SIAM J. COMPUT.
Vol. 40, No. 2, pp. 267–290

A COMBINATORIAL CONSTRUCTION OF ALMOST-RAMANUJAN GRAPHS USING THE ZIG-ZAG PRODUCT*

AVRAHAM BEN-AROYA† AND AMNON TA-SHMA†

Abstract. Reingold, Vadhan, and Wigderson [*Ann. of Math.* (2), 155 (2002), pp. 157–187] introduced the graph *zig-zag* product. This product combines a large and a small graph into one, such that the resulting graph inherits its size from the large graph, its degree from the small graph, and its spectral gap from both. Using this product, they gave a fully explicit combinatorial construction of D-regular graphs having spectral gap $1 - O(D^{-\frac{1}{3}})$. In the same paper, they posed the open problem of whether a similar graph product could be used to achieve the almost optimal spectral gap $1 - O(D^{-\frac{1}{2}})$. In this paper we propose a generalization of the zig-zag product that combines a large graph and several small graphs. The new product gives a better relation between the degree and the spectral gap of the resulting graph. We use the new product to give a fully explicit combinatorial construction of D-regular graphs having spectral gap $1 - D^{-\frac{1}{2}+o(1)}$.

Key words. expander graphs, zig-zag product, combinatorial construction

AMS subject classification. 05C50

DOI. 10.1137/080732651

1. Introduction.

Expander graphs are graphs of low degree and high connectivity. Pinsker [19] observed that random D-regular graphs have almost optimal *set expansion*; i.e., every not-too-large set S has close to $(D-1)|S|$ distinct neighbors. Finding *explicit* D-regular graphs with a large set expansion turned out to be a challenging task. A related task is that of finding explicit D-regular graphs with large *algebraic expansion*, i.e., graphs with a large gap between the first and second largest eigenvalues of the operator defined by the adjacency matrix of the graph. Most of the earlier work on the subject focused on constructing explicit D-regular graphs with good (and sometimes even almost optimal) algebraic expansion [14, 8, 11, 13, 15, 17]. Optimal algebraic expansion guarantees a set expansion factor of about $D/2$ rather than the optimal $D - O(1)$ [6, 3, 1, 12].

Ten years ago, Reingold, Vadhan, and Wigderson [21] gave another construction of algebraic expanders. Unlike previous constructions, their construction is combinatorial in nature and has an intuitive analysis that is based on elementary linear algebra. At the heart of this construction lies a graph product, named the zig-zag product. Following their work, Capalbo et al. [5] used a variant of the zig-zag product to explicitly construct D-regular graphs with set expansion close to D, improving over the $D/2$ factor that is achieved by graphs with almost optimal algebraic expansion. Also, in a seemingly different setting, Reingold [20] gave a log-space algorithm for undirected connectivity, settling a long-standing open problem by taking advantage of the simple combinatorial composition of the zig-zag product, among other things.

*Received by the editors August 13, 2008; accepted for publication (in revised form) November 19, 2010; published electronically March 17, 2011. A preliminary version appeared in STOC 2008. This work was supported by the European Commission IST project "Quantum Computer Science" QCS 25596.

http://www.siam.org/journals/sicomp/40-2/73265.html

†The Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel (abrhambe@tau.ac.il, amnon@tau.ac.il). The first author's work was supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. The second author's work was supported by Israel Science Foundation grant 1090/10.

Using the zig-zag product, [21] gave an expander construction with spectral gap $1-O(D^{-\frac{1}{4}})$. Another construction that appeared in [21] had an improved spectral gap of $1-O(D^{-\frac{1}{3}})$ by using a modified version of the zig-zag product. In the same paper, Reingold, Vadhan, and Wigderson posed the question of finding a variant of the zig-zag product that gives rise to constructions with almost optimal spectral gap $1-O(D^{-\frac{1}{2}})$. Bilu and Linial [4] gave a different iterative construction of algebraic expanders that is based on 2-lifts, with a close to optimal spectral gap $1-O(\log^{1.5}(D)\cdot D^{-\frac{1}{2}})$. Their construction, however, is only *mildly explicit*, meaning that, given N, one can build a graph G_N on N vertices in poly(N) time. Ultimately, we would like to find a *fully explicit* construction, meaning that, given a vertex $v \in V = [N]$ and an index $i \in [D]$, we can compute the ith neighbor of v in poly$(\log(N))$ time. The zig-zag construction and many other explicit constructions are fully explicit, and this stronger notion of explicitness is crucial for some applications.

Several works studied different aspects of the zig-zag product. Alon, Lubotzky, and Wigderson [2] showed, somewhat surprisingly, an algebraic interpretation of the zig-zag product over non-Abelian Cayley graphs. This led to new iterative constructions of Cayley expanders [16, 22], which were once again based on algebraic structures. While these constructions are not optimal, they contribute to our understanding of the power of the zig-zag product.

Our main result is a new variant of the zig-zag product that retains most of the properties of the standard zig-zag product while giving a better spectral gap. Specifically, we use the new variant of the zig-zag product to construct an explicit family of D-regular expanders with spectral gap $1-D^{-\frac{1}{2}+o(1)}$, thus nearly resolving the open problem of [21].

1.1. An intuitive description of the new product.

1.1.1. The zig-zag product. Let us review the zig-zag product of [21]. The purpose of the zig-zag operation is to decrease the degree of a graph without harming its spectral gap too much. This product, in turn, is based on the *replacement product*. The replacement product takes as input two graphs:

- The first graph, G_1, has N_1 vertices and is D_1-regular. We think of G_1 as being the "large" graph, with many vertices N_1 and a large degree D_1.
- The second graph, H, is the "small" graph. We require that it have $N_2 = D_1$ vertices; i.e., the number of vertices of H equals the degree of G_1.

Another perquisite of the product is that the edges of the graph G_1 are *labeled*, namely, that each vertex labels its D_1 edges, each with a unique number from $\{1,\ldots,D_1\}$. The ith edge leaving a vertex v is simply the edge that v labeled with label i.

The replacement product results in a graph with N_1N_2 vertices, where every vertex v of G_1 is replaced with a *cloud* of D_1 vertices $\{(v,i)\}_{i\in[D_1]}$. There is an "intercloud" edge between (v,i) and (w,j) if $e=(v,w)$ is an edge in G_1 and e is the ith edge leaving v and the jth edge leaving w. Besides these intercloud edges there are only edges that connect vertices within the same cloud. The "intracloud" edges inside each cloud are simply a copy of the edges of H. That is, for every cloud v there is an edge between (v,i) and (v,j) if (i,j) is an edge in H. Figure 1 illustrates the replacement product between the the 3-dimensional cube and the 3-cycle.

The zig-zag product graph corresponds to 3-step walks on the replacement product graph, where the first and last steps are intracloud edges and the middle step is an intercloud edge. Namely, the vertices of the zig-zag product graph $G_1 ⓩ H$ are the

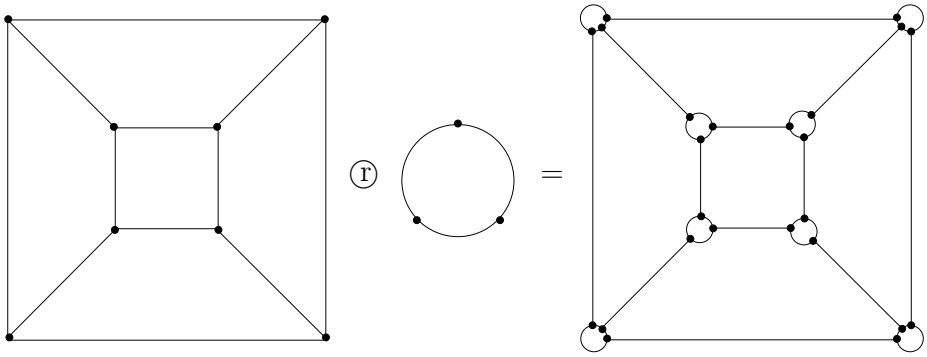

FIG. 1. *The replacement product between the cube and the* 3*-cycle.*

same as the those of the replacement product graph, and there is an edge between (v,i) and (w,j) if (w,j) can be reached from (v,i) by taking a 3-step walk: first an H step on the cloud of v, then an intercloud step from the cloud of v to the cloud of w, and finally an H step on the cloud of w. Thus, the number of vertices of $G_1 ⓩ H$ is $N_1 N_2$ and the degree is D_2^2; i.e., $G_1 ⓩ H$ inherits its size from the large graph and its degree from the small graph.

The main thing to analyze is the spectral gap of $G_1 ⓩ H$. Let us recall the definition of the spectral gap. Given a D-regular graph G we may view G as a Markov chain, where the states are the vertices of the graph, and the transition matrix of the chain is $A = \frac{1}{D}A'$, where A' is the adjacency matrix of the graph. If G has N vertices, then A is an $N \times N$ matrix. As in Markov chain theory, one may extend A to act on the whole vector space $\mathcal{V} = \mathbb{C}^N$ (rather than just on probability distributions) and then use linear algebra to analyze A. The *spectral gap* of G is the gap between the largest eigenvalue of A (which is 1 because G is regular) and the second largest eigenvalue (in absolute value) $\bar{\lambda}(A)$ of A. In the next section we give an overview of the proof (from [21]) that the second largest eigenvalue of $G_1 ⓩ H$ is small.

1.1.2. The analysis of the zig-zag product.

Entropy waves. Before attempting the formal algebraic analysis, let us consider a more intuitive, entropy-based analysis. If a graph is a good expander, then, given an entropy-deficient distribution (a distribution with not-too-high entropy) over its vertices, taking a random step over the graph results in a distribution with substantially more (Rényi) entropy.[1]

We shall consider two special distributions and we shall see that taking a random step over $G_1 ⓩ H$ when starting from either of these distributions substantially increases their entropy (assuming G_1 and H have this property). These two cases are representative in the sense that every distribution is essentially a linear combination of them, and therefore once we handle these cases we can handle any distribution.

Let us now describe the two distributions. Recall that a vertex $(v,i) \in [N_1] \times [N_2]$ of $G_1 ⓩ H$ is composed of two components; the first indicates the cloud in which the vertex resides, and the second corresponds to its position inside the cloud.

The first case we consider is a distribution $P = (P_1, P_2)$ that is entropy-deficient on the second component (meaning P_2 is entropy-deficient). In this case the first

[1] To see that, decompose the input distribution to a uniform component and a perpendicular component. Then, the expander keeps the uniform part in place while shrinking the perpendicular component.

H step adds entropy due to the fact that this step has the same effect of taking a step over H starting from the distribution P_2. The G_1 step is a permutation, i.e., it defines a bijection on the vertex set $[N_1] \times [N_2]$. As such it does not change the distribution's entropy. The second H step can never decrease entropy. Altogether, the final distribution has more entropy than P, and the amount of the added entropy is at least the amount that H adds to entropy-deficient distributions.

The second case is an entropy-deficient distribution $P = (P_1, P_2)$ that is "uniform over clouds"—a distribution that assigns the same probability to any two vertices inside the same cloud, i.e., for all $v \in [N_1]$ and $i, j \in [N_2]$, $P(v, i) = P(v, j)$. We call such a distribution a *parallel* distribution. Notice that since P is entropy-deficient, P_1 is entropy-deficient as well (since P_2 is uniform). Consider what happens when we apply the Markov chain defined by $G_1 ⓩ H$ on a parallel distribution:

- The first H step keeps P unchanged because H is a regular graph and hence it maps the uniform distribution on $[N_2]$ to itself.[2]
- The G_1 step is a permutation and does not change the entropy of P. As P is uniform over clouds, for any $v_1 \in [N_1]$ in the support of P_1, the conditional distribution $(P_2|P_1 = v)$ over the second component is uniform. Hence, the G_1 step maps any such cloud v_1 uniformly to the neighboring clouds, which are the clouds associated with the neighbors of v_1 in G_1. As G_1 is a good expander (and P_1 is entropy-deficient), the entropy of the first component of the distribution increases. Since the total entropy is unchanged, we conclude that the entropy in the second component decreases.
- Finally, as the second H step is applied when the second component is entropy-deficient, it must add entropy. Thus, the three steps together increase the entropy.

The algebraic analysis we describe next also works by analyzing two special cases—two orthogonal subspaces, each roughly corresponding to one of the above cases. In fact, the algebraic analysis also shows that it is sufficient to consider these two subspaces, as any vector can be decomposed to the sum of its two projections on them.

The algebraic analysis. Let $\tilde{H}$ denote the operator corresponding to an intracloud step and let $\dot{G}_1$ denote the transformation corresponding to an intercloud step. Namely, $\tilde{H}$ is the transition matrix of the subgraph of the replacement product graph that contains only the intracloud edges, and $\dot{G}_1$ is the transition matrix of the subgraph that contains all the intercloud edges. In particular, a linear combination of $\tilde{H}$ and $\dot{G}_1$ (with appropriate normalization according to the degrees) is the transition matrix of the replacement product graph. The transformation associated with the zig-zag product graph is $\tilde{H}\dot{G}_1\tilde{H}$, corresponding to an intracloud step followed by an intercloud and another intracloud steps.

The spectral gap of $G_1 ⓩ H$ is $1 - \bar{\lambda}$, where $\bar{\lambda}$ is the second largest eigenvalue of $\tilde{H}\dot{G}_1\tilde{H}$ in absolute value. We can write $\bar{\lambda}$ as

$$\bar{\lambda} = \max_{a,b \perp \mathbf{1}, \|a\|=\|b\|=1} |a^\dagger \tilde{H}\dot{G}_1\tilde{H}b|.$$

Our goal is to show that $\bar{\lambda}$ is small.

We consider vectors coming from two orthogonal subspaces. The first is the vector space $\mathcal{V}^{||}$ of all vectors a that $\tilde{H}$ keeps in place. A vector $a \in \mathbb{C}^{N_1 D_1}$ belongs to this

[2] One may interpret this as saying that the first H step is "wasted," and thus even though the degree of $G ⓩ H$ is D_2^2, the quality of the graph can be at most that of the best graph with degree D_2.

subspace if $a_{v,i} = a_{v,j}$ for all $v \in [N_1]$ and $i, j \in [N_2]$. We call such a vector a *parallel* vector. (Notice that every parallel distribution, when represented as a vector, is contained in $\mathcal{V}^{||}$.) The second vector space is the orthogonal complement of $\mathcal{V}^{||}$, denoted by $\mathcal{V}^{\perp}$. The vectors in $\mathcal{V}^{\perp}$ are called *perpendicular* vectors.

Now,

- if a is a parallel vector, then $\tilde{H}a = a$;
- if a is a perpendicular vector, then $\|\tilde{H}a\| \leq \bar{\lambda}(H)\|a\|$;
- if both a and b are parallel, then $a\dot{G}_1 b$ is essentially equivalent to the operation of G_1 on a and b.[3]

We need to analyze $a^{\dagger}\tilde{H}\dot{G}_1\tilde{H}b$ for arbitrary unit vectors a, b perpendicular to $\mathbf{1}$. We decompose a and b to their parallel and perpendicular components, denoting $a = a^{||} + a^{\perp}$, $b = b^{||} + b^{\perp}$, and we get four terms as follows:

- In the term $(a^{||})^{\dagger}\tilde{H}\dot{G}_1\tilde{H}b^{||}$, the operator $\dot{G}_1$ acts on parallel vectors (since they are unchanged by $\tilde{H}$) and is essentially identical to the operation of G. A simple analysis shows that this term contributes at most $\bar{\lambda}(G_1)$. In other words, for $b \in \mathcal{V}^{||}$, the parallel component of $\dot{G}_1 b$ is very small and therefore $\dot{G}_1 b$ lies almost entirely in $\mathcal{V}^{\perp}$.
- In the terms $(a^{||})^{\dagger}\tilde{H}\dot{G}_1\tilde{H}b^{\perp}$ and $(a^{\perp})^{\dagger}\tilde{H}\dot{G}_1\tilde{H}b^{||}$, one H step shrinks a perpendicular vector by a factor of at least $\bar{\lambda}(H)$. Thus, this term contributes at most $\bar{\lambda}(H)$.
- In the term $(a^{\perp})^{\dagger}\tilde{H}\dot{G}_1\tilde{H}b^{\perp}$, both H steps shrink $a^{\perp}$ and $b^{\perp}$ by a factor of at least $\bar{\lambda}(H)$. Hence, this term contributes at most $\bar{\lambda}(H)^2$.

Therefore, altogether, $\bar{\lambda}(\tilde{H}\dot{G}_1\tilde{H}) \leq \bar{\lambda}(G) + 2\bar{\lambda}(H) + \bar{\lambda}(H)^2$. A tighter analysis somewhat improves this bound.

The nonoptimality of the zig-zag product stems from the fact that the degree of the zig-zag graph is D_2^2, corresponding to *two* steps on H, while the guaranteed spectral gap is dominated by a term of magnitude $\bar{\lambda}(H)$, corresponding to a *single* step on H. Looking at the analysis, we see that this may happen, e.g., when a is parallel and b is perpendicular. In this case $\tilde{H}a = a$, so one H step is lost. Also, as b is perpendicular, $\tilde{H}$ shrinks it by a factor of $\bar{\lambda}(H)$. Now we are left with the inner product between $\dot{G}_1$ applied to a and some arbitrary perpendicular vector. This inner product can be very close to 1 and this means that the entire expression cannot be smaller than $\bar{\lambda}(H)$.

1.1.3. The *k*-step zig-zag product. In this paper we consider the variant of the zig-zag product, where we take k steps on H rather than just two steps. That is, we consider the graph whose transition matrix is $\tilde{H}\dot{G}_1\tilde{H}\ldots\tilde{H}\dot{G}_1\tilde{H}$ with k steps on H. How small is the second largest eigenvalue going to be? Analyzing each $\tilde{H}\dot{G}_1\tilde{H}$ term on its own, we see that the second largest eigenvalue is at most about $\bar{\lambda}(H)^{\lfloor k/2 \rfloor}$. Clearly, we must lose at least one H step, e.g., if we start with a parallel vector. Our goal is to find a variant of the construction where the second largest eigenvalue is at most about $\bar{\lambda}(H)^{k-1}$.

The problem. Let us consider what happens when we take three H steps. The operator we consider is $\tilde{H}\dot{G}_1\tilde{H}\dot{G}_1\tilde{H}$, and to bound the spectral gap we look at $a^{\dagger}\tilde{H}\dot{G}_1\tilde{H}\dot{G}_1\tilde{H}b$. We focus on the case where b is a parallel distribution.

- The first H step is lost (because b is parallel).

[3] Formally, the vectors a and b belong to $\mathbb{C}^{N_1 D_1}$, while G_1 acts on the space $\mathbb{C}^{N_1}$. However, in the introduction we choose to ignore this technical issue, so as not to obscure the ideas underlying the analysis.

- This is immediately followed by an intercloud step which propagates entropy from the second component (within the cloud) to the first component (the distribution over clouds). Equivalently, in algebraic notation, $\dot{G}_1\tilde{H}b$ lies almost entirely in $\mathcal{V}^{\perp}$.
- Next we apply a second H step which adds entropy (because the second component is entropy-deficient). Notice that $\tilde{H}\dot{G}_1\tilde{H}b$ lies almost entirely in $\mathcal{V}^{\perp}$, as $\mathcal{V}^{\perp}$ is invariant under $\tilde{H}$.
- Following that we apply $\dot{G}_1$ again. The first $\dot{G}_1$ application was applied on a parallel vector, and because of that we knew that it increases the entropy of the first component and decreases the entropy of the second component. However, now the $\dot{G}_1$ operator is applied on a vector (mostly) from $\mathcal{V}^{\perp}$, and therefore we have no guarantee on the output and, in particular, it is possible that the resulting vector $\dot{G}_1\tilde{H}\dot{G}_1\tilde{H}b$ lies in $\mathcal{V}^{||}$; i.e., applying $\dot{G}_1$ *increases* the entropy of the second component and *decreases* the entropy of the first component. In other words, we might have entropy flowing *backwards*. If this happens, then the final $\tilde{H}$ step is wasted again. Thus, we have three H steps, but only one is guaranteed to add entropy.

The first idea. We would like to make sure that entropy does not flow in the wrong direction. That is, our goal is to guarantee that whenever an H step does not add entropy, *all* of the following $\dot{G}_1$ applications move entropy from the second component (the distribution within the cloud) to the first component (the distribution over clouds). If we can guarantee that, then a single failure of an H step guarantees that all other H steps are successful.

When an H step fails, the distribution over the second component is close to uniform and contains about $\log(|V_2|)$ bits of entropy. To facilitate the above idea, we make the second component large enough such that $\log(|V_2|)$ bits of entropy suffice for a k-step random walk on G. For example, we can make the cloud size $|V_2|$ equal D_1^{4k}. The graph G_1 still has degree D_1, and so when the second component is uniform, it contains enough entropy for taking k independent steps on G_1.

Sure enough, now the size of V_2 is not the same as D_1 and we need to specify how to translate a cloud vertex (indexed by $[D_1]^{4k}$) to an edge label (indexed by $[D_1]$). For concreteness, let us assume that we take the edge label from the first $\log(D_1)$ bits of the cloud vertex. Now, all we need for the operator $\dot{G}_1$ to move entropy in the right direction is that the second component is uniform *only* on its first few bits.

Let us take a closer look at the situation. We start with a uniform distribution over the second component (because we are considering the case where $\tilde{H}$ fails) with about $4k\log(D_1)$ entropy. We apply $\dot{G}_1$, and up to $\log(D_1)$ entropy flows from the second component to the first component. Thus, there is still a lot of entropy left in the second component. We now apply $\tilde{H}$. Our goal is to guarantee that $\tilde{H}$ moves the entropy in the second component such that the first $\log(D_1)$ bits become close to uniform. If this happens, then the next $\dot{G}_1$ application moves more entropy from the second component to the first component and the entropy keeps flowing in the "right" direction.

A second problem. What does it take for the above idea to work? A simple probabilistic method argument shows that for any fixed distribution on the second component that has a lot of entropy, most small degree graphs H will indeed be good and make the first $\log(D_1)$ bits close to uniform.

Our problem is that we need to deal with more than just one fixed distribution. Instead, the distribution on the second component is determined by the action of $\tilde{H}\dot{G}_1$, and as $\dot{G}_1$ may correlate the first and second components, the distribution

on the second component may depend on the value of the first component. This is problematic to us because from our point of view D_1 and k are constants while N_1 is a growing parameter. Thus, it seems inevitable that for any graph H there exists some value of the first component for which H fails. Therefore, it seems that this approach is bound to fail.

A second idea. To solve the above problem we restrict ourselves to graphs G_1 of a special type. For example, let us assume for simplicity that the labeling in G_1 is such that if $e = (v, w)$ is the ith edge leaving v, then e is also the ith edge leaving w (the actual property we use is a bit weaker). In such graphs the operator $\dot{G}_1$ has no effect at all on the second component. Thus, in particular, the number of distributions we have to work with does not depend on N_1 and the probabilistic argument mentioned above works.

Indeed, for such nicely labeled graphs G_1, and using k different graphs H_i instead of the single graph H used above for all the k steps, one can easily show that random D_2-regular graphs $(H_1, \ldots, H_k)$ are good. Namely, if we start with a parallel vector (or, equivalently, if an H step fails), then the following G steps constantly move entropy from the second component to the first component, and later H steps are not wasted.

Redoing the analysis with algebraic notation. Let us now state the first problem above in algebraic notation. Starting with a vector $b \in \mathcal{V}^{||}$, we know that $\tilde{H}b \in \mathcal{V}^{||}$. Moreover, $\dot{G}_1\tilde{H}b$ and $\tilde{H}\dot{G}_1\tilde{H}b$ (mostly) belong to $\mathcal{V}^{\perp}$. The question is whether we can guarantee that $\dot{G}_1\tilde{H}_2\dot{G}_1\tilde{H}_1 b$ also (mostly) belongs to $\mathcal{V}^{\perp}$, in which case the next $\tilde{H}$ operator will work for us.

Recall that $\dot{G}_1\tilde{H}b$ mostly belongs to $\mathcal{V}^{\perp}$ due to the fact that for parallel unit vectors $a, b \in \mathcal{V}^{||}$, $a^{\dagger}\dot{G}_1 b$ behaves like the action of G_1 on a, b, and hence

$$|a^{\dagger}\dot{G}_1 b| \leq \bar{\lambda}(G_1). \tag{1}$$

In particular, $\dot{G}_1 b$ has only a very small parallel component and mostly belongs to $\mathcal{V}^{\perp}$. Our main technical lemma states that in our variant of the zig-zag product, for any parallel unit vectors $a, b \in \mathcal{V}^{||}$, $a^{\dagger}\dot{G}_1\tilde{H}\dot{G}_1 b$ behaves like the action of G_1^2 on a, b, and this implies

$$|a^{\dagger}\dot{G}_1\tilde{H}\dot{G}_1 b| \leq \bar{\lambda}(G_1)^2. \tag{2}$$

In particular, $\dot{G}_1\tilde{H}\dot{G}_1 b$ has only a very small parallel component and mostly belongs to $\mathcal{V}^{\perp}$. This is stated and proved in section 4.1. The above phenomenon generalizes to an arbitrary number of steps; i.e., for any parallel unit vectors $a, b \in \mathcal{V}^{||}$, $a^{\dagger}(\dot{G}_1\tilde{H})^{k-1}\dot{G}_1 b$ behaves like the action of G_1^{k-1} on a, b, and this implies $|a^{\dagger}(\dot{G}_1\tilde{H})^{k-1}\dot{G}_1 b| \leq \bar{\lambda}(G_1)^k$. Figure 2 illustrates the entire process that parallel vectors undergo.

Armed with that we go back to the zig-zag analysis. Doing it carefully, we get that composing G_1 (of degree D_1 and second eigenvalue λ_1) with k graphs H_i (each of degree D_2 and second eigenvalue λ_2) we get a new graph with degree D_2^k and second eigenvalue about $\lambda_2^{k-1} + \lambda_2^k + 2\lambda_1$. We can think of λ_1 as being arbitrarily small, as it can be decreased to any constant by increasing D_1 without affecting D_2 and the degree of the resulting graph (see subsection 1.1.4 for a more detailed explanation). One can interpret the above result as saying that $k-1$ out of the k steps worked for us!

1.1.4. An almost-Ramanujan expander construction. We now go back to the iterative expander construction of [21] and replace the zig-zag component there

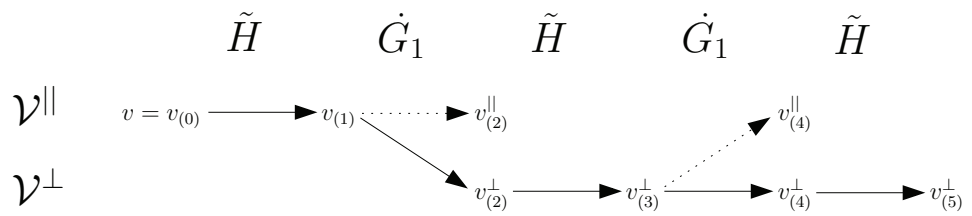

FIG. 2. *The action of a 3-step zig-zag on a parallel vector $v \in \mathcal{V}^{||}$. The process is composed of five steps, and $v_{(t)}$ denotes the vector after the tth step. Notice that $\tilde{H}$ sends perpendicular (parallel) vectors to perpendicular (parallel) vectors. Also, the parallel part of $\dot{G}_1 v_{(1)}$ is small by* (1). *Similarly, the parallel part of $\dot{G}_1 v^{\perp}_{(3)}$ is small by* (2). *These small parts are represented by dotted lines.*

with the k-step zig-zag product. We wish to construct D-regular graphs with second eigenvalue that is as close as possible to the optimal $\lambda_{\mathrm{Ram}}(D) = \frac{2\sqrt{D-1}}{D}$. For simplicity we start with the case where $D = D_2^k$ for some integer k (the general case is addressed in section 7). Doing the iterative construction we get a degree D expander by taking k steps over the graphs $\{H_i\}$, each of degree D_2 and over D_1^{4k} vertices. It will turn out that in this iterative construction we have the freedom of choosing any large enough D_1 without affecting the degree of the resulting graph. Taking D_1 large enough, the resulting second largest eigenvalue is about λ_2^{k-1}, where $\lambda_2 = \lambda_{\mathrm{Ram}}(D_2) = \frac{2\sqrt{D_2-1}}{D_2}$.

Comparing the second largest eigenvalue that we get with the optimal one, we see that the bound we get is roughly $2^{k-1} D_2^{-(k-1)/2}$, whereas the bound we would have liked to get is roughly $2D_2^{-k/2}$ (the optimal value for graphs with degree D_2^k). We do not achieve the optimal value for two reasons. First, we lose one application of H out of the k applications, and this loss amounts roughly to a $\sqrt{D_2}$ multiplicative factor. We also have a second loss of a 2^{k-1} multiplicative factor that corresponds to the fact that H^k is not optimal even when H is. Balancing these losses gives the following theorem.

THEOREM 1. *For every $D > 0$, there exists a fully explicit family of graphs $\{G_i\}$, with an increasing number of vertices, such that each G_i is D-regular and $\bar{\lambda}(G_i) \le D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})}$.*

1.2. Organization of the paper. In section 2 we give preliminary definitions. Section 3 contains the formal definition of the k-step zig-zag product. Section 4 gives the proof of the main statement regarding the k-step zig-zag product, assuming that good small graphs exist. The fact that such graphs exist is proven in section 5. In section 6 we use the product to give an iterative construction of expanders for degrees of a specific form. Finally, in section 7 we describe how to make the expander construction work for any degree.

2. Preliminaries.

Spectral gap. We associate a (directed or undirected) graph $G = (V, E)$ with its transition matrix, also denoted by G, i.e.,

$$G_{v,u} = \begin{cases} \frac{1}{\deg_{\mathrm{out}}(v)}, & (v,u) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

For a matrix G we denote by $s_i(G)$ the ith largest singular value of G. If the graph G is regular (i.e., $\deg_{\mathrm{in}}(v) = \deg_{\mathrm{out}}(v) = D$ for all $v \in V$), then $s_1(G) = 1$.

We define $\bar{\lambda}(G) = s_2(G)$. We say that a graph G is an (N, D, λ) graph if it is D-regular over N vertices and $\bar{\lambda}(G) \leq \lambda$. We sometimes omit the parameter N and say that G is a (D, λ) graph. If G is undirected, then the matrix G is Hermitian, G has an orthonormal eigenvector basis, and the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ are real. In this case,

$$\bar{\lambda}(G) = s_2(G) = \max\{\lambda_2, -\lambda_N\}.$$

We say that an undirected, D-regular graph G is *Ramanujan* if

$$\bar{\lambda}(G) \leq \lambda_{\mathrm{Ram}}(D) \stackrel{\mathrm{def}}{=} \frac{2\sqrt{D-1}}{D}.$$

Ramanujan graphs are essentially optimal algebraic expanders [18].

We can convert a directed graph G to an undirected graph U by undirecting the edges, i.e., $U = \frac{1}{2}[G+G^\dagger]$. If G is D-regular, then $\mathbf{1} \stackrel{\mathrm{def}}{=} \frac{1}{\sqrt{N}}(1, \ldots, 1)^t$ is an eigenvector of both G and $G^\dagger$. Therefore,

$$s_2(U) = \frac{1}{2}s_2(G+G^\dagger) = \frac{1}{2}\max_{u,v\perp\mathbf{1},\|u\|=\|v\|=1} |u^\dagger(G+G^\dagger)v| \leq \frac{1}{2}(s_2(G)+s_2(G^\dagger)) = s_2(G),$$

and it follows that U is an $(N, 2D, \lambda)$ graph.

Taking the transition matrix of a graph G and raising it to some power ℓ results in the transition matrix of another graph. This graph has the same set of vertices as G, and two vertices in this graph are connected if and only if there is a path of length ℓ between them in G.

FACT 2. *If G is an (N, D, λ) graph, then G^ℓ is an $(N, D^\ell, \lambda^\ell)$ graph.*

Similarly, taking the tensor product of the transition matrices of two graphs G_1 and G_2 results in the transition matrix of another graph. The set of vertices of this graph is the direct product of the sets of vertices of G_1 and G_2. In this graph, (v_1, v_2) is connected to (u_1, u_2) if and only if v_j is connected to u_j in G_j for $j = 1, 2$.

FACT 3. *If G_j is an (N_j, D_j, λ_j) graph for $j = 1, 2$, then $G_1 \otimes G_2$ is an $(N_1 \cdot N_2, D_1 \cdot D_2, \max\{\lambda_1, \lambda_2\})$ graph.*

Rotation maps. Following [21] we represent graphs using *rotation maps*, as we explain now. Let G be a directed D-regular graph $G = (V, E)$. Recall that $G^\dagger$ denotes the graph where the direction of each edge in E is reversed. We assume that the outgoing edges of G and $G^\dagger$ are labeled with D labels $\{1, \ldots, D\}$ such that for every $v \in V$, its D outgoing edges (either in G or in $G^\dagger$) are labeled with different labels. Let $v_G[i]$ denote the ith neighbor of v in G. We define the rotation map of G, $\mathrm{Rot}_G : V \times [D] \to V \times [D]$, by

$$\mathrm{Rot}_G(v, i) = (w, j) \iff v_G[i] = w \text{ and } w_{G^\dagger}[j] = v.$$

In other words, the ith neighbor of v in G is w, and the jth neighbor of w in the reversed graph $G^\dagger$ is v. Notice that if $\mathrm{Rot}_G(v, i) = (w, j)$, then $\mathrm{Rot}_{G^\dagger}(w, j) = (v, i)$.

The standard choice for the rotation maps of the graphs resulting from the operations of powering, tensoring, and undirecting is as follows.

Powering. If for all $j = 1, \ldots, \ell$ we have $\mathrm{Rot}_G(v_j, i_j) = (v_{j+1}, i'_j)$, then

$$\mathrm{Rot}_{G^\ell}(v_1, (i_1, \ldots, i_\ell)) = (v_{\ell+1}, (i'_\ell, \ldots, i'_1)). \tag{3}$$

Tensoring. If for $j = 1, 2$ we have $\mathrm{Rot}_{G_j}(v_j, i_j) = (u_j, i'_j)$, then

$$\mathrm{Rot}_{G_1\otimes G_2}((v_1, v_2), (i_1, i_2)) = ((u_1, u_2), (i'_1, i'_2)). \tag{4}$$

Undirecting. For a directed graph G we have

$$\text{Rot}_G(v,i) = (u,i') \Longrightarrow \text{Rot}_{\frac{1}{2}[G+G^\dagger]}(v,(b,i)) = (u,(1-b,i')). \quad (5)$$

[4]

We single out a special family of rotation functions.

DEFINITION 4. *A graph G is* locally invertible *if its rotation map is of the form $\text{Rot}_G(v,i) = (v[i],\phi(i))$ for some permutation $\phi : [d] \to [d]$. We say that ϕ is the local inversion function.*

In [20], a "π-consistently labeled graph" denotes a graph with local inversion π. Thus, a graph is locally invertible if and only if it is π-consistently labeled for some permutation π.[5]

A simple fact follows immediately from (3)–(5).

FACT 5. *If G_1, G_2 are locally invertible, then G_1^ℓ, $G_1 \otimes G_2$, and $\frac{1}{2}[G_1 + G_1^\dagger]$ are locally invertible.*

Miscellaneous notation. We often use vectors coming from a tensor vector space $\mathcal{V} = \mathcal{V}_1 \otimes \mathcal{V}_2$, as well as vertices coming from a product vertex set $V = V_1 \times V_2$. In such cases we use superscripts to indicate the universe a certain object resides in. For example, we denote vectors from $\mathcal{V}_1$ by $x^{(1)}$, $y^{(1)}$, etc. In particular, when $x \in \mathcal{V} = \mathcal{V}_1 \otimes \mathcal{V}_2$ is a product vector, $x^{(1)}$ denotes the $\mathcal{V}_1$ component, $x^{(2)}$ denotes the $\mathcal{V}_2$ component, and $x = x^{(1)} \otimes x^{(2)}$.

We denote by $\mathbf{1}_\mathcal{V}$ the all-ones vector over the vector space $\mathcal{V}$, normalized to have unit length. When the vector space is clear from the context we simply denote this vector by $\mathbf{1}$.

$\mathbb{S}_\Lambda$ denotes the symmetric group over Λ. $G_{N,D}$, for an even D, is the following distribution over D-regular, undirected graphs. First, uniformly choose $D/2$ permutations $\gamma_1, \ldots, \gamma_{D/2} \in \mathbb{S}_{[N]}$. Then, output the graph $G = (V = [N], E)$, whose edges are the *undirected* edges formed by the $D/2$ permutations.

Finally, for an n-dimensional vector x we let $|x|_1 = \sum_{i=1}^n |x_i|$ and $\|x\| = \sqrt{\langle x, x\rangle}$. We measure the distance between two distributions P, Q by $|P - Q|_1$. The operator norm of a linear operator L is $\|L\|_\infty = \max_{x:\|x\|=1} \|Lx\|$.

We will need the following claim, which states that if we average linear operators according to two statistically close distributions, we get essentially the same linear operator.

CLAIM 6. *Let P, Q be two distributions over Ω and let $\{\mathcal{L}_i\}_{i\in\Omega}$ be a set of linear operators over Λ, each with operator norm bounded by 1. Define $\mathcal{P} = \mathbb{E}_{x\sim P}[\mathcal{L}_x]$ and $\mathcal{Q} = \mathbb{E}_{x\sim Q}[\mathcal{L}_x]$. Then, for any $\tau, \xi \in \Lambda$,*

$$|\langle \mathcal{P}\tau, \xi\rangle - \langle \mathcal{Q}\tau, \xi\rangle| \le |P - Q|_1 \cdot \|\tau\| \cdot \|\xi\|.$$

Proof. First, notice that

$$\|\mathcal{P} - \mathcal{Q}\|_\infty \le \sum_x |P(x) - Q(x)| \cdot \|\mathcal{L}_x\|_\infty \le |P - Q|_1.$$

[4] Undirecting a graph doubles the degree. This is reflected in the addition of a bit $b \in \{0,1\}$ to the edge label.

[5] Perhaps a more appropriate name for "locally invertible graph" is "consistently labeled graph" (without the addition of the permutation π). However, the term "consistently labeled graph" is already used in the literature to denote a different property of the labeling of the edges [9]. An example of a graph that is consistently labeled, yet is not locally invertible, can be observed by taking the disjoint union of two graphs of the same degree that are locally invertible, each with a different inversion function.

Therefore, it follows that

$$|\langle \mathcal{P}\tau,\xi\rangle - \langle \mathcal{Q}\tau,\xi\rangle| = |\langle (\mathcal{P}-\mathcal{Q})\tau,\xi\rangle| \le \|\mathcal{P}-\mathcal{Q}\|_\infty \cdot \|\tau\| \cdot \|\xi\| \le |P-Q|_1 \cdot \|\tau\| \cdot \|\xi\|. \qquad \square$$

3. The k-step zig-zag product.

3.1. The product. The input to the product is

- an undirected graph $G_1 = (V_1 = [N_1], E_1)$ that is a (D_1, λ_1) graph. We assume that G_1 has a local inversion function $\phi = \phi_{G_1}$. That is, $\mathrm{Rot}_{G_1}(v^{(1)}, d_1) = (v^{(1)}[d_1], \phi_{G_1}(d_1))$;
- k undirected graphs $\bar{H} = (H_1, \ldots, H_k)$, where each H_i is an (N_2, D_2, λ_2) graph over the vertex set V_2.

In the replacement product (and also in the zig-zag product), the parameters are set such that the cardinality of V_2 equals the degree D_1 of G_1. An element $v_2 \in V_2$ is then interpreted as a label $d_1 \in [D_1]$. However, as explained in the introduction, we take larger graphs H_i with $V_2 = [D_1]^{4k}$. That is, we have D_1^{4k} vertices in V_2 rather than D_1 in the replacement product. Therefore, we need to explain how to map a vertex $v^{(2)} \in V_2 = [D_1]^{4k}$ to a label $d_1 \in [D_1]$ of G_1. For that we use a map $\pi : V_2 \to [D_1]$ that is *regular*, i.e., every element of $[D_1]$ has the same number of π preimages in V_2. For simplicity we fix one concrete such π as follows. For $j \in [4k]$ and $w = (w_1, \ldots, w_{4k}) \in [D_1]^{4k}$, we define π_j to be the projection of w on the jth coordinate, i.e., $\pi_j(w) = w_j$. The map π that we choose is $\pi = \pi_1$.

The graph $G_{\mathrm{new}} = G_1 \textcircled{z} \bar{H}$ that we construct is related to a k-step walk over this new replacement product. The vertices of G_{new} are $V_1 \times V_2$. The degree of the graph is D_2^k, and the edges are indexed by $\bar{i} = (i_1, \ldots, i_k) \in [D_2]^k$. We next define the rotation map $\mathrm{Rot}_{G_{\mathrm{new}}}$ of the new graph. For $v = (v^{(1)}, v^{(2)}) \in V_1 \times V_2$ and $\bar{i} = (i_1, \ldots, i_k) \in [D_2]^k$, $\mathrm{Rot}_{G_{\mathrm{new}}}(v, \bar{i})$ is defined as follows:

- We start the walk at $v = (v^{(1)}, v^{(2)}) = (v_0^{(1)}, v_0^{(2)})$.
- For $t = 1, \ldots, k$,
  - take one $H_t(\cdot, i_t)$ step on the second component. That is, the first component is left untouched, $v_{2t-1}^{(1)} = v_{2(t-1)}^{(1)}$, and we set $(v_{2t-1}^{(2)}, i_t') = \mathrm{Rot}_{H_t}(v_{2(t-1)}^{(2)}, i_t)$;
  - if $t < k$, we take one step on G_1 with $\pi_1(v_{2t-1}^{(2)})$ as the $[D_1]$ label to be used, i.e.,

$$v_{2t}^{(1)} = v_{2t-1}^{(1)}[\pi_1(v_{2t-1}^{(2)})].$$

    We also set $v_{2t}^{(2)} = \psi(v_{2t-1}^{(2)})$, where

$$\psi(v^{(2)}) = (\phi_{G_1}(\pi_1(v^{(2)})), \pi_2(v^{(2)}), \pi_3(v^{(2)}), \ldots, \pi_{4k}(v^{(2)})). \tag{6}$$

    Namely, for the first $[D_1]$ coordinate of the second component we use the local inversion function of G_1, and all other coordinates are left unchanged.

Finally, we specify

$$\mathrm{Rot}_{G_{\mathrm{new}}}(v, \bar{i}) = \left((v_{2k-1}^{(1)}, v_{2k-1}^{(2)}), (i_k', \ldots, i_1')\right). \tag{7}$$

It is straightforward to verify that $\mathrm{Rot}_{G_{\mathrm{new}}}$ is indeed a rotation map.

To summarize, we start with a locally invertible D_1-regular graph over N_1 vertices. We replace each degree D_1 vertex with a "cloud" of D_1^{4k} vertices, and map a cloud vertex to a D_1 instruction using π_1. We then take a $(2k-1)$-step walk, with alternating H and G_1 steps, over the resulting graph. Observe that the resulting graph is *directed* since, for instance, H_1 might be different from H_k. One can obtain an undirected graph simply by undirecting each edge. This transformation doubles the degree while retaining the spectral gap, as explained in section 2.

The following is immediate from the definition of the rotation map in (7).

FACT 7. *If, for $1 \le i \le k$, the graph H_i is locally invertible with the local inversion function ϕ_{H_i}, then $G_1 ⓩ (H_1, \ldots, H_k)$ is locally invertible with the local inversion function*

$$\phi(i_1, \ldots, i_k) = (\phi_{H_k}(i_k), \ldots, \phi_{H_1}(i_1)).$$

3.2. The linear operators. We want to express the k-step walk described in section 3.1 as a composition of linear operators. For $i \in \{1, 2\}$, we define a vector space $\mathcal{V}_i$ with $\dim(\mathcal{V}_i) = |V_i| = N_i$, and we identify an element $v^{(i)} \in V_i$ with a basis vector $\overrightarrow{v^{(i)}} \in \mathcal{V}_i$. Notice that

$$\left\{ \overrightarrow{v^{(1)}} \otimes \overrightarrow{v^{(2)}} \mid v^{(1)} \in V_1,\, v^{(2)} \in V_2 \right\}$$

is a basis for $\mathcal{V} = \mathcal{V}_1 \otimes \mathcal{V}_2$. On this basis we define the linear operators

$$\tilde{H}_i \left(\overrightarrow{v^{(1)}} \otimes \overrightarrow{v^{(2)}} \right) = \overrightarrow{v^{(1)}} \otimes \overrightarrow{H_i v^{(2)}}$$

and

$$\dot{G}_1 \left(\overrightarrow{v^{(1)}} \otimes \overrightarrow{v^{(2)}} \right) = \overrightarrow{v^{(1)}[\pi_1(v^{(2)})]} \otimes \overrightarrow{\psi(v^{(2)})},$$

where ψ is as defined in (6). Having this terminology, the transition matrix of the new graph G_{new} is the linear transformation on $\mathcal{V}$ defined by $\tilde{H}_k \dot{G}_1 \tilde{H}_{k-1} \dot{G}_1 \ldots \tilde{H}_2 \dot{G}_1 \tilde{H}_1$.

3.3. The action of the composition. Next we take advantage of the simple structure of locally invertible graphs, revealing how $\tilde{H}_k \dot{G}_1 \tilde{H}_{k-1} \dot{G}_1 \ldots \tilde{H}_2 \dot{G}_1 \tilde{H}_1$ correlates the first and second components. Fix a D_1-regular graph G_1 with local inversion function ϕ. As G_1 is D_1-regular, it can be represented as

$$G_1 = \frac{1}{D_1} \sum_{i=1}^{D_1} \mathcal{G}_i,$$

where $\mathcal{G}_i$ is the transition matrix of some permutation in $\mathbb{S}_{V_1}$. We can similarly decompose each graph H_i to a sum of D_2 permutations on V_2. We focus our attention on the case where the action of H_i is replaced with a single permutation, and the general case (where each H_i is a convex combination of D_2 permutations) follows by linearity.

LEMMA 8. *Assume that G_1 has a local inversion function ϕ that is extended to a permutation $\psi : V_2 \to V_2$ as in (6). Let $\gamma_1, \ldots, \gamma_\ell$ be ℓ permutations on V_2. Let Γ_i be the linear operator on $\mathcal{V}_2$ corresponding to the permutation γ_i and $\tilde{\Gamma}_i = I \otimes \Gamma_i$.*

For vertices $u^{(1)} \in V_1$, $u^{(2)} \in V_2$ *define* $w_0 = \overrightarrow{u^{(1)}} \otimes \overrightarrow{u^{(2)}}$ *and* $w_i = \tilde{\Gamma}_i \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 (\overrightarrow{u^{(1)}} \otimes \overrightarrow{u^{(2)}})$. *Then,* w_i *is a product vector, and*

$$w_i = \mathcal{G}_{\pi_1(q_{i-1}(u^{(2)}))} \dots \mathcal{G}_{\pi_1(q_0(u^{(2)}))} (\overrightarrow{u^{(1)}}) \otimes \overrightarrow{q_i(u^{(2)})},$$

where

$$q_0(u^{(2)}) = u^{(2)}, \tag{8}$$

$$q_i(u^{(2)}) = \gamma_i(\psi(q_{i-1}(u^{(2)}))). \tag{9}$$

Proof. We prove by induction. For $i = 0$, $w_0 = (\overrightarrow{u^{(1)}} \otimes \overrightarrow{u^{(2)}})$. The induction step follows immediately from the fact that $u_{i+1} = \tilde{\Gamma}_{i+1} \dot{G}_1 u_i$ and the definitions of $\tilde{\Gamma}_{i+1}$ and $\dot{G}_1$. ∎

We also capture from the proof the behavior $q_i(u^{(2)})$ of the second component values. We define the following.

DEFINITION 9. *Let* G_1 *be an undirected graph with local inversion function* ϕ *that is extended to a permutation* $\psi : V_2 \to V_2$ *as in* (6). *Let* $\bar{\gamma} = (\gamma_1, \dots, \gamma_\ell)$ *be a sequence of* ℓ *permutations over* V_2. *The permutation sequence* induced by $(\bar{\gamma}, \phi)$ *is* $\bar{q} = (q_0, \dots, q_\ell)$, *defined as in* (8) *and* (9).

COROLLARY 10. *Assume that* G_1 *has a local inversion function* ϕ. *Let* $\gamma_1, \dots, \gamma_\ell$ *be* ℓ *permutations on* V_2. *Let* Γ_i *be the linear operator on* $\mathcal{V}_2$ *corresponding to the permutation* γ_i *and* $\tilde{\Gamma}_i = I \otimes \Gamma_i$.

Then, there exists $\sigma \in \mathbb{S}_{V_2}$ *such that for any* $u^{(1)} \in V_1$ *and* $u^{(2)} \in V_2$

$$\dot{G}_1 \tilde{\Gamma}_\ell \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 (\overrightarrow{u^{(1)}} \otimes \overrightarrow{u^{(2)}}) = \mathcal{G}_{\pi_1(q_\ell(u^{(2)}))} \dots \mathcal{G}_{\pi_1(q_0(u^{(2)}))} (\overrightarrow{u^{(1)}}) \otimes \overrightarrow{\sigma(u^{(2)})},$$

where $(q_0, \dots, q_\ell)$ *is the permutation sequence induced by* $((\gamma_1, \dots, \gamma_\ell), \phi)$.

3.4. A condition guaranteeing good algebraic expansion. We say that $\bar{\gamma} = (\gamma_1, \dots, \gamma_\ell)$ is ε-*pseudorandom* with respect to ϕ if the distribution of the first $\log(D_1)$ bits in each of the $\ell + 1$ labels we encounter is ε-close to uniform. We define the following.

DEFINITION 11. *Let* G_1 *be an undirected graph with local inversion function* ϕ. *Let* $\bar{q}$ *be the permutations induced by* $(\bar{\gamma}, \phi)$. *We say that* $\bar{\gamma}$ *is* ε-pseudorandom *with respect to* ϕ *(or, equivalently,* ε*-pseudorandom with respect to* G_1*) if*

$$\left| \pi_1(q_0(U)) \circ \dots \circ \pi_1(q_\ell(U)) - U_{[D_1]^{\ell+1}} \right|_1 \le \varepsilon,$$

where $\pi_1(q_0(U)) \circ \dots \circ \pi_1(q_\ell(U))$ *is the distribution obtained by picking* $v^{(2)} \in V_2$ *uniformly at random and outputting* $(\pi_1(q_0(v^{(2)})), \dots, \pi_1(q_\ell(v^{(2)})))$, *and* $U_{[D_1]^{\ell+1}}$ *is the uniform distribution over* $[D_1]^{\ell+1}$.

Any D_2-regular graph H can be expressed as $H = \frac{1}{D_2} \sum_{j=1}^{D_2} \mathcal{H}_j$, where $\mathcal{H}_j$ is the transition matrix of a permutation $\gamma_j \in \mathbb{S}_{V_2}$. We now extend Definition 11 to a sequence of k D_2-regular graphs.

DEFINITION 12. *Let* G_1 *and* ϕ *be as above. Let* $\bar{H} = (H_1, \dots, H_k)$ *be a* k*-tuple of* D_2*-regular graphs over* V_2. *We say that* $\bar{H}$ *is* ε-pseudorandom *with respect to* ϕ *(or* G_1*) if we can express each graph* H_i *as* $H_i = \frac{1}{D_2} \sum_{j=1}^{D_2} \mathcal{H}_{i,j}$ *such that*

- $\mathcal{H}_{i,j}$ *is the transition matrix of a permutation* $\gamma_{i,j} \in \mathbb{S}_{V_2}$;
- *for any* $1 \le \ell_1 \le \ell_2 \le k$, $j_{\ell_1}, \dots, j_{\ell_2} \in [D_2]$, *the sequence* $\gamma_{\ell_1, j_{\ell_1}}, \dots, \gamma_{\ell_2, j_{\ell_2}}$ *is* ε*-pseudorandom with respect to* ϕ.

If, in addition, for each $i = 1, \ldots, k$ *we have* $\bar{\lambda}(H_i) \leq \lambda_{\mathrm{Ram}}(D_2) + \varepsilon$*, we say that* $\bar{H}$ *is* ε-good *with respect to* ϕ *(or* G_1*).*

Our main result states that whenever $\bar{H}$ is good with respect to G_1, the k-step zig-zag product does not lose much in the spectral gap. Formally, we have the following result.

THEOREM 13. *Let* $G_1 = (V_1 = [N_1], E_1)$ *be a* (D_1, λ_1) *locally invertible graph. Let* $\bar{H} = (H_1, \ldots, H_k)$ *be a sequence of* $(N_2 = D_1^{4k}, D_2, \lambda_2)$ *graphs that is* ε*-good with respect to* G_1*, and assume* $\lambda_2 \leq \frac{1}{2}$*. Then,* $G_{\mathrm{new}} = G_1 (z) \bar{H}$ *is an* $(N_1 \cdot N_2, D_2^k, f(\lambda_1, \lambda_2, \varepsilon, k))$ *graph for*

$$f(\lambda_1, \lambda_2, \varepsilon, k) = \lambda_2^{k-1} + 2(\varepsilon + \lambda_1) + \lambda_2^k.$$

Our ultimate goal is constructing a family of D-regular graphs with a spectral gap as large as we can for some given degree $D = D_2^k$. In section 6 we will present an iterative construction doing that using the k-step zig-zag product. It will turn out that in this iterative construction we have the freedom of choosing any large enough D_1, without affecting the degree of the resulting graph.

Thus, we may think of D_2 and k as fixed quantities while D_1 is unbounded. By choosing D_1 large enough, we can guarantee that λ_1 is negligible compared to λ_2^k. It will also turn out that for this choice of D_1, we can find $\bar{H}$ that is ε-good for ε which is negligible compared to λ_2^k. Thus the graph we construct has $\bar{\lambda} \approx \lambda_2^{k-1} + \lambda_2^k$. In other words, we do k zig-zag steps and almost all of them ($k-1$ out of k) "work" for us.

Thus, we are left with two tasks:

1. Prove Theorem 13, which is done in the following section.
2. Find an ε-good sequence $\bar{H}$. In fact, in section 5, we prove that almost all sequences are good.

4. A top-down view of the proof.

Proof of Theorem 13. G_{new} is a regular directed graph, and we wish to bound $s_2(G_{\mathrm{new}})$. Fix unit vectors $x, y \perp \mathbf{1}$ for which $s_2(G_{\mathrm{new}}) = \langle G_{\mathrm{new}} x, y \rangle$. As in the analysis of the zig-zag product, we decompose $\mathcal{V} = \mathcal{V}_1 \otimes \mathcal{V}_2$ to its parallel and perpendicular parts. The subspace $\mathcal{V}^{||}$ is defined by

$$\mathcal{V}^{||} = \mathrm{Span}\left\{ \overrightarrow{v^{(1)}} \otimes \mathbf{1} : v^{(1)} \in V_1 \right\},$$

and $\mathcal{V}^{\perp}$ is its orthogonal complement. For any vector $\tau \in \mathcal{V}$ we denote by $\tau^{||}$ and $\tau^{\perp}$ the projections of τ on $\mathcal{V}^{||}$ and $\mathcal{V}^{\perp}$, respectively. Notice that $\mathcal{V}^{||}$ is exactly the set of parallel vectors defined in the introduction, and $\mathcal{V}^{\perp}$ is the set of perpendicular vectors. Also notice that $v \in \mathcal{V}^{||}$ if and only if $v = v_1 \otimes \mathbf{1}$ for some $v_1 \in \mathcal{V}_1$. Recall that for $v \in \mathcal{V}^{||}$, $Hv = v$, while for $v \in \mathcal{V}^{\perp}$, $\|\tilde{H} v\| \leq \lambda_2 \|v\|$.

For the analysis we decompose not only $x_0 = x$ and $y_0 = y$, but also the vectors $x_1, \ldots, x_{k-1}$ and $y_1, \ldots, y_{k-1}$, where

$$x_i = \dot{G}_1 \tilde{H}_i x_{i-1}^{\perp} \quad \text{and} \quad y_i = \dot{G}_1 \tilde{H}_{k-i+1}\, y_{i-1}^{\perp}.$$

Observe that $\|x_i\| \leq \lambda_2^i \|x_0\|$ and $\|y_i\| \leq \lambda_2^i \|y_0\|$.

We now consider $y_0^{\dagger} \tilde{H}_k \dot{G}_1 \ldots \tilde{H}_2 \dot{G}_1 \tilde{H}_1 x_0$ and decompose $x_0 = x_0^{||} + x_0^{\perp}$. Focusing on $x_0^{\perp}$ we see that, by definition,

$$y_0^{\dagger} \tilde{H}_k \dot{G}_1 \ldots \tilde{H}_2 \dot{G}_1 \tilde{H}_1 x_0^{\perp} = y_0^{\dagger} \tilde{H}_k \dot{G}_1 \ldots \tilde{H}_3 \dot{G}_1 \tilde{H}_2 x_1.$$

We continue by decomposing $x_1, x_2, \ldots$, and eventually this results in

$$y_0^\dagger \tilde{H}_k \dot{G}_1 \ldots \tilde{H}_2 \dot{G}_1 \tilde{H}_1 x_0 = y_0^\dagger \tilde{H}_k x_{k-1}^\perp + \sum_{i=1}^{k} y_0^\dagger \tilde{H}_k \dot{G}_1 \ldots \tilde{H}_{i+1} \dot{G}_1 \tilde{H}_i x_{i-1}^{||}.$$

Doing the same decomposition on y_0 (and using the fact that both $\dot{G}_1$ and $\tilde{H}_j$ are Hermitian and so $(y_j^\perp)^\dagger \tilde{H}_{k-j} \dot{G}_1 = (\dot{G}_1 \tilde{H}_{k-j} y_j^\perp)^\dagger = y_{j+1}^\dagger$), we get

$$\begin{aligned} y_0^\dagger \tilde{H}_k \dot{G}_1 \ldots \tilde{H}_2 \dot{G}_1 \tilde{H}_1 x_0 = y_0^\dagger \tilde{H}_k x_{k-1}^\perp + \sum_{i=1}^{k} (y_{k-i}^\perp)^\dagger x_{i-1}^{||} + \sum_{i=1}^{k} (y_{k-i}^{||})^\dagger x_{i-1}^{||} \\ + \sum_{1 \le i < j \le k} (y_{k-j}^{||})^\dagger \dot{G}_1 \tilde{H}_{j-1} \ldots \tilde{H}_{i+1} \dot{G}_1 x_{i-1}^{||}. \end{aligned}$$

Now,

- $$\left| y_0^\dagger \tilde{H}_k x_{k-1}^\perp \right| \le \left\| \tilde{H}_k x_{k-1}^\perp \right\| \le \lambda_2 \left\| x_{k-1}^\perp \right\| \le \lambda_2 \|x_{k-1}\| \le \lambda_2 \lambda_2^{k-1} \|x_0\| = \lambda_2^k.$$

  In the terminology of Figure 2 (from the introduction), this term corresponds to a path that always stays in $\mathcal{V}^\perp$. The value such a term contributes is small because each application of $\tilde{H}$ shrinks the vector by a λ_2 factor.
- The term $\sum_{i=1}^{k} (y_{k-i}^\perp)^\dagger x_{i-1}^{||}$ is simply 0 because $\mathcal{V}^\perp \perp \mathcal{V}^{||}$.
- Each term in the sum $\left| \sum_{i=1}^{k} (y_{k-i}^{||})^\dagger x_{i-1}^{||} \right|$ corresponds to a path in Figure 2 that stays in $\mathcal{V}^\perp$, except for one step where it goes to $\mathcal{V}^{||}$. Such a term is small since all applications of $\tilde{H}$ except one shrink the vector by a λ_2 factor. A trivial argument thus bounds the sum by $k\lambda_2^{k-1}$. We can do slightly better, i.e.,

  $$\left| \sum_{i=1}^{k} (y_{k-i}^{||})^\dagger x_{i-1}^{||} \right| \le \sum_{i=1}^{k} \left\| y_{k-i}^{||} \right\| \cdot \left\| x_{i-1}^{||} \right\|,$$

  and in Lemma 16 we bound the right-hand side by λ_2^{k-1}.
- Finally, we are left with the term

  $$\sum_{1 \le i < j \le k} (y_{k-j}^{||})^\dagger \dot{G}_1 \tilde{H}_{j-1} \ldots \tilde{H}_{i+1} \dot{G}_1 x_{i-1}^{||}. \tag{10}$$

  To gain insight about this term, consider $k = 3$ and the term $y_0^{||} \dot{G}_1 \tilde{H}_2 \dot{G}_1 x_0^{||}$. This term corresponds to the weight of the path between $v_{(1)}$ and $v_{(4)}^{||}$ in Figure 2. The core of the analysis shows that this term is small.
  Formally, in Theorem 14 we show, assuming that $\bar{H}$ is good for G_1, that

  $$\left| (y_{k-j}^{||})^\dagger \dot{G}_1 \tilde{H}_{j-1} \ldots \tilde{H}_{i+1} \dot{G}_1 x_{i-1}^{||} \right| \le (\lambda_1^{j-i} + \varepsilon) \left\| y_{k-j}^{||} \right\| \left\| x_{i-1}^{||} \right\|.$$

  Therefore, the term in (10) is bounded by

$$\begin{aligned} \sum_{1 \le i < j \le k} (\lambda_1^{j-i} + \varepsilon) \left\| y_{k-j}^{||} \right\| \left\| x_{i-1}^{||} \right\| &= \sum_{t=1}^{k-1} (\lambda_1^t + \varepsilon) \sum_{i=1}^{k-t} \left\| y_{k-i-t}^{||} \right\| \left\| x_{i-1}^{||} \right\| \\ \le \sum_{t=1}^{k-1} (\lambda_1^t + \varepsilon) \lambda_2^{k-t-1} \le (\lambda_1 + \varepsilon) \sum_{t=1}^{k-1} \lambda_2^{k-t-1} &= (\lambda_1 + \varepsilon) \sum_{i=0}^{k-2} \lambda_2^i \le 2(\lambda_1 + \varepsilon), \end{aligned}$$

where the first inequality follows from Lemma 16 and the second uses the assumption $\lambda_2 \leq \frac{1}{2}$.

Altogether, $|y^\dagger G_{\text{new}} x| \leq \lambda_2^{k-1} + 2(\varepsilon + \lambda_1) + \lambda_2^k$, as desired. ∎

4.1. The action of the operator on parallel vectors. The action of our operator on parallel vectors is captured in the following theorem.

THEOREM 14. *For every $i, \ell \geq 1$ and $\tau, \xi \in \mathcal{V}^{||}$, $\tau, \xi \perp \mathbf{1}_V$,*

$$\left|\left\langle \dot{G}_1 \tilde{H}_{i+\ell} \dot{G}_1 \dots \tilde{H}_{i+1} \dot{G}_1 \tau, \xi \right\rangle\right| \leq (\lambda_1^{\ell+1} + \varepsilon)\|\tau\|\|\xi\|.$$

For the proof we need the following lemma. Informally, it states that the action of $\dot{G}_1 \tilde{H}_{i+\ell} \dot{G}_1 \dots \tilde{H}_{i+1} \dot{G}_1$ on $\mathcal{V}^{||}$ is essentially the same as the action of $G^{\ell+1}$ on $\mathcal{V}_1$.

LEMMA 15. *Suppose that $\bar{\gamma} = (\gamma_1, \dots, \gamma_\ell)$ is ε-pseudorandom with respect to G_1, and denote by $\tilde{\Gamma}_1, \dots, \tilde{\Gamma}_\ell$ the operators corresponding to $\gamma_1, \dots, \gamma_\ell$. Any $\tau, \xi \in \mathcal{V}^{||}$ can be written as $\tau = \tau^{(1)} \otimes \mathbf{1}_{\mathcal{V}_2}$ and $\xi = \xi^{(1)} \otimes \mathbf{1}_{\mathcal{V}_2}$. For any such τ, ξ*

$$\left|\left\langle \dot{G}_1 \tilde{\Gamma}_\ell \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 \tau, \xi \right\rangle - \left\langle G^{\ell+1} \tau^{(1)}, \xi^{(1)} \right\rangle\right| \leq \varepsilon \cdot \|\tau\| \cdot \|\xi\|.$$

Proof. G_1 is D_1-regular with local inversion function ϕ. We express $G_1 = \frac{1}{D_1} \sum_{i=1}^{D_1} \mathcal{G}_i$, where $\mathcal{G}_i$ is the transition matrix of some permutation in $\mathbb{S}_{V_1}$. We let $\bar{q} = (q_0, \dots, q_{k-1})$ be the permutations induced by $(\bar{\gamma}, \phi)$. By definition (and noting that $\mathbf{1}_{\mathcal{V}_2} = \frac{1}{\sqrt{N_2}} \sum_{v^{(2)} \in \mathcal{V}_2} v^{(2)}$) we get

$$\left\langle \dot{G}_1 \tilde{\Gamma}_\ell \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 \tau, \xi \right\rangle = \frac{1}{N_2} \left\langle \sum_{v^{(2)}, u^{(2)} \in V_2} \dot{G}_1 \tilde{\Gamma}_\ell \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 (\tau^{(1)} \otimes \overrightarrow{v^{(2)}}), \xi^{(1)} \otimes \overrightarrow{u^{(2)}} \right\rangle.$$

By Corollary 10,

$$\begin{aligned}
&\left\langle \dot{G}_1 \tilde{\Gamma}_\ell \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 \tau, \xi \right\rangle \\
&\quad = \frac{1}{N_2} \left\langle \sum_{v^{(2)}, u^{(2)} \in V_2} \mathcal{G}_{\pi_1(q_\ell(v^{(2)}))} \dots \mathcal{G}_{\pi_1(q_0(v^{(2)}))} (\tau^{(1)}) \otimes \overrightarrow{\sigma(v^{(2)})}, \xi^{(1)} \otimes \overrightarrow{u^{(2)}} \right\rangle \\
&\quad = \frac{1}{N_2} \sum_{v^{(2)}, u^{(2)} \in V_2} \left\langle \mathcal{G}_{\pi_1(q_\ell(v^{(2)}))} \dots \mathcal{G}_{\pi_1(q_0(v^{(2)}))} \tau^{(1)}, \xi^{(1)} \right\rangle \cdot \left\langle \overrightarrow{\sigma(v^{(2)})}, \overrightarrow{u^{(2)}} \right\rangle.
\end{aligned}$$

However, as σ is a permutation over V_2, for every $v^{(2)} \in V_2$ there is exactly one $u^{(2)}$ that does not vanish. Hence,

$$\begin{aligned}
\left\langle \dot{G}_1 \tilde{\Gamma}_\ell \dot{G}_1 \dots \tilde{\Gamma}_1 \dot{G}_1 \tau, \xi \right\rangle &= \frac{1}{N_2} \sum_{v^{(2)} \in V_2} \left\langle \mathcal{G}_{\pi_1(q_\ell(v^{(2)}))} \dots \mathcal{G}_{\pi_1(q_0(v^{(2)}))} \tau^{(1)}, \xi^{(1)} \right\rangle \\
&= \mathbb{E}_{z_1, \dots, z_\ell \sim \mathcal{Z}} \left[\left\langle \mathcal{G}_{z_\ell} \dots \mathcal{G}_{z_1} \tau^{(1)}, \xi^{(1)} \right\rangle \right],
\end{aligned}$$

where $\mathcal{Z}$ is the distribution on $[D_1]^\ell$ obtained by picking $v^{(2)}$ uniformly at random in V_2 and outputting $z_1, \dots, z_\ell$, where $z_i = \pi_1(q_i(v^{(2)}))$. Notice also that

$$G_1^k = \mathbb{E}_{z \in [D_1]^k} [\mathcal{G}_{z_\ell} \dots \mathcal{G}_{z_1}].$$

As $(\gamma_1, \ldots, \gamma_k)$ is ε-pseudorandom with respect to G_1, we know that $|\mathcal{Z} - U_{[D_1]^k}|_1 \le \varepsilon$. By Claim 6,

$$\begin{aligned}
&\left|\left\langle \dot{G}_1\tilde{\Gamma}_\ell\dot{G}_1\ldots\tilde{\Gamma}_1\dot{G}_1\tau,\xi\right\rangle - \left\langle G^{\ell+1}\tau^{(1)},\xi^{(1)}\right\rangle\right| \\
&= \left|\mathbb{E}_{z_1,\ldots,z_\ell\sim\mathcal{Z}}\left[\left\langle \mathcal{G}_{z_\ell}\ldots\mathcal{G}_{z_1}\tau^{(1)},\xi^{(1)}\right\rangle\right] - \mathbb{E}_{z\in[D_1]^k}\left[\left\langle \mathcal{G}_{z_\ell}\ldots\mathcal{G}_{z_1}\tau^{(1)},\xi^{(1)}\right\rangle\right]\right| \\
&\le \left|\mathcal{Z} - U_{[D_1]^k}\right|_1 \cdot \left\|\tau^{(1)}\right\| \cdot \left\|\xi^{(1)}\right\| \le \varepsilon\cdot\left\|\tau^{(1)}\right\|\cdot\left\|\xi^{(1)}\right\| = \varepsilon\cdot\|\tau\|\cdot\|\xi\|
\end{aligned}$$

(since $\|\tau\| = \|\tau^{(1)}\otimes\mathbf{1}\| = \|\tau^{(1)}\|\cdot\|\mathbf{1}\| = \|\tau^{(1)}\|$), and this completes the proof of Lemma 15. ∎

Having proved Lemma 15, we can now prove Theorem 14.

Proof of Theorem 14. Since $\bar{H}$ is ε-good with respect to G_1, we can express each H_i as $H_i = \frac{1}{D_2}\sum_{j=1}^{D_2}\mathcal{H}_{i,j}$ such that $\mathcal{H}_{i,j}$ is the transition matrix of a permutation $\gamma_{i,j}\in\mathbb{S}_{V_2}$ and each of the D_2^k sequences $\gamma_{1,j_1},\ldots,\gamma_{k,j_k}$ is ε-pseudorandom with respect to G_1. Let $\Gamma_{i,j}$ be the operator on $\mathcal{V}_2$ corresponding to the permutation $\gamma_{i,j}$ and let $\tilde{\Gamma}_{i,j} = I\otimes\Gamma_{i,j}$ be the corresponding operator on $\mathcal{V}_1\otimes\mathcal{V}_2$. Observe that

$$\left\langle \dot{G}_1\tilde{H}_{i+\ell}\dot{G}_1\ldots\tilde{H}_{i+1}\dot{G}_1\tau,\xi\right\rangle = \mathbb{E}_{j_1,\ldots,j_\ell\in[D_2]}\left[\left\langle \dot{G}_1\tilde{\Gamma}_{i+\ell,j_\ell}\dot{G}_1\ldots\tilde{\Gamma}_{i+1,j_1}\dot{G}_1\tau,\xi\right\rangle\right].$$

Thus, by Lemma 15,

$$\left|\left\langle \dot{G}_1\tilde{H}_{i+\ell}\dot{G}_1\ldots\tilde{H}_{i+1}\dot{G}_1\tau,\xi\right\rangle - \left\langle G^{\ell+1}\tau^{(1)},\xi^{(1)}\right\rangle\right| \le \varepsilon\cdot\|\tau\|\cdot\|\xi\|.$$

Since $\tau,\xi\perp\mathbf{1}$, we also have that $\tau^{(1)},\xi^{(1)}\perp\mathbf{1}$. Therefore,

$$\left|\left\langle G^{\ell+1}\tau^{(1)},\xi^{(1)}\right\rangle\right| \le \lambda_1^{\ell+1}\left\|\tau^{(1)}\right\|\left\|\xi^{(1)}\right\|.$$

The fact that $\|\tau\| = \|\tau^{(1)}\|$ and $\|\xi\| = \|\xi^{(1)}\|$ completes the proof. ∎

4.2. A lemma on partial sums. We conclude this section with a bound on

$$\sum_{i=1}^{k-t}\left\|y^{||}_{k-i-t}\right\|\cdot\left\|x^{||}_{i-1}\right\|.$$

The trivial bound is $(k-t)\lambda_2^{k-t-1}$ using the fact that $\|x_i\|,\|y_i\|\le\lambda_2^i$. Here we give a tighter bound.

LEMMA 16. *Let $t\ge 0$. Then*

$$\sum_{i=1}^{k-t}\left\|y^{||}_{k-i-t}\right\|\cdot\left\|x^{||}_{i-1}\right\| \le \lambda_2^{k-t-1}.$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{k-t}\left\|y^{||}_{k-i-t}\right\|\cdot\left\|x^{||}_{i-1}\right\| &= \lambda_2^{k-t-1}\sum_{i=1}^{k-t}\left\|\frac{y^{||}_{k-i-t}}{\lambda_2^{k-i-t}}\right\|\cdot\left\|\frac{x^{||}_{i-i}}{\lambda_2^{i-1}}\right\| \\
&\le \lambda_2^{k-t-1}\cdot\frac{1}{2}\left(\sum_{i=0}^{k-t-1}\left\|\frac{y^{||}_i}{\lambda_2^i}\right\|^2 + \sum_{i=0}^{k-t-1}\left\|\frac{x^{||}_i}{\lambda_2^i}\right\|^2\right).
\end{aligned}$$

Now we bound $\sum_{i=0}^{k-t-1} \|x_i^{||}/\lambda_2^i\|^2$, and the bound for the expression $\sum_{i=0}^{k-t-1} \|y_i^{||}/\lambda_2^i\|^2$ is similarly obtained. Denote

$$\Delta_\ell = \left\| \frac{x_\ell^\perp}{\lambda_2^\ell} \right\|^2 + \sum_{i=0}^{\ell} \left\| \frac{x_i^{||}}{\lambda_2^i} \right\|^2.$$

Then

$$\Delta_\ell = \left\| \frac{x_\ell}{\lambda_2^\ell} \right\|^2 + \sum_{i=0}^{\ell-1} \left\| \frac{x_i^{||}}{\lambda_2^i} \right\|^2 \le \left\| \frac{\lambda_2 x_{\ell-1}^\perp}{\lambda_2^\ell} \right\|^2 + \sum_{i=0}^{\ell-1} \left\| \frac{x_i^{||}}{\lambda_2^i} \right\|^2 = \Delta_{\ell-1}.$$

In particular, $\Delta_{k-t-1} \le \Delta_0 = \|x_0^{||}\|^2$. It follows that

$$\sum_{i=0}^{k-t-1} \left\| \frac{x_i^{||}}{\lambda_2^i} \right\|^2 \le \left\| x_0^{||} \right\|^2 - \left\| \frac{x_{k-t-1}^\perp}{\lambda_2^{k-t-1}} \right\|^2 \le \|x_0\|^2 = 1. \qquad \square$$

5. Almost any $\bar{H}$ is good.

5.1. A hypergeometric lemma.

We shall need the following tail estimate.

THEOREM 17 (see [10, Theorem 2.10]). *Let Ω be a universe and $S_1 \subseteq \Omega$ a fixed subset of size m_1. Let $S_2 \subseteq \Omega$ be a uniformly random subset of size m_2. Set $\mu = \mathbb{E}_{S_2}[|S_1 \cap S_2|] = \frac{m_1 m_2}{|\Omega|}$. Then, for every $\varepsilon > 0$,*

$$\Pr_{S_2}[|\,|S_1 \cap S_2| - \mu| \ge \varepsilon\mu] \le 2e^{-\frac{\varepsilon^2}{3}\mu}.$$

A simple generalization of this gives the following lemma.

LEMMA 18. *Let Ω be a universe and $S_1 \subseteq \Omega$ a fixed subset of size m. Let $S_2, \ldots, S_k \subseteq \Omega$ be uniformly random subsets of size m. Set $\mu_k = \mathbb{E}_{S_2,\ldots,S_k}[\,|\,S_1 \cap S_2 \cap \cdots \cap S_k\,|\,] = \frac{m^k}{|\Omega|^{k-1}}$. Then, for every $\varepsilon > 0$,*

$$\Pr_{S_2,\ldots,S_k}[(1-\varepsilon)^{k-1}\mu_k \le |S_1 \cap S_2 \cap \cdots \cap S_k| \le (1+\varepsilon)^{k-1}\mu_k] \ge 1 - 2(k-1)e^{-\frac{\varepsilon^2}{3}(1-\varepsilon)^{k-1}\mu_k}.$$

In particular, for $\varepsilon \le \frac{1}{4k}$,

$$\Pr_{S_2,\ldots,S_k}[|\,|S_1 \cap S_2 \cap \cdots \cap S_k| - \mu_k| \ge 2k\varepsilon\mu_k] \le 2ke^{-\frac{\varepsilon^2}{6}\mu_k}.$$

Proof (*by induction on k*). The case $k = 2$ follows from Theorem 17. Assume for k, and let us prove for $k+1$. Let $A = S_1 \cap \cdots \cap S_k \subseteq \Omega$. By the induction hypothesis we know that, except for probability $\delta_k = 2(k-1)e^{-\frac{\varepsilon^2}{3}(1-\varepsilon)^{k-1}\mu_k}$, the set A has size in the range $[(1-\varepsilon)^{k-1}\mu_k, (1+\varepsilon)^{k-1}\mu_k]$ for $\mu_k = \frac{m^k}{|\Omega|^{k-1}}$. When this happens, by Theorem 17, $|A \cap S_{k+1}|$ is in the range

$$\left[(1-\varepsilon)\frac{|A|m}{|\Omega|}, (1+\varepsilon)\frac{|A|m}{|\Omega|}\right] \subseteq [(1-\varepsilon)^k\mu_k, (1+\varepsilon)^k\mu_k]$$

except for probability

$$2e^{-\frac{\varepsilon^2}{3}\frac{|A|m}{|\Omega|}} \le 2e^{-\frac{\varepsilon^2}{3}(1-\varepsilon)^k\mu_{k+1}}.$$

Thus, $|A \cap S_{k+1}|$ is in the required range except for probability $\delta_k + 2e^{-\frac{\varepsilon^2}{3}(1-\varepsilon)^k\mu_{k+1}} \le 2ke^{-\frac{\varepsilon^2}{3}(1-\varepsilon)^k\mu_{k+1}}$, and this completes the proof. $\square$

5.2. Almost any $\bar{\gamma}$ is pseudorandom. The main lemma we prove in this section is the following.

LEMMA 19. *For every $\varepsilon \le \frac{1}{2}$, a sequence of uniformly random and independent permutations $(\gamma_1, \ldots, \gamma_{k-1})$ satisfies*

$$\Pr_{\gamma_1,\ldots,\gamma_{k-1}}[(\gamma_1,\ldots,\gamma_{k-1}) \text{ is not } \varepsilon\text{-pseudorandom with respect to } G_1] \le D_1^k \cdot 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})}.$$

Proof. Let $q_0, \ldots, q_{k-1} : V_2 \to V_2$ be the permutations induced by $(\bar{\gamma} = (\gamma_1, \ldots, \gamma_{k-1}), \psi)$, where ψ is as defined in (6). Let A denote the distribution $\pi_1(q_1(U)) \circ \cdots \circ \pi_1(q_k(U))$ and let $U_{D_1^k}$ denote the uniform distribution over $[D_1]^k$. Fix an arbitrary $\bar{r} = (r_1, \ldots, r_k) \in [D_1]^k$. We will show that

$$\Pr_{\gamma_1,\ldots,\gamma_{k-1}}[|A(\bar{r}) - U_{D_1^k}(\bar{r})| \ge \varepsilon D_1^{-k}] \le 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})}. \tag{11}$$

Therefore, using a simple union bound, the event $\exists \bar{r}\ |A(\bar{r}) - U_{D_1^k}(\bar{r})| \ge \varepsilon D_1^{-k}$ happens with probability at most $D_1^k \cdot 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})}$, and whenever it does not happen,

$$\left|A - U_{D_1^k}\right|_1 = \sum_{\bar{r}} |A(\bar{r}) - U_{D_1^k}(\bar{r})| \le D_1^k \cdot \max_{\bar{r}} \left\{|A(\bar{r}) - U_{D_1^k}(\bar{r})|\right\} \le \varepsilon.$$

We now prove the inequality in (11). Let $S_i = \{x \in V_2 \mid \pi_1(q_i(x)) = r_i\}$ for $1 \le i \le k$. Since q_i is a permutation and π_1 is a regular function, $|S_i| = \frac{|V_2|}{D_1}$. Also, for each i, q_i is a random permutation distributed uniformly in $\mathbb{S}_{V_2}$, and the permutations $\{q_i\}$ are independent. It follows that the sets $S_2, \ldots, S_k$ are random $\frac{|V_2|}{D_1}$ subsets of V_2, and they are independent as well.

By definition, $A(\bar{r}) = \frac{|S_1 \cap S_2 \cap \cdots \cap S_k|}{|V_2|}$. Also,

$$\mu_k = \mathbb{E}[|S_1 \cap S_2 \cap \cdots \cap S_k|] = \frac{(|V_2|/D_1)^k}{|V_2|^{k-1}} = \frac{|V_2|}{D_1^k} = D_1^{3k}$$

and $U_{D_1^k}(\bar{r}) = \frac{\mu_k}{|V_2|} = D_1^{-k}$. Thus, by Lemma 18, and setting $\zeta = \frac{\varepsilon}{2k} \le \frac{1}{4k}$,

$$\begin{aligned}\Pr_{\gamma_1,\ldots,\gamma_{k-1}}[|A(\bar{r}) - U_{D_1^k}(\bar{r})| \ge \varepsilon D_1^{-k}] &= \Pr_{S_2,\ldots,S_k}\left[\left|\frac{|S_1 \cap S_2 \cap \cdots \cap S_k|}{|V_2|} - \frac{\mu_k}{|V_2|}\right| \ge \frac{2k\zeta\mu_k}{|V_2|}\right] \\ &\le 2ke^{-\Omega(\zeta^2\mu_k)} = 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})},\end{aligned}$$

as stated in (11). ∎

5.3. The spectrum of random *D*-regular graphs. Friedman [7] proved the following theorem regarding the spectrum of random regular graphs. The distribution $G_{N,D}$ is described in section 2.

THEOREM 20 (see [7]). *For every $\delta > 0$ and for every even D, there exists a constant $c > 0$, independent of N, such that*

$$\Pr_{G \sim G_{N,D}}[\,\bar{\lambda}(G) > \lambda_{\mathrm{Ram}}(D) + \delta\,] \le c \cdot N^{-\lceil(\sqrt{D-1}+1)/2\rceil - 1}.$$

5.4. Almost any $\bar{H}$ is good.

THEOREM 21. *For every even $D_2 \geq 4$, integer $k \geq 3$, and $\varepsilon = D_2^{-k}$, there exists a constant B such that for every $D_1 \geq B$ there exists a sequence $\bar{H} = (H_1, \ldots, H_k)$ of $(N_2 = D_1^{4k}, D_2)$ graphs such that*

- *each H_i is locally invertible;*
- *$\bar{H}$ is ε-good with respect to any D_1-regular locally invertible graph.*

Proof. For a value D_1, let us randomly pick $\bar{H} = (H_1, \ldots, H_k)$ with each H_i sampled independently and uniformly from $G_{N_2=D_1^k, D_2}$. That is, let $\{\gamma_{i,j}\}_{i\in[k],\, j\in[D_2/2]}$ be a set of random permutations chosen uniformly and independently from $\mathbb{S}_{V_2}$. For $1 \leq i \leq k$, let H_i be the undirected graph over V_2 formed from the permutations $\{\gamma_{i,j}\}_{j\in[D_2/2]}$ and their inverses. We use the following labeling on the edges: we label the directed edge $(v, \gamma_{i,j}(v))$ with the label j and the edge $(v, \gamma_{i,j}^{-1}(v))$ with the label $D_2/2+j$ (recall that each edge needs to be labeled twice, once by each of its vertices). By definition, each H_i is locally invertible.

We show that for a large enough D_1 the probability that $\bar{H}$ is ε-pseudorandom with respect to any D_1-regular locally invertible graph is at least half, and therefore a good sequence exists. Fix a D_1-regular locally invertible graph G_1. We notice that the inverse of a uniform random permutation is also a uniform random permutation. Therefore, for every $j_1, \ldots, j_k \in [D_2/2]$ and for every $p_1, \ldots, p_k \in \{1, -1\}$, the k-tuple $\bar{\gamma} = (\gamma_{1,j_1}^{p_1}, \ldots, \gamma_{k,j_k}^{p_k})$ is uniform in $(\mathbb{S}_{|V_2|})^k$. Thus, by Lemma 19, $\bar{H}$ is not ε-pseudorandom with respect to G_1 with probability at most $k^2 \cdot D_2^k \cdot D_1^k \cdot 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})}$.[6] Taking $D_1 \geq D_2$, the error term is at most $\delta \stackrel{\text{def}}{=} D_1^{3k} e^{-\Omega(\frac{D_1^k}{k^2})}$.

There are only $D_1!$ local inversion functions over D_1 vertices (compared to the $N_2!$ permutations over V_2). We have seen that the probability a random $\bar{H}$ is bad for any of them is at most δ, and therefore the probability over $\bar{H}$ that it is bad for any of them is at most $D_1! \cdot \delta$. Taking D_1 large enough, this term is at most $\frac{1}{10}$.

Finally, by Theorem 20, the probability that there exists a graph H_i in $\bar{H}$ with $\bar{\lambda}(H_i) \geq \lambda_{\text{Ram}}(D_2) + \varepsilon$ is at most

$$k \cdot c \cdot |V_2|^{-\lceil(\sqrt{D_2-1}+1)/2\rceil-1} \leq k \cdot c \cdot |V_2|^{-1} = \frac{kc}{{D_1}^{4k}}$$

for some universal constant c independent of $|V_2|$ and therefore also independent of D_1. Taking D_1 large enough (depending on the unspecified constant c), this term also becomes smaller than $\frac{1}{10}$. Altogether, with probability at least $1/2$, $\bar{H}$ is ε-good with respect to any D_1-regular locally invertible graph. ∎

6. The iterative construction. In [21] an iterative construction of expanders was given, starting with constant-size expanders and constructing at each step larger constant-degree expanders. Each iteration is a sequence of tensoring (which makes the graph much larger, the degree larger, and the spectral gap the same), powering (which keeps the graph size the same and increases the spectral gap and the degree), and a zig-zag product (which reduces the degree back to what it should be without harming the spectral gap much). Here we follow the same strategy using the same sequence of tensoring, powering, and degree reduction, albeit using the k-step zig-zag product rather than the zig-zag product to reduce the degree. We do it for degrees D of the special form $D = 2D_2^k$.

[6] The D_2^k factor is for a union bound over all possible permutation sequences $\bar{\gamma}$; the k^2 factor is for a union bound over all possible consecutive subsequences $1 \leq \ell_1 \leq \ell_2 \leq k$.

We are given an arbitrary even number $D_2 \geq 4$ and an integer k. Our goal is to construct an infinite sequence of degree $D = 2D_2^k$-regular graphs $\{G_t\}$ with close to optimal spectral gap. Set $\varepsilon = D_2^{-k}$ and $\lambda_2 = \lambda_{\mathrm{Ram}}(D_2) + \varepsilon$. By Theorem 21, there exists some integer B such that for every even integer $D_1 \geq B$ there exists a sequence $\bar{H} = (H_1, \ldots, H_k)$ of $(N_2 = D_1^{4k}, D_2, \lambda_2)$ graphs that is ε-good with respect to D_1-regular locally invertible graphs. We take the first integer $m \geq 1$ such that $D^{4m} \geq B$ and we set $D_1 = D^{4m}$. We can verify that a given $\bar{H}$ is good in time depending only on D, D_1, D_2, and k, independent of N_1, and we find such a good sequence by brute force search.

We start with two constant-size, locally invertible graphs G_1 and G_2. G_1 is an (N_2, D, λ) graph, and G_2 is an (N_2^2, D, λ) graph, for $\lambda = \lambda_2^{k-1} + 2\lambda_2^k$. We find both graphs by a brute force search; the existence of these graphs follows from the existence of (N_2, D_2, λ_2) graphs guaranteed above. Now, for $t \geq 3$, define

- $G_t^{\mathrm{temp}} = (G_{\lfloor \frac{t-1}{2} \rfloor} \otimes G_{\lceil \frac{t-1}{2} \rceil})^{2m}$;
- $G_t = \frac{1}{2}\left[G_t^{\mathrm{temp}} ⓩ \bar{H} + (G_t^{\mathrm{temp}} ⓩ \bar{H})^{\dagger}\right]$.

THEOREM 22. *For every even $D_2 \geq 4$ and every $k \geq 50$, the family of undirected graphs $\{G_t\}$ is fully explicit, and each graph G_t is an (N_2^t, D, λ) graph.*

The theorem follows from the following two lemmas.

LEMMA 23. *For every even $D_2 \geq 4$ and every $k \geq 50$,*

- *for every $t \geq 1$, G_t is an (N_2^t, D, λ) undirected locally invertible graph;*
- *for every $t \geq 3$, G_t^{temp} is an $(N_2^{t-1}, D_1 = D^{4m}, \lambda^{2m})$ undirected locally invertible graph.*

Proof. The fact that G_t and G_t^{temp} are locally invertible follows by induction. Facts 5 and 7 guarantee that all the operations in the construction preserve the locality property.

The claims regarding the number of vertices and the degree follow by induction using Facts 2 and 3 and Theorem 13.

The only nontrivial part is proving the claim regarding the spectral gap of G_t and G_t^{temp}. For $t = 1, 2$, this follows from the way G_1 and G_2 were chosen. Let us assume for all $i \leq t$ and prove for $t + 1$.

Let α_t denote the second largest eigenvalue of G_t. Using the properties of tensoring, powering, and the induction hypothesis, the second largest eigenvalue of G_{t+1}^{temp} is at most λ^{2m}. By Theorem 13,

$$\alpha_{t+1} \leq \lambda_2^{k-1} + \lambda_2^k + 2(\lambda^{2m} + \varepsilon).$$

For $D_2 \geq 4$ and $k \geq 50$, and plugging $\lambda = \lambda_2^{k-1} + 2\lambda_2^k$, $\varepsilon = D_2^{-k} \leq \lambda_2^{2k}$, and $m \geq 1$, one can check that the above term is bounded by λ, as desired. ∎

LEMMA 24. *$\{G_t\}$ is a fully explicit family of graphs.*

Proof. To compute the rotation map of Rot_{G_t} on a given vertex and edge label, we make two calls for computing $\mathrm{Rot}_{G_t^{\mathrm{temp}} ⓩ \bar{H}}$. Each such call requires $k - 1$ calls to $\mathrm{Rot}_{G_t^{\mathrm{temp}}}$ and k calls to Rot_{H_i}. A call to $\mathrm{Rot}_{G_t^{\mathrm{temp}}}$ requires $O(m)$ calls to $\mathrm{Rot}_{G_{t'}}$ for $t' \leq \lceil \frac{t}{2} \rceil$. Altogether, we have $(km)^{O(\log t)} = \mathrm{poly}(t)$ calls to the rotation maps of the base graphs $G_1, G_2, H_1, \ldots, H_k$ (each of constant size). The number of vertices of G_t is $N_2^t = 2^{\Theta(t)}$; thus $\{G_t\}$ is fully explicit. ∎

The resulting eigenvalue is $\lambda \leq 2\lambda_2^{k-1} \leq \frac{2^k}{\sqrt{D}}$, whereas the best we can hope for is $\bar{\lambda}_{\mathrm{Ram}}(D) = \frac{2\sqrt{D-1}}{D}$. As explained in the introduction, our losses come from two different sources. First we lose one application of H out of the k different H applications, and this loss amounts roughly to a $\sqrt{D_2}$ multiplicative factor. We

also have a second loss of a 2^{k-1} multiplicative factor emanating from the fact that $\lambda_{\mathrm{Ram}}(D_2)^k \approx 2^{k-1}\lambda_{\mathrm{Ram}}(D_2^k)$. Balancing losses we roughly have $D = D_2^k$ and $D_2 = 2^k$, which is solved by $k = \log(D_2)$ and $D = 2^{\log^2(D_2)}$. Namely, our loss is about $2^k = 2^{\sqrt{\log(D)}}$. Formally, we have the following corollary.

COROLLARY 25. *Let D_2 be an arbitrary even number that is greater than 2, and let $D = 2D_2^{\log D_2}$. Then, there exists a fully explicit family of $(D, D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})})$ graphs.*

Proof. Set $k = \log D_2$ in the above construction. Clearly the resulting graphs are D-regular and fully explicit. Also, for every graph G in the family,

$$\bar{\lambda}(G) \le 2(\lambda_{\mathrm{Ram}}(D_2) + D_2^{-k})^{k-1} \le D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})}. \qquad \square$$

7. A construction for any degree. The construction in section 6 is applicable only when $D = 2D_2^{\log D_2}$ for some even $D_2 > 2$. Now we show how it can be used to construct graphs of arbitrary degree with about the same asymptotic spectral gap. In particular, this will prove Theorem 1.

Let D be an arbitrary integer, and say we wish to build an expander of even degree $2D$. (To construct a graph with an odd degree, we simply add another self-loop.) As in the previous section, we shall construct a directed graph of degree D and then we will undirect it. Set $D_2 = 2 \cdot \lceil 2^{\sqrt{\log D}} \rceil$ and let k be an integer such that $D_2^k \le D < D_2^{k+1}$ (k is about $\frac{\log D}{\log D_2}$). Ideally, we would like to do a k-step zig-zag between a large graph with some small spectral gap and a sequence of k degree D_2 graphs. This, however, will result in a degree D_2^k graph, and not degree D. So instead, we express the integer D in base D_2 and take care of the remainders by adding self-loops.

Formally, set $\lambda_2 = \lambda_{\mathrm{Ram}}(D_2) + D_2^{-k}$ and $\lambda_1 = \lambda_2^{k-1}$ and assume that D is large enough so that $k \ge 50$.

- Construct a locally invertible (N, D_1, λ_1) graph, G_1, where D_1 depends only on D. This can be done using Corollary 25.
- Find $\bar{H} = (H_1, \ldots, H_k)$ that is λ_1-good with respect to D_1-regular graphs, and where each H_i is a $(D_1^{4k}, D_2, \lambda_2)$ graph. (Such an $\bar{H}$ exists by Theorem 21.)

We express D in base D_2. Let $A_0 = D$, $A_{i+1} = \lfloor \frac{A_i}{D_2} \rfloor$, and $B_{i+1} = A_i \pmod{D_2}$. That is,

$$\forall 0 \le i \le k, \qquad A_i = A_{i+1} \cdot D_2 + B_{i+1}.$$

Notice that $D = A_0 > A_1 > \cdots > A_k \ge 1 > A_{k+1} = 0$ and $B_{k+1} = A_k$.

Define a sequence of directed graphs $\{Z_i\}$ by the following:

- Z_k is the graph with B_{k+1} self-loops.
- For $0 \le i \le k$, Z_i is the graph $\tilde{H}_{i+1}\dot{G}_1 Z_{i+1}$, with the addition of B_{i+1} self-loops.
- The output graph is $\frac{1}{2}(Z_0 + Z_0^\dagger)$.

Observe that $\deg(Z_k) = A_k$ and that for every $i < k$, $\deg(Z_i) = A_{i+1} \cdot D_2 + B_i = A_i$. In particular, $\deg(Z_0) = D$.

As always, we identify a graph with its transition matrix. The transition matrices of the graphs $\{Z_i\}$ are given by

$$Z_i = \begin{cases} I, & i = k, \\ (1 - \frac{B_{i+1}}{A_i})\tilde{H}_{i+1}\dot{G}_1 Z_{i+1} + \frac{B_{i+1}}{A_i} I, & 0 \le i < k. \end{cases}$$

For example, say $D = 1000$. We set $D_2 = 2 \cdot \lceil 2^{\sqrt{\log D}} \rceil = 18$ and express $1000 = 18 \cdot (3 \cdot 18 + 1) + 10$. We construct a degree 1000 graph by taking a k-step zig-zag *with self-loops* between $\dot{G}_1$ and $\bar{H}$, namely,

$$\frac{10}{1000} I + \frac{990}{1000} \tilde{H}_1 \dot{G}_1 \left(\frac{1}{990} I + \frac{989}{990} \tilde{H}_2 \dot{G}_1 \right).$$

We now bound $\bar{\lambda}(Z_0)$, and this proves Theorem 1.

CLAIM 26. $\bar{\lambda}(Z_0) \leq D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})}$.

Proof. Resolving the recursive formula for Z_0 we get

$$Z_0 = \sum_{i=0}^{k} \frac{B_{i+1}}{A_i} \left(\prod_{j=1}^{i} \left(1 - \frac{B_j}{A_{j-1}} \right) \tilde{H}_j \dot{G}_1 \right).$$

Since all the graphs here are regular (even though they are directed), they share the same first eigenvector and therefore we can apply the triangle inequality on s_2 to derive

$$\bar{\lambda}(Z_0) \leq \sum_{i=0}^{k} \frac{B_{i+1}}{A_i} \cdot \bar{\lambda} \left(\prod_{j=1}^{i} \left(1 - \frac{B_j}{A_{j-1}} \right) \tilde{H}_j \dot{G}_1 \right).$$

Now, since $B_i < D_2$ and $A_i \geq \frac{D}{D_2^{i+1}}$ for all $i = 0, \ldots, k$,

$$\bar{\lambda}(Z_0) \leq \sum_{i=0}^{k} \frac{D_2^{i+2}}{D} \cdot \bar{\lambda} \left(\prod_{j=1}^{i} \tilde{H}_j \dot{G}_1 \right). \tag{12}$$

Note that $\dot{G}_1$ is a unitary transformation; hence, for any X, $\bar{\lambda}(X\dot{G}_1) = \bar{\lambda}(X)$. By Theorem 13, for every i (the cases $i = 0, 1$ are trivial),

$$\bar{\lambda} \left(\prod_{j=1}^{i} \tilde{H}_j \dot{G}_1 \right) \leq \lambda_2^{i-1} + 4\lambda_1 + \lambda_2^i \leq 6\lambda_2^{i-1}.$$

Plugging this into (12) we get

$$\bar{\lambda}(Z_0) \leq \frac{6D_2^2}{\lambda_2 D} \sum_{i=0}^{k} D_2^i \cdot \lambda_2^i \leq O\left(\frac{D_2^3}{D} \cdot D_2^{k-1} \lambda_2^{k-1} \right) = O(D_2^3 \cdot D_2^{-k/2}),$$

which is $D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})}$ by our choice of D_2. ∎

Acknowledgments. We thank the anonymous referees for several useful suggestions that improved the presentation of the paper. We thank one of the anonymous referees for strengthening Theorem 21.

REFERENCES

[1] N. ALON, *Eigenvalues and expanders*, Combinatorica, 6 (1986), pp. 83–96.

[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, IEEE Press, Piscataway, NJ, 2001, pp. 630–637.

[3] N. Alon and V. Milman, λ_1, *isoperimetric inequalities for graphs, and superconcentrators*, J. Combin. Theory Ser. B, 38 (1985), pp. 73–88.
[4] Y. Bilu and N. Linial, *Lifts, discrepancy and nearly optimal spectral gap*, Combinatorica, 26 (2006), pp. 495–519.
[5] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson, *Randomness conductors and constant-degree lossless expanders*, in Proceedings of the 34th Annual ACM Symposium on Theory of Computing, 2002, pp. 659–668.
[6] J. Dodziuk, *Difference equations, isoperimetric inequality and transience of certain random walks*, Trans. Amer. Math. Soc., 284 (1984), pp. 787–794.
[7] J. Friedman, *A Proof of Alon's Second Eigenvalue Conjecture and Related Problems*, Mem. Amer. Math. Soc. 195, AMS, Providence, RI, 2008.
[8] O. Gabber and Z. Galil, *Explicit constructions of linear-sized superconcentrators*, J. Comput. System Sci., 22 (1981), pp. 407–420.
[9] S. Hoory and A. Wigderson, *Universal sequences for expander graphs*, Inform. Process. Lett., 46 (1993), pp. 67–69.
[10] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs*, John Wiley, New York, 2000.
[11] S. Jimbo and A. Maruoka, *Expanders obtained from affine transformations*, Combinatorica, 7 (1987), pp. 343–355.
[12] N. Kahale, *Eigenvalues and expansion of regular graphs*, J. ACM, 42 (1995), pp. 1091–1106.
[13] A. Lubotzky, R. Philips, and P. Sarnak, *Ramanujan graphs*, Combinatorica, 8 (1988), pp. 261–277.
[14] G. A. Margulis, *Explicit constructions of expanders*, Problemy Peredachi Informatsii, 9 (1973), pp. 71–80.
[15] G. A. Margulis, *Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators*, Problemy Peredachi Informatsii, 24 (1988), pp. 51–60.
[16] R. Meshulam and A. Wigderson, *Expanders in group algebras*, Combinatorica, 24 (2004), pp. 659–680.
[17] M. Morgenstern, *Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q*, J. Combin. Theory Ser. B, 62 (1994), pp. 44–62.
[18] A. Nilli, *On the second eigenvalue of a graph*, Discrete Math., 91 (1991), pp. 207–210.
[19] M. Pinsker, *On the complexity of a concentrator*, in Proceedings of the 7th International Teletraffic Conference, 1973, pp. 318/1–318/4.
[20] O. Reingold, *Undirected st-connectivity in log-space*, in Proceedings of the 37th Annual ACM Symposium on Theory of Computing, 2005, pp. 376–385.
[21] O. Reingold, S. Vadhan, and A. Wigderson, *Entropy waves, the zig-zag graph product, and new constant-degree expanders*, Ann. of Math. (2), 155 (2002), pp. 157–187.
[22] E. Rozenman, A. Shalev, and A. Wigderson, *Iterative construction of Cayley expander graphs*, Theory Comput., 2 (2006), pp. 91–120.