = \{|\lambda_{\mathbf{x}}| > \lambda\}$ satisfying $U_{0,0}^{\mathcal{E}_{\mathbf{x}}} \le p^{-d}$ as in (25). These functions are computable in time $\mathsf{poly}(|H|, p, \lambda^{-1})$ as required.

Finally, the pessimistic estimator for the failure event $\{\exists \mathbf{x} \ne \mathbf{0} : \lambda_{\mathbf{x}} > \lambda\}$ is now given by the union bound estimator

$$U_{i,j}(\alpha_1^1,\dots,\alpha_j^i) = \sum_{\mathbf{x}\ne\mathbf{0}} U_{i,j}^{\mathcal{E}_{\mathbf{x}}}(\alpha_1^1,\dots,\alpha_j^i) \ ,$$

where $U_{0,0} < 1$. Evaluating all $p^d - 1$ many functions $U_{i,j}^{\mathcal{E}_{\mathbf{x}}}$ incurs a factor of $p^d$ to the running time of the algorithm. This concludes the proof of Theorem 7.2.

7.2. **Low-degree expanders.** In order to construct sparse expanders via Theorem 7.1 we need to find groups that have irreducible representations of large dimension *over some finite field* $\mathbb{F}_p$. The dimensions of irreducible representations are often well-understood over (algebraically closed) fields of zero characteristic but not so in the case of finite fields. For example, Alon et al. [2] note that the sizes of the irreducible representations of $\mathbb{Z}_p$, $p$ prime, over $\mathbb{F}_2$ depend on an unsolved problem in number theory (Artin's conjecture on primitive roots).

Affine groups (Section 5.1.1) turn out to have irreducible representations of large dimension *over* $\mathbb{C}$. (In fact, it is shown in [7, 43] that the affine groups are among the few groups $G$ that have irreducible representations over $\mathbb{C}$ of dimension $d$ with $d(d+1) \ge |G|$.) We shall argue that these representations carry over to finite fields.

Meshulam & Wigderson [30] also consider affine groups but with a different approach. They want to write down highly *explicit* expanding orbits without resorting to the probabilistic existence proof of their generalization of Theorem 7.1, which they cannot derandomize. By contrast, we apply Theorem 7.1 and are content with the fact that the generators can be found in deterministic polynomial time due to our derandomization result of the previous section. Also, Meshulam & Wigderson work in characteristic $p = 2$, whereas we work in characteristic $p \ge 3$.

7.2.1. *A doubly transitive action of* $\mathrm{Aff}(\mathbb{F})$*.* Affine groups $\mathrm{Aff}(\mathbb{F}_q)$ were introduced in Section 5.1.1. An element $f \in \mathrm{Aff}(\mathbb{F}_q)$, considered as a function $\mathbb{F}_q \to \mathbb{F}_q$, acts naturally on $x \in \mathbb{F}_q$ as $f \cdot x := f(x)$. It can be easily checked that this action is *doubly transitive*: whenever $\{x, y\}$ and $\{x', y'\}$ are pairs of elements of $\mathbb{F}_q$ (i.e., $x \ne y$, $x' \ne y'$) then for some $f \in \mathrm{Aff}(\mathbb{F}_q)$ we have $f(x) = x'$ and $f(y) = y'$. We consider the permutation representation induced by this action, i.e., the $\mathrm{Aff}(\mathbb{F}_q)$-module $\mathbb{F}_p[\mathbb{F}_q]$ where the action of $\mathrm{Aff}(\mathbb{F}_q)$ on the basis elements $\mathbb{F}_q$ is extended linearly to all of $\mathbb{F}_p[\mathbb{F}_q]$.

We require that $p \nmid |\mathrm{Aff}(\mathbb{F}_q)| = q(q-1)$. By elementary number theory, there are infinitely many $q$ satisfying this requirement for every fixed $p \ge 3$.

We will apply the following well-known lemma, which we state in some generality.

**Lemma 7.4.** *Let $G$ act on $X$ doubly transitively and let $\mathbb{F}$ be a field of characteristic not dividing $|G|$. Then the $G$-module $\mathbb{F}[X]$ decomposes as $\mathbb{F}[X] \simeq \mathbb{F} \oplus M$, where $\mathbb{F}$ is the 1-dimensional trivial module and $M$ is simple.*

In text books (e.g. [10, p. 42], [21, p. 69] etc.), this lemma is usually proved using character theory in case $\mathbb{F}$ has characteristic zero—we presented essentially this argument in Section 3.3.3. Even though it is possible to transfer this result over to more general fields $\mathbb{F}$ (M. J. Collins, personal communication, July 2011), we choose to provide an elementary ad hoc proof of Lemma 7.4.

*Proof of Lemma 7.4.* Write $X = \{x_1, \dots, x_n\}$. The $G$-module $\mathbb{F}[X]$ decomposes as $T \oplus M$, where $T = \{a \sum_{i=1}^{n} \mathbf{x}_i \in \mathbb{F}[X] : a \in \mathbb{F}\}$ is the trivial representation and

$$M = \{\sum_{i=1}^{n} a_i \mathbf{x}_i \in \mathbb{F}[X] : \sum_{i=1}^{n} a_i = 0\} = \operatorname{span}_{\mathbb{F}}\{\mathbf{x}_i - \mathbf{x}_{i+1} : 1 \le i \le n-1\} \ .$$

In particular, $T \cap M = \{\mathbf{0}\}$, because of our assumption of $p$ not dividing $n = |X|$. It remains to show that $M$ is simple.

If $n = 2$, then $M$ is one-dimensional and we are done, so suppose $n \ge 3$. Let $\{\mathbf{0}\} \ne V \le M$ be a submodule. Pick some $\mathbf{v} = \sum_i a_i \mathbf{x}_i \in V$, $\mathbf{v} \ne \mathbf{0}$, and suppose (using double transitivity) that $a_1, a_2 \ne 0$. Let $G_i \le G$ be the stabilizer of $x_i$. Then because $G$ is doubly transitive, $G_i$ is transitive on $X \setminus \{x_i\}$ and we have,

$$\mathbf{w}_i := \Big(\frac{1}{|G_i|} \sum_{g \in G_i} \mathbf{g}\Big) \cdot \mathbf{v} = a_i \mathbf{x}_i - \frac{a_i}{n-1} \sum_{j \ne i} \mathbf{x}_j \in V \ .$$

(Here, $G_i$ acts transivitely on a set of size $n-1$ and thus $n-1$ divides $|G_i|$. Therefore, $p \nmid n-1$ as $p \nmid |G|$.) But now $a_1^{-1}\mathbf{w}_1 - a_2^{-1}\mathbf{w}_2 = \frac{n-2}{n-1}(\mathbf{x}_1 - \mathbf{x}_2) \in V$ generates $M$ so $V = M$ as required. □

7.2.2. *Construction of an $O(\log \log n)$-degree expander family.* Specializing to the case $X = \mathbb{F}_q$ we let $M_q \le \mathbb{F}_p[\mathbb{F}_q]$ denote the simple $\operatorname{Aff}(\mathbb{F}_q)$-module of dimension $q-1$ in Lemma 7.4. We can now construct an expander family out of soluble groups (of derived length 3) that has exponentially smaller degree than what is guaranteed by the Alon–Roichman theorem (Theorem 4.2).

**Theorem 7.5** ($O(\log \log n)$-degree expanders)**.** *Fix a prime $p \ge 3$. Let $q$ denote a prime power with $p \nmid q(q-1)$. Define, with the natural action,*

$$G_q := M_q \rtimes \operatorname{Aff}(\mathbb{F}_q) \ . \tag{27}$$

*There are generating sets $S_q \subseteq G_q$ such that $\{\mathcal{C}(G_q; S_q)\}_q$ is a family of expander graphs of degree $|S_q| = O(\log \log |G_q|)$.*

*Proof.* The groups $\operatorname{Aff}(\mathbb{F}_q)$ are semi-direct products of the form $\mathbb{F} \rtimes \mathbb{F}^\times$. Here, the abelian group $\mathbb{F}^\times$ has a 1/2-expanding generating set $T' \subseteq \mathbb{F}^\times$ of size $O(\log q)$ by Theorem 4.2. The $\mathbb{F}^\times$-orbit of every non-identity element in $\mathbb{F}$ is $\mathbb{F} \setminus \{0\}$ and this orbit generates $\mathbb{F}$ as a complete graph (a 1/2-expander). Thus, the generating set

$$T = \{\pm 1^{(|T'|)}\} \cup T' \subseteq \operatorname{Aff}(\mathbb{F}_q) \ , \qquad |T| = 3|T'| \ ,$$

is $\mu$-expanding for $\mu = g(1/2, 1/2, 2/3) < 1$ by Corollary 6.6.

Now, we apply Theorem 7.1 to see that $M_q$ has some constant number (depending on $p$) of elements $R = \{\mathbf{a}^1, \dots, \mathbf{a}^t\}$ whose orbits generate $M_q$ as a 1/2-expander. Set the multiplicity of each element in $R$ to be $|T|$ so that $|R| = t|T|$ and $\alpha = |R|/(|R| + |T|) = t/(t+1)$ is a constant. This allows us to conclude that $S_q = R \cup T$ is a $\lambda$-expanding generating set for $G_q$ where $\lambda = g(1/2, \mu, \alpha) < 1$ by Corollary 6.6.

Finally, the group $G_q$ has size $|G_q| = p^{q-1} q(q-1)$ so that

$$|S_q| = |R| + |T| = t|T| + |T| = O(\log q) = O(\log \log |G_q|) \ . \qquad \square$$

The attained degree $O(\log \log n)$ is the best possible for the family $\{G_q\}_q$ up to constant factors. This is because every $\lambda$-expanding generating set for $G_q$ implies by Proposition 5.1 a $\lambda$-expanding generating set of the same size for the quotient $G_q/M_q \simeq \mathrm{Aff}(\mathbb{F}_q)$, and we argued in Section 5.1.1 that $\mathrm{Aff}(\mathbb{F}_q)$ needs $\Omega_\lambda(\log|\mathrm{Aff}(\mathbb{F}_q)|) = \Omega_\lambda(\log\log|G_q|)$ generators to achieve $\lambda$-expansion.

We have derandomized every probabilistic existence proof that we have presented in this work. Hence, the following constructive refinement of Theorem 7.5 follows.

**Theorem 7.6.** *The generating sets $S_q \subseteq G_q$ in Theorem 7.5 can be found in time polynomial in $|G_q|$.*

## 8. Derandomized Squaring

The $O(\log\log n)$-degree expanders we constructed in Section 7.2 have their second largest eigenvalue bounded by $\lambda$, for *some* $\lambda < 1$. The methods we used in the construction—the wide replacement product in particular—do not allow us to choose $\lambda$ arbitrarily close to 0 (recall Example 3). By contrast, the majority of the results presented in this work are stated with respect to $\lambda > 0$ being a free variable, and only the multiplicative factors in the relevant upper and lower bounds depend on $\lambda$. A natural question arises:

**Question 3.** *How does the minimum size of a $\lambda$-expanding generating set depend on the particular choice of $\lambda \in (0,1)$?*

Using graph powering every $\mu$-expanding family $\{\mathcal{G}_i\}_{i\in\mathbb{N}}$ can be turned into a $\lambda$-expanding family $\{\mathcal{G}_i^k\}_{i\in\mathbb{N}}$ by choosing $k = \lceil \log_\mu(\lambda)\rceil = O(1)$ large enough. However, if $\{\mathcal{G}_i\}_{i\in\mathbb{N}}$ has unbounded degree, powering alters the asymptotic growth rate of the degree. Hence, generating set powering does not give a satisfactory answer to Question 3.

Many authors sidestep Question 3 by focusing on proving the existence of small (but unbounded) $\lambda$-expanding generating sets for *some* $\lambda < 1$, similarly to our $O(\log\log n)$-degree family. For example, [30, Thm 1.2] has the restriction $\lambda > 1/2$, and [19, Cor 1.9] is restricted by $\lambda > 3/4$.

To our knowledge Question 3 has not been asked explicitly nor answered implicitly in the literature.

In this section, we make the observation that the technique of *derandomized squaring*, due to Rozenman & Vadhan [40], can be applied in the context of Cayley graphs to give a clean answer to our question: the choice of $\lambda \in (0,1)$ only affects multiplicative factors in generating set size. More specifically, we will prove the following.

**Theorem 8.1.** *Let $\{\mathcal{C}(G_i;S_i)\}_{i\in\mathbb{N}}$ be a family of Cayley expander graphs. Then, for every $\lambda > 0$, there exist generating sets $S_i' \subseteq G_i$ of size $|S_i'| = O_\lambda(|S_i|)$ with $\lambda(G_i;S_i') \le \lambda$.*

In particular, this implies that [19, Cor 1.7] holds for *all* $0 < \epsilon < 1$ (here, $\epsilon = 1 - \lambda$ in our terminology) instead of *some* $\epsilon > 0$, and the restriction of $\epsilon < 1/4$ can be removed in [19, Cor 1.9]. Ultimately, our Theorem 8.1 says that the main question of Hadad [19, Question 1.5] does not depend on the parameter $\epsilon$ in an essential way.

*Remark* 8.1. The corresponding slight improvement to a result of Meshulam & Wigderson [30] is more subtle and technical. In [30, Thm 1.7] the authors consider the following inductively defined family of soluble groups (of unbounded derived length):

$$G_0 = \mathbb{Z}_2 \ , \qquad G_{i+1} = \mathbb{F}_{p_i}[G_i] \rtimes G_i \ ,$$

where $\{p_i\}_{i\ge1}$ are the odd primes and the semi-direct product is taken with respect to the natural $G_i$-module action. Using *(i)* the wide zig-zag product and *(ii)* generating set powering they construct generating sets $S_i \subseteq G_i$ of size

$$|S_i| = \exp(\exp(\Theta(i))) = O(\log^{(i-\log^* i)}|G_i|) \ . \tag{28}$$

If, instead, the construction is made using *(i)* the wide replacement product and *(ii)* Theorem 8.1, the sets $S_i$ would have size $\exp(\Theta(i))$. Even though this is an exponential improvement, the groups $G_i$ are so large that this optimization does not affect the asymptotics of (28) considerably.

8.1. **Definition of derandomized squaring.** In squaring a $d$-regular graph $\mathcal{G}$ the expansion improves as $\lambda \mapsto \lambda^2$ while the degree increases as $d \mapsto d^2$.

The graph $\mathcal{G}^2$ can be viewed as being obtained from $\mathcal{G}$ by placing a copy of $\tilde{\mathcal{K}}_d$ (the complete graph with self-loops; Section 2.2.1) on top of each neighbourhood $\Gamma_{\mathcal{G}}(v) = \{u \in V(\mathcal{G}) : \{u,v\} \in E(\mathcal{G})\}$, $v \in V(\mathcal{G})$. This point of view suggests the following intuition: $\mathcal{G}^2$ is good expander because $\mathcal{G}^2$ consists of copies of $\tilde{\mathcal{K}}_d$, and $\tilde{\mathcal{K}}_d$ is the perfect expander. However, we already know that we don't need quadratically many edges—as in $\tilde{\mathcal{K}}_d$—to achieve good expansion! Indeed, by replacing $\tilde{\mathcal{K}}_d$ in the above construction with an arbitrary expander graph $\mathcal{H}$ we arrive at the definition of derandomized graph squaring.

Rozenman & Vadhan [40] introduced derandomized squaring for general outregular directed graphs. To resolve the ambiguity in how a graph $\mathcal{H}$ is placed on the vertices $\Gamma_{\mathcal{G}}(v)$ Rozenman & Vadhan consider both one-way and two-way edge labellings (rotation maps; Section 6.1) for $\mathcal{H}$. Since we will only apply derandomized squaring to Cayley graphs and Cayley graphs carry a natural two-way labelling on their edges, we simplify the presentation by specializing the definitions and results of [40] to Cayley graphs.

**Definition 8.1 (Derandomized squaring, [40]).** Let $S \subseteq G$ be a generating set and let $\mathcal{H}$ be a $k$-regular graph on $V(\mathcal{H}) = S$. The *derandomized square* of $S$ by $\mathcal{H}$, denoted by $S \circledS \mathcal{H}$, is the symmetric set defined as

$$S \circledS \mathcal{H} := \{sr \in G : \{s, r^{-1}\} \in E(\mathcal{H})\} \ , \qquad |S \circledS \mathcal{H}| = k|S| \ . \tag{29}$$

8.2. **Derandomized squaring improves expansion.** Let $\{\mathcal{C}(G_i; S_i)\}_{i\in\mathbb{N}}$ be a family of Cayley expanders, possibly of unbounded degree, and let $\{\mathcal{H}_i\}_{i\in\mathbb{N}}$ be a family of $k$-regular expanders such that $V(\mathcal{H}_i) = S_i$. The next result shows that the family $\{\mathcal{C}(G_i; S_i \circledS \mathcal{H}_i)\}_{i\in\mathbb{N}}$ of degree $k|S_i|$ has improved expansion given that the $\mathcal{H}_i$ are good enough expanders.

**Theorem 8.2** ([40])**.** *If $\mathcal{C}(G;S)$ is an $(n,d,\lambda)$-graph and $\mathcal{H}$ is a $(d,k,\mu)$-graph, then $\mathcal{C}(G; S \circledS \mathcal{H})$ is an $(n, kd, f(\lambda,\mu))$-graph, where*

$$f(\lambda,\mu) := (1-\mu)\lambda^2 + \mu \ . \tag{30}$$

Note that $f$ is monotonically increasing in both coordinates. Also, if $\mathcal{H}$ is a good expander, i.e., $\mu \to 0$, then $f(\lambda,\mu) \to \lambda^2$, approximating graph squaring.

*Proof.* We have nothing to add to the original proof of [40]. Still, the proof is similar to our calculations in the previous sections so we repeat it here—as specialized to Cayley graphs—for the sake of completeness.

A random step in the graph $\mathcal{C}(G; S \circledS \mathcal{H})$ is obtained by the following process:

(1) Starting at a vertex $g \in G$ choose an element $s \in_{\mathsf{R}} S$ uniformly at random. We record this as saying that we are in *state* $(g, s)$.
(2) Go to state $(gs, s)$.
(3) Go to state $(gs, r)$, where $r^{-1} \in_{\mathsf{R}} \Gamma_{\mathcal{H}}(s)$.
(4) Go to state $(gsr, r)$.
(5) Arrive at $gsr \in G$.

We model the states above as the natural basis elements in $\mathbb{C}^{G\times S} \simeq \mathbb{C}^G \otimes \mathbb{C}^S$ so that probability distributions over the states can be thought of as vectors in this space.

The 1st step corresponds to a linear map $\mathbf{L} : \mathbb{C}^G \to \mathbb{C}^G \otimes \mathbb{C}^S$ that takes a basis vector $\mathbf{g} \in \mathbb{C}^G$ and maps it to the uniform distribution over states of the form

$\mathbf{g} \otimes \mathbf{s}$, $s \in S$, i.e., $\mathbf{Lg} = \frac{1}{d} \sum_{s \in S} \mathbf{g} \otimes \mathbf{s}$. In the 2nd step we apply the permutation map $\mathbf{g} \otimes \mathbf{s} \mapsto \mathbf{gs} \otimes \mathbf{s}$; call this map $\mathbf{G}$. In the 3rd step the $s$ in $\mathbf{g} \otimes \mathbf{s}$ takes a random step on $\mathcal{H}$ and gets sent to its inverse. This transformation is given by $\mathbf{H} := \mathbf{I} \otimes \mathbf{QA}(\mathcal{H})$, where $\mathbf{A}(\mathcal{H})$ is the random walk matrix of $\mathcal{H}$ and $\mathbf{Q} : \mathbf{s} \mapsto \mathbf{s^{-1}}$. In the 4th step we again apply $\mathbf{G}$. Finally, in the 5th step we project the state from $\mathbb{C}^G \otimes \mathbb{C}^S$ onto $\mathbb{C}^G$ via the map $\mathbf{P} : \mathbf{g} \otimes \mathbf{s} \mapsto \mathbf{g}$. (This is inverse to step 1 in that $\mathbf{PL} = \mathbf{I}$.) Thus, the random walk matrix $\mathbf{A} = \mathbf{A}(\mathcal{C}(G; S \circledS \mathcal{H}))$ is given by

$$\mathbf{A} = \mathbf{PGHGL} \ . \tag{31}$$

The mapping $\mathbf{QA}(\mathcal{H})$ has $\lambda(\mathbf{QA}(\mathcal{H})) \leq \lambda(\mathbf{A}(\mathcal{H})) \leq \mu$, so by Proposition 2.2 write $\mathbf{H} = (1-\mu)(\mathbf{I} \otimes \tilde{\mathbf{J}}) + \mu(\mathbf{I} \otimes \mathbf{E})$ for some $\mathbf{E}$ with $\|\mathbf{E}\|_2 \leq 1$. Substituting this into (31) we get

$$\begin{aligned} \mathbf{A} &= (1-\mu)\mathbf{PG}(\mathbf{I} \otimes \tilde{\mathbf{J}})\mathbf{GL} + \mu\mathbf{PG}(\mathbf{I} \otimes \mathbf{E})\mathbf{GL} \\ &=: (1-\mu)\mathbf{A}_1 + \mu\mathbf{A}_2 \ . \end{aligned}$$

But, $\mathbf{I} \otimes \tilde{\mathbf{J}} = \mathbf{LP}$ and $\mathbf{A}(\mathcal{C}(G;S)) = \mathbf{PGL}$ so that

$$\mathbf{A}_1 = \mathbf{PGL} \cdot \mathbf{PGL} = \mathbf{A}(\mathcal{C}(G;S))^2 \ ,$$

implying $\lambda(\mathbf{A}_1) \leq \lambda^2$. Finally, using the fact that $\|\mathbf{P}\|_2 = d^{1/2}$ and $\|\mathbf{L}\|_2 = d^{-1/2}$ we have $\|\mathbf{A}_2\|_2 \leq 1$ and consequently

$$\lambda(\mathbf{A}) \leq (1-\mu)\lambda^2 + \mu = f(\lambda, \mu) \ . \qquad \square$$

8.3. **Iterated squaring.** The basic idea of Rozenman & Vadhan [40] is to repeatedly apply Theorem 8.2 in order to bring the second largest eigenvalue of a graph close to 0. We explain their method and note that the construction can be carried out with only a linear increase in the degree, proving Theorem 8.1.

We need a family of constant-degree expanders in the construction. The following family can be obtained from Example 1 by an easy modification.

**Lemma 8.3.** *Let $\epsilon > 0$. There exists a family $\{\mathcal{H}_i\}_{i \geq 1}$ with $\mathcal{H}_i$ a $(d^i, d, \epsilon)$-graph, where $d = d(\epsilon)$.*

Our proof of Theorem 8.1 is essentially that of Theorem 7.1 in [40] with the restriction $\lambda \geq 1/2$ removed.

*Proof of Theorem 8.1.* Let $S \subseteq G$ be $\mu$-expanding for some $\mu < 1$. For $0 < \lambda < \mu$, we show how $S$ can be converted into a $\lambda$-expanding generating set of size $O_\lambda(|S|)$. We will assume that $\lambda < 1/2$ for computational convenience.

Let $\{\mathcal{H}_i\}_{i \geq 1}$ be a family of $(d^i, d, \lambda/2)$-graphs as in Lemma 8.3. We can assume without loss of generality that $|S| = d^j$ for some $j$ (if not, add identity elements to $S$ and replace $\mu$ by $1 - (1-\mu)/d$ if necessary; cf. Remark 2.2).

Define inductively

$$S_0 = S \quad \text{and} \quad S_{i+1} = S_i \circledS \mathcal{H}_{i+j} \ .$$

If we let $\mu_i = \lambda(G; S_i)$ we have that $\mu_0 \leq \mu$ and $\mu_{i+1} \leq f(\mu_i, \lambda/2)$ by Theorem 8.2. The function $g : x \mapsto f(x, \lambda/2)$ is strictly convex, has a fixed point $g(1) = 1$ and satisfies $g(\mu) < \mu$ and $g(\lambda) < \lambda$. Thus, $x - g(x)$ is strictly positive on the compact interval $[\lambda, \mu]$ so that it is bounded from below by some $\epsilon > 0$, i.e., $g(x) \leq x - \epsilon$ for $x \in [\lambda, \mu]$. From these facts it follows that $\mu_k \leq \lambda$ for some $k = O(1/\epsilon) = O_\lambda(1)$. Thus, $S_k$ is $\lambda$-expanding and $|S_k| = d^k|S| = O_\lambda(|S|)$. $\square$

## 9. Open Problems

In this dissertation we have focused on *known results* in theory of expansion in soluble groups. Consequently, our discussion does not really give rise to any new open problems in this area. We make only a couple of remarks:

(1) Our representation theoretic proof of the zig-zag theorem (Theorem 6.1) covers only the case of a semi-direct product by an abelian group. Removing this assumption by using a more involved representation theory would most probably render our approach unnecessarily complex as compared to the original proof of [2, 38].

(2) In Section 7.1 we asked whether the exponential factor $|\mathbb{F}^d| = p^d$ in Theorem 7.2 can be eliminated. Questions similar to this one have been raised in [2, 30]. A positive answer would entail applications in error correcting codes.

(3) Wigderson [45] has asked whether the property of a family of groups admitting bounded expanding generating sets can be characterized by a condition on the dimensions of the irreducible representations of the groups.

## References

[1] R. Ahlswede and A. Winter, *Strong converse for identification via quantum channels*, IEEE Transactions on Information Theory, 48 (2002), pp. 569–579.

[2] N. Alon, A. Lubotzky, and A. Wigderson, *Semi-direct product in groups and zig-zag product in graphs: Connections and applications*, in Proceedings of the 42nd Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, 2001, pp. 630–637.

[3] N. Alon and Y. Roichman, *Random Cayley graphs and expanders*, Random Structures Algorithms, 5 (1997), pp. 271–284.

[4] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., 3rd ed., 2008.

[5] S. Arora and B. Barak, *Computational Complexity: A Modern Approach*, Cambridge University Press, 2009.

[6] V. Arvind, P. Mukhopadhyay, and P. Nimbhorkar, *Erdős–Rényi sequences and deterministic construction of expanding Cayley graphs*, Electronic Colloquium on Computational Complexity (ECCC), 18 (2011), p. 81.

[7] Y. Berkovich, *Groups with few characters of small degrees*, Israel Journal of Mathematics, 110 (1999), pp. 325–332.

[8] M. Blum, R. M. Karp, O. Vornberger, C. H. Papadimitriu, and M. Yannakakis, *The complexity of testing whether a graph is a superconcentrator*, Information Processing Letters, 13 (1981), pp. 164–167.

[9] E. Breuillard, B. Green, and T. Tao, *Suzuki groups as expanders*, (2010). arXiv:1005.0782.

[10] P. Cameron, *Permutation Groups*, no. 45 in London Mathematical Society Student Texts, Cambridge University Press, 1999.

[11] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, *Harmonic Analysis on Finite Groups: Representation Theory, Gelfand Pairs and Markov Chains*, no. 108 in Cambridge studies in advanced mathematics, Cambridge University Press, 2008.

[12] F. R. K. Chung, *Spectral Graph Theory*, no. 92 in CBMS Conference on Recent Advances in Spectral Graph Theory, American Mathematical Society, 1997.

[13] G. P. Davidoff, P. Sarnak, and A. Valette, *Elementary number theory, group theory, and Ramanujan graphs*, no. 55 in London Mathematical Society Student Texts, Cambridge University Press, 2003.

[14] P. Diaconis, *Group Representations in Probability and Statistics*, vol. 11 of Institute of Mathematical Statistics Lecture Notes—Monograph Series, Institute of Mathematical Statistics, Hayward, CA, 1988.

[15] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transpositions*, Probability Theory and Related Fields, 57 (1981), pp. 159–179.

[16] P. Etingof, O. Golberg, S. Hensel, T. Liu, A. Schwendner, D. Vaintrob, and E. Yudovina, *Introduction to representation theory.* Lecture notes, February 2011. arXiv:0901.0827v5.

[17] J. Friedman, *A Proof of Alon's Second Eigenvalue Conjecture and Related Problems*, no. 109 in Memoirs of the American Mathematical Society, American Mathematical Society, 2008.

[18] C. Godsil and G. Royle, *Algebraic Graph Theory*, no. 207 in Graduate Texts in Mathematics, Springer, 2001.

[19] U. Hadad, *Kazhdan constants of group extensions*, International Journal of Algebra and Computation, 20 (2010), pp. 671–688.

[20] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulletin of the American Mathematical Society, 43 (2006), pp. 439–561.

[21] I. M. Isaacs, *Character Theory of Finite Groups*, Academic Press, 1976.

[22] G. D. James and M. W. Liebeck, *Representations and Characters of Groups*, Cambridge University Press, 2nd ed., 2001.

[23] D. L. Johnson, *Presentations of Groups*, no. 15 in London Mathematical Society Student Texts, Cambridge University Press, 2nd ed., 1997.

[24] M. Kassabov, *Symmetric groups and expander graphs*, Inventiones Mathematicae, 170 (2007), pp. 327–354.

[25] M. Kassabov, A. Lubotzky, and N. Nikolov, *Finite simple groups as expanders*, Proceedings of the National Academy of Sciences, 103 (2006), pp. 6116–6119.

[26] P.-S. Loh and L. J. Schulman, *Improved expansion of random Cayley graphs*, Discrete Mathematics and Theoretical Computer Science, 6 (2004), pp. 523–528.

[27] A. Lubotzky, *Discrete Groups, Expanding Graphs and Invariant Measures*, no. 125 in Progress in Mathematics, Birkhäuser, 1994.

[28] A. Lubotzky, *Expander graphs in pure and applied mathematics.* Notes prepared for the Colloquium Lectures at the Joint Annual Meeting of the American Mathematical Society (AMS) and the Mathematical Association of America (MAA), January 2011.

[29] A. Lubotzky and B. Weiss, *Groups and expanders*, in Expanding Graphs, J. Friedman, ed., vol. 10 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, 1993.

[30] R. Meshulam and A. Wigderson, *Expanders in group algebras*, Combinatorica, 24 (2004), pp. 659–680.

[31] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.

[32] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.

[33] I. Pak, *Random Cayley graphs with $O(\log|G|)$ generators are expanders*, in Proceedings of the 7th Annual European Symposium on Algorithms (ESA), Springer, 1999, pp. 521–526.

[34] M. S. Pinsker, *On the complexity of a concentrator*, in Proceedings of the 7th International Teletraffic Conference, 1973, pp. 318/1–318/4.

[35] P. Raghavan, *Probabilistic construction of deterministic algorithms: Approximating packing integer programs*, Journal of Computer and System Sciences, 37 (1988), pp. 130–143.

[36] O. Reingold, *Undirected connectivity in log-space*, Journal of the ACM, 55 (2008), pp. 17:1–17:24.

[37] O. Reingold, L. Trevisan, and S. Vadhan, *Pseudorandom walks on regular digraphs and the* $\mathsf{RL}$ *vs.* $\mathsf{L}$ *problem*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), ACM, 2006, pp. 457–466.

[38] O. Reingold, S. Vadhan, and A. Wigderson, *Entropy waves, the zig-zag graph product, and new constant-degree expanders*, The Annals of Mathematics, 155 (2002), pp. 157–187.

[39] E. Rozenman, A. Shalev, and A. Wigderson, *A new family of Cayley expanders (?)*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC), L. Babai, ed., ACM, 2004, pp. 445–454.

[40] E. Rozenman and S. Vadhan, *Derandomized squaring of graphs*, in Approximation, Randomization and Combinatorial Optimization: Algorithms and Techniques, C. Chekuri, K. Jansen, J. Rolim, and L. Trevisan, eds., no. 3624 in Lecture Notes in Computer Science, Springer, 2005, pp. 436–447.

[41] A. Russell and Z. Landau, *Random Cayley graphs are expanders: A simple proof of the Alon–Roichman theorem*, Electronic Journal of Combinatorics, 11 (2004), p. R62.

[42] J.-P. Serre, *Linear Representations of Finite Groups*, no. 42 in Graduate Texts in Mathematics, Springer, 1977.

[43] N. Snyder, *Groups with a character of large degree*, Proceedings of the American Mathematical Society, 136 (2008), pp. 1893–1903.

[44] S. Vadhan, *Pseudorandomness.* To appear in Foundations and Trends in Theoretical Computer Science, April 2011. Manuscript.

[45] A. Wigderson, *Expander graphs: applications and constructions.* Lecture at the Workshop on Expanders and Derandomization, Institut Henri Poincaré, March 2011.

[46] A. Wigderson and D. Xiao, *Derandomizing the Ahlswede–Winter matrix-valued Chernoff bound using pessimistic estimators, and applications*, Theory of Computing, 4 (2008), pp. 53–76.

[47] D. Xiao, *The evolution of expander graphs*, AB thesis, Harvard College, April 2003.