TEL AVIV UNIVERSITY
THE RAYMOND AND BEVERLY SACKLER FACULTY OF EXACT SCIENCES
THE BLAVATNIK SCHOOL OF COMPUTER SCIENCE

PSEUDORANDOMNESS AND QUANTUM INFORMATION

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY

AVRAHAM BEN AROYA

UNDER THE SUPERVISION OF
PROFESSOR ODED REGEV AND PROFESSOR AMNON TA-SHMA

SUBMITTED TO THE SENATE OF TEL AVIV UNIVERSITY
OCTOBER 2011

Acknowledgments

I would like to express my gratitude to my advisors, Oded Regev and Amnon Ta-Shma. I still remember the great class they gave together on quantum computing, which led me to pursue this topic in my graduate studies. Each of them had his unique and exciting teaching style, and it was clear that research was a passion for both of them (to the degree of altering the syllabus during the semester to incorporate new research results!). When I found out about the possibility of having two advisors, choosing both Oded and Amnon was the most natural choice. This choice was probably one of the best I have ever made. I have learned so much from each of them, and their advices during my studies were most invaluable. Most of all, I want to thank them for caring so much about me. They were everything and more that I could have asked for.

I wish to thank my other coauthors in results that appear in this thesis, Ronald de Wolf and Oded Schwartz.

I would like to thank my lab mates, Iftah Gamzu, Michal Moshkovitz, Ishay Haviv and Klim Efremenko for many interesting conversations and discussions, collaborations, UT games and generally being great companions.

Last but not least, I would like to express my deepest gratitude to my dear parents Eliyahu and Orna and my dear sister Tali. There are no words which can describe my love and appreciation for them. They stood by me and supported me during the best and the worst times I have had. I can honestly say that without their endless love and encouragement, completing this thesis would not have been possible. This thesis is dedicated to them.

Abstract

This thesis is concerned with pseudorandomness and its interplay with quantum information. Randomness is a very useful resource in computation, and for some computational problems, using randomness seems to allow savings in other resources (such as time and space). Despite its usefulness, it is yet not fully understood whether randomness is indeed *necessary* to achieve the aforementioned savings. The study of derandomization is concerned with eliminating or reducing the use of randomness in computation, without increasing the use of other resources.

As its name suggests, pseudorandomness deals with objects that are not truly random, but rather "random-like", where the definition of "random-like" depends on the type of object at hand. The list of pseudorandom objects of interest include pseudorandom generators, expander graphs, randomness extractors and error-correcting codes. The importance of studying these objects stems from the fact that, due to the pseudorandom properties they possess, they often play a central role in derandomization as well as in other applications.

Quantum computation is a model of computation built upon the principles of quantum mechanics. As in classical computation, there is a distinction between the actions performed by the computation, and the actual data that undergoes the computation. This data is what we refer to as quantum information. Unlike classical information, quantum information is very fragile. For example, almost any attempt to read it inherently causes some of the stored information to be lost.

In this thesis we study several problems regarding pseudorandom objects. In some of these problems we consider objects that either manipulate quantum information, or have some pseudorandom properties with respect to quantum side-information. Our results are divided into two categories:

Expanders. Expanders are graphs of low degree and high connectivity. These graphs possess several pseudorandom properties, and different constructions attempt to optimize different properties. In this thesis we focus on the algebraic property of expanders, i.e, the fact that the adjacency matrix of an expander has a large spectral gap. Our results concerning expanders are:

- We develop a new graph product, and use it to give a fully-explicit combinatorial con-

struction of expander graphs with near-optimal relation between their degree and their spectral gap.

- We introduce the notion quantum expanders, a generalization of classical expanders. We give two explicit constructions of quantum expanders: an algebraic construction and a combinatorial one. We demonstrate the usefulness of the notion by giving an application to quantum complexity theory.
- We use algebraic curves to give an improved explicit construction of small-bias sets. This immediately implies an improved explicit construction of Cayley expanders over the abelian group Z_2^k.

Extractors. Randomness extractors are functions that refine weak random sources. These objects have numerous applications in computer science. Consider the following scenario: suppose X is a uniformly distributed string and an adversary has limited side-information about X. Then, it is known that applying an appropriate extractor on X results in a distribution that is nearly uniform and, moreover, is almost completely unpredictable by the adversary. In this thesis we focus on this scenario, but where we assume the adversary is allowed to store (limited) quantum side-information about X. Extractors that can handle such adversaries are called quantum-proof extractors. Our results regarding quantum-proof extractors are:

- We prove a hypercontractive for matrix-valued functions. We use this inequality to prove that a certain XOR extractor is quantum-proof. This, in turn, implies bounds on quantum random access codes and a direct product theorem for one-way quantum communication complexity. We also use the inequality to derive a "non-quantum" proof of the fact that 2-query locally decodable codes require exponential length.
- We use ideas from classical extractor constructions (such as the use of condensers), to give new explicit constructions of quantum-proof extractors, improving upon previous constructions in some range of parameters.

Contents

Chapter 1

Introduction

This thesis is concerned with combinatorial pseudorandom objects and their interplay with quantum information. In what follows, we overview the problems considered in the thesis, and describe our contribution. Prior to delving into the details, let me mention that during my PhD studies I have also published the following results [22, 23, 16, 17], which are not included in this thesis.

- In [22] we study a generalized model of noise in quantum computing, and develop quantum error-correcting codes to cope with it.
- In [23] we show an efficient algorithm to compute an important norm on the space of super-operators, called the Diamond norm.
- Results [16, 17] describe methods to amplify the error tolerance of locally decodable codes.

Pseudorandomness. Randomness plays a central role in theoretical computer science. In particular, the rigorous study of some topics is meaningless without randomness. One example comes from the area of interactive proofs, which is concerned with protocols in which an all-powerful prover is interacting with a randomized and computationally bounded verifier. If the verifier is not allowed to use randomness then the entire interaction can be reduced to a single message sent by the prover, and the proof system becomes essentially non-interactive. Another example is cryptography. The whole idea of having secure communication which is immune to eavesdroppers is based on the fact that the communicating parties may keep some secret information, unknown to the adversary. However, in a deterministic setting, this is impossible.

There are other areas in which randomness might not be essential, yet still very useful. For instance, there are computational problems for which there are probabilistic algorithms which outperform any of the currently-known deterministic ones in terms of time or space. One such problem is the Polynomial Identity Testing problem, asking to decide whether two polynomials (represented

as arithmetic circuits) are identical. While this problem exhibits a simple probabilistic polynomial time algorithm, no efficient deterministic algorithm is currently known.

Another example of the usefulness of randomness comes from the study of constructions of combinatorial objects. It is often the case that for many combinatorial objects that possess some property of interest, the probabilistic method can be used to prove their existence. That is, usually one defines a probability distribution over a set of objects and then shows that with a nonzero probability an object with the required property is sampled from this distribution. Moreover, in most cases, the aforementioned nonzero probability is, in fact, extremely close to 1. This immediately gives a very simple randomized algorithm that constructs objects with the required property. However, for many of the most interesting and useful objects, the best known deterministic (i.e., explicit) constructions are inferior to their randomized counterparts (in terms of parameters).

The main question dealt with in derandomization is "when can the dependence on randomness be reduced or even eliminated?". This question is of major interest for several reasons. First it arises naturally in the study of efficient computations, since randomness is a resource and just like any other resource, we would our computations to use a little of it as possible. Furthermore, in some cases, such as in that of constructions of certain combinatorial objects, we need explicit constructions for the applications we have in mind. Sometimes, this is due to the fact that many of these combinatorial objects are used, in turn, to derandomize other algorithms. Finally, another reason for studying derandomization comes from "the real world". While in theory we would like to assume that our algorithms have access to an arbitrarily long string of uniformly random bits, in practice, such a physical source might be infeasible.

As its name suggests, pseudorandomness deals with objects that are not truly random, but rather "random-like", where the definition of "random-like" depends on the type of object at hand. The list of pseudorandom objects of interest include pseudorandom generators, expander graphs, randomness extractors, error-correcting codes and many others. These objects have found numerous applications beyond the original motivations for studying them. One of the greatest accomplishments of the theory of pseudorandomness is showing the close relations between them (see [139]).

The connection between derandomization and the aforementioned pseudorandom objects is two fold. On the one hand pseudorandom objects are now central components in derandomization, were for many derandomization problems, an appropriate off-the-shelf pseudorandom object will solve the problem at hand immediately. On the other hand, as explained above, obtaining explicit constructions of pseudorandom objects can be thought of as a derandomization problem.

In this thesis we obtain new results concerning two pseudorandom objects: expanders and extractors. Note that, as many pseudorandom objects are closely related to one another, some of our results can also be viewed as related to other objects (such as error-correcting codes and pseudorandom generators).

Some of our results are concerned with pseudorandom objects that deal with quantum information, which move to discuss next.

Quantum information. Classical information is usually modeled as strings of bits. A classical n-bit register can store any of the 2^n possible length n binary strings. Considering the vector space $\left(\mathbb{C}^2\right)^{\otimes n} \cong \mathbb{C}^{2^n}$, there is a natural bijection between the standard basis for this space and the set of n-bit strings. Thus, it is possible to interpret the value of an n-bit classical register as a vector in the standard basis of this vector space. A *quantum register* of length n is simply a register that can store *any* unit vector in $\left(\mathbb{C}^2\right)^{\otimes n}$. Any such unit vector is called a quantum (pure) state.

At first glance one may think that the capacity of a quantum register is infinite (as the set of quantum states is uncountable) and independent of n. However, the process of *reading* the value stored in a quantum register is very different than its classical counterpart. In the classical setting, reading the contents of a register simply means obtaining the data that was previously stored in it. (Or, in our interpretation, reading the corresponding basis vector that was stored in it.) In the quantum setting, the reading operation of a register, which is called *a measurement* is different in two aspects. First, we cannot obtain the exact vector we stored, but rather obtain "some" information about it. Second, the measurement *changes* the vector stored in the register. Suppose our quantum register holds the vector v. An example of a measurement is the measurement in the standard basis, which does the following:

- It returns a vector sampled from the standard basis $\{e_i\}$, where the probability of obtaining the vector e_i is $\| v \cdot e_i \|^2$. Observe that since v is a unit vector, this indeed defines a probability distribution.
- The quantum register changes its value to the returned vector.

It is clear that if our quantum register only holds vectors from the standard basis (i.e., classical states), then the above operation is deterministic and behaves exactly the same as the reading operation of a classical register. Thus, a quantum register can be seen as a powerful (yet delicate) generalization of a classical register.

In this thesis we deal with quantum information from two aspects. First, we study pseudorandom quantum transformations that *manipulate* quantum information. Second, we study classical pseudorandom objects that are *secure* against quantum side-information.

1.1 Expanders

The first part of this thesis is concerned with expander graphs. Expander graphs, or simply expanders, are graphs of low degree and high connectivity. It is clear that these two properties are in

contention, as the complete graph is optimal in terms of connectivity while the empty graph (which consists only of isolated nodes) is optimal in terms of degree. The original motivation for studying expanders was the design of sparse yet robust communication networks. However, they have found a myriad of applications in computer science and elsewhere. For a complete survey see [68].

There are several ways to measure the quality of expansion in a graph. One such way measures *set expansion*: given a not-too-large subset S of the vertices, it measures the size of the set $\Gamma(S)$ of neighbors of S, relative to the size of S. Another way is *(Rényi) entropic expansion*: given a distribution π on the vertices of the graph, it measures the amount of (Rényi) entropy added in $\pi' = G\pi$. This is closely related to measuring the *algebraic expansion* given by the spectral gap of the graph, i.e., the gap between the first and second largest eigenvalues of the operator defined by the adjacency matrix of the graph. Several works [44, 6, 3, 76] showed intimate connections between the different expansion measures.

1.1.1 A combinatorial construction of almost-Ramanujan graphs

Pinsker [114] was the first to observe that constant degree random graphs have almost-optimal set expansion. Explicitly finding such graphs turned out to be a major challenge. On the other hand, the algebraic measure of expansion led to to a series of explicit constructions based on algebraic structures, e. g., [98, 51, 74]. This line of research culminated in the works of Lubotzky, Phillips and Sarnak [96], Margulis [99], and Morgenstern [101] who explicitly constructed Ramanujan graphs, i. e., D-regular graphs achieving spectral gap of $1 - 2\sqrt{D-1}/D$.[1] Friedman [49] showed that random graphs are "almost Ramanujan" and Alon and Boppana (see [109]) showed Ramanujan graphs have almost the best possible algebraic expansion.

A decade ago, Reingold, Vadhan and Wigderson [120] 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. [35] 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 [119] gave a log-space algorithm for undirected connectivity, settling a long-standing open problem, by taking advantage, among other things, of the simple combinatorial composition of the Zig-Zag product.

Using the Zig-Zag product, [120] gave an expander construction with spectral gap $1 - O(D^{-\frac{1}{4}})$. Another construction that appeared in [120] 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 et al. posed the question

[1] Their constructions requires $D - 1$ to be a prime power.

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 [29] 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 i'th 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 et al. [5] showed, somewhat surprisingly, an algebraic interpretation of the Zig-Zag product over non-Abelian Cayley graphs. This lead to new iterative constructions of Cayley expanders [100, 123], 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.

In Chapter 2 we develop 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 [120].

The results of this chapter appear in [25]:

> A. Ben-Aroya and A. Ta-Shma, A combinatorial construction of almost-Ramanujan graphs using the Zig-Zag product, SIAM Journal on Computing, 40(2):267–290, 2011. Earlier version in STOC'08

1.1.2 Quantum expanders

One way to view a regular graph is by its transition matrix. This matrix maps any probability distribution π over the graph's vertices to the probability distribution obtained by choosing a vertex according to π and then taking a random step from the resulting vertex to a random adjacent vertex in the graph.

A graph is of low degree if its transition matrix can be written as the average of a few permutation matrices. A graph is a good algebraic expander if its transition matrix has a large spectral gap.

Each of the above notions has a natural and meaningful generalization in the quantum setting: The notion of a probability distribution is generalized to that of a density matrix (or mixed state). The transition matrix is, in turn, generalized to an admissible quantum transformation (or superoperator). We can say that a superoperator is of low degree if it can be expressed as the sum of a few

of unitary matrices, and we can analyze its spectral gap just like in the classical setting. Using these generalizations, we arrive to the notion of *quantum expanders*.

Unlike classical expanders, quantum expanders are not graphs, but rather transformations that allow one to manipulate quantum states in an interesting manner. This is explained in more detail in Chapter 3.

But why should we wish to pursue such a "quantization" of expanders in the first place? Our main reason is that since classical expanders are fundamental objects in computer science, we believe that their quantum counterparts should also be useful. We demonstrate that this is indeed the case by giving an application of quantum expanders to quantum complexity theory. Moreover, independent of our work, Hastings [65] gave a similar definition as well as another application of quantum expanders. Finally, looking back at a work of Ambainis and Smith [9] regarding quantum one-time pads, we find that, in fact, they have implicitly used a quantum expander (of non-constant degree). We think that these applications (and more, see Chapter 3) show that our definition is indeed a very natural and useful one. We hope that these interesting objects will find more applications in the future.

To summarize, in Chapter 3 we introduce a new object called a quantum expander, which generalize classical expanders in a natural way. We then go on to give two constructions of quantum expanders, one of which is fully-explicit. Finally, we give an application of quantum expanders to quantum statistical zero-knowledge.

The results of this chapter appear in [15]:

> A. Ben-Aroya, O. Schwartz, and A. Ta-Shma, Quantum expanders: motivation and constructions, Theory of Computing, 6(3):47–79, 2010. Earlier version in CCC'08.

1.1.3 Constructing Small-Bias Sets from Algebraic-Geometric Codes

Our last result about expanders is concerned with Cayley expanders over the abelian group Z_2^k. For a set $S \subseteq Z_2^k$, the the Cayley graph $C(Z_2^k, S)$ is a graph over the set Z_2^k. This graph contains an edge (g_1, g_2) if and only if $g_1 = g_2 \oplus s$ for some $s \in S$. Given k and ε, we are interested in finding a set S as small as possible such that the graph $C(Z_2^k, S)$ has spectral gap at least $1 - \varepsilon$.

An ε-biased set is a set $S \subseteq \{0,1\}^k$ that ε-fools every non-trivial linear function over the Boolean cube. In other words, S is ε-biased if for every non-empty subset $T \subseteq [k]$, the binary random variable $\bigoplus_{i \in T} s_i$, where s is sampled uniformly from S, has bias at most ε (here s_i denotes the ith bit of s). It is well known that a set S is ε-biased if and only if $C(Z_2^k, S)$ has spectral gap at least $1 - \varepsilon$ (see, e.g., [68, Proposition 11.7]).

In Chapter 4 we give an explicit construction of an ε-biased set $S \subseteq \{0,1\}^k$ of size $O\left(\frac{k}{\varepsilon^2 \log(1/\varepsilon)}\right)^{5/4}$. This improves upon previous explicit constructions when ε is roughly (ignoring logarithmic factors)

in the range $[k^{-1.5}, k^{-0.5}]$. The new ingredient in the construction is an algebraic-geometric code that is based on low-degree divisors whose degree is significantly smaller than the genus. Additionally, the chapter contains a discussion of the limits of our approach, based on a follow up work of Voloch [141].

The results of this chapter appear in [21]:

> A. Ben-Aroya and A. Ta-Shma, Constructing small-bias sets from algebraic-geometric codes, Proceedings of the 50th IEEE Symposium on Foundations of Computer Science (FOCS), pages 191–197, 2009.

1.2 Extractors

The second part of this thesis is concerned with randomness extractors. Originally, extractors were conceived to refine weak random sources, i.e., to transform distributions that contain some min-entropy to distributions which are close to uniform. These objects have found a wide variety of applications in theoretical computer science (see [127]).

Formally, an (n, k, ε) strong extractor E is a function $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ with the guarantee that for every distribution X over $\{0,1\}^n$ with min-entropy at least k, the distribution obtained by picking a seed y uniformly from $\{0,1\}^t$ and outputting $y \circ E(X,y)$ is ε-close to the uniform distribution over $\{0,1\}^m$. That is, the distribution $U_t \circ E(X, U_t)$ is ε-close to U_{t+m}, where U_ℓ denotes the uniform distribution over $\{0,1\}^\ell$.

Now, consider the case in which an adversary holds some quantum side-information about X. We model this side-information as a mixed state $\rho(x)$ defined for every x in the support of X. An extractor is secure against this side-information if the mixed state $U_t \circ E(X, U_t) \circ \rho(X)$ is ε-close to $U_{t+m} \circ \rho(X)$. For achieving security against such side-information, one has to impose some constraints on ρ (for otherwise it might be the case that $\rho(X)$ completely describe X). In this part of the thesis we focus on extractors that are secure against such side-information, where the constraints on ρ vary.

1.2.1 A Hypercontractive Inequality for Matrix-Valued Functions

One of the main tools in Fourier analysis on the Boolean cube is a hypercontractive inequality that is sometimes called the *Bonami-Beckner inequality*. In Chapter 5 we discuss a generalization of this inequality to matrix-valued functions, based on the work of Ball, Carlen, and Lieb [11].

We prefer to leave the part discussing the hypercontractive inequality to the chapter itself, as it does not concern the main topic of this thesis directly. Instead, we prefer to focus on the connection of our results to extractors. The connection, in fact, comes from our main application of this

inequality. We study an information-theoretic primitive called *k-out-of-n random access code*. This object allows encoding any n-bit string x into m qubits, in such a way that for any set $S \subseteq [n]$ of k indices, the k-bit substring x_S can be recovered with probability at least p by making an appropriate measurement on the encoding. Using the hypercontractive inequality we show that good k-out-of-n random access codes do not exist. More precisely, we show that if $m \ll n$ then the success probability p decays exponentially with k, i.e., $p \leq 2^{-\Omega(k)}$.

Let $F : \{0,1\}^n \times \binom{[n]}{k} \to \{0,1\}$ be a function defined by $F(x,S) = \bigoplus_{i \in S} x_i$, i.e., by taking the XOR of the k bits of x_S. The core of the proof for the impossibility result on k-out-of-n random access codes is to show that the function F is, in fact, an extractor which is secure against quantum side-information that is limited by storage.

More formally, call a function $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ an (n,k,b,ε) *strong extractor against quantum storage* if for any distribution X over $\{0,1\}^n$ with min-entropy at least k, and for any side-information ρ that maps every x in the support of X to a mixed state over b qubits, the mixed state $U_t \circ E(X, U_t) \circ \rho(X)$ is ε-close to $U_{t+m} \circ \rho(X)$. Using this terminology, what we show in Chapter 5 is that F is an (n,n,m,p) strong extractor against quantum storage, for $m \ll n$ and $p \leq 2^{-\Omega(k)}$. (In the chapter itself we prefer to use the notion of XOR random access codes, rather than discussing extractors, but the equivalence between the notions is immediate.)

Beside being interesting and natural objects in their own right, k-out-of-n random access codes have applications in communication complexity. Specifically, we use our bound on these codes to obtain a direct product theorem for one-way quantum communication complexity.

We also give an additional application of the hypercontractive inequality. We use it to derive a "non-quantum" proof of the result of Kerenidis and de Wolf [81] that 2-query locally decodable codes require exponential length.

To summarize, our results in Chapter 5 are:

- deriving the hypercontractive inequality (using the inequality of Ball, Carlen, and Lieb [11]);
- using it to show that the function F described above is a strong extractor against quantum storage;
- using the result on F to obtain a bound on k-out-of-n random access codes;
- obtaining a direct product theorem for one-way quantum communication complexity (using the bound on k-out-of-n random access codes);
- giving an alternative proof of the fact that error-correcting codes that are locally decodable with 2 queries require length exponential in the length of the encoded string.

The results of this chapter appear in [19]:

A. Ben-Aroya, O. Regev and R. de Wolf., A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs, Proceedings of the 49th IEEE Symposium on Foundations of Computer Science (FOCS), pages 477–486, 2008.

We mention that much after the publication of our results, our bound on random access codes was improved by De and Vidick [41].

1.2.2 Better short-seed quantum-proof extractors

The result in the previous discussed a specific XOR-extractor without caring much about the parameters it achieved (since it was merely used as a tool to study random access codes). In this section, however, we are more concerned with obtaining extractors that are secure against (various models of) quantum side-information, but also have good parameters, i.e., they use short seeds and output many bits. One motivation for explicitly constructing such extractors comes from the *privacy amplification problem*.

In this problem Alice and Bob share information that is only partially secret with respect to an eavesdropper Charlie. Their goal is to distill this information to a shorter string that is completely secret. The problem was introduced in [27, 26] for classical eavesdroppers. We are interested in a variant of the problem in which the eavesdropper is allowed to keep quantum information rather than just classical information. This variant was introduced by König, Maurer and Renner [85]. This situation naturally occurs in analyzing the security of some quantum key-distribution protocols [38] and in bounded-storage cryptography [88, 86].

The shared information between Alice and Bob is modeled as a shared string $x \in \{0,1\}^n$, sampled according a distribution X. The information of the eavesdropper is modeled as a mixed state, $\rho(x)$, which might correlated with x.

The privacy amplification problem can be solved by Alice and Bob, but only by using a (hopefully short) random seed y, which can be public. In this case, Alice and Bob can solve the problem by applying on their shared input x and the public random string y an extractor E that "can handle" the side-information $\rho(x)$.

We say that E is an *ε-strong extractor for a family of inputs* Ω, if for any distribution X and any quantum system ρ such that $(X;\rho) \in \Omega$, the distribution $Y \circ E(X,Y) \circ \rho$ is ε-close to $U \circ \rho$, where U denotes the uniform distribution. (See Section 6.1.2 for precise details.)

Clearly, no randomness can be extracted if, for every x, it is possible to recover x from the side information $\rho(x)$. We say the *conditional min-entropy* of X with respect to $\rho(X)$ is k, if an adversary holding the state $\rho(x)$ cannot guess the string x with probability higher than 2^{-k}. Roughly speaking, if one can extract k almost uniform bits from a source X in spite of the side

no. of truly random bits	no. of output bits	classical	quantum-proof
$O(n)$	$m = k - O(1)$	Pair-wise independence, [71]	✓[85]
$O(n - k + \log n)$	$m = n$	Fourier analysis, collision [43]	✓[48]
$\Theta(m)$	$m = k - O(1)$	Almost pair-wise ind., [128, 55]	✓, [133]
$O(\frac{\log^2 n}{\log(k)})$	$k^{1-\zeta}$	Designs, [135]	✓, [41]
$O(\log n)$	$m = \Omega(n)$	[110, 35]	✓, This thesis, $k > (\frac{1}{2} + \zeta)n$
$\log n + O(1)$	$m = k - O(1)$	Lower bound [110, 116]	✓

Table 1.1: Explicit quantum-proof (n, k, ε) strong extractors. To simplify parameters, the error ε is a constant.

information $\rho(X)$, then the state $X \circ \rho(X)$ is close to another state with conditional min-entropy at least k.[2] Thus, in a very concrete sense, the ultimate goal is finding extractors for sources with high conditional min-entropy.[3] We say E is a *quantum-proof* (n, k, ε) strong extractor if it extracts randomness from every input $(X; \rho)$ with conditional min-entropy at least k.

Not every classical extractor is quantum-proof, as was shown by Gavinsky et al. [53]. On the positive side, several well-known classical extractors are quantum-proof. Table 1.1 lists some of these constructions. We remark that the best explicit classical extractors [59, 46, 45] achieve significantly better parameters than those known to be quantum-proof.

A simpler adversarial model is the "bounded storage model" where the adversary may store a limited number of qubits, which was discussed in the previous section. The only advantage of the bounded storage model for extractors is that it simplifies the proofs, and allows us to achieve results which currently we cannot prove in the general model. Recall that E is an (n, k, b, ε) strong extractor against quantum storage if it extracts randomness from every pair $(X; \rho)$ for which X has at least k min-entropy and for every x, $\rho(x)$ is a mixed state with at most b qubits.

Our results also concern a slight generalization of the bounded storage model. We say E is a *quantum-proof* (n, f, k, ε) strong extractor for *flat distributions* if it extracts randomness from every input $(X; \rho)$ for which X is a flat distribution (meaning it is uniform over its support) with exactly f min-entropy and the conditional min-entropy is at least k. In Chapter 6 (specifically in Lemma 6.12) we prove the easy observation that any quantum-proof (n, f, k, ε) strong extractor for flat distributions is also a $(n, f, f - k, \varepsilon)$ strong extractor against quantum storage.

Our results, described in Chapter 6, are as follows. First, we show a generic reduction from the problem of constructing quantum-proof (n, f, k, ε) strong extractors for flat distributions to the

[2] Such a source is said to have conditional *smooth* min-entropy k.

[3] A simple argument shows an extractor for sources with high conditional min-entropy is also an extractor for sources with high conditional smooth min-entropy.

problem of constructing quantum-proof $((1+\alpha)f, f, k, \varepsilon)$ strong extractors for flat distributions, and a similar reduction for the bounded storage model. In other words, in our model the quantum adversary may have two types of information about the source: first, it may have some classical knowledge about it, reflected in the fact that the input x is taken from some classical flat distribution X, and second, it holds a quantum state that contains some information about the source. The reduction shows that without loss of generality we may assume the classical input distribution is almost uniform. The reduction uses a purely classical object called a *strong lossless condenser* and extends work done in [132] on extractors to quantum-proof extractors. This reduction holds for any setting of the parameters.

We then augment this with a simple construction that shows how to obtain a quantum-proof $((1+\alpha)f, f, k = (1-\beta)f, \varepsilon)$ strong extractor for flat distributions, provided that $\beta < \frac{1}{2}$. The argument here builds on work done in [110] on composition of extractors and extends it to quantum-proof extractors. Together, these two reductions give:

Theorem 1.1. *For any $\beta < \frac{1}{2}$ and $\varepsilon \geq 2^{-k^\beta}$, there exists an explicit quantum-proof $(n, k, (1-\beta)k, \varepsilon)$ strong extractor for flat sources $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ with seed length $t = O(\log n + \log(\frac{1}{\epsilon}))$ and output length $m = \Omega(k)$.*

Consequently,

Theorem 1.2. *For any $\beta < \frac{1}{2}$ and $\varepsilon \geq 2^{-k^\beta}$, there exists an explicit $(n, k, \beta k, \varepsilon)$ strong extractor against quantum storage, $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log(\frac{1}{\epsilon}))$ and output length $m = \Omega(k)$.*

This gives the first logarithmic seed length extractor against b quantum storage that works for every min-entropy k and extracts a constant fraction of the entropy, and it is applicable whenever $b = \beta k$ for $\beta < \frac{1}{2}$.

We would like to stress that in most practical applications, and in particular in cryptographic applications such as quantum key distribution, it is generally impossible to bound the *size* of the side information. For example, in quantum key distribution where extractors are used for privacy amplification, the conditional min-entropy of the source can be estimated by measuring the noise on the channel, whereas any estimate on the adversary's memory is an unproven assumption. Thus, an extractor proven to work only against quantum storage cannot be used in quantum key distribution protocols. We nevertheless feel that proving a result in the bounded storage model may serve as a first step towards solving the general question.

In fact, the second component in the above construction also works in the general quantum-proof setting. Specifically, this gives an extractor with seed length $t = O(\log n + \log(\frac{1}{\epsilon}))$ that extracts $\Omega(n)$ bits from any source with conditional min-entropy at least $(1-\beta)n$ for $\beta < \frac{1}{2}$.

Theorem 1.3. *For any $\beta < \frac{1}{2}$ and $\varepsilon \geq 2^{-n^{\beta}}$, there exists an explicit quantum-proof $(n, (1-\beta)n, \varepsilon)$ strong extractor $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log(\frac{1}{\epsilon}))$ and output length $m = \Omega(n)$.*

The results of this chapter appear in [24]:

A. Ben-Aroya and A. Ta-Shma, Better short-seed extractors against quantum knowledge, Theoretical Computer Science, To appear, 2011.

Chapter 2

A combinatorial construction of almost-Ramanujan graphs

Our main result in this chapter 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. We use the new variant of the Zig-Zag product to construct an explicit family of D–regular expanders with spectral gap $1 - D^{-\frac{1}{2}+o(1)}$.

2.1 Introduction

2.1.1 An intuitive description of the new product

The Zig-Zag product

Let us review the Zig-Zag product of [120]. 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 has $N_2 = D_1$ 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 i'th edge leaving a vertex v is simply the edge that v labeled with label i.

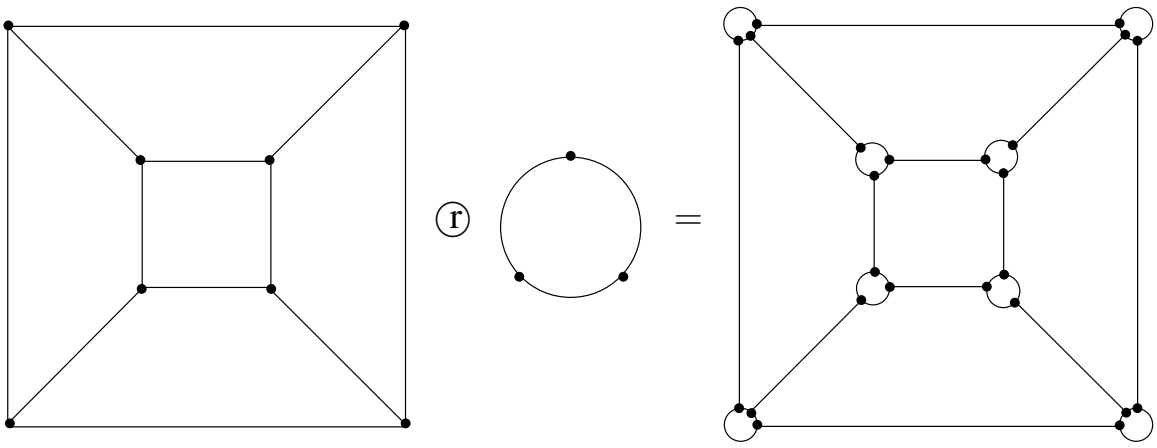


Figure 2.1: The replacement product between the cube and the 3-cycle

The replacement product results in a graph with $N_1 N_2$ 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 "inter-cloud" edge between (v,i) and (w,j) if $e=(v,w)$ is an edge in G_1 and e is the i'th edge leaving v and the j'th edge leaving w. Besides these inter-cloud edges there are only edges that connect vertices within the same cloud. The "intra-cloud" 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 2.1 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 intra-cloud edges and the middle step is an inter-cloud edge. Namely, the vertices of the Zig-Zag product graph $G_1 ⓩ H$ are the 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 inter-cloud 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 [120]) that

the second largest eigenvalue of $G_1 ⓩ H$ is small.

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 entropy. We shall consider two special distributions and we shall see that taking a random step over $G_1 ⓩ H$ when starting from any of these two distributions substantially increases their entropy. 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 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, that is, 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., $\forall v \in [N_1] \, \forall i, j \in [N_2] \;\; P(v, i) = P(v, j)$. 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.
- 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 intra-cloud step and $\dot{G}_1$ denote the transformation corresponding to an inter-cloud step. Namely, $\tilde{H}$ is the transition matrix of the subgraph of the replacement product graph that contains only the intra-cloud edges, and $\dot{G}_1$ is the transition matrix of the subgraph that contains all the inter-cloud edges. In particular, $\tilde{H} + \dot{G}_1$ 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 intra-cloud step followed by an inter-cloud and another intra-cloud 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 $\bar{\lambda}$ 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 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 $\left\|\tilde{H}a\right\| \leq \bar{\lambda}(H)\|a\|$.
- 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.[1]

[1] 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.

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 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 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 at 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 non-optimality 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 $\bar{\lambda}(H)$.

The k-step Zig-Zag product

In this chapter 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).

- This is immediately followed by an inter-cloud 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* 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 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 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 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 i'th edge leaving v then e is also the i'th 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

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

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

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

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 2.4.1. The above phenomena 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 the 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 2.2 illustrates the entire process that parallel vectors undergo.

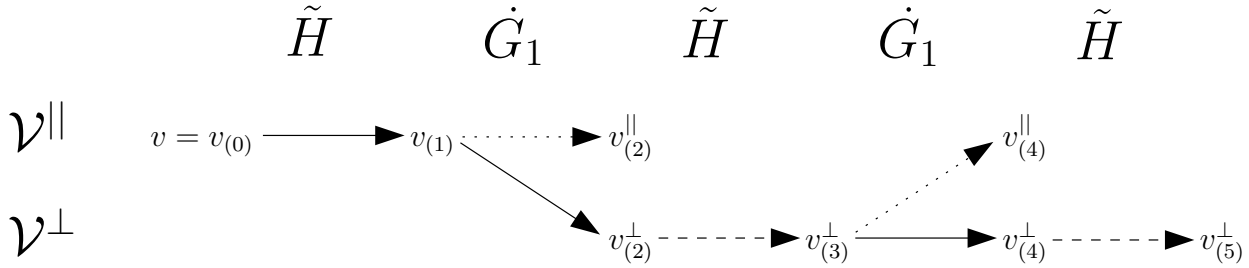


Figure 2.2: The action of a 3-step Zig-Zag on a parallel vector $v \in \mathcal{V}^{||}$. The process is composed of 5 steps and $v_{(t)}$ denotes the vector after the tth step.

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 be can decreased to any constant by increasing D_1 without affecting D_2 and the degree of the resulting graph. One can interpret the above result as saying that $k-1$ out of the k steps worked for us!

An almost-Ramanujan expander construction

We now go back to the iterative expander construction of [120] and replace the Zig-Zag component there 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 2.7). Doing the iterative construction we get a degree D expander, by taking k steps over the graphs $\{H_i\}$, each of degree D_2. Roughly speaking, the resulting second-largest eigenvalue is 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 to, roughly, a $\sqrt{D_2}$ multiplicative factor. We also have a second loss of 2^{k-1} multiplicative factor that corresponds to the fact that H^k is not optimal even when H is. Balancing these losses gives:

Theorem 2.1. *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 $\bar{\lambda}(G_i) \leq D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})}$.*

2.1.2 Organization of the chapter

In Section 2.2 we give preliminary definitions. Section 2.3 contains the formal definition of the k-step Zig-Zag product. Section 2.4 gives the proof of the main statement regarding the k-step Zig-Zag product, assuming good small graphs exist. The fact that such graphs exist is proven in Section 2.5. In Section 2.6 we use the product to give an iterative construction of expanders, for degrees of a specific form. Finally, in Section 2.7 we describe how to make the expander construction work for any degree.

2.2 Preliminaries

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

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

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

We define $\bar{\lambda}(G) = s_2(G)$. We say a graph G is a (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 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 $\lambda_1 \geq \ldots \geq \lambda_N$ are real. In this case,

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

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

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

Ramanujan graphs are essentially the optimal algebraic expanders [109].

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 $\mathbf{1} \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}}(1, \ldots, 1)^t$ 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 a $(N, 2D, \lambda)$ 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 2.2. 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 they v_j is connected to u_j in G_j, for $j = 1, 2$.

Fact 2.3. 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 [120] 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 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 i'th neighbor of v in G. We define the rotation map of G, $\mathrm{Rot}_G : V \times [D] \to V \times [D]$, by

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

In words, the i'th neighbor of v in G is w, and the j'th 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:

$$\forall 1 \leq j \leq \ell,\ \mathrm{Rot}_G(v_j, i_j) = (v_{j+1}, i'_j) \implies \mathrm{Rot}_{G^\ell}(v_1, (i_1, \ldots, i_\ell)) = \big(v_{\ell+1}, (i'_\ell, \ldots, i'_1)\big) \quad (2.1)$$

$$\forall 1 \leq j \leq 2,\ \mathrm{Rot}_{G_j}(v_j, i_j) = (u_j, i'_j) \implies \mathrm{Rot}_{G_1\otimes G_2}((v_1, v_2), (i_1, i_2)) = \big((u_1, u_2), (i'_1, i'_2)\big) \quad (2.2)$$

$$\text{For a directed graph } G,\ \ \mathrm{Rot}_G(v,i)=(u,i') \implies \mathrm{Rot}_{\frac{1}{2}[G+G^\dagger]}(v,(b,i))=\big(u,(1-b,i')\big)\ ^2 \tag{2.3}$$

We single out a special family of rotation functions:

Definition 2.4. A graph G is *locally invertible* if its rotation map is of the form $\mathrm{Rot}_G(v,i) = (v[i],\phi(i))$ for some permutation $\phi:[d]\to[d]$. We say that ϕ is the local inversion function.

In [119], 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.[3]

A simple fact following immediately from Equations (2.1)-(2.3).

Fact 2.5. 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 then $x^{(1)}$ 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 2.6. *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 $\tau,\xi\in\Lambda$,*

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

[3] 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 [69]. 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.

Proof: First, notice that

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

Therefore, it follows that

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

■

2.3 The k-step Zig-Zag product

2.3.1 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 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 a (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 pre-images 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, $v^{(1)}_{2t-1} = v^{(1)}_{2(t-1)}$ and we set $(v^{(2)}_{2t-1}, i'_t) = \mathrm{Rot}_{H_t}(v^{(2)}_{2(t-1)}, i_t)$.

– If $t < k$, we take one step on G_1 with $\pi_1(v^{(2)}_{2t-1})$ as the $[D_1]$ label to be used, i.e.,

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

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

$$\psi(v^{(2)}) \;=\; (\phi_{G_1}(\pi_1(v^{(2)})), \pi_2(v^{(2)}), \pi_3(v^{(2)}), \ldots, \pi_{4k}(v^{(2)})). \tag{2.4}$$

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

$$\mathrm{Rot}_{G_{\mathrm{new}}}(v, \bar{i}) = \left((v^{(1)}_{2k-1}, v^{(2)}_{2k-1}), (i'_k, \ldots, i'_1)\right). \tag{2.5}$$

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 2.2.

The following is immediate from the definition of the rotation map in Equation (2.5).

Fact 2.7. 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

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

2.3.2 The linear operators

We want to express the k-step walk described in Section 2.3.1 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 Equation (2.4). Having this terminology, the transition matrix of the new graph G_{new} is the linear transformation on $\mathcal{V}$ defined by $\tilde{H}_k \dot{G}_1 \tilde{H}_{k-1} \dot{G}_1 \dots \tilde{H}_2 \dot{G}_1 \tilde{H}_1$.

2.3.3 The action of the composition

Next we take advantage of the simple structure of locally invertible graphs, revealing how

$$\tilde{H}_k \dot{G}_1 \tilde{H}_{k-1} \dot{G}_1 \dots \tilde{H}_2 \dot{G}_1 \tilde{H}_1$$

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 2.8. *Assume G_1 has a local inversion function ϕ that is extended to a permutation $\psi : V_2 \to V_2$ as in Equation (2.4). 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 $\tilde{\Gamma}_i = I \otimes \Gamma_i$.*

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

$$q_0(u^{(2)}) = u^{(2)} \tag{2.6}$$

$$q_i(u^{(2)}) = \gamma_i(\psi(q_{i-1}(u^{(2)}))). \tag{2.7}$$

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:

Definition 2.9. 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 Equation (2.4). Let $\bar{\gamma} = (\gamma_1, \ldots, \gamma_\ell)$ be a sequence of ℓ permutations over V_2. The permutation sequence *induced by* $(\bar{\gamma}, \phi)$ is $\bar{q} = (q_0, \ldots, q_\ell)$ defined as in Equations (2.6) and (2.7).

Corollary 2.10. *Assume G_1 has a local inversion function ϕ. 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 $\tilde{\Gamma}_i = I \otimes \Gamma_i$.*

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

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

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

2.3.4 A condition guaranteeing good algebraic expansion

We say $\bar{\gamma} = (\gamma_1, \ldots, \gamma_\ell)$ 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:

Definition 2.11. 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 $\bar{\gamma}$ is *ε–pseudorandom* with respect to ϕ (or, equivalently, ε–pseudorandom with respect to G_1) if

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

where $\pi_1(q_0(U)) \circ \ldots \circ \pi_1(q_\ell(U))$ is the distribution obtained by picking $v^{(2)} \in V_2$ uniformly at random and outputting $(\pi_1(q_0(v^{(2)})), \ldots, \pi_1(q_\ell(v^{(2)})))$ 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 2.11 to a sequence of k D_2–regular graphs.

Definition 2.12. Let G_1 and ϕ be as above. Let $\bar{H} = (H_1, \ldots, H_k)$ be a k-tuple of D_2–regular graphs over V_2. We say $\bar{H}$ 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 $1 \leq \ell_1 \leq \ell_2 \leq k$, $j_{\ell_1}, \ldots, j_{\ell_2} \in [D_2]$, the sequence $\gamma_{\ell_1,j_{\ell_1}}, \ldots, \gamma_{\ell_2,j_{\ell_2}}$ 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 zigzag product does not lose much in the spectral gap. Formally,

Theorem 2.13. *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, $G_{\mathrm{new}} = G_1 ⓩ \bar{H}$ is a $(N_1 \cdot N_2, D_2^k, f(\lambda_1, \lambda_2, \varepsilon, k))$ graph for*

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

Given $D = D_2^k$ we wish to construct a D–regular graph with a spectral gap as large as we can. We have freedom in choosing the constant D_1 since it has no effect on the degree of the graph we construct. By choosing it to be 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 2.13, which is done in the following section.
2. Find an ε-good sequence $\bar{H}$. In fact, in Section 2.5, we prove that almost all sequences are good.

2.4 A top-down view of the proof

Proof of Theorem 2.13: 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 $v \in \mathcal{V}^{||}$ iff $v = v_1 \otimes \mathbf{1}$ for some $v_1 \in \mathcal{V}_1$.

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

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

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:

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

Now,

- $\left| y_0^{\dagger} \tilde{H}_k x_{k-1}^{\perp} \right| \leq \left\| \tilde{H}_k x_{k-1}^{\perp} \right\| \leq \lambda_2 \left\| x_{k-1}^{\perp} \right\| \leq \lambda_2 \|x_{k-1}\| \leq \lambda_2 \lambda_2^{k-1} \|x_0\| = \lambda_2^k.$
- Since $\mathcal{V}^{\perp} \perp \mathcal{V}^{||}$, the term $\sum_{i=1}^{k} (y_{k-i}^{\perp})^{\dagger} x_{i-1}^{||}$ is simply 0.
- The term
$$\left| \sum_{i=1}^{k} (y_{k-i}^{||})^{\dagger} x_{i-1}^{||} \right| \leq \sum_{i=1}^{k} \left\| y_{k-i}^{||} \right\| \cdot \left\| x_{i-1}^{||} \right\|$$
is bounded in Lemma 2.16 by λ_2^{k-1}.
- Finally, we are left with the term

$$\sum_{1\le i<j\le k} (y^{||}_{k-j})^\dagger \dot{G}_1 \tilde{H}_{j-1} \dots \tilde{H}_{i+1} \dot{G}_1 x^{||}_{i-1}. \tag{2.8}$$

In Theorem 2.14 we show that, assuming $\bar{H}$ is good for G_1,

$$\left|(y^{||}_{k-j})^\dagger \dot{G}_1 \tilde{H}_{j-1} \dots \tilde{H}_{i+1} \dot{G}_1 x^{||}_{i-1}\right| \;\le\; (\lambda_1^{j-i}+\varepsilon)\left\|y^{||}_{k-j}\right\|\left\|x^{||}_{i-1}\right\|.$$

Therefore, the term in Equation (2.8) is bounded by

$$\sum_{1\le i<j\le k} (\lambda_1^{j-i}+\varepsilon)\left\|y^{||}_{k-j}\right\|\left\|x^{||}_{i-1}\right\| \;=\; \sum_{t=1}^{k-1}(\lambda_1^t+\varepsilon)\sum_{i=1}^{k-t}\left\|y^{||}_{k-i-t}\right\|\left\|x^{||}_{i-1}\right\|$$
$$\le\; \sum_{t=1}^{k-1}(\lambda_1^t+\varepsilon)\lambda_2^{k-t-1} \;\le\; (\lambda_1+\varepsilon)\sum_{t=1}^{k-1}\lambda_2^{k-t-1} \;=\; (\lambda_1+\varepsilon)\sum_{i=0}^{k-2}\lambda_2^i \le 2(\lambda_1+\varepsilon),$$

where the first inequality follows from Lemma 2.16 and the second one uses the assumption $\lambda_2 \le \frac{1}{2}$.

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

2.4.1 The action of the operator on parallel vectors

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

Theorem 2.14. *For every $i, \ell \ge 1$ and $\tau, \xi \in \mathcal{V}^{||}$, $\tau, \xi \perp \mathbf{1}_V$,*

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

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 2.15. *Suppose $\bar{\gamma} = (\gamma_1, \dots, \gamma_\ell)$ 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:*

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

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 2.10,

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

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 $(\gamma_1, \dots, \gamma_k)$ is ε–pseudorandom with respect to G_1 we know that $\left| \mathcal{Z} - U_{[D_1]^k} \right|_1 \leq \varepsilon$. By Claim 2.6,

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

(since $\|\tau\| = \left\| \tau^{(1)} \otimes \mathbf{1} \right\| = \left\| \tau^{(1)} \right\| \cdot \|\mathbf{1}\| = \left\| \tau^{(1)} \right\|$) and this completes the proof of Lemma 2.15. ■

Having Lemma 2.15 we can prove Theorem 2.14.

Proof of Theorem 2.14: 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 $\gamma_{1,j_1}, \dots, \gamma_{k,j_k}$ 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 $\tilde{\Gamma}_{i,j} = I \otimes \Gamma_{i,j}$ be the corresponding operator on $\mathcal{V}_1 \otimes \mathcal{V}_2$.

Observe that:

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

Thus, by Lemma 2.15,

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

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 $\|\tau\| = \left\|\tau^{(1)}\right\|$ and $\|\xi\| = \left\|\xi^{(1)}\right\|$ completes the proof. ∎

2.4.2 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 2.16. *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 $\sum_{i=0}^{k-t-1} \left\| x_i^{||}/\lambda_2^i \right\|^2$ and the bound for the expression $\sum_{i=0}^{k-t-1} \left\| y_i^{||}/\lambda_2^i \right\|^2$ 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

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

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

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

■

2.5 Almost any $\bar{H}$ is good

2.5.1 A Hyper-Geometric lemma

We shall need the following tail estimate:

Theorem 2.17 ([73], 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 $\varepsilon > 0$,*

$$\Pr_{S_2}[|\, |S_1 \cap S_2| - \mu \big| \geq \varepsilon\mu] \;\leq\; 2e^{-\frac{\varepsilon^2}{3}\mu}.$$

A simple generalization of this gives:

Lemma 2.18. *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 $\mu_k = \mathbb{E}_{S_2,\ldots,S_k}[\,|\, S_1 \cap S_2 \ldots \cap S_k \,|\,] = \frac{m^k}{|\Omega|^{k-1}}$. Then for every $\varepsilon > 0$,*

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

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

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

Proof: By induction on k. The case $k = 2$ follows from Theorem 2.17. Assume for k, and let us prove for $k+1$. Let $A = S_1 \cap \ldots \cap S_k \subseteq \Omega$. 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 2.17, $|A \cap S_{k+1}|$ is in the range

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

except for probability

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

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

2.5.2 Almost any $\bar{\gamma}$ is pseudorandom

The main lemma we prove in this section is:

Lemma 2.19. *For every* $\varepsilon \leq \frac{1}{2}$, *a sequence of uniformly random and independent permutations* $(\gamma_1, \ldots, \gamma_{k-1})$ *satisfies*

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

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 Equation (2.4). Let A denote the distribution $\pi_1(q_1(U)) \circ \ldots \circ \pi_1(q_k(U))$ and $U_{D_1^k}$ 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:

$$\Pr_{\gamma_1,\ldots,\gamma_{k-1}}[|A(\bar{r}) - U_{D_1^k}(\bar{r})| \geq \varepsilon D_1^{-k}] \;\leq\; 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})}. \tag{2.9}$$

Therefore, using a simple union bound, the event $\exists \bar{r}\ |A(\bar{r}) - U_{D_1^k}(\bar{r})| \geq \varepsilon D_1^{-k}$ 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,

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

We now prove the inequality in Equation (2.9). Let $S_i = \{x \in V_2 \mid \pi_1(q_i(x)) = r_i\}$, for $1 \leq i \leq k$. 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, $A(\bar{r}) = \frac{|S_1 \cap S_2 \ldots \cap S_k|}{|V_2|}$. Also,

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

and $U_{D_1^k}(\bar{r}) = \frac{\mu_k}{|V_2|} = D_1^{-k}$. Thus, by Lemma 2.18, and setting $\zeta = \frac{\varepsilon}{2k} \leq \frac{1}{4k}$,

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

as stated in Equation (2.9). ∎

2.5.3 The spectrum of random D-regular graphs

Friedman [49] proved the following theorem regarding the spectrum of random regular graphs. The distribution $G_{N,D}$ is described in Section 2.2.

Theorem 2.20 ([49])**.** *For every $\delta > 0$ and for every even D, there exists a constant $c > 0$, independent of N, such that*

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

2.5.4 Almost any $\bar{H}$ is good

Theorem 2.21. *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 $G_{N_2=D_1^k,D_2}$. I.e., 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 $\bar{H}$ 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 2.19, $\bar{H}$ is not ε–pseudorandom with respect to G_1 with probability at most $k^2 \cdot D_2^k \cdot D_1^k \cdot 2ke^{-\Omega(\frac{\varepsilon^2 D_1^{3k}}{k^2})}$. [4] 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 2.20, the probability that there exists a graph H_i in $\bar{H}$ with $\bar{\lambda}(H_i) \geq \lambda_{\mathrm{Ram}}(D_2) + \varepsilon$ 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. ∎

2.6 The iterative construction

In [120] 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

[4] 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 sub-sequences $1 \leq \ell_1 \leq \ell_2 \leq k$.

tensoring (which makes the graph much larger, the degree larger and the spectral gap the same), powering (which keeps the graph size the same, increases the spectral gap and the degree) and a Zig-Zag product (that 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 we use the k-step zigzag product rather than the Zig-Zag product to reduce the degree. We do it for degrees D of the special form $D = 2D_2^k$.

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 2.21, 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 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 a (N_2, D, λ) graph, and G_2 is a (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:

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

Theorem 2.22. *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 2.23. *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 2.5 and 2.7 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 2.2 and 2.3 and Theorem 2.13.

The only non-trivial 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 2.13,

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

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 2.24. *$\{G_t\}$ 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 to 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 rotations 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 $\bar{\lambda}_{\mathrm{Ram}}(D) = \frac{2\sqrt{D-1}}{D}$. 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 to, roughly, $\sqrt{D_2}$ multiplicative factor. We also have a second loss of 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,

Corollary 2.25. *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,

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

∎

2.7 A construction for any degree

The construction in Section 2.6 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 2.1.

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 $D_2^k \leq D < D_2^{k+1}$ (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 $k \geq 50$.

- 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 2.25.
- 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 $\bar{H}$ exists by Theorem 2.21.)

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

$$\forall 0 \leq i \leq k \quad A_i = A_{i+1} \cdot D_2 + B_{i+1}.$$

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

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

- Z_k is the graph with B_{k+1} self loops.
- For $0 \leq i \leq k$, 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:

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

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 2.1.

Claim 2.26. $\bar{\lambda}(Z_0) \leq D^{-\frac{1}{2}+O(\frac{1}{\sqrt{\log D}})}$.

Proof: Resolving the recursive formula for Z_0 we get

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

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:

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

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

$$\bar{\lambda}(Z_0) \;\leq\; \sum_{i=0}^{k} \frac{D_2^{i+2}}{D} \cdot \bar{\lambda} \left(\prod_{j=1}^{i} \tilde{H}_j \dot{G}_1 \right). \tag{2.10}$$

Note that $\dot{G}_1$ is a unitary transformation, hence for any X, $\bar{\lambda}(X\dot{G}_1) = \bar{\lambda}(X)$. By Theorem 2.13, 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 Equation (2.10) we get,

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

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

Chapter 3

Quantum expanders

In this chapter, we introduce a new object called a quantum expander, which generalize classical expanders in a natural way. We then go on to give two constructions of quantum expanders, one of which is fully-explicit. Finally, we give an application of quantum expanders to quantum statistical zero-knowledge.

3.1 Introduction

The algebraic definition of expansion views a regular graph $G = (V, E)$ as a linear operator on a Hilbert space $\mathcal{V}$ of dimension $|V|$. In this view an element $v \in V$ is identified with a basis vector $|v\rangle \in \mathcal{V}$, and a distribution π on V corresponds to the vector $|\pi\rangle = \sum_{v \in V} \pi(v) |v\rangle$. The action of G on $\mathcal{V}$ is the action of the normalized adjacency matrix $A : \mathcal{V} \to \mathcal{V}$, where the normalization factor is the degree of G, and therefore A maps probability distributions to probability distributions. This mapping corresponds to taking a random walk on G. Specifically, if one takes a random walk on G starting at time 0 with the distribution π_0 on, then the distribution on the vertices at time k is $A^k |\pi_0\rangle$. Viewing G as a linear operator allows one to consider the action of A on arbitrary vectors in $\mathcal{V}$, not necessarily corresponding to distributions over V. While such vectors have no combinatorial interpretation, they are crucial for understanding the spectrum of A; none of the non-trivial eigenvectors of A correspond to probability distributions. To summarize: a D-regular expander $G = (V, E)$ is a linear transformation $A : \mathcal{V} \to \mathcal{V}$ that can be implemented by a classical circuit and maps probability distributions to probability distributions. It is a good expander if it has a *large spectral gap* and a *small degree*.

We now want to extend the definition of D-regular expanders to linear operators that map *quantum states* to *quantum states*. A general quantum state is a *density matrix*, which is a trace 1, positive semidefinite operator, i. e., an operator of the form $\rho = \sum p_v |\psi_v\rangle\langle\psi_v|$, where $0 \leq p_v \leq 1$,

$\sum p_v = 1$, and $\{\psi_v\}$ is an orthonormal basis of $\mathcal{V}$. Notice that $\rho \in L(\mathcal{V}) \triangleq \mathrm{Hom}(\mathcal{V}, \mathcal{V})$.

Among the set of *admissible* quantum transformations $E : L(\mathcal{V}) \to L(\mathcal{V})$, which are those implementable by quantum circuits (allowing both unitary operations and measurements), are those given by the following definition.

Definition 3.1. A superoperator $E : L(\mathcal{V}) \to L(\mathcal{V})$ is a *D-regular admissible superoperator* if

$$E = \frac{1}{D} \sum_{d=1}^{D} E_d \,,$$

where, for each $d \in [D]$, $E_d(X) = U_d X U_d^\dagger$ for some unitary transformation U_d over $\mathcal{V}$.

Note that this definition generalizes the classical one: any D-regular graph can be viewed as a sum of D permutations, each corresponding to a unitary transformation. In fact many classical expander constructions explicitly use this property [120, 35]. The definition is also intuitive in a more basic sense. Unitary transformations (or permutations in the classical setting) are those transformations that do not change the entropy of a state. An operator has small degree if it can never add much entropy to the state it acts upon. Specifically, a degree D operator can never add more than $\log(D)$ entropy. Such a view is almost explicit in the work of Capalbo et al. [35], where they view expanders as entropy conductors.

It is clear that all of the singular values of a D-regular admissible super-operator $E : L(\mathcal{V}) \to L(\mathcal{V})$ are at most 1, and that the completely mixed state $\tilde{I} = I/|V|$ is an eigenvector of any such E, with corresponding eigenvalue 1. We say that such a super-operator E has a $1 - \overline{\lambda}$ spectral gap if all the remaining singular values of E are smaller than $\overline{\lambda}$. This is analogous to the way regular, directed expanders are defined, where the regularity implies that the largest eigenvalue is 1, and furthermore this eigenvalue is obtained with the normalized all-ones vector (that corresponds to the uniform distribution). The spectral gap requires that all other singular values are bounded by $\overline{\lambda}$.

Definition 3.2. A D-regular admissible superoperator $E : L(\mathcal{V}) \to L(\mathcal{V})$ is $\overline{\lambda}$–expanding if:

- The eigenspace of E corresponding to the eigenvalue 1 is the one-dimensional space spanned by $\tilde{I}$.
- For any $B \in L(\mathcal{V})$ orthogonal to $\tilde{I}$, it holds that $\| E(B) \|_2 \leq \overline{\lambda} \| B \|_2$.

We also say that a superoperator $E : L(\mathcal{V}) \to L(\mathcal{V})$ is a $(\dim(\mathcal{V}), D, \overline{\lambda})$ *quantum expander* if it is D-regular and $\overline{\lambda}$–expanding, and that it is *explicit* if it can be implemented by a quantum circuit of size polynomial in $\log(\dim(\mathcal{V}))$. We sometimes omit the dimension and say that E is a $(D, \overline{\lambda})$ quantum expander.

The orthogonality in the above definition is with respect to the *Hilbert-Schmidt inner product*, defined as $\langle A, B\rangle = \mathrm{Tr}(A^\dagger B)$, and the norm is the one induced by this inner product: $\| B \|_2 = \sqrt{\langle B, B\rangle}$.

Definition 3.2 implies that D-regular quantum expanders can never add more than $\log(D)$ entropy to the state they act on, but always add entropy to states that are far away from the completely-mixed state. This definition can be generalized to the more general class of superoperators that can be expressed as the sum of D Kraus operators, but for simplicity we work only with D-regular admissible superoperators. A similar definition was independently given by Hastings [65].

3.1.1 Quantum expander constructions

In this chapter we give two quantum expander constructions. We give a brief review of all the currently known constructions in the order in which they appeared. All of the constructions are essentially based on classical expanders, with a twist allowing them to work in the quantum setting as well.

The first construction was already implicit in the work of Ambainis and Smith [9] on state randomization:

Theorem 3.3 ([9])**.** *For every $\overline{\lambda} > 0$, there exists an explicit $\left(N, O\left(\log^2(N)/\overline{\lambda}^2\right), \overline{\lambda}\right)$ quantum expander.*

Their quantum expander is based on a Cayley expander over the Abelian group $\mathbb{Z}_2^n$. The main drawback of Cayley graphs over Abelian groups is that [84, 7] showed that such an approach cannot yield constant degree expanders. Indeed, this is reflected in the $\log^2 N$ term in Theorem 3.3. There are constant degree, Ramanujan Cayley graphs, i. e., Cayley graphs that achieve the best possible relationship between the degree and the spectral gap, and in fact the construction in [96] and [99] is such, but they are built over non-Abelian groups.

In order to work with general groups, we describe (in Section 3.3.2) a natural way to lift a Cayley graph $G = (V, E)$ into a corresponding quantum superoperator T. However, the analysis shows that the spectral gap of T is 0, and more specifically, T has $|V|$ eigenspaces each of dimension $|V|$, with eigenvalues $\vec{\lambda} = (\lambda_1 = 1, \ldots, \lambda_{|V|})$, where $\vec{\lambda}$ is the spectrum of the Cayley graph.

Our first construction starts with the constant degree Ramanujan expander presented in [96]. This expander is a Cayley graph over the non-Abelian group $\mathrm{PGL}(2, q)$. We build from it a quantum expander as follows: we take two steps on the classical expander graph (by applying the superoperator T twice), with a basis change between the two steps. The basis change is a carefully chosen refinement of the Fourier transform that maps the standard basis $|g\rangle$ to the basis of the irreducible, invariant subspaces of $\mathrm{PGL}(2, q)$. Intuitively, in the Abelian case this basis change corresponds to

dealing with both the bit and the phase degrees of freedom, and is similar to the construction of quantum error correcting codes by first applying a classical code in the standard basis and then in the Fourier basis. However, this intuition is not as clear in the non-Abelian case. Furthermore, in the non-Abelian case not every Fourier transform ensures that the construction works. In this work we single out a natural algebraic property we need from the underlying group that is sufficient for the existence of a good basis change, and we prove that $\mathrm{PGL}(2,q)$ has this property. This results in a construction of a $(D = O(1/\overline{\lambda}^4), \overline{\lambda})$ quantum expander. We describe this construction in detail in Section 3.3.

This construction is not explicit in the sense that it uses the Fourier transform over $\mathrm{PGL}(2,q)$, which is not known to have an efficient implementation. (See [90] for a non-trivial, but still not fast enough, algorithm.) We mention that there are also explicit, constant degree (non-Ramanujan) Cayley expanders over the symmetric group S_n and the alternating group A_n [79]. Also, there is an efficient implementation of the Fourier transform over S_n [12]. We do not know, however, whether S_n (or A_n) respects our additional property.

Following the publication of this construction (given first in [20]), Hastings [66] showed, using elegant techniques, that quantum expanders cannot be better than Ramanujan, i. e., cannot have spectral gap better than $1 - 2\sqrt{D-1}/D$. Hastings also showed that taking D random unitaries gives an almost-Ramanujan expander. This settles the parameters that can be achieved with a *non-explicit* construction. However, Hastings' work does not give an explicit construction, because a random unitary is a highly non-explicit object.

The second construction presented in this chapter adapts the classical Zig-Zag construction [120] to the quantum world. The construction is iterative, starts with a good quantum expander of constant size (that is found with a brute force search), and then builds quantum expanders for larger spaces by repeatedly applying tensoring (which makes the space larger at the expense of the spectral gap), squaring (that improves the spectral gap at the expense of the degree) and a Zig-Zag operation that reduces the degree back to that of the constant-size expander. We again work by lifting the classical operators working over $\mathcal{V}$ to quantum operators working over $L(\mathcal{V})$, and we adapt the analysis along similar lines. The main issue is generalizing and analyzing the Zig-Zag product. Remarkably, this translation works smoothly and gives the desired quantum expanders with almost the same proof applied over $L(\mathcal{V})$ rather than $\mathcal{V}$. The construction gives explicit, constant degree quantum expanders with a constant spectral gap. We describe this construction in detail in Section 3.4.

Two other explicit constructions of quantum expanders were published in [62] and [57] shortly after our work first appeared. In [57] it was shown how the expander of Margulis [98] can be twisted to the quantum setting, and in [62] it was shown how any classical Cayley expander can be converted to a quantum expander, provided the underlying group has an efficient quantum Fourier transform and a large irreducible representation. Applying this recipe to the Cayley expanders over S_n of [79]

results in another construction of explicit, constant degree quantum expanders. One advantage of our explicit construction is that it achieves a much better relation between the spectral gap and the degree compared to that of the other explicit constructions [98, 62].

The Zig-Zag construction we describe in this chapter gives a natural, iterative quantum expander with parameters that are as good as our first construction. However, the Zig-Zag construction is explicit whereas the first construction is not yet explicit (because we do not have an efficient implementation for the Fourier transform of $\mathrm{PGL}(2, q)$). We nevertheless decided to include the first construction. First, we believe it describes a natural approach, and this can be seen from the various other quantum expander constructions that are based on Cayley graphs. Also, the first construction is appealing in that it has only two stages, and each stage naturally corresponds to a well-known Cayley graph. Finally, and more importantly, in the classical setting there are algebraic constructions of Ramanujan expanders (as opposed to combinatorial constructions). Therefore, we believe our first construction has the potential of being improved to a construction of a quantum Ramanujan expander.

3.1.2 Applications of quantum expanders

Classical expanders have become well-known and fundamental objects in mathematics and computer science. This is due to the many applications these objects have found and to the intimate relations they have with other central notions in computational complexity.

While quantum expanders are a natural generalization of classical expanders, they have only recently been defined and it is yet to be seen whether they will be as useful as their classical counterparts. Thus far, the following list of applications has been identified.

- Quantum one-time pads. Ambainis and Smith [9] implicitly used quantum expanders to construct short quantum one-time pads. Loosely speaking, they showed how two parties sharing a random bit string of length $n + O(\log n)$ can communicate an n qubit state such that any eavesdropper cannot learn much about the transmitted state. A subsequent work [42] showed how to remove the $O(\log n)$ term.

- Hastings [65] gave an application from physics. Using quantum expanders, he showed that there exist gapped one-dimensional systems for which the entropy between a given subvolume and the rest of the system is exponential in the correlation length.

- Recently, Hastings and Harrow [64] used specialized quantum expanders (called tensor product expanders) to approximate t-designs as well as to attack a certain open question regarding the Solovay-Kitaev gate approximation.

- In this work we use the quantum expanders constructed by Ambainis and Smith [9] in order to show the problem Quantum Entropy Difference problem (QED) is QSZK-complete.

Let us now elaborate on the last application.

Watrous [142] defined the complexity class of quantum statistical zero knowledge languages (QSZK). QSZK is the class of all languages that have a quantum interactive proof system, along with an efficient simulator. The simulator produces transcripts that, for inputs in the language, are statistically close to the correct ones (for the precise details see [142, 143]). Watrous defined the Quantum State Distinguishability promise problem ($\text{QSD}_{\alpha,\beta}$):

Input: Quantum circuits Q_0, Q_1.
Accept: If $\| \tau_{Q_0} - \tau_{Q_1} \|_{\text{tr}} \geq \beta$.
Reject: If $\| \tau_{Q_0} - \tau_{Q_1} \|_{\text{tr}} \leq \alpha$.

Here, the notation τ_Q denotes the mixed state obtained by running the quantum circuit Q on the initial state $|0^n\rangle$ and tracing out the non-output qubits,[1] and $\| A \|_{\text{tr}} = \text{Tr}\,|A|$ is the quantum analogue of the classical ℓ_1-norm (and so in particular $\| \rho_1 - \rho_2 \|_{\text{tr}}$ is the quantum analogue of the classical variational distance of two probability distributions).

In [142], Watrous showed $\text{QSD}_{\alpha,\beta}$ is complete for honest-verifier-QSZK (QSZK_{HV}) when $0 \leq \alpha < \beta^2 \leq 1$. He further showed that QSZK_{HV} is closed under complement, that any problem in QSZK_{HV} has a 2-message proof system and a 3-message public-coin proof system, and also that QSZK $\subseteq$ PSPACE. Subsequently, in [143], he showed that $\text{QSZK}_{\text{HV}} = \text{QSZK}$.

The above results have classical analogues. However, in the classical setting there is another canonical complete promise problem, the Entropy Difference problem (ED). There is a natural quantum analogue to ED, the Quantum Entropy Difference problem (QED), that we now define:

Input: Quantum circuits Q_0, Q_1.
Accept: If $S(\tau_{Q_0}) - S(\tau_{Q_1}) \geq \frac{1}{2}$.
Reject: If $S(\tau_{Q_1}) - S(\tau_{Q_0}) \geq \frac{1}{2}$.

Here, $S(\rho)$ is the von Neumann entropy of the mixed state ρ (see Section 3.2). The problem QED is very natural from a physical point of view. It corresponds to the following task: we are given two mixed states, each given by a quantum circuit generating it, and we are asked to decide which mixed state has more entropy. This problem is, in particular, as hard[2] as approximating the amount of entropy in a given mixed state (when again the mixed state is given by a circuit generating it).

We prove that QED is QSZK–complete. The proof follows the classical intuition, which uses classical expanders to convert high entropy states to the completely mixed state, while keeping

[1] Here we assume that a quantum circuit also designates a set of output qubits.

[2] Under Turing reductions.

low-entropy states entropy-deficient. Indeed, our proof is an adaptation of the classical proof to quantum entropies, but it crucially depends on the use of quantum expanders replacing the classical expanders used in the classical proof.

The proof requires an explicit quantum expander with a near-optimal *entropy loss* (see Section 3.5.1). As it turns out, the only expander that we currently know of that satisfies this property is the Ambainis-Smith expander. (Indeed it is of non-constant degree but this turns out to be irrelevant in this case.) Using it we obtain that QED is QSZK–complete.

This result implies that it is not likely that one can estimate quantum entropies in BQP. Furthermore, a common way of measuring the amount of entanglement between registers A and B in a pure state ψ is by the von Neumann entropy of $\mathrm{Tr}_B(|\psi\rangle\langle\psi|)$ [115]. Now suppose we are given two circuits Q_1 and Q_2, both acting on the same initial pure-state $|0^n\rangle$, and we want to know which circuit produces more entanglement between A and B. Our result shows that this problem is QSZK–complete. As before, this also shows that the problem of *estimating* the amount of entanglement between two registers in a given pure-state is QSZK–hard under Turing reductions and hence unlikely to be in BQP.

The remainder of this chapter is organized as follows. After the preliminaries (Section 3.2), we give our first construction and its analysis in Section 3.3. In Section 3.4 we describe the Zig-Zag construction. Finally, Section 3.5 is devoted to proving the completeness of QED in QSZK.

3.2 Preliminaries

For any finite-dimensional Hilbert space $\mathcal{V}$, we write $L(\mathcal{V})$ to denote the set of linear operators over $\mathcal{V}$. The set $L(\mathcal{V})$ is also a Hilbert space, equipped with the inner-product $\langle A, B\rangle = \mathrm{Tr}(A^\dagger B)$ and the norm $\| A \|_2 = \sqrt{\langle A, A\rangle}$.

Let $P = (p_1, \ldots, p_m)$ be a vector with real values $p_i \geq 0$.

- The *Shannon entropy* is $H(P) = \sum_{i=1}^m p_i \log \frac{1}{p_i}$.
- The *min-entropy* is $H_\infty(P) = \min_i \log \frac{1}{p_i}$.
- The *Rényi entropy* is $H_2(P) = \log \frac{1}{\mathrm{Col}(P)}$, where $\mathrm{Col}(P) = \sum p_i^2$ is the collision probability of the distribution defined by $\mathrm{Col}(P) = \Pr_{x,y}[x = y]$ when x, y are sampled independently from P.

(We write $\log(\cdot)$ to denote the base 2 logarithm, and $\ln(\cdot)$ to denote the natural logarithm.)

We have analogous definitions for density matrices. For a density matrix ρ, let $\alpha = (\alpha_1, \ldots, \alpha_N)$ be its set of eigenvalues. Since ρ is a density matrix, all these eigenvalues are non-negative and their sum is 1. Thus we can view α as a classical probability distribution.

- The *von Neumann entropy* of ρ is $S(\rho) = H(\alpha)$.
- The *min-entropy* of ρ is $H_\infty(\rho) = H_\infty(\alpha)$.
- The *Rényi entropy* of ρ is $H_2(\rho) = H_2(\alpha)$. The analogue of the collision probability is simply $\mathrm{Tr}(\rho^2) = \sum_i \alpha_i^2 = \|\rho\|_2^2$.

We remark that for any ρ, $H_\infty(\rho) \leq H_2(\rho) \leq S(\rho)$.

The *statistical difference* between two classical distributions $P = (p_1, \ldots, p_m)$ and $Q = (q_1, \ldots, q_m)$ is

$$\mathrm{SD}(P, Q) = \frac{1}{2} \sum_{i=1}^{m} |p_i - q_i| ,$$

i. e., half the ℓ_1 norm of $P - Q$. This is generalized to the quantum setting by defining the trace-norm of a matrix $X \in L(\mathcal{V})$ to be $\| X \|_{\mathrm{tr}} = \mathrm{Tr}(|X|)$, where $|X| = \sqrt{X^\dagger X}$, and by defining the *trace distance* between density matrices ρ and σ to be $\frac{1}{2} \| \rho - \sigma \|_{\mathrm{tr}}$.

3.3 Quantum expanders from non-Abelian Cayley graphs

The construction we present in this section constructs a quantum expander by first taking a step on a non-Abelian Cayley expander followed by a Fourier transform and another step on the non-Abelian Cayley expander. It is similar in spirit to the construction of good quantum error correcting codes given by first encoding the input word with a good classical code, then applying a Fourier transform and then encoding it again with a classical code. Technically the analysis here is more complicated because we use a Fourier transform over a non-Abelian group.

We begin this section with some necessary representation theory background. We then describe the construction and we conclude with its analysis.

3.3.1 Representation theory background

We survey some basic elements of representation theory. For complete accounts, consult the books of Serre [126] or Fulton and Harris [61].

A *representation* ρ of a finite group G is a homomorphism $\rho : G \to \mathrm{GL}(\mathcal{V})$, where $\mathcal{V}$ is a (finite-dimensional) vector space over $\mathbb{C}$ and $\mathrm{GL}(\mathcal{V})$ denotes the group of invertible linear operators on $\mathcal{V}$. Fixing a basis for $\mathcal{V}$, each $\rho(g)$ may be realized as a $d \times d$ matrix over $\mathbb{C}$, where d is the dimension of $\mathcal{V}$. As ρ is a homomorphism, for any $g, h \in G$, $\rho(gh) = \rho(g)\rho(h)$ (the second product being matrix multiplication). The *dimension* d_ρ of the representation ρ is d, the dimension of $\mathcal{V}$.

We say that two representations $\rho_1 : G \to \mathrm{GL}(\mathcal{V})$ and $\rho_2 : G \to \mathrm{GL}(\mathcal{W})$ of a group G are *isomorphic* when there is a linear isomorphism of the two vector spaces $\phi : \mathcal{V} \to \mathcal{W}$ so that for all $g \in G$, $\phi\rho_1(g) = \rho_2(g)\phi$. In this case, we write $\rho_1 \cong \rho_2$.

We say that a subspace $\mathcal{W} \subseteq \mathcal{V}$ is an *invariant subspace* of a representation $\rho : G \to \mathrm{GL}(\mathcal{V})$ if $\rho(g)\mathcal{W} \subseteq \mathcal{W}$ for all $g \in G$. The zero subspace and the subspace $\mathcal{V}$ are always invariant. If no nonzero proper subspaces are invariant, the representation is said to be *irreducible*. Up to isomorphism, a finite group has a finite number of irreducible representations; we let $\widehat{G}$ denote this collection of representations.

If $\rho : G \to \mathrm{GL}(\mathcal{V})$ is a representation, $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$, and each $\mathcal{V}_i$ is an invariant subspace of ρ, then $\rho(g)$ defines two linear representations $\rho_i : G \to \mathrm{GL}(\mathcal{V}_i)$ such that $\rho(g) = \rho_1(g) + \rho_2(g)$. We then write $\rho = \rho_1 \oplus \rho_2$. Any representation ρ can be written as $\rho = \rho_1 \oplus \rho_2 \oplus \cdots \oplus \rho_k$, where each ρ_i is irreducible. In particular, there is a basis in which every matrix $\rho(g)$ is block diagonal, the ith block corresponding to the ith representation in the decomposition. While this decomposition is not, in general, unique, the *number* of times a given irreducible representation appears in this decomposition (up to isomorphism) depends only on the original representation ρ.

The *group algebra* $\mathbb{C}[G]$ of a group G is a vector space of dimension $|G|$ over $\mathbb{C}$, with an orthonormal basis $\{|g\rangle \mid g \in G\}$ and multiplication

$$\sum a_g |g\rangle \cdot \sum b_{g'} |g'\rangle = \sum_{g,g'} a_g b_{g'} |g \cdot g'\rangle \ .$$

This algebra is in bijection with the set $\{f : G \to \mathbb{C}\}$ with the bijection being $f \to \sum_g f(g)|g\rangle$. The inner product in $\mathbb{C}[G]$ translates to the familiar inner product $\langle f, h\rangle = \sum_g \overline{f(g)}h(g)$. The *regular representation* $\rho_{\mathrm{reg}} : G \to \mathrm{GL}(\mathbb{C}[G])$ is defined by $\rho_{\mathrm{reg}}(s) : |g\rangle \mapsto |sg\rangle$, for any $g \in G$. Notice that $\rho_{\mathrm{reg}}(s)$ is a permutation matrix for any $s \in G$.

An interesting fact about the regular representation is that it contains every irreducible representation of G. In particular, if $\rho_1, \ldots, \rho_k$ are the irreducible representations of G with dimensions $d_{\rho_1}, \ldots, d_{\rho_k}$, then

$$\rho_{\mathrm{reg}} = d_{\rho_1}\rho_1 \oplus \cdots \oplus d_{\rho_k}\rho_k \ ,$$

that is, the regular representation contains each irreducible representation ρ exactly d_ρ times.

The *Fourier transform* over G is the unitary transformation F defined by:

$$F|g\rangle \ = \ \sum_{\rho \in \widehat{G}} \sum_{1 \le i,j \le d_\rho} \sqrt{\frac{d_\rho}{|G|}} \rho_{i,j}(g) |\rho, i, j\rangle \ ,$$

where $\rho_{i,j}(g)$ is the (i,j)-th entry of $\rho(g)$ in some predefined basis. In general one has freedom in

choosing a basis for each invariant subspace. In this chapter we choose an *arbitrary* basis, and later fix this choice by using special properties of the group G.

Fact 3.4. The Fourier transform block-diagonalizes the regular representation, i. e.,

$$F\rho_{\mathrm{reg}}(g)F^{\dagger} = \sum_{\rho\in\widehat{G}} \sum_{1\le i,i',j\le d_\rho} \rho_{i,i'}(g) \left|\rho,i,j\right\rangle \left\langle\rho,i',j\right| .$$

This means that when we represent $\rho_{\mathrm{reg}}(g)$ in the basis given by F, we get a block diagonal matrix, with an invariant subspace of dimension d_ρ for each $\rho \in \widehat{G}$, and with $\rho(g)$ as the values of that block.

3.3.2 The construction

Fix an arbitrary (Abelian or non-Abelian) group G of order N, and a subset Γ of group elements closed under inversion. The *Cayley graph* $C(G,\Gamma)$ associated with Γ is a graph over N vertices, each corresponding to an element of G. This graph contains an edge (g_1, g_2) if and only if $g_1 = g_2\gamma$ for some $\gamma \in \Gamma$. The graph $C(G,\Gamma)$ is a regular undirected graph of degree $|\Gamma|$.

We associate with the graph $C(G,\Gamma)$ the linear operator M over $\mathbb{C}[G]$ whose matrix representation agrees with the normalized adjacency matrix of $C(G,\Gamma)$, i. e.,

$$M = \frac{1}{|\Gamma|} \sum_{\gamma\in\Gamma, x\in G} |x\gamma\rangle\langle x| \ .^{3}$$

(The normalization is such that the operator norm is 1.)

Notice that M is a real and symmetric operator, and therefore diagonalizes with real eigenvalues. We denote by $\lambda_1 \ge \cdots \ge \lambda_N$ the eigenvalues of M with orthonormal eigenvectors $v_1, \ldots, v_N$. As $C(G,\Gamma)$ is regular, we have $\lambda_1 = 1$ and $\overline{\lambda} = \max_{i>1} |\lambda_i| \le 1$.

We define the superoperator $T : L(\mathbb{C}[G]) \to L(\mathbb{C}[G])$ that corresponds to randomly taking one step on the Cayley graph $C(G,\Gamma)$. More precisely, this superoperator describes the process whereby a register R of dimension $|\Gamma|$ is initialized to $\left|\overline{0}\right\rangle$ and the following steps are taken. First,

[3] In our definition the generators act from the right. Sometimes the Cayley graph is defined with left action, i. e., g_1 is connected to g_2 if and only if $g_1 = \gamma g_2$. However, note that if we define the invertible linear transformation P that maps the basis vector $|g\rangle$ to the basis vector $\left|g^{-1}\right\rangle$, then $PMP^{-1} = PMP$ maps x to

$$\frac{1}{|\Gamma|}\sum_{\gamma} \left|(x^{-1}\gamma)^{-1}\right\rangle = \frac{1}{|\Gamma|}\sum_{\gamma} \left|\gamma^{-1}x\right\rangle = \frac{1}{|\Gamma|}\sum_{\gamma} |\gamma x\rangle$$

and so the right action is M and the left action is PMP^{-1}, and therefore they are similar and in particular have the same spectrum.

a transformation H is performed on R that maps $|0\rangle$ to

$$\frac{1}{\sqrt{|\Gamma|}}\sum_{\gamma\in\Gamma}|\gamma\rangle\ ,$$

yielding, for an input state ρ, the state

$$\frac{1}{|\Gamma|}\rho\otimes\sum_{\gamma,\gamma'\in\Gamma}|\gamma\rangle\langle\gamma'|\,.$$

Then, the unitary transformation $Z : |g,\gamma\rangle \to |g\gamma,\gamma\rangle$ is applied, and finally the register R is discarded. In more algebraic terms,

$$T(\rho)=\mathrm{Tr}_R\Big[Z(I\otimes H)(\rho\otimes|\overline{0}\rangle\langle\overline{0}|)(I\otimes H)Z^\dagger\Big]\,.$$

We note that the transformation Z is a permutation over the standard basis, and is classically easy to compute in both directions, and therefore has an efficient quantum circuit.

We also need the notion of a *good basis change*. We say a unitary transformation U is a good basis change if for any $g_1\neq e$ (where e denotes the identity element of G) and any g_2 it holds that

$$\mathrm{Tr}\big(U\rho_{\mathrm{reg}}(g_1)U^\dagger\rho_{\mathrm{reg}}(g_2)\big)=0\,. \tag{3.1}$$

The quantum expander is then defined as

$$E(\rho)=T(UT(\rho)U^\dagger)\,.$$

Lemma 3.5. *If U is a good basis change then E is a $(|\Gamma|^2,\overline{\lambda})$ quantum expander for $\overline{\lambda}$ as defined as above.*

The fact that E is $|\Gamma|^2$-regular is immediate and the rest of this section is devoted to proving the claimed spectral gap.

Lubotzky et al. [96] described a constant degree Ramanujan Cayley graph over $\mathrm{PGL}(2,q)$, with degree $|\Gamma|$ and second-largest eigenvalue $\overline{\lambda}$ satisfying $\overline{\lambda}^2\leq 4/|\Gamma|$. In Section 3.3.5 we show how to modify the Fourier transform for $\mathrm{PGL}(2,q)$ to obtain a good basis change, and by plugging this basis change into Lemma 3.5 we obtain a $(16/\overline{\lambda}^4,\overline{\lambda})$ quantum expander. The construction is not explicit as it is yet unknown how to efficiently implement the quantum Fourier transform for $\mathrm{PGL}(2,q)$.

3.3.3 The analysis

First, we fully identify the spectrum of T. We view any eigenvector $v_i \in \mathbb{C}^N$ (of M) as an element of $\mathbb{C}[G]$, $|v_i\rangle = \sum_g v_i(g)\,|g\rangle$. We also define a linear transformation $\mathsf{Diag} : \mathbb{C}[G] \to L(\mathbb{C}[G])$ by $\mathsf{Diag}\,|g\rangle = |g\rangle\langle g|$. Denote

$$\mu_{i,g} \;=\; \rho_{\mathrm{reg}}(g)(\mathsf{Diag}\,|v_i\rangle) \;=\; \sum_{x\in G} v_i(x)\,|gx\rangle\langle x|\ .$$

Then it is easy to see that these matrices form a set of eigenvectors of T.

Lemma 3.6. *The vectors $\{\mu_{i,g} \mid i = 1,\ldots,N, g \in G\}$ form an orthonormal basis of $L(\mathbb{C}[G])$, and $\mu_{i,g}$ is an eigenvector of T with eigenvalue λ_i.*

Proof: Notice that $T(|g_1\rangle\langle g_2|) = \mathrm{Tr}_R\big[\frac{1}{|\Gamma|}\sum_{\gamma_1,\gamma_2} Z\,|g_1,\gamma_1\rangle\langle g_2,\gamma_2|\,Z^\dagger\big] = \frac{1}{|\Gamma|}\sum_\gamma |g_1\gamma\rangle\langle g_2\gamma|$. Now,

$$\begin{aligned} T(\mu_{i,g}) \;&=\; \frac{1}{|\Gamma|}\sum_{x,\gamma} v_i(x)\,|gx\gamma\rangle\langle x\gamma| = \rho_{\mathrm{reg}}(g)\frac{1}{|\Gamma|}\sum_{x,\gamma} v_i(x)\,|x\gamma\rangle\langle x\gamma| \\ &=\; \rho_{\mathrm{reg}}(g)\mathsf{Diag}\Big(\sum_x v_i(x)M\,|x\rangle\Big) = \rho_{\mathrm{reg}}(g)\mathsf{Diag}(M\,|v_i\rangle) = \lambda_i\rho_{\mathrm{reg}}(g)\mathsf{Diag}(|v_i\rangle) = \lambda_i\mu_{i,g}\,. \end{aligned}$$

To verify orthonormality, notice that $\mathrm{Tr}(\mu_{i,g_1}\mu^\dagger_{i',g_2}) = 0$ for every choice of $g_1 \neq g_2$, as each entry (k,ℓ) must be zero for at least one of the matrices. If $g_1 = g_2 = g$ then

$$\mathrm{Tr}(\mu_{i,g}\mu^\dagger_{i',g}) = \langle v_{i'}|v_i\rangle = \delta_{i,i'}\,.$$

As the number of vectors $\{\mu_{i,g}\}$ is N^2, they form an orthonormal basis for $L(\mathbb{C}[G])$. ∎

We decompose the space $L(\mathbb{C}[G])$ into three perpendicular spaces:

$$\begin{array}{lcll} \mathrm{Span}\,\{\mu_{1,e}\}\,, & & & \\ W & = & \mathrm{Span}\,\{\mu_{1,g} \mid g \in G, g \neq e\}\,, & \text{and} \\ \mu^\perp & = & \mathrm{Span}\,\{\mu_{i,g} \mid i \neq 1, g \in G\}\,. & \end{array}$$

We also denote $\mu^{||} = \mathrm{Span}\,\{\mu_{1,e}\} + W = \mathrm{Span}\,\{\mu_{1,g} \mid g \in G\}$. Notice that $T(\mu^{||}) = \mu^{||}$ and $T(\mu^\perp) = \mu^\perp$.

Claim 3.7. *If $\rho \in W$ and U is a good basis change then $U\rho U^\dagger \in \mu^\perp$.*

Proof: The set $\{\rho_{\mathrm{reg}}(g) \mid g \in G\}$ is an orthonormal basis for $\mu^{||}$ and hence $\{\rho_{\mathrm{reg}}(g) \mid g \in G,\ g \neq e\}$ is an orthonormal basis for W. Therefore, it is enough to verify that $\mathrm{Tr}(U\rho_{\mathrm{reg}}(g_1)U^\dagger\rho_{\mathrm{reg}}(g_2)^\dagger) = 0$ for any $g_1 \neq e$ and for any g_2. Given that $\rho_{\mathrm{reg}}(g_2)^\dagger = \rho_{\mathrm{reg}}(g_2^{-1})$, this follows directly from (3.1). ∎

Thus, intuitively speaking, we have a win-win situation when E is applied to ρ. If ρ is in $\mu^{\perp}$, then the first application of T shrinks its norm, while if ρ is in W, then the first application of T keeps it unchanged, the basis change maps it to $\mu^{\perp}$, and the last T application shrinks its norm. Indeed, we are now ready to prove Lemma 3.5, namely, that if U is a good basis change then E is a $(|\Gamma|^2, \overline{\lambda})$ quantum expander.

Proof of Lemma 3.5: The regularity of E is clear from its definition. Fix any $X \in L(\mathbb{C}[G])$ that is perpendicular to $\tilde{I} = \mu_{1,e}$, and write $X = X^{||} + X^{\perp}$ for $X^{||} \in W$ and $X^{\perp} \in \mu^{\perp}$. We have

$$E(X) = T(\sigma^{||} + \sigma^{\perp}) \,,$$

where $\sigma^{||} = UT(X^{||})U^{\dagger}$ and $\sigma^{\perp} = UT(X^{\perp})U^{\dagger}$. Observe the following. First, $T(X^{||}) \in W$, so by Claim 3.7, $\sigma^{||} \perp \mu^{||}$. Also, $T(X^{||}) \perp T(X^{\perp})$ (as T preserves both $\mu^{||}$ and $\mu^{\perp}$), and therefore $\sigma^{||} \perp \sigma^{\perp}$. Moreover, by Lemma 3.6 we know T is normal.

By Lemma 3.8 (stated and proved below) we see that

$$\begin{aligned} \|E(X)\|_2^2 &= \|T(\sigma^{||} + \sigma^{\perp})\|_2^2 \leq \overline{\lambda}^2 \|\sigma^{||}\|_2^2 + \|\sigma^{\perp}\|_2^2 \\ &= \overline{\lambda}^2 \|T(X^{||})\|_2^2 + \|T(X^{\perp})\|_2^2 \leq \overline{\lambda}^2 \|X^{||}\|_2^2 + \overline{\lambda}^2 \|X^{\perp}\|_2^2 = \overline{\lambda}^2 \|X\|_2^2 \end{aligned}$$

as required. ∎

We are left to prove the following lemma.

Lemma 3.8. *Let T be a normal linear operator with eigenspaces $\mathcal{V}_1, \ldots, \mathcal{V}_n$ and corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ in descending* absolute *value. Suppose u and w are vectors such that*

$$u \in \mathrm{Span}\{\mathcal{V}_2, \ldots, \mathcal{V}_n\}$$

and $w \perp u$ (and where w does not necessarily belong to $\mathcal{V}_1$). Then

$$\|T(u+w)\|_2^2 \leq |\lambda_2|^2 \|u\|_2^2 + |\lambda_1|^2 \|w\|_2^2 \,.$$

Proof: Let $\{v_j\}$ be an eigenvector basis for T with eigenvalues δ_j (from the set $\{\lambda_1, \ldots, \lambda_n\}$). Writing $u = \sum_j \alpha_j v_j$ and $w = \beta v + \sum_j \beta_j v_j$ with $v_j \in \mathrm{Span}\{\mathcal{V}_2, \ldots, \mathcal{V}_n\}$ and $v \in \mathcal{V}_1$, we get:

$$\begin{aligned} \|T(u+w)\|_2^2 &= \left\| \lambda_1 \beta v + \sum_j \delta_j (\alpha_j + \beta_j) v_j \right\|_2^2 \leq |\lambda_1|^2 |\beta|^2 + |\lambda_2|^2 \sum_j |\alpha_j + \beta_j|^2 \\ &= |\lambda_1|^2 |\beta|^2 + |\lambda_2|^2 \Big(\sum_j |\alpha_j|^2 + \sum_j |\beta_j|^2 + \langle u|w\rangle + \langle w|u\rangle \Big) \leq |\lambda_2|^2 \|u\|_2^2 + |\lambda_1|^2 \|w\|_2^2 \,. \end{aligned}$$

■

3.3.4 A sufficient condition that guarantees a good basis change

So far we have reduced the problem of constructing a quantum expander to that of finding a Cayley graph $C(G,\Gamma)$ and a good basis change for G. We now concentrate on the problem of finding a good basis change for a given group G, and show that if G respects some general condition then one can efficiently construct a good basis change from G from its Fourier transform.

A basic fact of representation theory states that $\sum_{\rho\in\widehat{G}} d_\rho^2 = |G|$. Equivalently, for any group G there is a bijection between

$$\left\{(\rho,i,j)\,\middle|\,\rho\in\widehat{G}, 1\le i,j\le d_\rho\right\}$$

and G. Finding such a natural bijection is a fundamental problem both in mathematics (where it is equivalent to describing the invariant subspaces of the regular representation of G) and in computer science (where it is a main step towards implementing a fast Fourier transform). Indeed, this question was extensively studied. For example, the "Robinson-Schensted" algorithm [122, 125] is a mapping from pairs (P,T) of standard shapes (a shape corresponds to an irreducible representation of S_n, and its dimension is the number of valid fillings of that shape) to S_n.

Here we require more from such a mapping.

Definition 3.9. Let f be a bijection from $\left\{(\rho,i,j) \mid \rho\in\widehat{G}, 1\le i,j\le d_\rho\right\}$ to G. We say that f is a *product mapping* if, for every $\rho\in\widehat{G}$,

$$f(\rho,i,j) \;=\; f_1(\rho,i)\cdot f_2(\rho,j) \tag{3.2}$$

for some choice of functions $f_1(\rho,\cdot), f_2(\rho,\cdot) : [d_\rho]\to G$.

The Robinson-Schensted mapping is *not* a product mapping. However, S_n has a product mapping for $n \le 6$, and we think it is a natural question whether product mappings for S_n exist for all n. For some groups it is easy to find a product mapping. For example, in any Abelian group all irreducible representations are of dimension one and so we can define $f_1(\rho,i) = e$ and $f_2(\rho,j) = f(\rho,1,1)$.

Another easy example is the dihedral group D_m of rotations and reflections of a regular polygon with m sides. Its generators are r, the rotation element, and s, the reflection element. This group has $2m$ elements and the defining relations are $s^2 = 1$ and $srs = r^{-1}$. We shall argue this group has a product mapping for odd m (although it is true for even m as well). The dihedral group has $(m-1)/2$ representations $\{\rho_\ell\}$ of dimension two and two representations $\{\tau_1,\tau_2\}$ of dimension one (see [126, Section 5.3]). A product mapping in this case can be given by defining $f(\rho,i,j)$ as

follows:

$$f(\rho,i,j) \quad = \quad \begin{cases} 1 & \text{if } \rho=\tau_1, i=j=1, \\ s & \text{if } \rho=\tau_2, i=j=1, \\ r^{2(\ell-1)+i}s^j & \text{if } \rho=\rho_\ell. \end{cases} \tag{3.3}$$

We now show that if G has a product mapping then G has a good basis change:

Lemma 3.10. *Let G be a group that has a product mapping f, and let F be the Fourier transform over G, that is*

$$F\left|g\right\rangle = \sum_{\rho\in\widehat{G}}\sum_{1\leq i,j\leq d_\rho}\sqrt{\frac{d_\rho}{|G|}}\rho_{i,j}(g)\left|\rho,i,j\right\rangle \ .$$

Define the unitary mapping

$$S:\left|\rho,i,j\right\rangle \mapsto \omega_{d_\rho}^{ij}\left|f(\rho,i,j)\right\rangle \ ,$$

where ω_{d_ρ} is a primitive root of unity of order d_ρ, and set U to be the unitary transformation $U=SF$. Then U is a good basis change.

Proof: Fix $g_1\neq e$ and g_2. If $g_2=e$ then

$$\mathrm{Tr}\left(U\rho_{\mathrm{reg}}(g_1)U^\dagger\rho_{\mathrm{reg}}(g_2)\right)=\mathrm{Tr}\left(U\rho_{\mathrm{reg}}(g_1)U^\dagger\right)=\mathrm{Tr}\left(\rho_{\mathrm{reg}}(g_1)\right)=0\,,$$

where the last equality follows from the assumption that $g_1\neq e$.

We are left with the case $g_2\neq e$. By Fact 3.4, it holds that

$$\mathrm{Tr}\left(SF\rho_{\mathrm{reg}}(g_1)F^\dagger S^\dagger\rho_{\mathrm{reg}}(g_2)\right) \quad = \quad \sum_{\rho\in\widehat{G}}\sum_{1\leq i,i'\leq d_\rho}\rho_{i,i'}(g_1)\,\mathrm{Tr}\left(\sum_{j=1}^{d_\rho}S\left|\rho,i,j\right\rangle\left\langle\rho,i',j\right|S^\dagger\sum_x\left|g_2x\right\rangle\left\langle x\right|\right) \ .$$

Therefore, it suffices to show that for any ρ,i,i' we have

$$\mathrm{Tr}\left(\sum_{j=1}^{d_\rho}S\left|\rho,i,j\right\rangle\left\langle\rho,i',j\right|S^\dagger\sum_x\left|g_2x\right\rangle\left\langle x\right|\right)=0\,.$$

Fix $\rho\in\widehat{G}$ and $i,i'\in\{1,\ldots,d_\rho\}$. Because f is a product mapping, $f(\rho,i,j)=f_1(\rho,i)\cdot f_2(\rho,j)$ for some choice of functions f_1,f_2. Denote $h_i=f_1(\rho,i)$ and $t_j=f_2(\rho,j)$. The sum we need to calculate can be written as:

$$\sum_{j=1}^{d_\rho}\sum_x\omega_{d_\rho}^{ij-i'j}\,\mathrm{Tr}\left(\left|h_it_j\right\rangle\left\langle h_{i'}t_j\,\middle|\,g_2x\right\rangle\left\langle x\right|\right) \quad = \quad \sum_{j=1}^{d_\rho}\omega_{d_\rho}^{ij-i'j}\sum_x\left\langle x\,\middle|\,h_it_j\right\rangle\left\langle h_{i'}t_j\,\middle|\,g_2x\right\rangle$$

$$= \sum_{j=1}^{d_\rho} \omega_{d_\rho}^{(i-i')j} \left\langle g_2 \,|\, h_{i'} h_i^{-1} \right\rangle ,$$

where the last equality follows from the observation that the sum over x yields a non-zero value if and only if $x = h_i t_j$ and $h_{i'} t_j = g_2 x$. This happens if and only if $h_i t_j = g_2^{-1} h_{i'} t_j$, or equivalently $g_2 = h_{i'} h_i^{-1}$. However, when $g_2 = h_{i'} h_i^{-1}$, we obtain the sum $\sum_{j=1}^{d_\rho} \omega_{d_\rho}^{(i-i')j}$, and because $g_2 \neq e$ it follows that $i \neq i'$. Hence the expression is zero, as required. ∎

3.3.5 PGL$(2, q)$ has a product bijection

The group PGL$(2, q)$ is the group of all 2×2 invertible matrices over $\mathbb{F}_q$ modulo the group center. This group has $(q-3)/2$ irreducible representations of dimension $q+1$, $(q-1)/2$ irreducible representations of dimension $q-1$, 2 irreducible representations of dimension q and 2 irreducible representations of dimension 1 (see [61, Section 5.2] and [2]). We let ρ_x^d denote the xth irreducible representation of dimension d.

We look for a bijection from G to the irreducible representations of G. Our approach is to use a tower of subgroups,

$$G_3 = G > G_2 = D_{2q} > G_1 = Z_q > G_0 = \{e\} ,$$

with G_2 and G_1 defined as follows. The group G_2 is generated by the equivalence classes of

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and of} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} .$$

This group is a dihedral subgroup of G with $2q$ elements, i. e., D_q. The first matrix is the reflection, denoted by s, and the second is the rotation, denoted by r. This group has a cyclic subgroup $G_1 \cong Z_q$ (the group generated by r).

Let $T_2 = \{t_1, \ldots, t_\ell\}$ be a transversal for G_2 with

$$\ell = \frac{|G|}{|G_2|} = \frac{(q-1)(q+1)}{2} .$$

For each $\rho \in \widehat{G}$ we let $f_1(\rho, i) \in \{t_1, \ldots, t_\ell\}$ define a coset of G_2, and let $f_2(\rho, j) \in G_2$ define an element in G_2 as follows. The representations of dimension $q+1$ take the first $(q-3)(q+1)/2$ cosets:

$$f_1(\rho_x^{q+1}, i) = \begin{cases} r^{i-1} & \text{if } i = 1, \ldots, q, \\ s & \text{if } i = q+1, \end{cases}$$

$$f_2(\rho_x^{q+1}, j) = t_{(x-1)(q+1)+j},$$

for all $x = 1, \ldots, \frac{q-1}{2} - 1$ and $i, j = 1, \ldots, q+1$. We match them with representations of dimension $q-1$:

$$\begin{aligned} f_1(\rho_x^{q-1}, i) &= sr^i, \\ f_2(\rho_x^{q-1}, j) &= t_{(x-1)(q-1)+j}\,, \end{aligned}$$

for all $x = 1, \ldots, (q-1)/2$ and $i, j = 1, \ldots, q-1$. Notice that so far we have covered the first

$$\frac{(q-3)(q+1)}{2} = \frac{(q-1)(q-1)}{2} - 2$$

cosets without repetitions. Two cosets are partially covered with dimension $q-1$ representations (in each coset $q-1$ elements are covered). We put the dimension 1 representation into these cosets:

$$\begin{aligned} f_1(\rho_x^1, 1) &= s, \\ f_2(\rho_x^1, 1) &= t_{\frac{(q-3)(q+1)}{2}+x}\,, \end{aligned}$$

for $x = 1, 2$. Finally, we fill all the remaining gaps with dimension q representations. The first two fill the partially full cosets, and the rest fill each coset in pairs. Notice that here we use the fact that $G_1 < G_2$. The function f_2 returns an element in the traversal set of G_1 and f_1 returns an element of G_1:

$$\begin{aligned} f_1(\rho_x^q, i) &= r^i, \\ f_2(\rho_x^q, j) &= \begin{cases} t_{\frac{(q-3)(q+1)}{2}+x} & \text{if } j = q, \\ s^{x-1} t_{\frac{(q-1)(q-1)}{2}+j} & \text{otherwise,} \end{cases} \end{aligned}$$

for $x = 1, 2$. One can verify that this product mapping is a bijection as desired.

3.4 The Zig-Zag construction

We now present our second construction of quantum expanders, following the iterative construction of Reingold et al. [120]. We already discussed their construction in Chapter 2. However, to allow this chapter to be more self contained, and since its description is quite short, we recall it here as well.

The starting point of the construction of [120] is a good expander of constant size, which can

be found by an exhaustive search. Then, they construct a series of expanders with an increasing number of vertices by applying a sequence of three basic transformations: tensoring (that squares the number of vertices at the expense of a worse ratio between the spectral gap and the degree), squaring (that improves the spectral gap) and the Zig-Zag product (that reduces the degree to its original size). These three transformations are repeated iteratively, resulting in a good constant-degree expander over many vertices.

The first two transformations have natural counterparts in the quantum setting. For ease of notation, we denote by $T(\mathcal{V})$ the set of superoperators on $L(\mathcal{V})$ (that is, $T(\mathcal{V}) = L(L(\mathcal{V}))$). We also denote by $U(\mathcal{V})$ the set of unitary operators in $L(\mathcal{V})$.

- **Squaring:** For a superoperator $G \in T(\mathcal{V})$ we denote by G^2 the superoperator given by $G^2(X) = G(G(X))$ for any $X \in L(\mathcal{V})$.

- **Tensoring:** For superoperators $G_1 \in T(\mathcal{V}_1)$ and $G_2 \in T(\mathcal{V}_2)$ we denote by $G_1 \otimes G_2$ the superoperator given by $(G_1 \otimes G_2)(X \otimes Y) = G_1(X) \otimes G_2(Y)$ for any $X \in L(\mathcal{V}_1), Y \in L(\mathcal{V}_2)$.

In order to define the quantum Zig-Zag product we first recall the classical Zig-Zag product. We have two *graphs* G_1 and G_2. The graph G_1 is a D_1-regular graph over N_1 vertices and the graph G_2 is a D_2-regular graph over $N_2 = D_1$ vertices. We first define the *replacement product* graph, which has $V_1 \times V_2$ as its set of vertices. We refer to the set of vertices $\{v\} \times V_2$ as the *cloud* of v. The replacement product has a copy of G_2 on each cloud, and also *inter-cloud* edges between (v, i) and (w, j) if the i-th neighbor of v is w and the j-th neighbor of w is v in G_1. Thus, the replacement product has the same connected components as the original graph but a much lower degree ($D_2 + 1$ instead of D_1). The Zig-Zag product graph $G_1 \textcircled{z} G_2$ has the same set of vertices as the replacement product, but has an edge between $x = (v, a)$ and $x' = (v', a')$ if and only if in the replacement product graph there is a three step walk from x to x' that first takes a cloud edge, then an inter-cloud edge, and then again a cloud edge. Thus, the graph $G_1 \textcircled{z} G_2$ is D_2^2-regular.

We now define the quantum Zig-Zag transformation. Let $G_1 \in T(\mathcal{H}_{N_1})$ be an N_2-regular operator and $G_2 \in T(\mathcal{H}_{N_2})$, where $\mathcal{H}_N$ denotes the Hilbert space of dimension N. As G_1 is D_1-regular, it can be expressed as

$$G_1(X) = \frac{1}{D_1} \sum_d U_d X U_d^\dagger$$

for some unitaries $U_d \in U(\mathcal{H}_{N_1})$. We lift the ensemble $\{U_d\}$ to a superoperator $\dot{U} \in L(\mathcal{H}_{N_1} \otimes \mathcal{H}_{D_1})$ defined by $\dot{U}(|a\rangle \otimes |b\rangle) = U_b |a\rangle \otimes |b\rangle$. We also define $\dot{G}_1 \in T(\mathcal{H}_{N_1} \otimes \mathcal{H}_{D_1})$ by $\dot{G}_1(X) = \dot{U} X \dot{U}^\dagger$. The superoperator $\dot{G}_1$ corresponds to the inter-cloud edges in the replacement product. We are now

ready to define the quantum Zig-Zag product.

Definition 3.11. Let G_1, G_2 be as above. The Zig-Zag product of G_1 and G_2, denoted by $G_1 \textcircled{z} G_2 \in T(\mathcal{H}_{N_1} \otimes \mathcal{H}_{D_1})$, is defined to be $(G_1 \textcircled{z} G_2)X = (I \otimes G_2^\dagger)\dot{G}_1(I \otimes G_2)X$.

Remark 3.12. Notice that formally $G_1 \textcircled{z} G_2$ depends on the Kraus decomposition of G_1 and the notation should have reflected this. However, we fix this decomposition once and use this simpler notation.

Finally we explain how to find a base quantum operator H that is a (D^8, D, λ) quantum expander. Its existence follows from the following result of Hastings [66].

Theorem 3.13. *There exists an integer D_0 such that for every $D > D_0$ there exists a (D^8, D, λ) quantum expander for $\lambda = 4\sqrt{D-1}/D$.*

Remark 3.14. Hastings actually shows the stronger result that, for any D, there exist a

$$\left(D^8, D, \left(1 + O(D^{-16/15} \log D)\right) \frac{2\sqrt{D-1}}{D}\right)$$

quantum expander.

We use an exhaustive search over a net $S \subset U(\mathcal{H}_{D^8})$ of unitary matrices to find such a quantum expander. The set S has the property that for any unitary matrix $U \in U(\mathcal{H}_{D^8})$ there exists some $V_U \in S$ such that

$$\sup_{\| X \| = 1} \left\| UXU^\dagger - V_U X V_U^\dagger \right\| \leq \lambda \,.$$

It is not hard to verify that indeed such S exists, with size depending only on D and λ. Moreover, we can find such a set in time depending only on D and λ.[4] Suppose that

$$G(X) = \frac{1}{D} \sum_{i=1}^{D} U_i X U_i^\dagger \,.$$

is a (D^8, D, λ) quantum expander, and denote by G' the superoperator

$$G'(X) = \frac{1}{D} \sum_{i=1}^{D} V_{U_i} X V_{U_i}^\dagger \,.$$

[4] One way to see this is using the Solovay-Kitaev theorem (see, e. g., [39]). The theorem assures us that, for example, the set of all the quantum circuits of length $O(\log^4 \epsilon^{-1})$ generated only by Hadamard and Tofolli gates gives an ϵ-net of unitaries. The accuracy of the net is measured differently in the Solovay-Kitaev theorem, but it can be verified that the accuracy measure we use here is roughly equivalent.

For $X \in L(\mathcal{H}_{D^8})$ orthogonal to $\tilde{I}$, it holds that

$$\| G'(X) \| = \left\| \frac{1}{D} \sum_{i=1}^{D} V_{U_i} X V_{U_i}^{\dagger} \right\| \leq \| G(X) \| + \lambda \| X \| \leq 2\lambda \| X \| .$$

Hence, G' is a $(D^8, D, 8\sqrt{D-1}/D)$ quantum expander. This implies that a brute force search over the net finds a good base superoperator H in time that depends only on D and λ.

Remark 3.15. We can actually get an eigenvalue bound of $(1+\epsilon)2\sqrt{D-1}/D$ for an arbitrary small ϵ at the expense of increasing D_0, using the better bound in Remark 3.14.

Given all these ingredients we define an iterative process as in [120], composed of a series of superoperators. The first two superoperators are $G_1 = H^2$ and $G_2 = H \otimes H$. For every $t > 2$ we define

$$G_t = \left(G_{\lceil \frac{t-1}{2} \rceil} \otimes G_{\lfloor \frac{t-1}{2} \rfloor} \right)^2 \textcircled{z} H .$$

Theorem 3.16. *For every $t > 0$, G_t is an explicit (D^{8t}, D^2, λ_t) quantum expander with $\lambda_t = \lambda + O(\lambda^2)$, where the constant in the O notation is an absolute constant.*

Thus, G_t is a constant degree, constant gap quantum expander, as desired.

3.4.1 The analysis

Tensoring and squaring are easy to analyze, and the following proposition is immediate from the definitions of these operations.

Proposition 3.17. *If G is a (N, D, λ) quantum expander then G^2 is a (N, D^2, λ^2) quantum expander. If G_1 is a (N_1, D_1, λ_1) quantum expander and G_2 is a (N_2, D_2, λ_2) quantum expander then $G_1 \otimes G_2$ is a $(N_1 \cdot N_2, D_1 \cdot D_2, \max(\lambda_1, \lambda_2))$ quantum expander.*

We are left to analyze is the quantum Zig-Zag product.

Theorem 3.18. *If G_1 is a (N_1, D_1, λ_1) quantum expander and G_2 is a (D_1, D_2, λ_2) quantum expander then $G_1 \textcircled{z} G_2$ is a $(N_1 \cdot D_1, D_2^2, \lambda_1 + \lambda_2 + \lambda_2^2)$ quantum expander.*

With the above two claims, the proof of Theorem 3.16 is identical to the one in [120] and is omitted. In order to prove Theorem 3.18 we claim the following.

Proposition 3.19. *For any $X, Y \in L(\mathcal{H}_{N_1} \otimes \mathcal{H}_{D_1})$ such that X is orthogonal to the identity operator we have*

$$\left| \langle (G_1 \textcircled{z} G_2) X, Y \rangle \right| \leq f(\lambda_1, \lambda_2) \| X \| \cdot \| Y \| ,$$

where $f(\lambda_1, \lambda_2) = \lambda_1 + \lambda_2 + \lambda_2^2$.

Theorem 3.18 follows from this proposition: for a given X orthogonal to $\tilde{I}$ we let $Y = (G_1 \textcircled{z} G_2)X$ and plug X and Y into the proposition. We see that $\| (G_1 \textcircled{z} G_2)X \| \leq f(\lambda_1, \lambda_2) \| X \|$ as required.

The proof of Proposition 3.19 is an adaptation of the proof in [120]. The main difference is that the classical proof works over the Hilbert space $\mathcal{V}$ whereas the quantum proof works over $L(\mathcal{V})$. Remarkably, the same intuition works in both cases.

Proof of Proposition 3.19: We first decompose the space $L(\mathcal{H}_{N_1} \otimes \mathcal{H}_{D_1})$ into

$$\begin{aligned} W^{||} &= \mathrm{Span}\{\sigma \otimes \widetilde{I} \mid \sigma \in L(\mathcal{H}_{N_1})\} \quad \text{and} \\ W^{\perp} &= \mathrm{Span}\{\sigma \otimes \tau \mid \sigma \in L(\mathcal{H}_{N_1}), \tau \in L(\mathcal{H}_{D_1}), \langle \tau, \widetilde{I} \rangle = 0\}\,. \end{aligned}$$

Next, we write X as $X = X^{||} + X^{\perp}$, where $X^{||} \in W^{||}$ and $X^{\perp} \in W^{\perp}$, and similarly $Y = Y^{||} + Y^{\perp}$. By definition,

$$|\langle (G_1 \textcircled{z} G_2)X, Y \rangle| = |\langle \dot{G_1}(I \otimes G_2)(X^{||} + X^{\perp}), (I \otimes G_2)(Y^{||} + Y^{\perp}) \rangle|\,.$$

Using linearity and the triangle inequality (and the fact that $I \otimes G_2$ acts trivially on $W^{||}$), we get

$$\begin{aligned} |\langle (G_1 \textcircled{z} G_2)X, Y \rangle| \;\leq\; & |\langle \dot{G_1} X^{||}, Y^{||} \rangle| + |\langle \dot{G_1} X^{||}, (I \otimes G_2) Y^{\perp} \rangle| + \\ & |\langle \dot{G_1}(I \otimes G_2) X^{\perp}, Y^{||} \rangle| + |\langle \dot{G_1}(I \otimes G_2) X^{\perp}, (I \otimes G_2) Y^{\perp} \rangle|\,. \end{aligned}$$

In the last three terms we have $I \otimes G_2$ acting on an operator from $W^{\perp}$. As expected, when this happens the quantum expander G_2 shrinks the norm of the operator.

Claim 3.20. *For any $Z \in W^{\perp}$ it holds that $\| (I \otimes G_2)Z \| \leq \lambda_2 \| Z \|$.*

Proof: The matrix Z can be written as $Z = \sum_i \sigma_i \otimes \tau_i$, where each τ_i is perpendicular to $\tilde{I}$ and $\{\sigma_i\}$ is an orthogonal set. Hence,

$$\| (I \otimes G_2)Z \|^2 = \Big\| \sum_i \sigma_i \otimes G_2(\tau_i) \Big\|^2 = \sum_i \| \sigma_i \otimes G_2(\tau_i) \|^2 \leq \sum_i \lambda_2^2 \| \sigma_i \otimes \tau_i \|^2 = \lambda_2^2 \| Z \|^2\,.$$

■

To bound the first term, we observe that on inputs from $W^{||}$ the operator $\dot{G_1}$ mimics the operation of G_1 with a random seed.

Claim 3.21. *For any $A, B \in W^{||}$ such that $\langle A, \tilde{I} \rangle = 0$, it holds that $|\langle \dot{G_1}(A), B \rangle| \leq \lambda_1 \| A \| \cdot \| B \|$.*

Proof: Any choice of $A, B \in W^{||}$ may be written as

$$A = \sigma \otimes \tilde{I} = \frac{1}{D_1} \sum_i \sigma \otimes |i\rangle\langle i| \ ,$$
$$B = \eta \otimes \tilde{I} = \frac{1}{D_1} \sum_i \eta \otimes |i\rangle\langle i| \ .$$

Moreover, as A is perpendicular to the identity operator, it follows that σ is perpendicular to the identity operator on the space $L(\mathcal{H}_{N_1})$. This means that applying G_1 on σ will shrinks its norm by at a factor of least λ_1.

Considering the inner product

$$\begin{aligned}
|\langle \dot{G_1} A, B\rangle| &= \frac{1}{D_1^2} \left| \sum_{i,j} \mathrm{Tr}\left(\left((U_i \sigma U_i^\dagger) \otimes |i\rangle\langle i| \right) (\eta \otimes |j\rangle\langle j|)^\dagger \right) \right| \\
&= \frac{1}{D_1^2} \left| \sum_{i,j} \mathrm{Tr}\left((U_i \sigma U_i^\dagger \eta^\dagger) \otimes |i\rangle \langle i\,|\, j\rangle \langle j| \right) \right| \\
&= \frac{1}{D_1^2} \left| \sum_{i} \mathrm{Tr}\left((U_i \sigma U_i^\dagger \eta^\dagger) \otimes |i\rangle\langle i| \right) \right| \\
&= \frac{1}{D_1^2} \left| \sum_{i} \mathrm{Tr}\left(U_i \sigma U_i^\dagger \eta^\dagger \right) \right| \\
&= \frac{1}{D_1} \left| \mathrm{Tr}\left(\left(\frac{1}{D_1} \sum_i U_i \sigma U_i^\dagger \right) \eta^\dagger \right) \right| \\
&= \frac{1}{D_1} |\langle G_1(\sigma), \eta\rangle| \leq \frac{\lambda_1}{D_1} \| \sigma \| \cdot \| \eta \| = \lambda_1 \| A \| \cdot \| B \| \ ,
\end{aligned}$$

where the inequality follows from the expansion property of G_1 (and Cauchy-Schwartz). ■

With the above claims in hand we see that

$$|\langle (G_1 \textcircled{z} G_2) X, Y\rangle| \ \leq \ (p_X p_Y \lambda_1 + p_X q_Y \lambda_2 + p_Y q_X \lambda_2 + q_X q_Y \lambda_2^2) \| X \| \cdot \| Y \| \ , \quad (3.4)$$

where

$$p_X = \frac{\| X^{||} \|}{\| X \|} \qquad \text{and} \qquad q_X = \frac{\| X^{\perp} \|}{\| X \|} \, ,$$

and similarly

$$p_Y = \frac{\| Y^{||} \|}{\| Y \|} \qquad \text{and} \qquad q_Y = \frac{\| X^{\perp} \|}{\| Y \|} \, .$$

Notice that $p_X^2 + q_X^2 = p_Y^2 + q_Y^2 = 1$. It is easy to see that $p_X p_Y, q_X q_Y \leq 1$. Also, by Cauchy-Schwartz, $p_X q_Y + p_Y q_X \leq 1$. Therefore, the right hand side of equation (3.4) is upper bounded by the quantity $f(\lambda_1, \lambda_2) \parallel X \parallel \cdot \parallel Y \parallel$. ■

3.4.2 Explicitness

Recall that a D-regular superoperator

$$E(X) = \frac{1}{D} \sum_i U_i X U_i^\dagger$$

is said to be *explicit* if it can be implemented by an efficient quantum circuit. Now we need a slight refinement of this definition: we say that E is *label-explicit* if each U_i has an efficient implementation. It can be checked that the squaring, tensoring and Zig-Zag operations map label-explicit transformations to label-explicit transformations. Also, our base superoperator is label-explicit (since it is defined over a constant size space). Therefore, the construction is label-explicit (and therefore explicit).

3.5 The complexity of estimating entropy

In this section we show that the language QED is QSZK–complete. The proof that QSD $\leq$ QED is standard, and is described in Subsection 3.5.4.

The more challenging direction is the proof that QED is in QSZK, or equivalently that QED $\leq$ QSD. In the classical setting this reduction is proved using extractors. Some parts of our proof of this reduction, for the quantum setting, are also standard. We define the problem QEA (Quantum Entropy Approximation) as follows:

Input: Quantum circuit Q, $t \geq 0$.
Accept: If $S(\tau_Q) \geq t + \frac{1}{2}$.
Reject: If $S(\tau_Q) \leq t - \frac{1}{2}$.

QEA is the problem of comparing the entropy of a given quantum circuit to some *known* threshold t (whereas QED compares two quantum circuits with unknown entropies). One immediately sees that

$$\text{QED}(Q_0, Q_1) = \bigvee_{t=1}^{\max\{\mathsf{out}_1, \mathsf{out}_2\}} [((Q_0, t) \in \text{QEA}_Y) \wedge ((Q_1, t) \in \text{QEA}_N)] \ ,$$

where out_i is the number of output qubits of Q_i.

A standard classical reduction can be easily adapted to the quantum setting to show that QEA $\in$ QSZK implies that QED $\in$ QSZK. We describe this part in Section 3.6. Thus, it is sufficient to prove that QEA $\in$ QSZK. We now focus on this part and the use of quantum expanders in the proof.

The classical reduction from EA to SD (where EA is like QEA but with the input being a classical circuit) uses *extractors*. An extractor is a function of the form $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, and we say that such a function is a (k, ϵ) extractor if, for every distribution X on $\{0,1\}^n$ that has min-entropy k, the distribution $E(X, U_d)$ obtained by sampling $x \in X$, $y \in \{0,1\}^d$ and outputting $E(x, y)$ is ϵ–close to uniform.

We begin with the classical intuition why EA reduces to SD. We are given a classical circuit C and we want to distinguish between the cases where the distribution it defines has substantially more than t entropy or substantially less than t entropy. First assume that the distribution is flat, i. e., all elements that have a non-zero probability in the distribution have equal probability. In such a case we can apply an extractor to the n output bits of C, hashing it to about t output bits. If the distribution C defines has high entropy, it also has high min-entropy (because for flat distributions entropy is the same as min-entropy) and therefore the output of the extractor is close to uniform. If, on the other hand, the entropy is less than $t - d - 1$, where d is the extractor's seed length, then even after applying the extractor the output distribution has at most $t - 1$ bits of entropy, and therefore it must be "far away" from uniform. Hence, we get a reduction to $\overline{\text{SD}}$.

There are, of course, a few gaps to fill in. First, the distribution C defines is not necessarily flat. This is solved in the classical case by taking many independent copies of the circuit C, which makes the output distribution "close" to "nearly-flat." A simple analysis shows that this flattening works also in the quantum setting (this is Lemma 3.27). Also, we need to amplify the gap we have between $t + 1/2$ and $t - 1/2$ to a gap larger than d (the seed length). This, again, is solved by taking many independent copies of C, given that $S(C^{\otimes q}) = qS(C)$.

This section is organized as follows. We first discuss quantum extractors. We then prove the quantum flattening lemma, and prove that QEA $\leq \overline{\text{QSD}}$ through the use of quantum extractors. Together with the closure of QSZK under Boolean formulas, which is proved in Section 3.6, we have that QED $\in$ QSZK. We conclude this section with a proof that QSD $\leq$ QED, using a simple quantum adaptation of the classical proof.

3.5.1 Quantum extractors

Definition 3.22. A superoperator $T : L(\mathcal{H}_N) \to L(\mathcal{H}_N)$ is a (k, d, ϵ) *quantum extractor* if:

- The superoperator T is 2^d-regular.
- For every density matrix $\rho \in L(\mathcal{H}_N)$ with $H_\infty(\rho) \geq k$, it holds that $\left\| T(\rho) - \tilde{I} \right\|_{\text{tr}} \leq \epsilon$.

We say T is explicit if T can be implemented by a quantum circuit of size polynomial in $\log(N)$. The *entropy loss* of T is $k + d - \log(N)$.

In the classical world balanced extractors are closely related to expanders (see, e. g., [55]). This generalizes to the quantum setting, as we now prove.

Lemma 3.23. *Suppose $T : L(\mathcal{H}_N) \to L(\mathcal{H}_N)$ is a $(N = 2^n, D = 2^d, \overline{\lambda})$ quantum expander. Then for every $t > 0$, T is also a $(k = n - t, d, \epsilon)$ quantum extractor with $\epsilon = 2^{t/2} \cdot \overline{\lambda}$. The entropy loss of T is $k + d - n = d - t$.*

Proof: The superoperator T has a one-dimensional eigenspace W_1 with eigenvalue 1, spanned by the unit eigenvector $v_1 = \frac{1}{\sqrt{N}} I$. Our input ρ is a density matrix, and therefore

$$\langle \rho, v_1 \rangle = \frac{1}{\sqrt{N}} \operatorname{Tr}(\rho) = \frac{1}{\sqrt{N}} .$$

In particular, $\rho - \tilde{I} = \rho - \frac{1}{\sqrt{N}} v_1$ is perpendicular to W_1. It follows that

$$\|T(\rho) - \tilde{I}\|_2^2 = \|T(\rho - \tilde{I})\|_2^2 \;\le\; \overline{\lambda}^2 \|\rho - \tilde{I}\|_2^2 \;\le\; \overline{\lambda}^2 \|\rho\|_2^2 ,$$

where we have used

$$\|\rho - \tilde{I}\|_2^2 = \|\rho\|_2^2 - 2 \operatorname{Tr}(\tilde{I}\rho) + \|\tilde{I}\|_2^2 = \|\rho\|_2^2 - \frac{1}{N} \le \|\rho\|_2^2 .$$

Given that $H_2(\rho) \ge H_\infty(\rho) \ge k = n - t$ we see that $\|T(\rho) - \tilde{I}\|_2^2 \le \overline{\lambda}^2 2^{-(n-t)}$. By the Cauchy-Schwartz inequality, it follows that

$$\left\| T(\rho) - \tilde{I} \right\|_{\mathrm{tr}} \le \sqrt{N} \|T(\rho) - \tilde{I}\|_2 \le \epsilon ,$$

which completes the proof. ∎

Corollary 3.24. *For every $n, t, \epsilon \ge 0$ there exists an explicit $(n - t, d, \epsilon)$ quantum extractor $T : L(\mathcal{H}_{2^n}) \to L(\mathcal{H}_{2^n})$, where*

1. *$d = 2(t + 2\log\left(\frac{1}{\epsilon}\right)) + O(1)$ and the entropy loss is $t + 4\log\left(\frac{1}{\epsilon}\right) + O(1)$, or*
2. *$d = t + 2\log\left(\frac{1}{\epsilon}\right) + 2\log(n) + O(1)$ and the entropy loss is $2\log(n) + 2\log\left(\frac{1}{\epsilon}\right) + O(1)$.*

The first bound on d is achieved using the Zig-Zag quantum expander of Theorem 3.16, and the second bound is achieved using the explicit construction of Ambainis and Smith [9] cited in Theorem 3.3.

One natural generalization of Definition 3.22 is to superoperators of the form $T : L(\mathcal{H}_N) \to L(\mathcal{H}_M)$ where $N = 2^n$ is not necessarily equal to $M = 2^m$. That is, such a superoperator T may map a large Hilbert space $\mathcal{H}_N$ to a much smaller Hilbert space $\mathcal{H}_M$. In the classical case this corresponds to hashing a large universe $\{0,1\}^n$ to a much smaller universe $\{0,1\}^m$. We suspect that unlike the classical case, no non-trivial unbalanced quantum extractors exist when $M < N/2$. Specifically, we suspect that all (k, d, ϵ) quantum extractors $T : L(\mathcal{H}_N) \to L(\mathcal{H}_M)$ with $k = n-1$ and $d < n-1$ must have error ϵ close to 1.

3.5.2 A flattening lemma

We first recall the classical flattening lemma that appears, e. g., in [138, Section 3.4.3].

Lemma 3.25. *Let $\lambda = (\lambda_1, \ldots, \lambda_M)$ be a distribution, let q be a positive integer, and let $\otimes^q \lambda$ denote the distribution composed of q independent copies of λ. Suppose that $\lambda_i \geq \Delta$ for all i. Then for every $\epsilon > 0$, the distribution $\otimes^q \lambda$ is ϵ-close to some distribution σ such that*

$$H_\infty(\sigma) \geq qH(\lambda) - O\left(\log\left(\frac{1}{\Delta}\right)\sqrt{q\log\left(\frac{1}{\epsilon}\right)}\right).$$

One can prove a similar lemma for density matrices.

Lemma 3.26. *Let ρ be a density matrix whose eigenvalues are $\lambda = (\lambda_1, \ldots, \lambda_M)$ and let q a positive integer. Suppose that for all i, $\lambda_i \geq \Delta$. Then for every $\epsilon > 0$, $\rho^{\otimes q}$ is ϵ-close to some density matrix σ such that*

$$H_\infty(\sigma) \geq qS(\rho) - O\left(\log\left(\frac{1}{\Delta}\right)\sqrt{q\log\left(\frac{1}{\epsilon}\right)}\right).$$

Lemma 3.26 follows directly from Lemma 3.25 because $S(\rho) = H(\lambda)$ and the vector of eigenvalues of $\rho^{\otimes q}$ equals $\otimes^q \lambda$.

We also need a way to deal with density matrices that may have arbitrarily small eigenvalues. This is really just a technicality as extremely small eigenvalues hardly affect the von Neumann entropy.

Lemma 3.27. *Let ρ be a density matrix of rank 2^m, let $\epsilon > 0$ and let q be a positive integer. Then $\rho^{\otimes q}$ is 2ϵ-close to a density matrix σ, such that*

$$H_\infty(\sigma) \geq qS(\rho) - O\left(m + \log\left(\frac{q}{\epsilon}\right)\right)\sqrt{q\log\left(\frac{1}{\epsilon}\right)}.$$

To prove this lemma, we will make use of the following fact [107, Box 11.2].

Fact 3.28 (Fannes' inequality)**.** Suppose ρ and σ are density matrices over a Hilbert space of dimension d. Suppose further that the trace distance between them satisfies $t = \| \rho - \sigma \|_{\mathrm{tr}} \leq 1/e$. Then

$$|S(\rho) - S(\sigma)| \leq t(\ln d - \ln t)\,.$$

Proof of Lemma 3.27: Let $\rho = \sum_{i=1}^{2^m} \lambda_i \, |v_i\rangle\langle v_i|$ be the spectral decomposition of ρ. Let

$$A = \left\{ i \;\middle|\; \lambda_i < \frac{\epsilon}{q 2^m} \right\}$$

denote the set of indices of "light" eigenvalues and define $\rho_0 = \sum_{i \notin A} \lambda_i \, |v_i\rangle\langle v_i|$. Observe that

$$\left\| \rho - \frac{\rho_0}{\mathrm{Tr}(\rho_0)} \right\|_{\mathrm{tr}} \leq \frac{\epsilon}{q}\,.$$

The eigenvalues of the density matrix $\rho_0 / \mathrm{Tr}(\rho_0)$ are all at least $\epsilon/(q2^m)$. Hence, by Lemma 3.26, it holds that $(\rho_0/\mathrm{Tr}(\rho_0))^{\otimes q}$ is ϵ-close to a density matrix σ such that

$$H_\infty(\sigma) \geq q \cdot S((\rho_0/\mathrm{Tr}(\rho_0))) - O\left(m + \log\left(\frac{q}{\epsilon}\right)\right)\sqrt{q \log\left(\frac{1}{\epsilon}\right)}\,.$$

Notice that

$$\left\| \rho^{\otimes q} - \left(\frac{\rho_0}{\mathrm{Tr}(\rho_0)}\right)^{\otimes q} \right\|_{\mathrm{tr}} \leq q \left\| \rho - \frac{\rho_0}{\mathrm{Tr}(\rho_0)} \right\|_{\mathrm{tr}} \leq \epsilon\,,$$

and therefore $\| \rho^{\otimes q} - \sigma \|_{\mathrm{tr}} \leq 2\epsilon$. By Fact 3.28,

$$\left| S\left(\frac{\rho_0}{\mathrm{Tr}(\rho_0)}\right) - S(\rho) \right| \leq \frac{\epsilon}{q}\left(m + \log\left(\frac{q}{\epsilon}\right)\right)\,.$$

Thus,

$$H_\infty(\sigma) \geq q \cdot S(\rho) - O\left(m + \log\left(\frac{q}{\epsilon}\right)\right)\sqrt{q \log\left(\frac{1}{\epsilon}\right)}\,,$$

which completes the proof. ∎

3.5.3 QEA $\leq \overline{\text{QSD}}$

We follow the outline of the classical reduction described at the beginning of the section. Let (Q, t) be an input to QEA, where Q is a quantum circuit with n input qubits and m output qubits. We consider the circuit $Q^{\otimes q}$ for $q = \mathrm{poly}(n)$ to be specified later, and we let E be a (qt, d, ϵ) quantum

extractor operating on qm qubits, where

$$d = q(m-t) + 2\log(1/\epsilon) + \log(qm) + O(1)\,,$$

and where $\epsilon = 1/\mathrm{poly}(n)$ is to be specified later. Such an extractor E exists by Corollary 3.24. We then let $\xi = E(\tau_Q^{\otimes q})$ and $\tilde{I} = 2^{-qm} I$, and take the output of the reduction to be $(\xi, \tilde{I})$.

To prove the correctness of the reduction, consider first a NO-instance $(Q,t) \in \mathrm{QEA}_N$. This implies

$$S(\xi) \leq S(\tau_Q^{\otimes q}) + d \leq q(t-0.5) + d\,.$$

We fix the parameters such that

$$\frac{q}{2} \geq 2\log\Big(\frac{1}{\epsilon}\Big) + \log(qm) + O(1) \tag{3.5}$$

and then $S(\xi) \leq qm - 1$. However, for any density matrix ρ over n qubits and $\epsilon > 0$, if $S(\rho) \leq (1-\epsilon)n$ then

$$\left\| \rho - \frac{1}{2^n} I \right\|_{\mathrm{tr}} \geq \epsilon - \frac{1}{2^n}\,.$$

It follows that

$$\left\| \xi - \tilde{I} \right\|_{\mathrm{tr}} \geq \frac{1}{qm} - \frac{1}{2^{qm}} \triangleq \beta$$

as required.

Now assume $(Q,t) \in \mathrm{QEA}_Y$. By Lemma 3.27, $\tau_Q^{\otimes q}$ is 2ϵ-close to a density matrix σ such that

$$\begin{aligned} H_\infty(\sigma) &\geq qS(\rho) - O\left(m + \log\left(\frac{q}{\epsilon}\right)\right)\sqrt{q\log\left(\frac{1}{\epsilon}\right)} \\ &\geq q\Big(t+\frac{1}{2}\Big) - O\left(m + \log\left(\frac{q}{\epsilon}\right)\right)\sqrt{q\log\left(\frac{1}{\epsilon}\right)}\,, \end{aligned}$$

and $\left\| \xi - \tilde{I} \right\|_{\mathrm{tr}} \leq \left\| E(\sigma) - \tilde{I} \right\|_{\mathrm{tr}} + 2\epsilon$. We set the parameters such that $H_\infty(\sigma)$ is larger than qt, that is,

$$\frac{q}{2} \geq O\left(m + \log\big(\frac{q}{\epsilon}\big)\right)\sqrt{q\log\left(1/\epsilon\right)}\,. \tag{3.6}$$

Now, by the quantum extractor property we obtain $\left\| \sigma - \tilde{I} \right\|_{\mathrm{tr}} \leq \epsilon$. Therefore, $\left\| \xi - \tilde{I} \right\|_{\mathrm{tr}} \leq 3\epsilon \triangleq \alpha$.

We set q and ϵ^{-1} large enough (but still polynomial in n, e. g., $\epsilon = \Theta(m^{-10})$ and $q = \Theta(m^4)$) such that the constraints (3.5) and (3.6) are satisfied and also that $\alpha \leq \beta^2$. Watrous [142] showed $\overline{\mathrm{QSD}_{\alpha,\beta}} \in \mathrm{QSZK}$ for these values of α, β.

3.5.4 QSD $\leq$ QED

Watrous [142] showed that $\text{QSD}_{\alpha,\beta}$ is QSZK-complete, even with parameters $\alpha = w(n)$ and $\beta = 1 - w(n)$ where n is the size of the input and $w(n)$ is a function smaller than any inverse polynomial in n. Assume we are given an input to $\text{QSD}_{\alpha,\beta}$, namely, two quantum circuits Q_0, Q_1, and construct quantum circuits Z_0 and Z_1 as follows. The circuit Z_1 outputs

$$\frac{1}{2}\left(|0\rangle\langle 0| \otimes \tau_{Q_0} + |1\rangle\langle 1| \otimes \tau_{Q_1}\right),$$

and the circuit Z_0 is the same as Z_1 except that the first register is traced out. The output of Z_0 is therefore $(1/2)(\tau_{Q_0} + \tau_{Q_1})$.

First consider the case where τ_{Q_0} and τ_{Q_1} are α close to each other, i. e., Q_0 and Q_1 produce almost the same mixed state. In this case $\tau_{Z_0} \approx \tau_{Q_0}$ whereas

$$\tau_{Z_1} \approx \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \otimes \tau_{Q_0},$$

and therefore τ_{Z_1} has about one bit of entropy more than τ_{Z_0}. On the other hand, when τ_{Q_0} and τ_{Q_1} are very far from each other, $\tau_{Z_0} = (1/2)(\tau_{Q_0} + \tau_{Q_1})$ contains about the same amount of entropy as

$$\tau_{Z_1} = \frac{1}{2}|0\rangle\langle 0| \otimes \tau_{Q_0} + \frac{1}{2}|1\rangle\langle 1| \otimes \tau_{Q_1}.$$

Formally, to estimate the entropy of τ_{Z_1} one can use the joint-entropy theorem (see [107, Theorem 11.8]) to get that $S(\tau_{Z_1}) = 1 + (1/2)(S(\tau_{Q_0}) + S(\tau_{Q_1}))$. When τ_{Q_0} and τ_{Q_1} are α close to each other, Fannes' inequality (Fact 3.28) tells us that $S(\tau_{Z_0})$ is close to

$$\frac{1}{2}(S(\tau_{Q_0}) + S(\tau_{Q_1})) \leq S(\tau_{Z_1}) - 0.9.$$

When τ_{Q_0} and τ_{Q_1} are β far from each other, there exists a measurement that distinguishes the two with probability $(1+\beta)/2$, so by [8, Lemma 3.2] we have

$$S(\tau_{Z_0}) \geq \frac{1}{2}[S(\tau_{Q_0}) + S(\tau_{Q_1})] + \left(1 - H\left(\frac{1+\beta}{2}\right)\right) \geq S(\tau_{Z_1}) - 0.1.$$

The reduction from $\text{QSD}_{\alpha,\beta}$ to QED is therefore as follows. Given an input (Q_0, Q_1) to $\text{QSD}_{\alpha,\beta}$ we reduce it to the pair of circuits $(O_0 = Z_0 \otimes Z_0 \otimes C, O_1 = Z_1 \otimes Z_1)$ where C outputs a qubit in the completely mixed state. If $(Q_0, Q_1) \in (\text{QSD}_{\alpha,\beta})_Y$ then

$$S(\tau_{O_0}) = S(\tau_{Z_0 \otimes Z_0 \otimes C}) = 2S(\tau_{Z_0}) + 1 \leq 2S(\tau_{Z_1}) - 0.8 < S(\tau_{O_1}),$$

whereas if $(Q_0, Q_1) \in (\text{QSD}_{\alpha,\beta})_N$ then

$$S(\tau_{O_0}) = S(\tau_{Z_0 \otimes Z_0 \otimes C}) = 2S(\tau_{Z_0}) + 1 \geq 2S(\tau_{Z_1}) + 0.8 = S(\tau_{O_1}) + 0.8\,.$$

3.6 Closure under Boolean formulas

We have observed that one can express QED as a formula in QEA, namely,

$$\text{QED}(Q_0, Q_1) = \bigvee_{t=1}^{\max\{\mathsf{out}_1, \mathsf{out}_2\}} \left[((Q_0, t) \in \text{QEA}_Y) \wedge ((Q_1, t) \in \text{QEA}_N)\right]\,,$$

where out_i is the number of output qubits of Q_i. In the classical setting it is known that SZK is closed under Boolean formulas. We now briefly explain why the same holds for QSZK, and refer the reader to [124] for further details. We first define what closure under Boolean formulas means. For a promise problem Π, the *characteristic function* of Π is the map $\chi_\Pi : \{0,1\}^* \to \{0,1,\star\}$ given by

$$\chi_\Pi(x) = \begin{cases} 1 & \text{if } x \in \Pi_Y, \\ 0 & \text{if } x \in \Pi_N, \\ \star & \text{otherwise.} \end{cases}$$

A *partial assignment* to variables $v_1, \ldots, v_k$ is a k-tuple $\overline{a} = (a_1, \ldots, a_k) \in \{0,1,\star\}^k$. For a propositional formula ϕ on variables $v_1, \ldots, v_k$ the evaluation $\phi(\overline{a})$ is recursively defined as follows:

$$v_i(\overline{a}) = a_i\,, \qquad (\neg\phi)(\overline{a}) = \begin{cases} 1 & \text{if } \phi(\overline{a}) = 0, \\ 0 & \text{if } \phi(\overline{a}) = 1, \\ \star & \text{otherwise,} \end{cases}$$

$$(\phi \wedge \psi)(\overline{a}) = \begin{cases} 1 & \text{if } \phi(\overline{a}) = 1 \text{ and } \psi(\overline{a}) = 1, \\ 0 & \text{if } \phi(\overline{a}) = 0 \text{ or } \psi(\overline{a}) = 0, \\ \star & \text{otherwise,} \end{cases} \qquad (\phi \vee \psi)(\overline{a}) = \begin{cases} 1 & \text{if } \phi(\overline{a}) = 1 \text{ or } \psi(\overline{a}) = 1, \\ 0 & \text{if } \phi(\overline{a}) = 0 \text{ and } \psi(\overline{a}) = 0, \\ \star & \text{otherwise.} \end{cases}$$

Notice that, e. g., $0 \wedge \star = 0$ even though one of the inputs is "undefined" in Π. This is because one has the evaluation $a \wedge 0 = 0$, irrespective of the value of a. For any promise problem Π, we define a new promise problem $\Phi(\Pi)$, with m instances of Π as input, as follows:

$$\Phi(\Pi)_Y \quad = \quad \{(\phi, x_1, \ldots, x_m) \mid \phi(\chi_\Pi(x_1), \ldots, \chi_\Pi(x_m)) = 1\}\,,$$

$$\Phi(\Pi)_N \quad = \quad \{(\phi, x_1, \ldots, x_m) \mid \phi(\chi_\Pi(x_1), \ldots, \chi_\Pi(x_m)) = 0\}\,.$$

If one can solve $\Phi(\Pi)$ then one can solve any Boolean formula over Π.

Theorem 3.29. *For any promise problem $\Pi \in$ QSZK we have $\Phi(\Pi) \in$ QSZK.*

The proof is identical to the classical proof in [124] except for straightforward adaptations (replacing the variational distance with the trace distance, using the closure of QSZK under complement, using the polarization lemma for QSD, etc.) and we sketch it here for completeness.

Proof: As QSD is QSZK-complete, Π reduces to QSD, inducing a reduction from $\Phi(\Pi)$ to $\Phi(\text{QSD})$. Thus, it suffice to show that $\Phi(\text{QSD})$ reduces to QSD. To this end, let $w = (\phi, (X_0^1, X_1^1), \ldots, (X_0^m, X_1^m))$ be an instance of $\Phi(\text{QSD})$. By applying De Morgan's Laws, we may assume that the only negations in ϕ are applied directly to the variables. (Note that De Morgan's Laws still hold in our extended Boolean algebra.) By the polarization lemma [142] and by the closure of QSZK under complementation [142], we can construct pairs of circuits $(Y_0^1, Y_1^1), \ldots, (Y_0^m, Y_1^m)$ and $(Z_0^1, Z_1^1), \ldots, (Z_0^m, Z_1^m)$ in polynomial time such that:

$$\begin{aligned}
(X_0^i, X_1^i) \in \text{QSD}_Y \quad &\Rightarrow \quad \left\| \tau_{Y_0^i} - \tau_{Y_1^i} \right\|_{\text{tr}} \geq 1 - \frac{1}{3|\phi|} \quad \text{and} \quad \left\| \tau_{Z_0^i} - \tau_{Z_1^i} \right\|_{\text{tr}} \leq \frac{1}{3|\phi|}\,, \\
(X_0^i, X_1^i) \in \text{QSD}_N \quad &\Rightarrow \quad \left\| \tau_{Y_0^i} - \tau_{Y_1^i} \right\|_{\text{tr}} \leq \frac{1}{3|\phi|} \quad \text{and} \quad \left\| \tau_{Z_0^i} - \tau_{Z_1^i} \right\|_{\text{tr}} \geq 1 - \frac{1}{3|\phi|}\,.
\end{aligned}$$

The reduction outputs the pair of circuits $(\text{BuildCircuit}(\phi, 0), \text{BuildCircuit}(\phi, 1))$, where BuildCircuit is described by the following recursive procedure:

BuildCircuit(ψ, b)

1. If $\psi = v_i$, output Y_b^i.
2. if $\psi = \neg v_i$, output Z_b^i.
3. If $\psi = \zeta \vee \mu$, output $\text{BuildCircuit}(\zeta, b) \otimes \text{BuildCircuit}(\mu, b)$.
4. If $\psi = \zeta \wedge \mu$, output

$$\frac{1}{2}(\text{BuildCircuit}(\zeta, 0) \otimes \text{BuildCircuit}(\mu, b)) + \frac{1}{2}(\text{BuildCircuit}(\zeta, 1) \otimes \text{BuildCircuit}(\mu, 1-b))\,.$$

Notice that the number of recursive calls equals the number of sub-formula of ϕ, and therefore the procedure runs in time polynomial in $|\psi|$ and $|X_i^j|$, i. e., polynomial in its input length.

We now turn to proving the correctness of this reduction. The correctness will follow from the claim below, wherein we define

$$\Delta(\zeta) \;=\; \frac{1}{2}\Big\| (\text{BuildCircuit}(\zeta,0) - \text{BuildCircuit}(\zeta,1))\,|0\rangle \Big\|_{\text{tr}}$$

for each sub-formula ζ of ϕ.

Claim 3.30. *Let $\overline{a} = (\chi_{QSD}(X_0^1, X_1^1), \ldots, \chi_{QSD}(X_0^m, X_1^m))$. For every sub-formula ψ of ϕ, we have:*

$$\begin{aligned} \psi(\overline{a}) = 1 \quad &\Rightarrow \quad \Delta(\psi) \geq 1 - \frac{|\psi|}{3|\phi|}\,, \\ \psi(\overline{a}) = 0 \quad &\Rightarrow \quad \Delta(\psi) \leq \frac{|\psi|}{3|\phi|}\,. \end{aligned}$$

Proof: The proof is by induction on the sub-formulas ψ of ϕ, and we note that it clearly holds for atomic sub-formulas. The remaining two cases are as follows.

Case 1: $\psi = \zeta \vee \mu$. If $\psi(\overline{a}) = 1$ then either $\zeta(\overline{a}) = 1$ or $\mu(\overline{a}) = 1$. Without loss of generality assume $\zeta(\overline{a}) = 1$. In this case we have for any $i \in \{0,1\}$ that BuildCircuit$(\zeta, i) = \mathcal{E}\,(\text{BuildCircuit}(\psi, i))$, where $\mathcal{E}$ is the quantum operation tracing out the registers associated with the μ sub-formula. Thus, by induction,

$$\Delta(\psi) \geq \Delta(\zeta) \geq 1 - \frac{|\zeta|}{3|\phi|} \geq 1 - \frac{|\psi|}{3|\phi|}\,.$$

If $\psi(\overline{a}) = 0$, then both $\zeta(\overline{a}) = \mu(\overline{a}) = 0$.

Using

$$\begin{aligned} \| \rho_0 \otimes \rho_1 - \sigma_0 \otimes \sigma_1 \|_{\text{tr}} \;&\leq\; \| \rho_0 \otimes \rho_1 - \sigma_0 \otimes \rho_1 \|_{\text{tr}} + \| \sigma_0 \otimes \rho_1 - \sigma_0 \otimes \sigma_1 \|_{\text{tr}} \\ &=\; \| \rho_0 - \sigma_0 \|_{\text{tr}} + \| \rho_1 - \sigma_1 \|_{\text{tr}}\,, \end{aligned}$$

we obtain

$$\Delta(\psi) \leq \Delta(\zeta) + \Delta(\mu) \leq \frac{|\zeta|}{3|\phi|} + \frac{|\mu|}{3|\phi|} \leq \frac{|\psi|}{3|\phi|}\,.$$

Case 2: $\psi = \zeta \wedge \mu$. Using

$$\begin{aligned} &\frac{1}{2}\left\| \frac{1}{2}[\rho_0 \otimes \sigma_0 + \rho_1 \otimes \sigma_1] - \frac{1}{2}[\rho_0 \otimes \sigma_1 + \rho_1 \otimes \sigma_0] \right\|_{\text{tr}} \\ =\; &\frac{1}{4} \| (\rho_0 - \rho_1) \otimes (\sigma_0 - \sigma_1) \|_{\text{tr}} \;=\; \frac{1}{4} \| \rho_0 - \rho_1 \|_{\text{tr}} \| \sigma_0 - \sigma_1 \|_{\text{tr}}\,, \end{aligned}$$

we obtain $\Delta(\psi) = \Delta(\zeta) \cdot \Delta(\mu)$. If $\psi(\overline{a}) = 1$, then, by induction,

$$\Delta(\psi) \geq \left(1 - \frac{|\zeta|}{3|\phi|}\right)\left(1 - \frac{|\mu|}{3|\phi|}\right) > 1 - \frac{|\zeta| + |\mu|}{3|\phi|} \geq 1 - \frac{|\psi|}{3|\phi|}\,.$$

If $\psi(\overline{a}) = 0$, then, without loss of generality, we may assume $\zeta(\overline{a}) = 0$. By induction we have

$$\Delta(\psi) = \Delta(\zeta) \cdot \Delta(\mu) \leq \Delta(\zeta) \leq \frac{|\zeta|}{3|\phi|} \leq \frac{|\psi|}{3|\phi|}\,.$$

Thus, the claim has been proved. ■

Let $A_b = \text{BuildCircuit}(\phi, b)$. By the above claim, if $w \in \Phi(\text{QSD})_Y$ then $\| \tau_{A_0} - \tau_{A_1} \|_{\text{tr}} \geq 2/3$ and if $w \in \Phi(\text{QSD})_N$ then $\| \tau_{A_0} - \tau_{A_1} \|_{\text{tr}} \leq 1/3$. This completes the proof of the theorem. ■

Chapter 4

Constructing Small-Bias Sets from AG Codes

In this chapter we use algebraic-geometric codes to give an explicit construction of an ϵ-biased set $S \subseteq \{0,1\}^k$ of size $O\left(\frac{k}{\epsilon^2 \log(1/\epsilon)}\right)^{5/4}$. This improves upon previous explicit constructions when ϵ is roughly (ignoring logarithmic factors) in the range $[k^{-1.5}, k^{-0.5}]$. Additionally, we discuss of the limits of our approach, based on a follow up work of Voloch [141].

4.1 Introduction

As discussed in introduction to this thesis, explicitly constructing pseudorandom objects is an intriguing challenge in computer science. Often, it is easy to verify that a random object satisfies the required pseudorandom property with high probability, while it is difficult to pin down such an explicit object.

In most cases it is believed (and sometimes proven) that a random object is nearly optimal. There are, however, rare cases in which explicit constructions outperform naive random constructions. Perhaps the most remarkable example of this type is that of Algebraic-Geometric codes (AG codes). In the seminal work of Tsfasman et al. [137] it was shown that there are Algebraic-Geometric codes over constant size alphabets that lie above the Gilbert-Varshamov bound, a bound that was believed to be optimal at the time.

The important case of *binary* error correcting codes is still open. The Gilbert-Varshamov bound gives the best known (explicit or non-explicit) codes to date. Finding an explicit construction that attains this bound is an open problem as well. The above statements also apply if we restrict ourselves to codes with distance close to half, which is a case of special interest.

Another closely related question is that of finding an $[n, k, (\frac{1}{2} - \epsilon)n]_2$ binary code, in which the

relative weight of every non-zero codeword is in the range $[\frac{1}{2}-\epsilon, \frac{1}{2}+\epsilon]$. Such codes are called *ϵ-balanced* and they are closely related to emphϵ-biased sets. Recall that an ϵ-biased set is a set $S \subseteq \{0,1\}^k$ such that for every non-empty subset $T \subseteq [k]$, the binary random variable $\bigoplus_{i \in T} s_i$, where s is sampled uniformly from S, has bias at most ϵ. It turns out that ϵ-biased sets are just ϵ-balanced codes in a different guise: the columns of a matrix whose rows generate an ϵ-balanced code form an ϵ-biased set, and vise versa. In terms of parameters, an $[n,k]_2$ ϵ-balanced code is equivalent to an ϵ-biased set $S \subseteq \{0,1\}^k$ of size n.

The status of ϵ-balanced codes is similar to that of $[n,k,(\frac{1}{2}-\epsilon)n]_2$ codes. In both cases the probabilistic method gives non-explicit $[n,k]_2$ ϵ-balanced codes with $n = O(\frac{k}{\epsilon^2})$, whereas the best lower bound is $n = \Omega(\frac{k}{\epsilon^2 \log(\frac{1}{\epsilon})})$. For a discussion of these bounds see [4, Section 7].

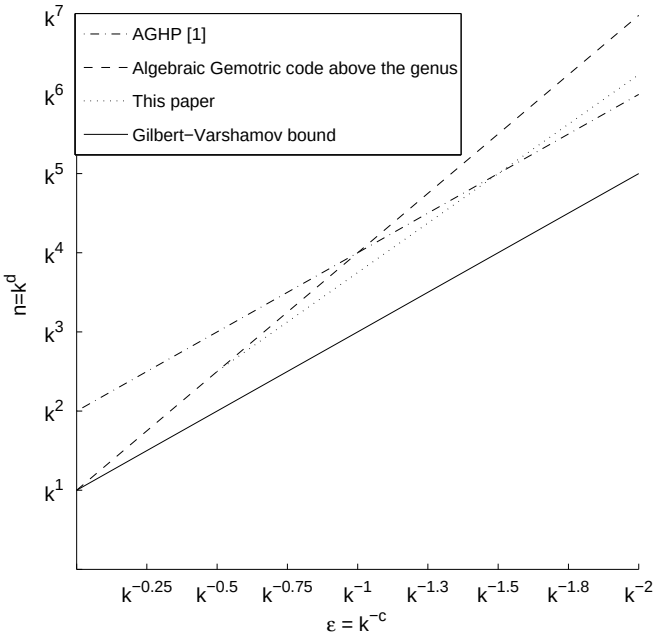
There are several *explicit* constructions of such codes. Naor and Naor [104] give a construction with $n = k \cdot \mathrm{poly}(\epsilon^{-1})$. Alon et al. [4] have the incomparable bound $n = O(\frac{k^2}{\epsilon^2 \log^2(k/\epsilon)})$. Concatenating Algebraic-Geometric codes with the Hadamard code gives $n = O(\frac{k}{\epsilon^3 \log(\frac{1}{\epsilon})})$. In this chapter we show an explicit construction of an $[n,k]_2$ ϵ-balanced code with $n = O(\frac{k}{\epsilon^2 \log(\frac{1}{\epsilon})})^{5/4}$, which improves upon previous explicit constructions when ϵ is roughly (ignoring logarithmic factors) in the range of $k^{-1.5} \leq \epsilon \leq k^{-0.5}$ (see Figure 4.1).

The construction is simple and can be described by elementary means. We first take a finite field $\mathbb{F}_q$ of the appropriate size. We then carefully choose a subset A of $\mathbb{F}_q \times \mathbb{F}_q$. The elements in the ϵ-biased set are indexed by pairs $((a,b),c) \in A \times \mathbb{F}_q$. For each $((a,b),c) \in A \times \mathbb{F}_q$ the corresponding element is the bit vector $\left(\langle (a^i b^j), c\rangle_2\right)_{i,j}$, where (i,j) range over all integers i,j whose sum is bounded by an appropriately chosen parameter and the inner product is of the binary representation of the elements in $\mathbb{F}_q$. The analysis of the construction relies on Bézout's Theorem.

To put the construction in context, we need to move to algebraic function fields terminology. AG codes are *evaluation* codes where a certain set of *evaluation functions* is evaluated at a chosen set of *evaluation points*. The space of evaluation functions used is a vector space (this is the reason we get a linear error correcting code) and is determined by a *divisor* G. We explain what a divisor is and other terminology in Section 4.3, and for the time being continue with an intuitive discussion. We denote the code associated with a divisor G by $C(G)$.

The code $C(G)$ has the following parameters. The *length* of the code is the number of evaluation points and is denoted by $N = N(F)$ (F is the algebraic function field). The distance of the code is $N - \deg(G)$ ($\deg(G)$ is the degree of G, we explain what it is in Section 4.3). The dimension of the code, $\dim(G)$, is the dimension of the vector space of evaluation functions. When the "degree" of G is larger than the *genus* (we explain what the genus is in Section 4.3), the Riemann-Roch Theorem [129, Thm I.5.17] tells us exactly what the dimension $\dim(G)$ is, and it turns out to be $\deg(G) - g + 1$. This almost matches the Singleton bound, except for a loss of g. Thus, our goal is to

Figure 4.1: Constructions of ϵ-biased sets for $\epsilon = k^{-c}$



get as many evaluation points while keeping the genus small. Indeed, a lot of research was done on the best possible ratio between the length of the code $N(F)$ and the genus. The bottom line of this research, roughly speaking, is that $N(F)$ can be larger than the genus by at most a multiplicative $\sqrt{q} - 1$ factor and this is essentially optimal.

A simple check shows that when $\deg(G)$ is larger than the genus, an AG code concatenated with Hadamard cannot give ϵ-balanced codes with n better than $O(\frac{k}{\epsilon^3 \log(\frac{1}{\epsilon})})$. In contrast, our construction takes as an outer code an AG code $C(G)$ where $\deg(G)$ is much smaller than the genus, and we show that this leads to a better code.

A natural question is whether the ϵ-balanced codes we achieve are the best binary codes one can achieve using this approach. We do not know the answer to this question. When $\deg(G)$ is smaller than the genus, one cannot use the Riemann-Roch Theorem, and estimating $\dim(G)$ is often a challenging task. Furthermore, $\dim(G)$ now depends on G itself, and not just on its degree as before. However, we can formulate the question as follows. The important thing to us is not the best possible ratio between the number of rational points $N(F)$ and the genus. Instead, we are interested in the best possible ratio between $N(F)$ and $\deg(G)$, where G is a *low-degree* divisor having a *large dimension*.

Following our work Felipe Voloch [141] used a variant of Castelnuovo's bound to show our approach cannot lead to error correcting codes approaching the Gilbert-Varshamov bound. We

show that a careful analysis of Voloch's argument imply that all dimension k, ϵ-balanced codes built using our approach must have length $n = \Omega(\frac{k}{\epsilon^{2.5}\log^2(\epsilon)})$.

The rest of the chapter is organized as follows. In Section 4.2 we describe the construction and its analysis using Bézout's Theorem. Section 4.3 contains a description of the same construction in algebraic function fields terminology. In Subsections 4.3.1 and 4.3.1 we give the necessary background on algebraic function fields and geometric Goppa codes. Finally, in Section 4.4 we analyze the limits of our approach based on Voloch's work.

4.2 A self-contained elementary description of the construction

We first recall the definition of an ϵ-biased set:

Definition 4.1. A set $S \subseteq \{0,1\}^k$ is *ϵ-biased* if for every nonempty $T \subseteq [k]$,

$$\frac{1}{|S|}\Big|\sum_{s\in S}(-1)^{\sum_{i\in T} s_i}\Big| \leq \epsilon.$$

The construction: Given k and ϵ, let $p = 2^\ell$ be a power of 2 in the range $\left[\frac{1}{2}\left(\frac{k}{\epsilon^2}\right)^{1/4}, \left(\frac{k}{\epsilon^2}\right)^{1/4}\right]$. That is, $\frac{1}{16}\frac{k}{\epsilon^2} \leq p^4 \leq \frac{k}{\epsilon^2}$. Define $q = p^2$ and $r = \epsilon p^3$. Let $\mathbb{F}_q$ denote the finite field with q elements and $\mathbb{F}_p$ its subfield with p elements. Consider the vector space of bivariate polynomials over $\mathbb{F}_q$ with total degree at most $r/(p+1)$:

$$V = \left\{\phi \in \mathbb{F}_q[x,y] \ : \ \deg(\phi) \leq \frac{r}{p+1}\right\} = \mathrm{Span}\left\{x^i y^j \ : \ i+j \leq \frac{r}{p+1}\right\}.$$

We denote the dimension of this space (over $\mathbb{F}_q$) by k'. It follows that

$$k' = \Omega(\frac{r^2}{p^2}) = \Omega(\frac{\epsilon^2 p^6}{p^2}) = \Omega(\epsilon^2 p^4) = \Omega(k).$$

Let $A \subseteq \mathbb{F}_q \times \mathbb{F}_q$ be the set of roots of the polynomial $y^p + y - x^{p+1}$. The ϵ-biased set over k' bits that we construct is

$$S = \left\{\left(\left\langle \mathrm{bin}(a^i b^j), \mathrm{bin}(c)\right\rangle_2\right)_{i+j\leq \frac{r}{p+1}} \ : \ (a,b) \in A \ \text{ and } \ \mathrm{c} \in \mathbb{F}_\mathrm{q}\right\},$$

where $\mathrm{bin} : \mathbb{F}_q \to \mathbb{Z}_2^{2\ell}$ is any isomorphism between the additive group of $\mathbb{F}_q$ and the vector space $\mathbb{Z}_2^{2\ell}$ and $\langle\cdot,\cdot\rangle_2$ denotes inner product over $\mathbb{Z}_2^{2\ell}$.

The analysis: The following claim will be used to bound the size of S.

Claim 4.2. *The cardinality of A is p^3.*

Proof: The trace function $\mathrm{Tr}(y) = y^p + y$ maps $\mathbb{F}_q$ to $\mathbb{F}_p$. We claim that for every $\alpha \in \mathbb{F}_p$, the number of solutions in $\mathbb{F}_q$ to $\mathrm{Tr}(y) = \alpha$ is p. To see this, observe that Tr is a linear function. Hence, the set of solutions to $\mathrm{Tr}(y) = 0$ is a subgroup of $\mathbb{F}_q$ that has at most p elements. For every $\alpha \in \mathbb{F}_p$, the set of solutions to $\mathrm{Tr}(y) = \alpha$ is either empty or a coset of this subgroup. As every element of $\mathbb{F}_q$ is in one of these cosets, it must be the case that for every $\alpha \in \mathbb{F}_p$ there are exactly p solutions.

The norm function $\mathrm{N}(x) = x^{p+1}$ also maps $\mathbb{F}_q$ to $\mathbb{F}_p$. Thus, for every $\alpha \in \mathbb{F}_q$ there are exactly p values $\beta \in \mathbb{F}_q$ such that $\mathrm{Tr}(\beta) = \mathrm{N}(\alpha)$. Therefore, $|A| = p^3$. ■

We want to apply Bézout's Theorem on the bivariate polynomial $y^p + y - x^{p+1}$. However, we first need to show it is irreducible. We need Eisenstein's Criterion for irreducibility:

Theorem 4.3 (Eisenstein's Criterion [91, Thm 3.1])**.** *Let U be a unique factorization ring and let K be its field of fractions. Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial of degree $n \geq 1$ in $U[x]$. Let ρ be a prime of U, and assume:*

- $a_n \neq 0 \pmod{\rho}$
- *For every $i < n$, $a_i = 0 \pmod{\rho}$*
- $a_0 \neq 0 \pmod{\rho^2}$.

Then $f(x)$ is irreducible in $K[x]$.

With that we conclude:

Claim 4.4. *The polynomial $y^p + y - x^{p+1}$ is irreducible over $\mathbb{F}_q$.*

Proof: This follows from Eisenstein's Criterion. The unique factorization ring we consider is $U = \mathbb{F}_q[y]$. The prime element we use is $\rho = y$. The leading coefficient is -1 and $-1 \neq 0 \pmod{y}$. Every other coefficient except the last is 0, hence it is $0 \pmod{y}$. The last coefficient is also $0 \pmod{y}$. Finally, since $p \geq 2$, $y^p = 0 \pmod{y^2}$ but $y \neq 0 \pmod{y^2}$, hence $y^p + y \neq 0 \pmod{y^2}$. Therefore the univariate polynomial (in x) is irreducible over the field of fractions. As one of the coefficients is -1, it follows that the bivariate polynomial is irreducible over the field $\mathbb{F}_q$ (see [91, Thm 2.3]). ■

We are now ready to recall Bézout's Theorem and apply it prove S is indeed ϵ-biased.

Theorem 4.5 (Bézout's Theorem [50, Section 5.3])**.** *Suppose ϕ and ψ are two bivariate polynomials over some field. If ϕ and ψ have more than $\deg(\phi) \cdot \deg(\psi)$ common roots then they have a common factor.*

Theorem 4.6. *For every k and ϵ such that $\epsilon < \frac{1}{\sqrt{k}}$, S is an ϵ-biased set over $k' = \Omega(k)$ bits of size $O\left(\frac{k}{\epsilon^2}\right)^{5/4}$.*

Proof: By Claim 4.2, $|S| = |A| \cdot q = p^5 = O\left(\frac{k}{\epsilon^2}\right)^{5/4}$.

Let $T \subseteq [k']$ be some non-empty set. We identify $[k']$ with the set $\left\{(i,j) \ : \ i+j \leq \frac{r}{p+1}]\right\}$ and T with the corresponding subset.

Let $s \in S$ be an element specified by the pair $((a,b),c) \in A \times \mathbb{F}_q$. Then,

$$\sum_{(i,j)\in T} s_{(i,j)} = \sum_{(i,j)\in T} \left\langle \mathrm{bin}(a^i b^j), \mathrm{bin}(c) \right\rangle_2 = \left\langle \mathrm{bin}\Big(\sum_{(i,j)\in T} a^i b^j \Big), \mathrm{bin}(c) \right\rangle_2 .$$

The polynomial $\phi_T = \sum_{(i,j)\in T} x^i y^j$ is a non-zero polynomial. Clearly, for any (a,b) which is not a root of ϕ_T, the inner-product will be unbiased when ranging over c (i.e. exactly half of the values for c will make the inner product 0). From the assumption $\epsilon < \frac{1}{\sqrt{k}}$ it follows that $\deg(\phi_T) < p+1$, since

$$\frac{\deg(\phi_T)}{p+1} \leq \frac{r}{(p+1)^2} < \epsilon p \leq k^{1/4}\sqrt{\epsilon} < 1.$$

Hence, by Claim 4.4 it follows that ϕ_T and $y^p + y - x^{p+1}$ have no common factors. Therefore, by Bézout's theorem we conclude that the number of roots of ϕ_T that are in A is at most $\frac{r}{p+1} \cdot (p+1) = r$, and,

$$\frac{1}{|S|}\Big| \sum_{s\in S} (-1)^{\sum_{i\in T} s_i} \Big| \leq \frac{r}{|A|} = \epsilon.$$

■

Remark 4.7. The above construction can be improved to an ϵ-biased set of size $O\left(\frac{k}{\epsilon^2 \log(1/\epsilon)}\right)^{5/4}$ for every k and ϵ such that $\frac{\epsilon}{\sqrt{\log(1/\epsilon)}} < \frac{1}{\sqrt{k}}$. To achieve this we choose $p = \Theta\left(\frac{k}{\epsilon^2 \log(1/eps)}\right)^{1/4}$. We then observe that instead of taking a basis for V over $\mathbb{F}_q$, we can actually afford to take a basis over $\mathbb{F}_2$. Finally we need to use the fact that by the constraints we have on ϵ, it follows that $\log(1/\epsilon) = \Theta(\log(p))$. When we restate the construction in algebraic function fields terminology, we also include this improvement.

4.3 Restating the construction in AG terminology

Without putting the above construction in the proper context, it may appear coincidental. We now describe the general framework of algebraic-geometric codes and explain why the above construction fits into this framework.

4.3.1 Algebraic-Geometry

We recall a few notions from the theory of algebraic function fields. A detailed exposition of the subject can be found, e.g., in [129].

$\mathbb{F}_q$ denotes the finite field with q elements. $\mathbb{F}_q(x)$, where x is transcendental over $\mathbb{F}_q$, is the *rational* function field, and it contains all rational functions in x with coefficients in $\mathbb{F}_q$. $F/\mathbb{F}_q$ is an algebraic function field, if F is a finite algebraic extension of $\mathbb{F}_q(x)$.

A *place* P of $F/\mathbb{F}_q$ is a maximal ideal of some valuation ring O of the function field. We denote by O_P the valuation ring that corresponds to the place P. We denote by v_P the *discrete valuation* that corresponds to the valuation ring O_P. Therefore, we can write P and O_P as

$$P = \{x \in F \ : \ v_P(x) > 0\} \quad \text{and} \quad \mathrm{O_P} = \{\mathrm{x} \in \mathrm{F} \ : \ \mathrm{v_P(x)} \geq 0\}.$$

Since P is a maximal ideal, $F_P = O_P/P$ is a field. In fact, it is a finite field [129, Proposition I.1.14]. For every $x \in O_P$, $x(P)$ denotes $x(\bmod P)$ and is an element of F_P. The degree of a place P is defined to be $\deg(P) = [F_P : \mathbb{F}_q]$. In particular, if a place is of degree 1 then F_P is isomorphic to $\mathbb{F}_q$. $\mathcal{P}_F$ is the set of places of F. $N(F)$ is the number places of *degree* 1 (also called *rational points*) in $F/\mathbb{F}_q$ and is always finite.

$\mathcal{D}_F$ is the free abelian group over the places of F. A *divisor* is an element in this group, i.e., it is a sum $G = \sum_{P \in \mathcal{P}_F} n_P P$ with $n_P \in \mathbb{Z}$ and where $n_P \neq 0$ for only a finite number of places. We also denote $v_P(G) = n_P$. The *degree* of the divisor $\sum_P n_P P$ is defined to be $\sum_P n_P \cdot \deg(P)$, and it is always finite. We say $G_1 \geq G_2$ if G_1 is component-wise larger than G_2, i.e., $v_P(G_1) \geq v_P(G_2)$ for any place P. The *support* of a divisor G is $\mathrm{Supp}(G) = \{P \in \mathcal{P}_F \ : \ v_P(G) \neq 0\}$.

Each element $0 \neq x \in F$ is associated with two divisors. The first is called the *principal divisor* of x and it is defined by

$$(x) = \sum_P v_P(x) P.$$

The degree of a principal devisor is always 0. The second is the *pole divisor* of x and it is defined by

$$(x)_\infty = \sum_{P : v_P(x) < 0} -v_P(x) P.$$

If $x \in F \setminus \mathbb{F}_q$ then $\deg((x)_\infty) = [F : \mathbb{F}_q(x)]$.

For a divisor G, we define the *Riemann-Roch space* is

$$\mathcal{L}(G) = \{x \in F \ : \ (x) \geq -G\} \cup \{0\}.$$

We define the dimension of G by $\dim(G) = \dim \mathcal{L}(G)$ and we use the two notations interchangeably. The fact that the degree of each principal divisor is 0 implies that if $\deg(G) < 0$ then $\dim(\mathcal{L}(G)) = 0$.

Geometric Goppa Codes

A Goppa code is specified by a triplet (F, Y, G), where $F/\mathbb{F}_q$ is a function field, $Y = \{P_1, \ldots, P_n\}$ is a set of places of degree 1 and G is an arbitrary divisor with no support over any place in Y. Notice that for any $x \in \mathcal{L}(G)$, $v_{P_i}(x) \geq 0$ and therefore $x \in O_{P_i}$ and $x(P_i) \in \mathbb{F}_q$. The triplet (F, Y, G) specifies the code:

$$C(Y; G) = \{(x(P_1), \ldots, x(P_n)) \ : \ x \in \mathcal{L}(G)\} \subseteq {\mathbb{F}_q}^n.$$

Claim 4.8 ([129, Cor II.2.3]). *If* $\deg(G) < n$ *then* $C(Y; G)$ *is an* $[n, \dim(\mathcal{L}(G)), n - \deg(G)]$ *linear code over* $\mathbb{F}_q$.

We want the gap between $\dim(\mathcal{L}(G))$ and $\deg(G)$ to be small. It turns out that for any function field $F/\mathbb{F}_q$ there exists a constant $g \in \mathbb{N}$, such that for any divisor $G \in \mathcal{D}_F$, $\deg(G) - \dim(\mathcal{L}(G)) \leq g - 1$. The minimal integer with this property is called the *genus* of $F/\mathbb{F}_q$. The Riemann-Roch Theorem says that:

Theorem 4.9 ([129, Thm I.5.17]). *If* $\deg(G) \geq 2g - 1$ *then* $\dim(\mathcal{L}(G)) = \deg(G) - g + 1$.

This, in particular, allows one to easily compute the dimension of the code when $\deg(G) > 2g - 2$. The only remaining question is whether there are function fields with a large number $N = N(F)$ of rational points, and a small genus g. This is addressed in:

Theorem 4.10 (Hasse-Weil bound [129, Thm V.2.3]). *Let* $F/\mathbb{F}_q$ *be a function field of genus* g. *Then, the number* N *of places of degree one satisfies* $N \leq (q + 1) + 2\sqrt{q}g$.

The Drinfeld-Vlăduţ bound tells us that when g tends to infinity, the bound can be strengthened by about a factor of 2, and roughly speaking, $N \leq g(\sqrt{q} - 1)$. This is tight for prime power squares q, and several explicit constructions meet the bound (see [52, Chapter 1]).

In this chapter we look at divisors G whose degree is smaller than the genus. Much less is known about such small-degree divisors. In this regime, $\dim(\mathcal{L}(G))$ depends on the divisor G itself, and

not only on its degree, as is the case when $\deg(G) > 2g - 2$. For some special algebraic function fields the vector space $\mathcal{L}(G)$ (and therefore also its dimension) is known in full. We talk more about this below.

4.3.2 Concatenating AG codes with Hadamard

We concatenate an outer code with the Hadamard code. If the outer code is an $[n_1, k_1, d]_q$ code and q is a power of two, then concatenating it with the $[q, \log(q), q/2]_2$ Hadamard code gives an $[n = n_1 q, k = k_1 \log q]_2$ code that is $\epsilon = \frac{n_1 - d}{n_1}$ balanced, because non-zero symbols in the outer code expand by the concatenation to perfectly balanced blocks.[1]

Using a $[q, k_1, q - k_1 + 1]_q$ Reed-Solomon code as the outer code, one gets an $[n = q^2, k = k_1 \log q]_2$ code that is $\epsilon < \frac{k_1}{q}$ balanced. Rearranging parameters, this gives an $[n, k]_2$ ϵ-balanced code with

$$n = O((\frac{k}{\epsilon \log(\frac{k}{\epsilon})})^2). \tag{4.1}$$

This is one of the constructions in [4].

Taking the outer code to be a $[N, \dim(G), N - \deg(G)]_q$ AG code $C(Y; G)$ over $\mathbb{F}_q$ and concatenating it with Hadmard, we get a $[n = Nq, k = \dim(G) \log q]_2$ code that is $\epsilon = \frac{\deg(G)}{N}$ balanced. We can choose an AG code which uses a curve of genus g with $N \geq g\sqrt{q}$ rational points. Picking the divisor G to be of degree $\deg(G) \geq 2g$ and setting $q = \frac{1}{\epsilon^2}$ results in

$$N = \frac{\deg(G)}{\epsilon} \approx \frac{\dim(G) + g}{\epsilon} \approx \frac{k}{\epsilon \log(\frac{1}{\epsilon})},$$

where the second equality follows from the Riemann-Roch Theorem. Thus, we get an ϵ-balanced code of length

$$n = Nq = O(\frac{k}{\epsilon^3 \log(\frac{1}{\epsilon})}). \tag{4.2}$$

In fact, if one takes an AG code over $\mathbb{F}_q$ with large genus $g \geq \sqrt{q}$ then

$$N \geq \frac{\dim(G)}{\epsilon} = \frac{k}{\epsilon \log q} \qquad \text{and} \qquad q = \Omega(\frac{1}{\epsilon^2})$$

and Equation (4.2) is tight. Taking an AG code with a small genus $g \leq \sqrt{q}$ is essentially equivalent to taking a Reed-Solomon outer code and cannot be better (up to constant factors) than Equa-

[1] If q is a power of 2, then the resulting concatenated code is linear. Concatenation is well defined even when q is not a power of 2. In such a case we embed $\mathbb{F}_q$ into $\mathbb{F}_2{}^{\lceil \log q \rceil}$ using any one-to-one mapping. The resulting (non-linear) code has essentially the same dimension and distance as in the previous case - the only difference is a small loss due to the fact that $2^{\lceil \log q \rceil}$ is slightly larger than q. From now on we will discuss the simpler case where q is a power of two, keeping in mind that everything also holds for arbitrary q.

tion (4.1). In what follows, we show one can improve on both bounds when the AG code has degree much smaller than the genus.

So we now turn our attention to the case where $\deg(G) \leq 2g-1$. In this case $\dim \mathcal{L}(G)$ depends on the divisor G and not just its degree. One special case is the case where $G = rQ$, $r \in \mathbb{N}$ and Q is a place of degree 1. For any such r, $\dim \mathcal{L}(rQ)$ is either equal to $\dim \mathcal{L}((r-1)Q)$ or to $\dim \mathcal{L}((r-1)Q)+1$. In the former case r is said to be a *gap number* of Q. Weierstrass Gap Theorem [129, Thm I.6.7] says that for any place Q there are exactly $g = \mathrm{genus}(F/\mathbb{F}_q)$ gap numbers, and they are all in the range $[1, 2g-1]$.

The non-gap numbers (also called *pole numbers*) form a semigroup of $\mathbb{N}$ (i.e. a set that is closed under addition). This semigroup is sometimes referred to as the *Weierstrass semigroup* of Q. We say a semi-group S is generated by a set of elements $\{g_i\}$, if each $g_i \in S$ and, furthermore, every element $s \in S$ can be expressed as $s = \sum a_i g_i$ with $a_i \in \mathbb{N}$.

The structure of the Weierstrass semigroup is crucial to our construction. We know that there are exactly g elements of this semigroup in the range $[1, 2g]$. If these elements are too concentrated on the upper side of the range then the behavior of $\dim \mathcal{L}(rQ)$ will be very similar to the case where $r > 2g-1$. Thus, our goal is to find a function field F that has many places of degree 1, say, $N(F) \geq \Omega(g\sqrt{q})$, while at the same time F has a degree 1 place Q with a "good" Weierstrass semigroup.

4.3.3 The Construction

Let p be a prime power and $q = p^2$. The Hermitian function field over $\mathbb{F}_q$ (see [129, Lemma VI.4.4]) can be represented as the extension field $\mathbb{F}_q(x,y)$ of the rational function field $\mathbb{F}_q(x)$ with $y^p + y = x^{p+1}$. This function field has $1+p^3$ places of degree one. First, there is the common pole Q_∞ of x and y. Moreover, for each pair $(\alpha, \beta) \in \mathbb{F}_q$ with $\beta^p + \beta = \alpha^{p+1}$ there is a unique place $P_{\alpha,\beta}$ of degree one such that $x(P_{\alpha,\beta}) = \alpha$ and $y(P_{\alpha,\beta}) = \beta$ and we already saw there are p^3 such points. The genus of the Hermitian function field is $g = p(p-1)/2$.

For the outer code we take the Goppa code $C_r = C(Y, G = rQ_\infty)$, where Y is the set of all degree 1 places $P_{\alpha,\beta}$ mentioned above and $r = \epsilon p^3$. The Weierstrass semigroup of G is generated by p and $p+1$, and a basis for $\mathcal{L}(G) = \mathcal{L}(rQ_\infty)$ is

$$\big\{x^i y^j \ : \ j \leq p-1 \ \text{ and } \ ip + j(p+1) \leq r\big\}.$$

The dimension of the code is

$$\big|\{(i,j) \ : \ j \leq p-1 \ \text{ and } \ ip + j(p+1) \leq r\}\big|.$$

We can now see the similarity between this construction and the one in Section 4.2. The parameter r will be chosen such that the constraint $j \leq p-1$ will be nullified. Therefore, both use evaluations of low degree bivariate polynomials over the same set of p^3 points.[2]

Theorem 4.11. *For every k and every ϵ such that $\frac{\epsilon}{\sqrt{\log(1/\epsilon)}} \leq \frac{1}{\sqrt{k}}$, there exists an explicit $[n, \Omega(k)]_2$ code that is ϵ-balanced, with $n = O\left(\frac{k}{\epsilon^2 \log(1/\epsilon)}\right)^{5/4}$.*

Proof: For a given k and ϵ, let

$$p \in \left[\frac{1}{2}\left(\frac{k}{\epsilon^2 \log(1/\epsilon)}\right)^{1/4}, \left(\frac{k}{\epsilon^2 \log(1/\epsilon)}\right)^{1/4}\right]$$

be a power of two. It can verified that $\frac{1}{16p^4} \leq \epsilon \leq \frac{1}{p}$ as

$$\begin{aligned} \frac{1}{p} &\geq \left(\frac{\epsilon^2 \log(1/\epsilon)}{k}\right)^{1/4} \geq \left(\epsilon^2 \log(1/\epsilon) \cdot \frac{\epsilon^2}{\log(\frac{1}{\epsilon})}\right)^{1/4} = \epsilon \quad \text{and} \\ \epsilon &= \epsilon^2 \cdot \frac{1}{\epsilon} \geq \epsilon^2 \cdot \log(\frac{1}{\epsilon}) \geq \frac{k}{16p^4} \geq \frac{1}{16p^4}, \end{aligned}$$

and so $\log(1/\epsilon) = \Theta(\log(p))$.

Let $r = \epsilon p^3$ and let $\mathbb{F}_q$ be the field with $q = p^2$ elements. Let F denote the Hermitian function field over $\mathbb{F}_q$ and let Y denote its set of places of degree 1, excluding Q_∞. This implies that $|Y| = p^3$. Define the divisor G to be $G = rQ_\infty$. Since $r \leq p^2$,

$$\dim \mathcal{L}(rQ_\infty) \geq (\frac{r}{2(p+1)})^2 = \Omega(\epsilon^2 p^4) = \Omega(\frac{k}{\log(p)}).$$

By Claim 4.8, the Goppa code that is obtained from the triplet (F, Y, G) is a

$$[p^3, \Omega(\frac{k}{\log(p)}), p^3 - r]_{p^2}$$

code. Concatenating this code with Hadamard gives a $[p^5, \Omega(k)]_2$ code that is ϵ-balanced (since $\frac{r}{p^3} = \epsilon$). Now, by our choice of p, it follows that

$$\frac{k}{\epsilon^2 \log(\frac{1}{\epsilon})} = \Theta(p^4)$$

[2]The only slight difference is that in this construction we take all bivariate polynomials with bounded *weighted* total degree. However, the weight is nearly identical for both variables and so this does not affect much the parameters of the construction.

and therefore $n = p^5 = O((\frac{k}{\epsilon^2 \log(\frac{1}{\epsilon})})^{5/4})$ as desired. ■

4.4 The approach limits

As explained in Section 4.3.1, the genus measures the maximal loss in dimension compared to the degree. The Drinfeld-Vlăduţ bound implies that the number of evaluation points (which is bounded by the number of degree one places $N(F)$) is at most $O(g\sqrt{q})$ when $N(F) \gg q$. In Section 4.3.2 we saw this implies that when $\deg(G) > 2g$, concatenating the best AG code $C(Y;G)$ with Hadamard cannot give ϵ-balanced codes of dimension k and length $n = O(\frac{k}{\epsilon^3 \log(1/\epsilon)})$.

Our construction shows substantially better results are possible when $\deg(G) \ll g$. Namely, we show that there exists a code $C(Y;G)$ with $\deg(G) \ll g$ such that when this code is concatenated with Hadamard, it gives a dimension k, ϵ-balanced code of length $n = O((\frac{k}{\epsilon^2 \log(1/\epsilon)})^{5/4})$. It is therefore natural to ask what are the limits of our approach. More concretely we ask what are the best codes one can construct concatenating an AG code with a Hadamard code? Let us state the question precisely. We look at constructions of the following structure:

- An outer AG code $C = C(Y;G)$, defined by an algebraic function field $F/\mathbb{F}_q$, a set of rational points Y and a divisor $G \in \mathcal{D}_F$ with no support over any place in Y.
- An inner Hadamard code.

In the analysis we view C as a $[|Y|, \dim(G), |Y| - \deg(G)]_q$ code, and then the concatenated code has parameters $[|Y|q, \dim(G)\log(q), \frac{1}{2} - \frac{\deg(G)}{|Y|}]_2$. Notice that it may be the case that C has better distance than the so-called *designated* distance, but as far as we are concerned the analysis does not take advantage of that, and we take the distance to be $|Y| - \deg(G)$.

In this section we prove:

Theorem 4.12. *Any ϵ-balanced $[n,k]_2$ code that is constructed and analyzed as above, must have*

$$n \geq \Omega\left(\frac{k}{\epsilon^2} \cdot \min\left\{\frac{k}{\log^2(\frac{k}{\epsilon})}, \frac{1}{\sqrt{\epsilon}\log(\frac{k}{\epsilon})}\right\}\right).$$

For the proof we need definitions and theorems about finite extensions of algebraic functions fields. Specifically, for an extension F of a function field F', we use the following notation:

- A place $P \in \mathcal{P}_F$ lying over a place $P' \in \mathcal{P}_{F'}$, denoted by $P|P'$, see [129, Def III.1.3],
- The ramification index of P over P', denoted by $e(P|P')$, see [129, Def III.1.5],
- The conorm of a divisor $G' \in \mathcal{D}_{F'}$, denoted by $\mathrm{Con}_{F/F'}(G')$, see [129, Def III.1.8].

For more details we refer the reader to [129, Chapter III].

4.4.1 AG theorems about degree vs. dimension

It turns out the above question boils down to the question of whether there are function fields with many rational points (compared to the genus) and with low-degree divisors (of degree much smaller than the genus) of high dimension. We start by presenting two AG theorems relating degree to dimension in the small degree regime (when the degree is smaller than the genus).

The first argument we present shows any divisor with non-trivial dimension must have degree at least $N(F)/(q+1)$. The argument was shown to us by Henning Stichtenoth [130].

Lemma 4.13. *Let $F/\mathbb{F}_q$ be a function field and $G \in \mathcal{D}_F$ a divisor with $\dim(\mathcal{L}(G)) > 1$. Then $N(F) \leq \deg(G) \cdot (q+1)$.*

Proof: As $\dim(\mathcal{L}(G)) > 1$, there exists some $x \in F \setminus \mathbb{F}_q$ such that $(x) \geq -G$. Fix any such x. In particular, $\deg((x)_\infty) \leq \deg(G)$. Also, by [129, Thm I.4.11], $\deg(x)_\infty = [F : \mathbb{F}_q(x)]$. We may view F as a finite extension over the rational function field $\mathbb{F}_q(x)$. Every place of degree 1 of F lies above some place of degree 1 of $\mathbb{F}_q(x)$. There are exactly $q+1$ places of degree 1 of $\mathbb{F}_q(x)$, and each one of them may split to at most $[F : \mathbb{F}_q(x)]$ places of degree 1 of F (by the fundamental equality, [129, Thm III.1.11]). Altogether, $N(F) \leq (q+1)[F : \mathbb{F}_q(x)] = (q+1)\deg(x)_\infty \leq (q+1)\deg(G)$. ∎

Remark 4.14. Lemma 4.13 only uses the fact that G is non-trivial. We wonder if one can strengthen the lemma for divisors G of high dimension. In particular, is it true that if $\dim(\mathcal{L}(G)) > \ell$ then $N(F) \leq \frac{\deg(G)\cdot(q+1)}{f(\ell)}$ for some function f that goes to infinity with ℓ ?

We now move to the second theorem. For a set $S \subseteq F$, where F is a field, let $\mathrm{Closure}(S)$ denote the minimal subfield of F that contains S. Following our work, Voloch [141] showed, based on Castelnuovo bound, that:

Theorem 4.15 ([141, based on Castelnuovo bound])**.** *Let K be an arbitrary field. Let F/K be a function field of genus g. Let $G \in \mathcal{D}_F$ be a divisor with degree $d+1$ and dimension $\ell+2$ such that $\mathrm{Closure}(\mathcal{L}(G)) = F$. Let $m = d \operatorname{div} \ell$ and $r = d \bmod \ell$. Then*

$$g \leq m(m-1)\ell + m(2r+1),$$

and, in particular, $g \leq m(m+1)\ell$.

Using Theorem 4.15 requires an assumption on the AG code, namely, that the closure of the Riemann space of the divisor used to define the code is the entire function field F. The following

lemma, based on private communication with Voloch, shows that this assumption is inessential when analyzing the rate versus distance problem.

Lemma 4.16. *Let K be a finite field, F/K be a function field, $G \in \mathcal{D}_F$ is a divisor. Let C be the Goppa code of length n, dimension k and designated relative distance δ, specified by some triplet (F, Y, G). Define a function field $F' = \mathrm{Closure}(\mathcal{L}(G))$. Then there exists a Goppa code C' defined by a triplet (F', Y', G'), of length $n' \leq n$, dimension k and designated relative distance $\delta' \geq \delta$, such that $\mathrm{Closure}(\mathcal{L}(G')) = F'$.*

Proof: We first define the new triplet (F', Y', G').

- We already have that $F' = \mathrm{Closure}(\mathcal{L}(G))$.
- Next let $B = \{s_1, \ldots, s_k\}$ be a basis for $\mathcal{L}(G)$. Define,

$$G'' = \sum_{P' \in \mathcal{P}_{F'}} \max_i \{-v_{P'}(s_i)\} \cdot P'.$$

  We would like to exchange G'' with an equivalent divisor that has no support over places of degree 1. By the Weak Approximation Theorem [129, Thm I.3.1] there exists $z \in F/K$ such that for every place P of degree 1, $v_P(z) = -v_P(G)$ and we let

$$G' = G'' + (z).$$

- Define a set $Y' \subset \mathcal{P}_{F'}$ by

$$Y' = \left\{ P' \in \mathcal{P}_{F'} \ : \ \exists P \in Y \ \text{ such that } \ P | P' \right\}.$$

  Observe that since Y consists only of places of degree 1 this is also true for Y' (see [129, Proposition III.1.6]).

Consider the Goppa code C' defined by the triplet (F', Y', G'). Notice that Y' does not intersect G' because Y' contains only degree 1 places and G' has no support over degree 1 places. We will prove:

- The dimension of C' is the same as C, i.e., $\dim(\mathcal{L}'_F(G')) = k$.
- The length of C' is at most the length of C, i.e., $n' = |Y'| \leq |Y| = n$.

- The designated relative distance of C' is at least as good as in C, i.e.,

$$\delta' = 1 - \frac{\deg(G')}{|Y'|} \geq \delta.$$

For the proof we will show:

Claim 4.17.

$$G \geq \mathrm{Con}_{F/F'}(G'').$$

With that we can prove the three assertions above about C':

Dimension: Since $\mathcal{L}_{F'}(G'') \subseteq \mathcal{L}_F(\mathrm{Con}_{F/F'}(G''))$, it follows that

$$\dim(\mathcal{L}_{F'}(G'')) \leq \dim(\mathcal{L}_F(\mathrm{Con}_{F/F'}(G''))).$$

Thus, by Claim 4.17,

$$\dim(\mathcal{L}_F(G)) \geq \dim(\mathcal{L}_F(\mathrm{Con}_{F/F'}(G''))) \geq \dim(\mathcal{L}_{F'}(G'')) \geq |B| = \dim(\mathcal{L}_F(G)),$$

and therefore

$$\dim(\mathcal{L}_{F'}(G'')) = \dim(\mathcal{L}_F(G)) = k.$$

The claim follows since $\dim(\mathcal{L}_{F'}(G')) = \dim(\mathcal{L}_{F'}(G''))$ by [129, Lemma I.4.6].

Length: $|Y| \geq |Y'|$, since every place of Y lies over exactly one place of Y', see [129, Proposition III.1.7].

Designated distance: By Claim 4.17, $\deg(G) \geq \deg(\mathrm{Con}_{F/F'}(G''))$. By [129, Cor III.1.13],

$$\deg(\mathrm{Con}_{F/F'}(G'')) = [F : F'] \cdot \deg(G'').$$

Since $\deg(G'') = \deg(G')$ it follows that

$$\deg(G) \geq [F : F'] \cdot \deg(G').$$

Also, since every place P' can split to at most $[F : F']$ places in F/K we have

$$|Y| \leq |Y'| \cdot [F : F'].$$

Altogether,

$$\delta' = 1 - \frac{\deg(G')}{|Y'|} \geq 1 - \frac{\deg(G)}{[F : F'] \cdot |Y'|} \geq 1 - \frac{\deg(G)}{|Y|} = \delta.$$

■

We are left with proving Claim 4.17.

Proof of Claim 4.17: By definition, $B \subseteq \mathcal{L}(G'')$, and therefore

$$F' = \text{Closure}(\mathcal{L}(G)) = \text{Closure}(B) \subseteq \text{Closure}(\mathcal{L}(G'')),$$

and $\text{Closure}(\mathcal{L}(G'')) = F'$.

Also, for any $P|P'$ (where $P' \in \mathcal{P}_{F'}$ and $P \in \mathcal{P}_F$) and for any i,

$$e(P|P') \cdot v_{P'}(s_i) = v_P(s_i) \geq -v_P(G),$$

where the last inequality is simply because $s_i \in \mathcal{L}(G)$. Therefore

$$v_{P'}(G'') = \max\{-v_{P'}(s_i)\} = \max\left\{-\frac{v_P(s_i)}{e(P|P')}\right\} \leq \frac{v_P(G)}{e(P|P')},$$

and the claim follows from the definition of the conorm. ■

4.4.2 The bound

We are now ready to prove Theorem 4.12.

Proof of Theorem 4.12: Assume a code is obtained by concatenating the AG code specified by the triplet $(F/\mathbb{F}_q, Y, G)$ with the Hadamard code. Let $\ell = \dim(G)$ and $d = \deg(G)$. The AG code $C(Y; G)$ is a $[|Y|, \ell]_q$ code, with designated distance $|Y| - d$. The concatenated code is therefore a

$$[n = |Y| \cdot q, k = \ell \log(q)]_2$$

code which is ϵ-balanced for

$$\epsilon = \frac{d}{|Y|}.$$

By Lemma 4.16 we can assume without loss of generality that $\text{Closure}(\mathcal{L}(G)) = F$.

There are two extreme cases that we handle separately:

Large base field: If the base field size q is too large the theorem is trivially true. Namely, If $q > \frac{k}{\epsilon^3}$ we are done because $n \geq q > \frac{k}{\epsilon^3}$. We can therefore assume without loss of generality that

$$\begin{aligned} q &\leq \frac{k}{\epsilon^3}, \text{ and,} \\ \log(q) &= O(\log(\frac{k}{\epsilon})). \end{aligned} \tag{4.3}$$

Few evaluation points: If the number of evaluation points is about the field size, we are essentially in the Reed-Solomon case and we are done. Specifically, if $|Y| \leq 4q$ then

$$4n = |Y| \cdot 4q \geq |Y|^2 = \frac{d^2}{\epsilon^2} \geq \frac{\ell^2}{\epsilon^2} = \frac{k^2}{\epsilon^2 \log^2(q)} = \Omega\left(\frac{k^2}{\epsilon^2 \log^2(\frac{k}{\epsilon})}\right),$$

and we are done. We can therefore assume without loss of generality that

$$|Y| > 4q.$$

This also implies that $\sqrt{q} < g$ since by Theorem 4.10, $|Y| \leq N(F) \leq q + 1 + 2g\sqrt{q}$. We can therefore conclude (again, by Theorem 4.10) that

$$N(F) \leq 4g\sqrt{q}. \tag{4.4}$$

Let $m = d \operatorname{div} \ell \geq 1$. By Theorem 4.15, $g \leq 2m^2\ell$, and by Eq (4.4), $N(F) \leq 8m^2\ell\sqrt{q}$. Thus,

$$n = |Y| \cdot q = \frac{N(F) \cdot |Y| \cdot q}{N(F)} \geq \frac{N(F) \cdot |Y| \cdot q}{8m^2\ell\sqrt{q}}.$$

Substituting $m \leq d/\ell$ and $d = \epsilon|Y|$ we see that

$$n \geq \frac{N(F) \cdot \sqrt{q} \cdot \ell}{8\epsilon^2|Y|}.$$

Substituting $\ell = \frac{k}{\log(q)}$ and using $N(F) > |Y|$,

$$n \geq \frac{\sqrt{q} \cdot k}{8\epsilon^2 \log(q)} = \Omega\left(\frac{\sqrt{q} \cdot k}{\epsilon^2 \log(\frac{k}{\epsilon})}\right),$$

where the last equality follows from Eq (4.3). To finish the argument notice that by Lemma 4.13, $N(F) \leq d(q+1)$. This implies $\frac{d}{\epsilon} = |Y| \leq d(q+1)$ and $\epsilon \geq \frac{1}{q+1}$, hence, $n = \Omega(\frac{k}{\epsilon^{2.5}\log(\frac{k}{\epsilon})})$. ∎

4.4.3 An open problem

Can one strengthen the above lower bound to match the parameters given in Section 4.3? More specifically we ask whether it is possible to get a concatenated code with $n = \tilde{O}(\frac{k}{\epsilon^{2.5}})$, where the $\tilde{O}$ notation is used to hide poly-logarithmic factors in q (or equivalently in k and ϵ). We know the following:

- $n = \tilde{O}(\frac{k^2}{\epsilon^2})$ implies $N(F) = \tilde{\Omega}(qm^2)$. (We already saw that in the proof of Theorem 4.12.)
- A similar calculation shows $n = \tilde{O}(\frac{k}{\epsilon^{2.5}})$ implies $N(F) = \tilde{\Omega}(q^{2/3}m^{5/3}\ell)$.

We also know two upper bounds on $N(F)$, namely:

- $N(F) = \tilde{O}(qm\ell)$ (follows from $N(F) \leq d(q+1)$), and,
- $N(F) = \tilde{O}(q^{1/2}m^2\ell)$ (since we can assume $N(F) \geq 2(q+1)$, as explained in the proof of Theorem 4.12).

Solving the constraints we get $m = \tilde{\Theta}(\sqrt{q})$. We thus see that the approach can lead to codes with $n = \tilde{O}(\frac{k}{\epsilon^{2.5}})$ if and only if the following question has a positive answer:

Open Problem 4.18. Given a prime power q and an integer $d = \tilde{O}(q)$ is there an algebraic function field $F/\mathbb{F}_q$ with $\tilde{\Omega}(q^2)$ places of degree one, and a divisor G such that $\deg(G) = d$ and $\dim(G) \geq \tilde{O}(\frac{d}{\sqrt{q}})$.

One might suspect such a high dimension, low-degree divisor does not exist. However, Theorem 4.15 and Lemma 4.13 are not strong enough to disprove it. We remark that the lower bound could be improved, if Lemma 4.13 could be strengthened to use the high-dimension of G, as suggested in Remark 4.14.

Chapter 5

A Hypercontractive Inequality for Matrix-Valued Functions

Our results in this chapter are:

- deriving a hypercontractive inequality for matrix-valued functions, using the inequality of Ball, Carlen, and Lieb [11];
- using it to show the function $F : \{0,1\}^n \times \binom{[n]}{k} \to \{0,1\}$ defined by $F(x,S) = \bigoplus_{i \in S} x_i$ is a strong extractor against quantum storage;
- using the result on F to obtain a bound on k-out-of-n random access codes;
- obtaining a direct product theorem for one-way quantum communication complexity (using the bound on k-out-of-n random access codes);
- giving an alternative proof of the fact that error-correcting codes that are locally decodable with 2 queries require length exponential in the length of the encoded string.

5.1 Introduction

5.1.1 A hypercontractive inequality for matrix-valued functions

Fourier analysis of real-valued functions on the Boolean cube has been widely used in the theory of computing. Applications include analyzing the influence of variables on Boolean functions [77], probabilistically-checkable proofs and associated hardness of approximation [63], analysis of threshold phenomena [78], noise stability [102, 111], voting schemes [113], learning under the uniform distribution [95, 97, 72, 103], communication complexity [117, 82, 53], etc.

One of the main technical tools in this area is a hypercontractive inequality that is sometimes called the *Bonami-Beckner inequality* [31, 14], though its history would also justify other names (see Lecture 16 of [112] for some background and history). For a fixed $\rho \in [0,1]$, consider the linear operator T_ρ on the space of all functions $f : \{0,1\}^n \to \mathbb{R}$ defined by

$$(T_\rho(f))(x) = \mathbb{E}_y[f(y)],$$

where the expectation is taken over y obtained from x by negating each bit independently with probability $(1-\rho)/2$. In other words, the value of $T_\rho(f)$ at a point x is obtained by averaging the values of f over a certain neighborhood of x. One important property of T_ρ for $\rho < 1$ is that it has a "smoothing" effect: any "high peaks" present in f are smoothed out in $T_\rho(f)$. The hypercontractive inequality formalizes this intuition. To state it precisely, define the p-norm of a function f by $\|f\|_p = (\frac{1}{2^n}\sum_x |f(x)|^p)^{1/p}$. It is not difficult to prove that the norm is nondecreasing with p. Also, the higher p is, the more sensitive the norm becomes to peaks in the function f. The hypercontractive inequality says that for certain $q > p$, the q-norm of $T_\rho(f)$ is upper bounded by the p-norm of f. This exactly captures the intuition that $T_\rho(f)$ is a smoothed version of f: even though we are considering a higher norm, the norm does not increase. More precisely, the hypercontractive inequality says that as long as $1 \le p \le q$ and $\rho \le \sqrt{(p-1)/(q-1)}$, we have

$$\|T_\rho(f)\|_q \le \|f\|_p. \tag{5.1}$$

The most interesting case for us is when $q = 2$, since in this case one can view the inequality as a statement about the Fourier coefficients of f, as we describe next. Let us first recall some basic definitions from Fourier analysis. For every $S \subseteq [n]$ (which by some abuse of notation we will also view as an n-bit string) and $x \in \{0,1\}^n$, define $\chi_S(x) = (-1)^{x \cdot S}$ to be the parity of the bits of x indexed by S. The *Fourier transform* of a function $f : \{0,1\}^n \to \mathbb{R}$ is the function $\widehat{f} : \{0,1\}^n \to \mathbb{R}$ defined by

$$\widehat{f}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x).$$

The values $\widehat{f}(S)$ are called the *Fourier coefficients* of f. The coefficient $\widehat{f}(S)$ may be viewed as measuring the correlation between f and the parity function χ_S. Since the functions χ_S form an orthonormal basis of the space of all functions from $\{0,1\}^n$ to $\mathbb{R}$, we can express f in terms of its Fourier coefficients as

$$f = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S. \tag{5.2}$$

Using the same reasoning we obtain Parseval's identity,

$$\|f\|_2 = \left(\sum_{S \subseteq [n]} \widehat{f}(S)^2 \right)^{1/2}.$$

The operator T_ρ has a particularly elegant description in terms of the Fourier coefficients. Namely, it simply multiplies each Fourier coefficient $\widehat{f}(S)$ by a factor of $\rho^{|S|}$:

$$T_\rho(f) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) \chi_S.$$

The higher $|S|$ is, the stronger the Fourier coefficient $\widehat{f}(S)$ is "attenuated" by T_ρ. Using Parseval's identity, we can now write the hypercontractive inequality (5.1) for the case $q = 2$ as follows. For every $p \in [1, 2]$,

$$\left(\sum_{S \subseteq [n]} (p-1)^{|S|} \widehat{f}(S)^2 \right)^{1/2} \leq \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} |f(x)|^p \right)^{1/p}. \tag{5.3}$$

This gives an upper bound on a weighted sum of the squared Fourier coefficients of f, where each coefficient is attenuated by a factor $(p-1)^{|S|}$. We are interested in generalizing this hypercontractive inequality to *matrix-valued* functions. Let $\mathcal{M}$ be the space of $d \times d$ complex matrices and suppose we have a function $f : \{0,1\}^n \to \mathcal{M}$. For example, a natural scenario where this arises is in quantum information theory, if we assign to every $x \in \{0,1\}^n$ some m-qubit *density matrix* $f(x)$ (so $d = 2^m$). We define the Fourier transform $\widehat{f}$ of a matrix-valued function f exactly as before:

$$\widehat{f}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) \chi_S(x).$$

The Fourier coefficients $\widehat{f}(S)$ are now also $d \times d$ matrices. An equivalent definition is by applying the standard Fourier transform to each i, j-entry separately: $\widehat{f}(S)_{ij} = \widehat{f(\cdot)_{ij}}(S)$. This extension of the Fourier transform to matrix-valued functions is quite natural, and has also been used in, e.g., [106, 48].

Our main tool, which we prove in Section 5.3, is an extension of the hypercontractive inequality to matrix-valued functions. For $M \in \mathcal{M}$ with singular values $\sigma_1, \ldots, \sigma_d$, we define its (normalized Schatten) p-norm as $\|M\|_p = (\frac{1}{d} \sum_{i=1}^d \sigma_i^p)^{1/p}$.

Theorem 5.1. *For every $f : \{0,1\}^n \to \mathcal{M}$ and $1 \leq p \leq 2$,*

$$\left(\sum_{S \subseteq [n]} (p-1)^{|S|} \left\| \widehat{f}(S) \right\|_p^2 \right)^{1/2} \leq \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} \|f(x)\|_p^p \right)^{1/p}.$$

This is the analogue of Eq. (5.3) for matrix-valued functions, with p-norms replacing absolute values. The case $n = 1$ can be seen as a geometrical statement that extends the familiar parallelogram law in Euclidean geometry and is closely related to the notion of uniform convexity. This case was first proven for certain values of p by Tomczak-Jaegermann [134] and then in full generality by Ball, Carlen, and Lieb [11]. Among its applications are the work of Carlen and Lieb on fermion fields [36], and the more recent work of Lee and Naor on metric embeddings [92].

To the best of our knowledge, the general case $n \geq 1$ has not appeared before.[1] Its proof is not difficult, and follows by induction on n, similar to the proof of the usual hypercontractive inequality.[2] Although one might justly regard Theorem 5.1 as a "standard" corollary of the result by Ball, Carlen, and Lieb, such "tensorized inequalities" tend to be extremely useful (see, e.g., [30, 58]) and we believe that the matrix-valued hypercontractive inequality will have more applications in the future.

5.1.2 Application: k-out-of-n random access codes

Our main application of Theorem 5.1 is for the following information-theoretic problem. Recall that a k-out-of-n quantum random access code is a way to encode an n-bit string x into m qubits, in such a way that for any set $S \subseteq [n]$ of k indices, the k-bit substring x_S can be recovered with probability at least p by making an appropriate measurement on the encoding. We are allowed to use probabilistic encodings here, so the encoding need not be a function mapping x to a fixed quantum pure state.

We are interested in the tradeoff between the length m of the quantum random access code, and the success probability p. Clearly, if $m \geq n$ then we can just use the identity encoding to obtain $p = 1$. If $m < n$ then by Holevo's theorem [67] our encoding will be "lossy", and p will be less than 1. The case $k = 1$ was first studied by Ambainis et al. [8], who showed that if p is bounded away from 1/2, then $m = \Omega(n/\log n)$. Nayak [105] subsequently strengthened this bound to $m \geq (1 - H(p))n$, where $H(\cdot)$ is the binary entropy function. This bound is optimal up to an

[1] A different generalization of the Bonami-Beckner inequality was given by Borell [32]. His generalization, however, is an easy corollary of the Bonami-Beckner inequality and is therefore relatively weak (although it does apply to any Banach space, and not just to the space of matrices with the Schatten p-norm).

[2] We remark that Carlen and Lieb's proof in [36] also uses induction and has some superficial resemblance to the proof given here. Their induction, however, is on the *dimension* of the matrices (or more precisely, the number of fermions), and moreover leads to an entirely different inequality.

additive $\log n$ term both for classical and quantum encodings. The intuition of Nayak's proof is that, for average i, the encoding only contains $m/n < 1$ bits of information about the bit x_i, which limits our ability to predict x_i given the encoding.

Now suppose that $k > 1$, and m is much smaller than n. Clearly, for predicting one specific bit x_i, with i uniformly chosen, Nayak's result applies, and we will have a success probability that is bounded away from 1. But intuitively this should apply to each of the k bits that we need to predict. Moreover, these k success probabilities should not be very correlated, so we expect an overall success probability that is exponentially small in k. Nayak's proof does not generalize to the case $k \gg 1$ (or at least, we do not know how to do it). The reason it fails is the following. Suppose we probabilistically encode $x \in \{0,1\}^n$ as follows: with probability 1/4 our encoding is x itself, and with probability 3/4 our encoding is the empty string. Then the average length of the output (and hence the entropy or amount of information in the encoding) is only $n/4$ bits, or 1/4 bit for an average x_i. Yet from this encoding one can predict *all of* x with success probability $1/4$! Hence, if we want to prove our intuition, we should make use of the fact that the encoding is always confined to a 2^m-dimensional space (a property which the above example lacks). Arguments based on von Neumann entropy, such as the one of [105], do not seem capable of capturing this condition (however, a *min-entropy* argument recently enabled König and Renner to prove a closely related but incomparable result, see below). The new hypercontractive inequality offers an alternative approach—in fact the only alternative approach to entropy-based methods that we are aware of in quantum information. Applying the inequality to the matrix-valued function that gives the encoding implies $p \leq 2^{-\Omega(k)}$ if $m \ll n$. More precisely:

Theorem 5.2. *For any $\eta > 2\ln 2$ there exists a constant C_η such that if n/k is large enough then for any k-out-of-n quantum random access code on m qubits, the success probability satisfies*

$$p \leq C_\eta \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\eta m}{n}}\right)^k.$$

In particular, the success probability is exponentially small in k if $m/n < 1/(2\ln 2) \approx 0.721$. Notice that for very small m/n the bound on p gets close to 2^{-k}, which is what one gets by guessing the k-bit answer randomly. We also obtain bounds if k is close to n, but these are a bit harder to state. Luckily, in all our applications we are free to choose a small enough m. We note that in contrast to Nayak's approach, our proof does not use the strong subadditivity of von Neumann entropy.

We mention that much after the publication of our results in [18, 19] our bound on random access codes was improved by De and Vidick [41].

The classical case. We now give a few comments regarding the special case of classical (probabilistic) m-bit encodings. First, in this case the encodings are represented by diagonal matrices. For such matrices, the base case $n = 1$ of Theorem 5.1 can be derived directly from the Bonami-Beckner inequality, without requiring the full strength of the Ball-Carlen-Lieb inequality (see [11] for details). Alternatively, one can derive Theorem 5.2 in the classical case directly from the Bonami-Beckner inequality by conditioning on a fixed m-bit string of the encoding (this step is already impossible in the quantum case) and then analyzing the resulting distribution on $\{0,1\}^n$. This proof is very similar to the one we give in Section 5.4 (and in fact slightly less elegant due to the conditioning step) and we therefore omit the details.

Interestingly, in the classical case there is a simpler argument that avoids Bonami-Beckner altogether. This argument was used in [140] and was communicated to us by the authors of that paper. We briefly sketch it here. Suppose we have a classical (possibly randomized) m-bit encoding that allows to recover any k-bit set with probability at least p using a (possibly randomized) decoder. By Yao's minimax principle, there is a way to fix the randomness in both the encoding and decoding procedures, such that the probability of succeeding in recovering all k bits of a randomly chosen k-set from an encoding of a uniformly random $x \in \{0,1\}^n$ is at least p. So now we have deterministic encoding and decoding, but there is still randomness in the input x. Call an x "good" if the probability of the decoding procedure being successful on a random k-tuple is at least $p/2$ (given the m-bit encoding of that x). By Markov's inequality, at least a $p/2$-fraction of the inputs x are good. Now consider the following experiment. Given the encoding of a uniform x, we take $\ell = 100n/k$ uniformly and independently chosen k-sets and apply the decoding procedure to all of them. We then output an n-bit string with the "union" of all the answers we received (if we received multiple contradictory answers for the same bit, we can put either answer there), and random bits for the positions that are not in the union. With probability $p/2$, x is good. Conditioned on this, with probability at least $(p/2)^\ell$ all our decodings are correct. Moreover, except with probability $2^{-\Omega(n)}$, the union of our ℓ k-sets is of size at least $0.9n$. The probability of guessing the remaining $n/10$ bits right is $2^{-n/10}$. Therefore the probability of successfully recovering all of x is at least $(p/2) \cdot ((p/2)^\ell - 2^{-\Omega(n)}) \cdot 2^{-n/10}$. A simple counting argument shows that this is impossible unless $p \leq 2^{-\Omega(k)}$ or m is close to n. This argument does not work for quantum encodings, of course, because these cannot just be reused (a quantum measurement changes the state).

The König-Renner result. Independently but subsequent to our work (which first appeared on the arxiv preprint server in May 2007), König and Renner [86] used sophisticated quantum information theoretic arguments to show a result with a similar flavor to ours. Each of the results is tuned for different scenarios. In particular, the results are incomparable, and our applications to direct product theorems do not follow from their result, nor do their applications follow from our result. We briefly

describe their result and explain the distinction between the two.

Let $X = X_1, \ldots, X_n$ be classical random variables, not necessarily uniformly distributed or even independent. Suppose that each $X_i \in \{0,1\}^b$. Suppose further that the "smooth min-entropy of X relative to a quantum state ρ" is at least some number h (see [86] for the precise definitions, which are quite technical). If we randomly pick r distinct indices $i_1, \ldots, i_r$, then intuitively the smooth min-entropy of $X' = X_{i_1}, \ldots, X_{i_r}$ relative to ρ should not be much smaller than hr/n. König and Renner show that if b is larger than n/r then this is indeed the case, except with probability exponentially small in r. Note that they are picking b-bit blocks $X_{i_1}, \ldots, X_{i_r}$ instead of individual bits, but this can also be viewed as picking (not quite uniformly) $k = rb$ bits from a string of nb bits.

On the one hand, the constants in their bounds are essentially optimal, while ours are a factor $2 \ln 2$ off from what we expect they should be. Also, while they need very few assumptions on the random variables $X_1, \ldots, X_n$ and on the quantum encoding, we assume the random variables are uniformly distributed bits, and our quantum encoding is confined to a 2^m-dimensional space. We can in fact slightly relax both the assumption on the input and the encoding, but do not discuss these relaxations since they are of less interest to us. Finally, their result still works if the indices $i_1, \ldots, i_r$ are not sampled uniformly, but are sampled in some randomness-efficient way. This allows them to obtain efficient key-agreement schemes in a cryptographic model where the adversary can only store a bounded number of quantum bits.

On the other hand, our result works even if only a small number of bits is sampled, while theirs only kicks in when the number of bits being sampled ($k = rb$) is at least the square-root of the total number of bits nb. This is not very explicit in their paper, but can be seen by observing that the parameter $\kappa = n/(rb)$ on page 8 and in Corollary 6.19 needs to be at most a constant (whence the assumption that b is larger than n/r). So the total number of bits is $nb = O(rb^2) = O(r^2b^2) = O(k^2)$. Since we are interested in small as well as large k, this limitation of their approach is significant. A final distinction between the results is in the length of the proof. While the information-theoretic intuition in their paper is clear and well-explained, the details get to be quite technical, resulting in a proof which is significantly longer than ours.

5.1.3 Application: Direct product theorem for one-way quantum communication complexity

Our result for k-out-of-n random access codes has the flavor of a direct product theorem: the success probability of performing a certain task on k instances (i.e., k distinct indices) goes down exponentially with k. In Section 5.5, we use this to prove a new strong direct product theorem for one-way communication complexity.

Consider the 2-party Disjointness function: Alice receives input $x \in \{0,1\}^n$, Bob receives input $y \in \{0,1\}^n$, and they want to determine whether the sets represented by their inputs are disjoint, i.e. whether $x_i y_i = 0$ for all $i \in [n]$. They want to do this while communicating as few qubits as possible (allowing some small error probability, say 1/3). We can either consider one-way protocols, where Alice sends one message to Bob who then computes the output; or two-way protocols, which are interactive. The quantum communication complexity of Disjointness is fairly well understood: it is $\Theta(n)$ qubits for one-way protocols [34], and $\Theta(\sqrt{n})$ qubits for two-way protocols [33, 70, 1, 118].

Now consider the case of k independent instances: Alice receives inputs $x_1, \ldots, x_k$ (each of n bits), Bob receives $y_1, \ldots, y_k$, and their goal is to compute all k bits $\mathrm{DISJ}_n(x_1, y_1), \ldots, \mathrm{DISJ}_n(x_k, y_k)$. Klauck et al. [83] proved an optimal direct product theorem for *two-way* quantum communication: every protocol that communicates fewer than $\alpha k\sqrt{n}$ qubits (for some small constant $\alpha > 0$) will have a success probability that is exponentially small in k. Surprisingly, prior to our work no strong direct product theorem was known for the usually simpler case of *one-way* communication. In Section 5.5 we derive such a theorem from our k-out-of-n random access code lower bound: if $\eta > 2\ln 2$, then every one-way quantum protocol that sends fewer than kn/η qubits will have success probability at most $2^{-\Omega(k)}$.

These results can straightforwardly be generalized to get a bound for all functions in terms of their *VC-dimension*. If f has VC-dimension d, then any one-way quantum protocol for computing k independent copies of f that sends kd/η qubits, has success probability $2^{-\Omega(k)}$. For simplicity, Section 5.5 only presents the case of Disjointness. Finally, by the work of Beame et al. [13], such direct product theorems imply lower bounds on *3-party* protocols where the first party sends only one message. We elaborate on this in Section 5.6.

5.1.4 Application: Locally decodable codes

A locally decodable error-correcting code (LDC) $C : \{0,1\}^n \to \{0,1\}^N$ encodes n bits into N bits, in such a way that each encoded bit can be recovered from a noisy codeword by a randomized decoder that queries only a small number q of bit-positions in that codeword. Such codes have applications in a variety of different complexity-theoretic and cryptographic settings; see for instance Trevisan's survey and the references therein [136]. The main theoretical issue in LDCs is the tradeoff between q and N. The best known constructions of LDCs with constant q have a length N that is sub-exponential in n but still superpolynomial [146, 47]. On the other hand, the only superpolynomial *lower* bound known for general LDCs is the tight bound $N = 2^{\Omega(n)}$ for $q = 2$ due to Kerenidis and de Wolf [81] (generalizing an earlier exponential lower bound for *linear* codes by [54]). Rather surprisingly, the proof of [81] relied heavily on techniques from quantum information theory: despite being a result purely about classical codes and classical decoders, the quantum

perspective was crucial for their proof. In particular, they show that the two queries of a classical decoder can be replaced by one quantum query, then they turn this quantum query into a random access code for the encoded string x, and finally invoke Nayak's lower bound for quantum random access codes.

In Section 5.7 we reprove an exponential lower bound on N for the case $q = 2$ without invoking any quantum information theory: we just use classical reductions, matrix analysis, and the hypercontractive inequality for matrix-valued functions. Hence it is a classical (non-quantum) proof as asked for by Trevisan [136, Open question 3 in Section 3.6]. It should be noted that this new proof is still quite close in spirit (though not terminology) to the quantum proof of [81]. This is not too surprising given the fact that the proof of [81] uses Nayak's lower bound on random access codes, generalizations of which follow from the hypercontractive inequality. We discuss the similarities and differences between the two proofs in Section 5.7.

We feel the merit of this new approach is not so much in giving a partly new proof of the known lower bound on 2-query LDCs, but in its potential application to codes with more than 2 queries. Recently Efremenko [47] building on Yekhanin [146] constructed 3-query LDCs with $N = 2^{2^{\sqrt{\log n \log\log n}}}$. For $q = 3$, the best known lower bounds on N are slightly less than n^2 [80, 81, 145]. Despite considerable effort, this gap still looms large. Our hope is that our approach can be generalized to 3 or more queries. Specifically, what we would need is a generalization of tensors of rank 2 (i.e., matrices) to tensors of rank q; an appropriate tensor norm; and a generalization of the hypercontractive inequality from matrix-valued to tensor-valued functions.

5.2 Preliminaries

Norms: Recall that we define the p-norm of a d-dimensional vector v by

$$\|v\|_p = \left(\frac{1}{d}\sum_{i=1}^{d} |v_i|^p\right)^{1/p}.$$

We extend this to matrices by defining the (normalized Schatten) p-norm of a matrix $A \in \mathbb{C}^{d\times d}$ as

$$\|A\|_p = \left(\frac{1}{d}\mathrm{Tr}|A|^p\right)^{1/p}.$$

This is equivalent to the p-norm of the vector of singular values of A. For diagonal matrices this definition coincides with the one for vectors. For convenience we defined all norms to be under the normalized counting measure, even though for matrices this is nonstandard. The advantage of the normalized norm is that it is nondecreasing with p. We also define the *trace norm* $\|A\|_{\mathrm{tr}}$ of a matrix

A as the sum of its singular values, hence we have $\|A\|_{\mathrm{tr}} = d\|A\|_1$ for any $d \times d$ matrix A.

Quantum states: An m-qubit *pure state* is a superposition $|\phi\rangle = \sum_{z\in\{0,1\}^m} \alpha_z |z\rangle$ over all classical m-bit states. The α_z's are complex numbers called *amplitudes*, and $\sum_z |\alpha_z|^2 = 1$. Hence a pure state $|\phi\rangle$ is a unit vector in $\mathbb{C}^{2^m}$. Its complex conjugate (a row vector with entries conjugated) is denoted $\langle\phi|$. The inner product between $|\phi\rangle = \sum_z \alpha_z|z\rangle$ and $|\psi\rangle = \sum_z \beta_z|z\rangle$ is the dot product $\langle\phi| \cdot |\psi\rangle = \langle\phi|\psi\rangle = \sum_z \alpha_z^*\beta_z$. An m-qubit *mixed state* (or *density matrix*) $\rho = \sum_i p_i|\phi_i\rangle\langle\phi_i|$ corresponds to a probability distribution over m-qubit pure states, where $|\phi_i\rangle$ is given with probability p_i. The eigenvalues $\lambda_1, \ldots, \lambda_d$ of ρ are non-negative reals that sum to 1, so they form a probability distribution. If ρ is pure then one eigenvalue is 1 while all others are 0. Hence for any $p \geq 1$, the maximal p-norm is achieved by pure states:

$$\|\rho\|_p^p = \frac{1}{d}\sum_{i=1}^{d} \lambda_i^p \leq \frac{1}{d}\sum_{i=1}^{d} \lambda_i = \frac{1}{d}. \tag{5.4}$$

A k-outcome *positive operator-valued measurement* (POVM) is given by k positive semidefinite operators $E_1, \ldots, E_k$ with the property that $\sum_{i=1}^k E_i = I$. When this POVM is applied to a mixed state ρ, the probability of the ith outcome is given by the trace $\mathrm{Tr}(E_i\rho)$. The following well known fact gives the close relationship between trace distance and distinguishability of density matrices:

Fact 5.3. The best possible measurement to distinguish two density matrices ρ_0 and ρ_1 has bias $\frac{1}{2}\|\rho_0 - \rho_1\|_{\mathrm{tr}}$.

Here "bias" is defined as twice the success probability, minus 1. We refer to Nielsen and Chuang [108] for more details.

5.3 The hypercontractive inequality for matrix-valued functions

Here we prove Theorem 5.1. The proof relies on the following powerful inequality by Ball et al. [11] (they state this inequality for the usual unnormalized Schatten p-norm, but both statements are clearly equivalent).

Lemma 5.4. *([11, Theorem 1]) For any matrices A, B and any $1 \leq p \leq 2$, it holds that*

$$\left(\left\|\frac{A+B}{2}\right\|_p^2 + (p-1)\left\|\frac{A-B}{2}\right\|_p^2\right)^{1/2} \leq \left(\frac{\|A\|_p^p + \|B\|_p^p}{2}\right)^{1/p}.$$

Theorem 5.1. For any $f : \{0,1\}^n \to \mathcal{M}$ and for any $1 \le p \le 2$,

$$\left(\sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{f}(S)\right\|_p^2\right)^{1/2} \le \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|f(x)\|_p^p\right)^{1/p}.$$

Proof: By induction. The case $n = 1$ follows from Lemma 5.4 by setting $A = f(0)$ and $B = f(1)$, and noting that $(A + B)/2$ and $(A - B)/2$ are exactly the Fourier coefficients $\widehat{f}(0)$ and $\widehat{f}(1)$.

We now assume the lemma holds for n and prove it for $n + 1$. Let $f : \{0,1\}^{n+1} \to \mathcal{M}$ be some matrix-valued function. For $i \in \{0,1\}$, let $g_i = f|_{x_{n+1}=i}$ be the function obtained by fixing the last input bit of f to i. We apply the induction hypothesis on g_0 and g_1 to obtain

$$\left(\sum_{S\subseteq[n]} (p-1)^{|S|} \|\widehat{g_0}(S)\|_p^2\right)^{1/2} \le \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|g_0(x)\|_p^p\right)^{1/p}$$
$$\left(\sum_{S\subseteq[n]} (p-1)^{|S|} \|\widehat{g_1}(S)\|_p^2\right)^{1/2} \le \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|g_1(x)\|_p^p\right)^{1/p}.$$

Take the L_p average of these two inequalities: raise each to the pth power, average them and take the pth root. We get

$$\left(\frac{1}{2}\sum_{i\in\{0,1\}} \left(\sum_{S\subseteq[n]} (p-1)^{|S|} \|\widehat{g_i}(S)\|_p^2\right)^{p/2}\right)^{1/p} \le \left(\frac{1}{2^{n+1}}\sum_{x\in\{0,1\}^n} \left(\|g_0(x)\|_p^p + \|g_1(x)\|_p^p\right)\right)^{1/p} \tag{5.5}$$
$$= \left(\frac{1}{2^{n+1}}\sum_{x\in\{0,1\}^{n+1}} \|f(x)\|_p^p\right)^{1/p}.$$

The right-hand side is the expression we wish to lower bound. To bound the left-hand side, we need the following inequality (to get a sense of why this holds, consider the case where $q_1 = 1$ and $q_2 = \infty$).

Lemma 5.5 (Minkowski's inequality, [60, Theorem 26])**.** *For any $r_1 \times r_2$ matrix whose rows are given by $u_1, \ldots, u_{r_1}$ and whose columns are given by $v_1, \ldots, v_{r_2}$, and any $1 \le q_1 < q_2 \le \infty$,*

$$\left\|\left(\|v_1\|_{q_2}, \ldots, \|v_{r_2}\|_{q_2}\right)\right\|_{q_1} \ge \left\|\left(\|u_1\|_{q_1}, \ldots, \|u_{r_1}\|_{q_1}\right)\right\|_{q_2},$$

i.e., the value obtained by taking the q_2-norm of each column and then taking the q_1-norm of the

results, is at least that obtained by first taking the q_1-norm of each row and then taking the q_2-norm of the results.

Consider now the $2^n \times 2$ matrix whose entries are given by

$$c_{S,i} = 2^{n/2} \left\| (p-1)^{|S|/2} \widehat{g_i}(S) \right\|_p$$

where $i \in \{0,1\}$ and $S \subseteq [n]$. The left-hand side of (5.5) is then

$$\begin{aligned} \left(\frac{1}{2} \sum_{i \in \{0,1\}} \left(\frac{1}{2^n} \sum_{S \subseteq [n]} c_{S,i}^2 \right)^{p/2} \right)^{1/p} &\geq \left(\frac{1}{2^n} \sum_{S \subseteq [n]} \left(\frac{1}{2} \sum_{i \in \{0,1\}} c_{S,i}^p \right)^{2/p} \right)^{1/2} \\ &= \left(\sum_{S \subseteq [n]} (p-1)^{|S|} \left(\frac{\|\widehat{g_0}(S)\|_p^p + \|\widehat{g_1}(S)\|_p^p}{2} \right)^{2/p} \right)^{1/2}, \end{aligned}$$

where the inequality follows from Lemma 5.5 with $q_1 = p$, $q_2 = 2$. We now apply Lemma 5.4 to deduce that the above is lower bounded by

$$\left(\sum_{S \subseteq [n]} (p-1)^{|S|} \left(\left\| \frac{\widehat{g_0}(S) + \widehat{g_1}(S)}{2} \right\|_p^2 + (p-1) \left\| \frac{\widehat{g_0}(S) - \widehat{g_1}(S)}{2} \right\|_p^2 \right) \right)^{1/2} = \left(\sum_{S \subseteq [n+1]} (p-1)^{|S|} \left\| \widehat{f}(S) \right\|_p^2 \right)^{1/2}$$

where we used $\widehat{f}(S) = \frac{1}{2}(\widehat{g_0}(S) + \widehat{g_1}(S))$ and $\widehat{f}(S \cup \{n+1\}) = \frac{1}{2}(\widehat{g_0}(S) - \widehat{g_1}(S))$ for any $S \subseteq [n]$. ■

5.4 Bounds for k-out-of-n quantum random access codes

In this section we prove Theorem 5.2. Recall that a k-out-of-n random access code allows us to encode n bits into m qubits, such that we can recover any k-bit substring with probability at least p. We now define this notion formally. In fact, we consider a somewhat weaker notion where we only measure the success probability for a random k subset, and a random input $x \in \{0,1\}^n$. Since we only prove impossibility results, this clearly makes our results stronger.

Definition 5.6. A k-out-of-n quantum random access code on m qubits with success probability p (for short (k,n,m,p)-QRAC), is a map

$$f : \{0,1\}^n \to \mathbb{C}^{2^m \times 2^m}$$

that assigns an m-qubit density matrix $f(x)$ to every $x \in \{0,1\}^n$, and a quantum measurement $\{M_{S,z}\}_{z\in\{0,1\}^k}$ to every set $S \in \binom{[n]}{k}$, with the property that

$$\mathbb{E}_{x,S}[\mathrm{Tr}(M_{S,x_S} \cdot f(x))] \geq p,$$

where the expectation is taken over a uniform choice of $x \in \{0,1\}^n$ and $S \in \binom{[n]}{k}$, and x_S denotes the k-bit substring of x specified by S.

In order to prove Theorem 5.2, we introduce another notion of QRAC, which we call *XOR-QRAC*. Here, the goal is to predict the XOR of the k bits indexed by S (as opposed to guessing all the bits in S). Since one can always predict a bit with probability $\frac{1}{2}$, it is convenient to define the *bias* of the prediction as $\varepsilon = 2p - 1$ where p is the probability of a correct prediction. Hence a bias of 1 means that the prediction is always correct, whereas a bias of -1 means that it is always wrong. The advantage of dealing with an XOR-QRAC is that it is easy to express the best achievable prediction bias without any need to introduce measurements. Namely, if $f : \{0,1\}^n \to \mathbb{C}^{2^m \times 2^m}$ is the encoding function, then the best achievable bias in predicting the XOR of the bits in S (over a random $\{0,1\}^n$) is exactly half the trace distance between the average of $f(x)$ over all x with the XOR of the bits in S being 0 and the average of $f(x)$ over all x with the XOR of the bits in S being 1. Using our notation for Fourier coefficients, this can be written simply as $\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}$.

Definition 5.7. A k-out-of-n XOR quantum random access code on m qubits with bias ε (for short (k,n,m,ε)-XOR-QRAC), is a map

$$f : \{0,1\}^n \to \mathbb{C}^{2^m \times 2^m}$$

that assigns an m-qubit density matrix $f(x)$ to every $x \in \{0,1\}^n$ and has the property that

$$\mathbb{E}_{S\sim\binom{[n]}{k}}\left[\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}\right] \geq \varepsilon.$$

Our new hypercontractive inequality allows us to easily derive the following key lemma:

Lemma 5.8. *Let $f : \{0,1\}^n \to \mathbb{C}^{2^m \times 2^m}$ be any mapping from n-bit strings to m-qubit density matrices. Then for any $0 \leq \delta \leq 1$, we have*

$$\sum_{S\subseteq[n]} \delta^{|S|}\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}^2 \leq 2^{2\delta m}.$$

Proof: Let $p = 1 + \delta$. On one hand, by Theorem 5.1 and Eq. (5.4) we have

$$\sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{f}(S)\right\|_p^2 \leq \left(\frac{1}{2^n} \sum_{x\in\{0,1\}^n} \left\|f(x)\right\|_p^p\right)^{2/p} \leq \left(\frac{1}{2^n} \cdot 2^n \cdot \frac{1}{2^m}\right)^{2/p} = 2^{-2m/p}.$$

On the other hand, by norm monotonicity we have

$$\sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{f}(S)\right\|_p^2 \quad \geq \quad \sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{f}(S)\right\|_1^2 = 2^{-2m} \sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{f}(S)\right\|_{\mathrm{tr}}^2.$$

By rearranging we have

$$\sum_{S\subseteq[n]} (p-1)^{|S|} \left\|\widehat{f}(S)\right\|_{\mathrm{tr}}^2 \leq 2^{2m(1-1/p)} \leq 2^{2m(p-1)},$$

as required. ∎

The following is our main theorem regarding XOR-QRAC. In particular it shows that if $k = o(n)$ and $m/n < 1/(2\ln 2) \approx 0.721$, then the bias will be exponentially small in k.

Theorem 5.9. *For any (k, n, m, ε)-XOR-QRAC we have the following bound on the bias*

$$\varepsilon \leq \left(\frac{(2e\ln 2)m}{k}\right)^{k/2} \binom{n}{k}^{-1/2}.$$

In particular, for any $\eta > 2\ln 2$ there exists a constant C_η such that if n/k is large enough then for any (k, n, m, ε)-XOR-QRAC,

$$\varepsilon \leq C_\eta \left(\frac{\eta m}{n}\right)^{k/2}.$$

Proof: Apply Lemma 5.8 with $\delta = \frac{k}{(2\ln 2)m}$ and only take the sum on S with $|S| = k$. This gives

$$\mathbb{E}_{S\sim\binom{[n]}{k}} \left[\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}^2\right] \leq 2^{2\delta m} \delta^{-k} \binom{n}{k}^{-1} = \left(\frac{(2e\ln 2)m}{k}\right)^k \binom{n}{k}^{-1}.$$

The first bound on ε now follows by convexity (Jensen's inequality). To derive the second bound, approximate $\binom{n}{k}$ using Stirling's approximation $n! = \Theta(\sqrt{n}(n/e)^n)$:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \Theta\left(\sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{k}\right)^k \left(1 + \frac{k}{n-k}\right)^{n-k}\right).$$

Now use the fact that for large enough n/k we have $(1 + k/(n-k))^{(n-k)/k} > (2e\ln 2)/\eta$, and

notice that the factor $\sqrt{n/k(n-k)} \geq \sqrt{1/k}$ can be absorbed by this approximation. ■

We now derive Theorem 5.2 from Theorem 5.9.

Proof of Theorem 5.2: Consider a (k, n, m, p)-QRAC, given by encoding function f and measurements $\{M_{T,z}\}_{z\in\{0,1\}^k}$ for all $T \in \binom{[n]}{k}$. Define $p_T(w) = \mathbb{E}_x \left[\Pr[z \oplus x_T = w]\right]$ as the distribution on the "error vector" $w \in \{0,1\}^k$ of the measurement outcome $z \in \{0,1\}^k$ when applying $\{M_{T,z}\}$. By definition, we have that $p \leq \mathbb{E}_T[p_T(0^k)]$.

Now suppose we want to predict the parity of the bits of some set S of size at most k. We can do this as follows: uniformly pick a set $T \in \binom{[n]}{k}$ that contains S, measure $f(x)$ with $\{M_{T,z}\}$, and output the parity of the bits corresponding to S in the measurement outcome z. Note that our output is correct if and only if the bits corresponding to S in the error vector w have even parity. Hence the bias of our output is

$$\beta_S = \mathbb{E}_{T:T\supseteq S}\left[\sum_{w\in\{0,1\}^k} p_T(w)\chi_S(w)\right] = 2^k\,\mathbb{E}_{T:T\supseteq S}\left[\widehat{p_T}(S)\right].$$

(We slightly abuse notation here by viewing S both as a subset of T and as a subset of $[k]$ obtained by identifying T with $[k]$.) Notice that β_S can be upper bounded by the best-achievable bias $\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}$.

Consider the distribution $\mathcal{S}$ on sets S defined as follows: first pick j from the binomial distribution $B(k, 1/2)$ and then uniformly pick $S \in \binom{[n]}{j}$. Notice that the distribution on pairs (S, T) obtained by first choosing $S \sim \mathcal{S}$ and then choosing a uniform $T \supseteq S$ from $\binom{[n]}{k}$ is identical to the one obtained by first choosing uniformly T from $\binom{[n]}{k}$ and then choosing a uniform $S \subseteq T$. This allows us to show that the average bias β_S over $S \sim \mathcal{S}$ is at least p, as follows:

$$\begin{aligned}
\mathbb{E}_{S\sim\mathcal{S}}\left[\beta_S\right] &= 2^k\mathbb{E}_{S\sim\mathcal{S},T\supseteq S}\left[\widehat{p_T}(S)\right] \\
&= 2^k\mathbb{E}_{T\sim\binom{[n]}{k},S\subseteq T}\left[\widehat{p_T}(S)\right] \\
&= \mathbb{E}_{T\sim\binom{[n]}{k}}\left[\sum_{S\subseteq T}\widehat{p_T}(S)\right] \\
&= \mathbb{E}_{T\sim\binom{[n]}{k}}\left[p_T(0^k)\right] \geq p,
\end{aligned}$$

where the last equality follows from Eq. (5.2). On the other hand, using Theorem 5.9 we obtain

$$\begin{aligned}
\mathbb{E}_{S\sim\mathcal{S}}\left[\beta_S\right] &\leq \mathbb{E}_{S\sim\mathcal{S}}\left[\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}\right] \\
&= \frac{1}{2^k}\sum_{j=0}^{k}\binom{k}{j}\mathbb{E}_{S\sim\binom{[n]}{j}}\left[\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}\right]
\end{aligned}$$

$$\leq \frac{1}{2^k} \sum_{j=0}^{k} \binom{k}{j} C_\eta \left(\frac{\eta m}{n}\right)^{j/2}$$
$$= C_\eta \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\eta m}{n}}\right)^k,$$

where the last equality uses the binomial theorem. Combining the two inequalities completes the proof. ∎

5.5 Direct product theorem for one-way quantum communication

The setting of communication complexity is by now well-known, so we will not give formal definitions of protocols etc., referring to [89, 144] instead. Consider the n-bit Disjointness problem in 2-party communication complexity. Alice receives n-bit string x and Bob receives n-bit string y. They interpret these strings as subsets of $[n]$ and want to decide whether their sets are disjoint. In other words, $\mathrm{DISJ}_n(x, y) = 1$ if and only if $x \cap y = \emptyset$. Let $\mathrm{DISJ}_n^{(k)}$ denote k independent instances of this problem. That is, Alice's input is a k-tuple $x_1, \ldots, x_k$ of n-bit strings, Bob's input is a k-tuple $y_1, \ldots, y_k$, and they should output all k bits: $\mathrm{DISJ}_n^{(k)}(x_1, \ldots, x_k, y_1, \ldots, y_k) = \mathrm{DISJ}_n(x_1, y_1), \ldots, \mathrm{DISJ}_n(x_k, y_k)$. The trivial protocol where Alice sends all her inputs to Bob has success probability 1 and communication complexity kn. We want to show that if the total one-way communication is much smaller than kn qubits, then the success probability is exponentially small in k. We will do that by deriving a random access code from the protocol's message.

Lemma 5.10. *Let $\ell \leq k$. If there is a c-qubit one-way communication protocol for $\mathrm{DISJ}_n^{(k)}$ with success probability σ, then there is an ℓ-out-of-kn quantum random access code of c qubits with success probability $p \geq \sigma\,(1 - \ell/k)^\ell$.*

Proof: Consider the following one-way communication setting: Alice has a kn-bit string x, and Bob has ℓ distinct indices $i_1, \ldots, i_\ell \in [kn]$ chosen uniformly from $\binom{[kn]}{\ell}$ and wants to learn the corresponding bits of x.

In order to do this, Alice sends the c-qubit message corresponding to input x in the $\mathrm{DISJ}_n^{(k)}$ protocol. We view x as consisting of k disjoint blocks of n bits each. The probability (over the choice of Bob's input) that $i_1, \ldots, i_\ell \in [kn]$ are in ℓ different blocks is

$$\prod_{i=0}^{\ell-1} \frac{kn - in}{kn - i} \geq \left(\frac{kn - \ell n}{kn}\right)^\ell = \left(1 - \frac{\ell}{k}\right)^\ell.$$

If this is the case, Bob chooses his Disjointness inputs $y_1, \ldots, y_k$ as follows. If index i_j is somewhere in block $b \in [k]$, then he chooses y_b to be the string having a 1 at the position where i_j is, and

0s elsewhere. Note that the correct output for the b-th instance of Disjointness with inputs x and $y_1, \ldots, y_k$ is exactly $1 - x_{i_j}$. Now Bob completes the protocol and gets a k-bit output for the k-fold Disjointness problem. A correct output tells him the ℓ bits he wants to know (he can just disregard the outcomes of the other $k - \ell$ instances). Overall the success probability is at least $\sigma(1 - \ell/k)^\ell$. Therefore, the random access code that encodes x by Alice's message proves the lemma. ∎

Combining the previous lemma with our earlier upper bound on p for ℓ-out-of-kn quantum random access codes (Theorem 5.2), we obtain the following upper bound on the success probability σ of c-qubit one-way communication protocols for $\mathrm{DISJ}_n^{(k)}$. For every $\eta > 2\ln 2$ there exists a constant C_η such that:

$$\sigma \leq 2p(1 - \ell/k)^{-\ell} \leq 2C_\eta \left(\left(\frac{1}{2} + \frac{1}{2} \sqrt{\frac{\eta(c + O(k + \log(kn)))}{kn}} \right) \left(\frac{k}{k - \ell} \right) \right)^\ell .$$

Choosing ℓ a sufficiently small constant fraction of k (depending on η), we obtain a strong direct product theorem for one-way communication:

Theorem 5.11. *For any $\eta > 2\ln 2$ the following holds: for any large enough n and any k, every one-way quantum protocol for $\mathrm{DISJ}_n^{(k)}$ that communicates $c \leq kn/\eta$ qubits, has success probability $\sigma \leq 2^{-\Omega(k)}$ (where the constant in the $\Omega(\cdot)$ depends on η).*

The above strong direct product theorem (SDPT) bounds the success probability for protocols that are required to compute *all* k instances correctly. We call this a *zero-error* SDPT. What if we settle for a weaker notion of "success", namely getting a $(1 - \varepsilon)$-fraction of the k instances right, for some small $\varepsilon > 0$? An *ε-error SDPT* is a theorem to the effect that even in this case the success probability is exponentially small. An ε-error SDPT follows from a zero-error SDPT as follows. Run an ε-error protocol with success probability p ("success" now means getting $1 - \varepsilon$ of the k instances right), guess up to εk positions and change them. With probability at least p, the number of errors of the ε-error protocol is at most εk, and with probability at least $1/\sum_{i=0}^{\varepsilon k} \binom{k}{i}$ we now have corrected all those errors. Since $\sum_{i=0}^{\varepsilon k} \binom{k}{i} \leq 2^{kH(\varepsilon)}$ (see, e.g., [75, Corollary 23.6]), we have a protocol that computes all instances correctly with success probability $\sigma \geq p2^{-kH(\varepsilon)}$. If we have a zero-error SDPT that bounds $\sigma \leq 2^{-\gamma k}$ for some $\gamma > H(\varepsilon)$, then it follows that p must be exponentially small as well: $p \leq 2^{-(\gamma - H(\varepsilon))k}$. Hence Theorem 5.11 implies:

Theorem 5.12. *For any $\eta > 2\ln 2$ there exists an $\varepsilon > 0$ such that the following holds: for every one-way quantum protocol for $\mathrm{DISJ}_n^{(k)}$ that communicates $c \leq kn/\eta$ qubits, its probability to compute at least a $(1 - \varepsilon)$-fraction of the k instances correctly is at most $2^{-\Omega(k)}$.*

5.6 3-party NOF communication complexity of Disjointness

Some of the most interesting open problems in communication complexity arise in the "number on the forehead" (NOF) model of multiparty communication complexity, with applications ranging from bounds on proof systems to circuit lower bounds. Here, there are ℓ players and ℓ inputs $x_1, \ldots, x_\ell$. The players want to compute some function $f(x_1, \ldots, x_\ell)$. Each player j sees all inputs *except* x_j. In the ℓ-party version of the Disjointness problem, the ℓ players want to figure out whether there is an index $i \in [n]$ where all ℓ input strings have a 1. For any constant ℓ, the best known upper bound is linear in n [56].

While the case $\ell = 2$ has been well-understood for a long time, the first polynomial lower bounds for $\ell \geq 3$ were shown only very recently. Lee and Shraibman [94], and independently Chattopadhyay and Ada [37], showed lower bounds of the form $\Omega(n^{1/(\ell+1)})$ on the classical communication complexity for constant ℓ. This becomes $\Omega(n^{1/4})$ for $\ell = 3$ players.

Stronger lower bounds can be shown if we limit the kind of interaction allowed between the players. Viola and Wigderson [140] showed a lower bound of $\Omega(n^{1/(\ell-1)})$ for the *one-way* complexity of ℓ-player Disjointness, for any constant ℓ. In particular, this gives $\Omega(\sqrt{n})$ for $\ell = 3$.[3] An intermediate model was studied by Beame et al. [13], namely protocols where Charlie first sends a message to Bob, and then Alice and Bob are allowed two-way communication between each other to compute $\mathrm{DISJ}_n(x_1, x_2, x_3)$. This model is weaker than full interaction, but stronger than the one-way model. Beame et al. showed (using a direct product theorem) that any protocol of this form requires $\Omega(n^{1/3})$ bits of communication.[4]

Here we strengthen these two 3-player results to *quantum* communication complexity, while at the same time slightly simplifying the proofs. These results will follow easily from two direct product theorems: the one for two-way communication from [83], and the new one for one-way communication that we prove here. Lee, Schechtman, and Shraibman [93] have recently extended their $\Omega(n^{1/(\ell+1)})$ classical lower bound to ℓ-player quantum protocols. While that result holds for a stronger communication model than ours (arbitrary point-to-point quantum messages), their bound for $\ell = 3$ is weaker than ours ($\Omega(n^{1/4})$ vs $\Omega(n^{1/3})$).

5.6.1 Communication-type $C \to (B \leftrightarrow A)$

Consider 3-party Disjointness on inputs $x, y, z \in \{0,1\}^n$. Here Alice sees x and z, Bob sees y and z, and Charlie sees x and y. Their goal is to decide if there is an $i \in [n]$ such that $x_i = y_i = z_i = 1$.

Suppose we have a 3-party protocol P for Disjointness with the following "flow" of communication. Charlie sends a message of c_1 classical bits to Alice and Bob (or just to Bob, it doesn't really

[3] Actually, this bound for the case $\ell = 3$ was already known earlier; see [10].

[4] Their conference paper had an $\Omega(n^{1/3}/\log n)$ bound, but the journal version [13] managed to get rid of the $\log n$.

matter), who then exchange c_2 *qu*bits and compute Disjointness with bounded error probability. Our lower bound approach is similar to the one of Beame et al. [13], the main change being our use of stronger direct product theorems. Combining the (0-error) two-way quantum strong direct product theorem for Disjointness from [83] with the argument from the end of our Section 5.5, we have the following ε-error strong direct product theorem for k instances of 2-party Disjointness:

Theorem 5.13. *There exist constants $\varepsilon > 0$ and $\alpha > 0$ such that the following holds: for every two-way quantum protocol for $\mathrm{DISJ}_n^{(k)}$ that communicates at most $\alpha k\sqrt{n}$ qubits, its probability to compute at least an $(1-\varepsilon)$-fraction of the k instances correctly, is at most $2^{-\Omega(k)}$.*

Assume without loss of generality that the error probability of our initial 3-party protocol P is at most half the ε of Theorem 5.13. View the n-bit inputs of protocol P as consisting of t consecutive blocks of n/t bits each. We will restrict attention to inputs $z = z_1 \dots z_t$ where one z_i is all-1, and the other z_j are all-0. Note that for such a z, we have $\mathrm{DISJ}_n(x,y,z) = \mathrm{DISJ}_{n/t}(x_i, y_i)$. Fixing z thus reduces the 3-party Disjointness on (x,y,z) to 2-party Disjointness on a smaller instance (x_i, y_i). Since Charlie does not see input z, his c_1-bit message is independent of z. Now by going over all t possible z's, and running their 2-party protocol t times starting from Charlie's message, Alice and Bob obtain a protocol P' that computes t independent instances of 2-party Disjointness, namely on each of the t inputs $(x_1,y_1),\dots,(x_t,y_t)$. This P' uses at most tc_2 qubits of communication. For every x and y, it follows from linearity of expectation that the expected number of instances where P' errs, is at most $\varepsilon t/2$ (expectation taken over Charlie's message, and the t-fold Alice-Bob protocol). Hence by Markov's inequality, the probability that P' errs on more than εt instances, is at most 1/2. Then for every x, y there exists a c_1-bit message m_{xy} such that P', when given that message to start with, with probability at least 1/2 correctly computes $1-\varepsilon$ of all t instances.

Now replace Charlie's c_1-bit message by a uniformly random message m. Alice and Bob can just generate this by themselves using shared randomness. This gives a new 2-party protocol P''. For each x, y, with probability 2^{-c_1} we have $m = m_{xy}$, hence with probability at least $\frac{1}{2}2^{-c_1}$ the protocol P'' correctly computes $1-\varepsilon$ of all t instances of Disjointness on n/t bits each. Choosing $t = O(c_1)$ and invoking Theorem 5.13 gives a lower bound on the communication in P'': $tc_2 = \Omega(t\sqrt{n/t})$. Hence $c_2 = \Omega(\sqrt{n/c_1})$. The overall communication of the original 3-party protocol P is

$$c_1 + c_2 = c_1 + \Omega(\sqrt{n/c_1}) = \Omega(n^{1/3})$$

(the minimizing value is $t = n^{1/3}$).

This generalizes the bound of Beame et al. [13] to the case where we allow Alice and Bob to send each other qubits. Note that this bound is tight for our restricted set of z's, since Alice and Bob know z and can compute the 2-party Disjointness on the relevant (x_i, y_i) in $O(\sqrt{n^{2/3}}) = O(n^{1/3})$

qubits of two-way communication without help from Charlie, using the optimal quantum protocol for 2-party Disjointness [1].

5.6.2 Communication-type $C \to B \to A$

Now consider an even more restricted type of communication: Charlie sends a classical message to Bob, then Bob sends a quantum message to Alice, and Alice computes the output. We can use a similar argument as before, dividing the inputs into $t = O(n^{1/2})$ equal-sized blocks instead of $O(n^{1/3})$ equal-sized blocks. If we now replace the two-way SDPT (Theorem 5.13) by the new one-way SDPT (Theorem 5.12), we obtain a lower bound of $\Omega(\sqrt{n})$ for 3-party bounded-error protocols for Disjointness of this restricted type.

Remark. If Charlie's message is quantum as well, then the same approach works, except we need to reduce the error of the protocol to $\ll 1/t$ at a multiplicative cost of $O(\log t) = O(\log n)$ to both c_1 and c_2 (Charlie's one quantum message needs to be reused t times). This worsens the two communication lower bounds to $\Omega(n^{1/3}/\log n)$ and $\Omega(\sqrt{n}/\log n)$ qubits, respectively.

5.7 Lower bounds on locally decodable codes

When analyzing locally decodable codes, it will be convenient to view bits as elements of $\{\pm 1\}$ instead of $\{0,1\}$. Formally, a locally decodable code is defined as follows.

Definition 5.14. $C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a (q, δ, ε)-*locally decodable code* (LDC) if there is a randomized decoding algorithm A such that

1. For all $x \in \{\pm 1\}^n$, $i \in [n]$, and $y \in \{\pm 1\}^N$ with Hamming distance $d(C(x), y) \leq \delta N$, we have $\Pr[A^y(i) = x_i] \geq 1/2 + \varepsilon$. Here $A^y(i)$ is the random variable that is A's output given input i and oracle y.

2. A makes at most q queries to y, non-adaptively.

In Section 5.8 we show that such a code implies the following: For each $i \in [n]$, there is a set M_i of at least $\delta\varepsilon N/q^2$ disjoint tuples, each of at most q elements from $[N]$, and a sign $a_{i,Q} \in \{\pm 1\}$ for each $Q \in M_i$, such that

$$\mathbb{E}_x[a_{i,Q} x_i \prod_{j \in Q} C(x)_j] \geq \frac{\varepsilon}{2^q},$$

where the expectation is uniformly over all $x \in \{\pm 1\}^n$. In other words, the parity of each of the tuples in M_i allows us to predict x_i with non-trivial bias (averaged over all x).

Kerenidis and de Wolf [81] used quantum information theory to show the lower bound $N = 2^{\Omega(\delta\varepsilon^2 n)}$ on the length of 2-query LDCs. Using the new hypercontractive inequality, we can prove a similar lower bound. Our dependence on ε and δ is slightly worse, but can probably be improved by a more careful analysis.

Theorem 5.15. *If $C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a $(2, \delta, \varepsilon)$-LDC, then $N = 2^{\Omega(\delta^2\varepsilon^4 n)}$.*

Proof: Define $f(x)$ as the $N \times N$ matrix whose (i, j)-entry is $C(x)_i C(x)_j$. Since $f(x)$ has rank 1 and its N^2 entries are all $+1$ or -1, its only non-zero singular value is N. Hence $\|f(x)\|_p^p = N^{p-1}$ for every x.

Consider the $N \times N$ matrices $\widehat{f}(\{i\})$ that are the Fourier transform of f at the singleton sets $\{i\}$:

$$\widehat{f}(\{i\}) = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x) x_i.$$

We want to lower bound $\left\|\widehat{f}(\{i\})\right\|_p$.

With the above notation, each set M_i consists of at least $\delta\varepsilon N/4$ disjoint pairs of indices.[5] For simplicity assume $M_i = \{(1, 2), (3, 4), (5, 6), \ldots\}$. The 2×2 submatrix in the upper left corner of $f(x)$ is

$$\begin{pmatrix} 1 & C(x)_1 C(x)_2 \\ C(x)_1 C(x)_2 & 1 \end{pmatrix}.$$

Since $(1, 2) \in M_i$, we have $\mathbb{E}_x[C(x)_1 C(x)_2 x_i a_{i,(1,2)}] \in [\varepsilon/4, 1]$. Hence the 2×2 submatrix in the upper left corner of $\widehat{f}(\{i\})$ is

$$\begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}$$

for some a with $|a| \in [\varepsilon/4, 1]$. The same is true for each of the first $\delta\varepsilon N/4$ 2×2 diagonal blocks of $\widehat{f}(\{i\})$ (each such 2×2 block corresponds to a pair in M_i). Let P be the $N \times N$ permutation matrix that swaps rows 1 and 2, swaps rows 3 and 4, etc. Then the first $\delta\varepsilon N/2$ diagonal entries of $F_i = P\widehat{f}(\{i\})$ all have absolute value in $[\varepsilon/4, 1]$.

The $\|\cdot\|_p$ norm is *unitarily invariant*: $\|UAV\|_p = \|A\|_p$ for every matrix A and unitaries U, V. Note the following lemma, which is a special case of [28, Eq. (IV.52) on p. 97]. We include its proof for completeness.

Lemma 5.16. *Let $\|\cdot\|$ be a unitarily-invariant norm on the set of $d \times d$ complex matrices. If A is a matrix and $\mathrm{diag}(A)$ is the matrix obtained from A by setting its off-diagonal entries to 0, then $\|\mathrm{diag}(A)\| \leq \|A\|$.*

[5] Actually some of the elements of M_i may be singletons. Dealing with this is a technicality that we will ignore here in order to simplify the presentation.

Proof: We will step-by-step set the off-diagonal entries of A to 0, without increasing its norm. We start with the off-diagonal entries in the dth row and column. Let D_d be the diagonal matrix that has $D_{d,d} = -1$ and $D_{i,i} = 1$ for $i < d$. Note that $D_d A D_d$ is the same as A, except that the off-diagonal entries of the dth row and column are multiplied by -1. Hence $A' = (A + D_d A D_d)/2$ is the matrix obtained from A by setting those entries to 0 (this doesn't affect the diagonal). Since D_d is unitary and every norm satisfies the triangle inequality, we have

$$\|A'\| = \|(A + D_d A D_d)/2\| \leq \frac{1}{2}(\|A\| + \|D_d A D_d\|) = \|A\|.$$

In the second step, we can set the off-diagonal entries in the $(d-1)$st row and column of A' to 0, using the diagonal matrix D_{d-1} which has a -1 only on its $(d-1)$st position. Continuing in this manner, we set all off-diagonal entries of A to zero without affecting its diagonal, and without increasing its norm. ∎

Using this lemma, we obtain

$$\left\|\widehat{f}(\{i\})\right\|_p = \|F_i\|_p \geq \|\mathrm{diag}(F_i)\|_p \geq \left(\frac{1}{N}(\delta\varepsilon N/2)(\varepsilon/4)^p\right)^{1/p} = (\delta\varepsilon/2)^{1/p}\varepsilon/4.$$

Using the hypercontractive inequality (Theorem 5.1), we have for any $p \in [1,2]$

$$n(p-1)(\delta\varepsilon/2)^{2/p}(\varepsilon/4)^2 \leq \sum_{i=1}^{n}(p-1)\left\|\widehat{f}(\{i\})\right\|_p^2 \leq \left(\frac{1}{2^n}\sum_x \|f(x)\|_p^p\right)^{2/p} = N^{2(p-1)/p}.$$

Choosing $p = 1 + 1/\log N$ and rearranging implies the result. ∎

Let us elaborate on the similarities and differences between this proof and the quantum proof of [81]. On the one hand, the present proof makes no use of quantum information theory. It only uses the well known version of LDCs mentioned after Definition 5.14, some basic matrix analysis, and our hypercontractive inequality for matrix-valued functions. On the other hand, the proof may still be viewed as a translation of the original quantum proof to a different language. The quantum proof defines, for each x, a $\log(N)$-qubit state $|\phi(x)\rangle$ which is the uniform superposition over the N indices of the codeword $C(x)$. It then proceeds in two steps: (1) by viewing the elements of M_i as 2-dimensional projectors in a quantum measurement of $|\phi(x)\rangle$, we can with good probability recover the parity $C(x)_j C(x)_k$ for a random element (j,k) of the matching M_i. Since that parity has non-trivial correlation with x_i, the states $|\phi(x)\rangle$ form a quantum random access code: they allow us to recover each x_i with decent probability (averaged over all x); (2) the quantum proof then invokes Nayak's linear lower bound on the number of qubits of a random access code to conclude

$\log N = \Omega(n)$. The present proof mimics this quantum proof quite closely: the matrix $f(x)$ is, up to normalization, the density matrix corresponding to the state $|\phi(x)\rangle$; the fact that matrix $\widehat{f}(\{i\})$ has fairly high norm corresponds to the fact that the parity produced by the quantum measurement has fairly good correlation with x_i; and finally, our invocation of Theorem 5.1 replaces (but is not identical to) the linear lower bound on quantum random access codes. We feel that by avoiding any explicit use of quantum information theory, the new proof holds some promise for potential extensions to codes with $q \geq 3$.

5.8 Massaging locally decodable codes to a special form

In this section we justify the special decoding-format of LDCs claimed after Definition 5.14. First, it will be convenient to switch to the notion of a *smooth code*, introduced by Katz and Trevisan [80].

Definition 5.17. $C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a (q, c, ε)-*smooth code* if there is a randomized decoding algorithm A such that

1. A makes at most q queries, non-adaptively.
2. For all $x \in \{\pm 1\}^n$ and $i \in [n]$ we have $\Pr[A^{C(x)}(i) = x_i] \geq 1/2 + \varepsilon$.
3. For all $x \in \{\pm 1\}^n$, $i \in [n]$, and $j \in [N]$, the probability that on input i algorithm A queries index j is at most c/N.

Note that smooth codes only require good decoding on codewords $C(x)$, not on y that are close to $C(x)$. Katz and Trevisan [80, Theorem 1] established the following connection:

Theorem 5.18 ([80])**.** *A (q, δ, ε)-LDC is a $(q, q/\delta, \varepsilon)$-smooth code.*

Proof: Let C be a (q, δ, ε)-LDC and A be its q-query decoder. For each $i \in [n]$, let $p_i(j)$ be the probability that on input i, algorithm A queries index j. Let $H_i = \{j \mid p_i(j) > q/(\delta N)\}$. Then $|H_i| \leq \delta N$, because A makes no more than q queries. Let B be the decoder that simulates A, except that on input i it does not make queries to $j \in H_i$, but instead acts as if those bits of its oracle are 0. Then B does not query any j with probability greater than $q/(\delta N)$. Also, B's behavior on input i and oracle $C(x)$ is the same as A's behavior on input i and the oracle y that is obtained by setting the H_i-indices of $C(x)$ to 0. Since y has distance at most $|H_i| \leq \delta N$ from $C(x)$, we have $\Pr[B^{C(x)}(i) = x_i] = \Pr[A^y(i) = x_i] \geq 1/2 + \varepsilon$. ■

A converse to Theorem 5.18 also holds: a (q, c, ε)-smooth code is a $(q, \delta, \varepsilon - c\delta)$-LDC, because the probability that the decoder queries one of δN corrupted positions is at most $(c/N)(\delta N) = c\delta$. Hence LDCs and smooth codes are essentially equivalent, for appropriate choices of the parameters.

Theorem 5.19 ([80]). *Suppose $C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a (q, c, ε)-smooth code. Then for every $i \in [n]$, there exists a set M_i, consisting of at least $\varepsilon N/(cq)$ disjoint sets of at most q elements of $[N]$ each, such that for every $Q \in M_i$ there exists a function $f_Q : \{\pm 1\}^{|Q|} \to \{\pm 1\}$ with the property*

$$\mathbb{E}_x[f_Q(C(x)_Q)x_i] \geq \varepsilon.$$

Here $C(x)_Q$ is the restriction of $C(x)$ to the bits in Q, and the expectation is uniform over all $x \in \{\pm 1\}^n$.

Proof: Fix some $i \in [n]$. Without loss of generality we assume that to decode x_i, the decoder picks some set $Q \subseteq [N]$ (of at most q indices) with probability $p(Q)$, queries those bits, and then outputs a *random variable* (not yet a function) $f_Q(C(x)_Q) \in \{\pm 1\}$ that depends on the query-answers. Call such a Q "good" if

$$\Pr_x[f_Q(C(x)_Q) = x_i] \geq 1/2 + \varepsilon/2.$$

Equivalently, Q is good if

$$\mathbb{E}_x[f_Q(C(x)_Q)x_i] \geq \varepsilon.$$

Now consider the hypergraph $H_i = (V, E_i)$ with vertex-set $V = [N]$ and edge-set E_i consisting of all good sets Q. The probability that the decoder queries some $Q \in E_i$ is $p(E_i) := \sum_{Q \in E_i} p(Q)$. If it queries some $Q \in E_i$ then $\mathbb{E}_x[f_Q(C(x)_Q)x_i] \leq 1$, and if it queries some $Q \notin E_i$ then $\mathbb{E}_x[f_Q(C(x)_Q)x_i] < \varepsilon$. Since the overall probability of outputting x_i is at least $1/2 + \varepsilon$ for every x, we have

$$2\varepsilon \leq \mathbb{E}_{x,Q}[f_Q(C(x)_Q)x_i] < p(E_i) \cdot 1 + (1 - p(E_i))\varepsilon = \varepsilon + p(E_i)(1 - \varepsilon),$$

hence

$$p(E_i) > \varepsilon/(1 - \varepsilon) \geq \varepsilon.$$

Since C is smooth, for every $j \in [N]$ we have

$$\sum_{Q \in E_i : j \in Q} p(Q) \leq \sum_{Q : j \in Q} p(Q) = \Pr[A \text{ queries } j] \leq \frac{c}{N}.$$

A *matching* of H_i is a set of disjoint $Q \in E_i$. Let M_i be a matching in H_i of maximal size. Our goal is to show $|M_i| \geq \varepsilon N/(cq)$. Define $T = \cup_{Q \in M_i} Q$. This set T has at most $q|M_i|$ elements, and intersects each $Q \in E_i$ (otherwise M_i would not be maximal). We now lower bound the size of M_i

as follows:

$$\varepsilon < p(E_i) = \sum_{Q:Q\in E_i} p(Q) \overset{(*)}{\leq} \sum_{j\in T} \sum_{Q\in E_i : j\in Q} p(Q) \leq \frac{c|T|}{N} \leq \frac{cq|M_i|}{N},$$

where $(*)$ holds because each $Q \in E_i$ is counted exactly once on the left and at least once on the right (since T intersects each $Q \in E_i$). Hence $|M_i| \geq \varepsilon N/(cq)$. It remains to turn the random variables $f_Q(C(x)_Q)$ into fixed values in $\{\pm 1\}$; it is easy to see that this can always be done without reducing the correlation $\mathbb{E}_x[f_Q(C(x)_Q)x_i]$. ■

The previous theorem establishes that the decoder can just pick a uniformly random element $Q \in M_i$, and then continue as the original decoder would on those queries, at the expense of reducing the average success probability by a factor 2. In principle, the decoder could output any function of the $|Q|$ queried bits that it wants. We now show (along the lines of [81, Lemma 2]) that we can restrict attention to parities (or their negations), at the expense of decreasing the average success probability by another factor of 2^q.

Theorem 5.20. *Suppose $C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a (q, c, ε)-smooth code. Then for every $i \in [n]$, there exists a set M_i, consisting of at least $\varepsilon N/(cq)$ disjoint sets of at most q elements of $[N]$ each, such that for every $Q \in M_i$ there exists an $a_{i,Q} \in \{\pm 1\}$ with the property that*

$$\mathbb{E}_x[a_{i,Q} x_i \prod_{j\in Q} C(x)_j] \geq \frac{\varepsilon}{2^q}.$$

Proof: Fix $i \in [n]$ and take the set M_i produced by Theorem 5.19. For every $Q \in M_i$ we have

$$\mathbb{E}_x[f_Q(C(x)_Q)x_i] \geq \varepsilon.$$

We would like to turn the functions $f_Q : \{\pm 1\}^{|Q|} \to \{\pm 1\}$ into parity functions. Consider the Fourier transform of f_Q: for $S \subseteq [|Q|]$ and $z \in \{\pm 1\}^{|Q|}$, define parity function $\chi_S(z) = \prod_{j\in S} z_j$ and Fourier coefficient $\widehat{f_Q}(S) = \frac{1}{2^{|Q|}} \sum_z f_Q(z)\chi_S(z)$. Then we can write

$$f_Q = \sum_S \widehat{f_Q}(S)\chi_S.$$

Using that $\widehat{f_Q}(S) \in [-1, 1]$ for all S, we have

$$\varepsilon \leq \mathbb{E}_x[f_Q(C(x)_Q)x_i] = \sum_S \widehat{f_Q}(S)\mathbb{E}_x[x_i\chi_S(C(x)_Q)] \leq \sum_S \left|\mathbb{E}_x[x_i\chi_S(C(x)_Q)]\right|.$$

Since the right-hand side is the sum of $2^{|Q|}$ terms, there exists an S with $|\mathbb{E}_x[x_i\chi_S(C(x)_Q)]| \geq \frac{\varepsilon}{2^{|Q|}}$.

Defining $a_{i,Q} = \text{sign}(\mathbb{E}_x[x_i\chi_S(C(x)_Q)]) \in \{\pm 1\}$, we have

$$\mathbb{E}_x[a_{i,Q}x_i\prod_{j\in S}C(x)_j] = |\mathbb{E}_x[x_i\chi_S(C(x)_Q)]| \geq \frac{\varepsilon}{2^{|Q|}} \geq \frac{\varepsilon}{2^q}.$$

The theorem follows by replacing each Q in M_i by the set S just obtained from it. ∎

Combining Theorems 5.18 and 5.20 gives the decoding-format claimed after Definition 5.14.

Chapter 6

Better short-seed quantum-proof extractors

In this chapter we give the following constructions of extractors:

Theorem 1.1. For any $\beta < \frac{1}{2}$ and $\epsilon \geq 2^{-k^\beta}$, there exists an explicit quantum-proof $(n, k, (1-\beta)k, \epsilon)$ strong extractor for flat sources $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ with seed length $t = O(\log n + \log \epsilon^{-1})$ and output length $m = \Omega(k)$.

Theorem 1.2. For any $\beta < \frac{1}{2}$ and $\epsilon \geq 2^{-k^\beta}$, there exists an explicit $(n, k, \beta k, \epsilon)$ strong extractor against quantum storage, $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log \epsilon^{-1})$ and output length $m = \Omega(k)$.

Theorem 1.3. For any $\beta < \frac{1}{2}$ and $\epsilon \geq 2^{-n^\beta}$, there exists an explicit quantum-proof $(n, (1-\beta)n, \epsilon)$ strong extractor $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log \epsilon^{-1})$ and output length $m = \Omega(n)$.

6.1 Preliminaries

Distributions. A distribution D on Λ is a function $D : \Lambda \to [0,1]$ such that $\sum_{a \in \Lambda} D(a) = 1$. We denote by $x \sim D$ sampling x according to the distribution D. Let U_t denote the uniform distribution over $\{0,1\}^t$. We measure the distance between two distributions with the variational distance $|D_1 - D_2|_1 = \frac{1}{2} \sum_{a \in \Lambda} |D_1(a) - D_2(a)|$. The distributions D_1 and D_2 are ϵ-*close* if $|D_1 - D_2|_1 \leq \epsilon$.

The min-entropy of D is denoted by $H_\infty(D)$ and is defined to be

$$H_\infty(D) = \min_{a : D(a) > 0} -\log(D(a)).$$

If $H_\infty(D) \geq k$ then for all a in the support of D it holds that $D(a) \leq 2^{-k}$. A distribution is *flat* if it is uniformly distributed over its support. Every distribution D with $H_\infty(D) \geq k$ can be expressed as a convex combination $\sum \alpha_i D_i$ of flat distributions $\{D_i\}$, each with min-entropy at least k. We sometimes abuse notation and identify a set X with the flat distribution that is uniform over X.

If X is a distribution over Λ_1 and $f : \Lambda_1 \to \Lambda_2$ then $f(X)$ denotes the distribution over Λ_2 obtained by sampling x from X and outputting $f(x)$. If X_1 and X_2 are *correlated* distributions we denote their joint distribution by $X_1 \circ X_2$. If X_1 and X_2 are *independent* distributions we replace $\circ$ by $\times$ and write $X_1 \times X_2$.

Mixed states. A pure state is a vector in some Hilbert space. A general quantum system is in a *mixed state* — a probability distribution over pure states. Let $\{p_i, |\phi_i\rangle\}$ denote the mixed state where the pure state $|\phi_i\rangle$ occurs with probability p_i. The behavior of the mixed state $\{p_i, |\phi_i\rangle\}$ is completely characterized by its *density matrix* $\rho = \sum_i p_i |\phi_i\rangle\langle\phi_i|$, in the sense that two mixed states with the same density matrix have the same behavior under any physical operation. Notice that a density matrix over a Hilbert space $\mathcal{H}$ belongs to $\mathrm{Hom}(\mathcal{H}, \mathcal{H})$, the set of linear transformation from $\mathcal{H}$ to $\mathcal{H}$. Density matrices are positive semi-definite operators and have trace 1.

The *trace distance* between density matrices ρ_1 and ρ_2 is $\|\rho_1 - \rho_2\|_{\mathrm{tr}} = \frac{1}{2}\sum_i |\lambda_i|$, where $\{\lambda_i\}$ are the eigenvalues of $\rho_1 - \rho_2$. The trace distance coincides with the variational distance when ρ_1 and ρ_2 are classical states (ρ is classical if it is diagonal in the standard basis). Similarly to probability distributions, the density matrices ρ_1 and ρ_2 are *ϵ-close* if the trace distance between them is at most ϵ.

A positive operator valued measure (POVM) is the most general formulation of a measurement in quantum computation. A POVM on a Hilbert space $\mathcal{H}$ is a collection $\{F_i\}$ of positive semi-definite operators $F_i : \mathrm{Hom}(\mathcal{H}, \mathcal{H}) \to \mathrm{Hom}(\mathcal{H}, \mathcal{H})$ that sum-up to the identity transformation, i.e., $F_i \succeq 0$ and $\sum F_i = I$. Applying a POVM $F = \{F_i\}$ on a density matrix ρ results in the distribution $F(\rho)$ that outputs i with probability $\mathrm{Tr}(F_i\rho)$.

A Boolean measurement $\{F, I - F\}$ *ϵ-distinguishes* ρ_1 and ρ_2 if $|\mathrm{Tr}(F\rho_1) - \mathrm{Tr}(F\rho_2)| \geq \epsilon$.

We shall need the following facts regarding the trace distance.

Fact 6.1. If $\|\rho_1 - \rho_2\|_{\mathrm{tr}} = \delta$ then there exists a Boolean measurement that δ-distinguishes ρ_1 and ρ_2.

Fact 6.2. If ρ_1 and ρ_2 are ϵ-close then $\mathcal{E}(\rho_1)$ and $\mathcal{E}(\rho_2)$ are ϵ-close, for any physically realizable transformation $\mathcal{E}$.

6.1.1 Min-entropy

To define the notion of quantum-proof extractors we first need the notion of quantum encoding of classical states.

Definition 6.3. Let X be a distribution over some set Λ.

- An *encoding* of X is a collection $\rho = \{\rho(x)\}_{x\in\Lambda}$ of density matrices.
- An encoding ρ is a *b-storage encoding* if $\rho(x)$ is a mixed state over b qubits, for all $x \in \Lambda$.
- An encoding is *classical* if $\rho(x)$ is classical for all x.

The average encoding is denoted by $\bar{\rho}_X = \mathbb{E}_{x\sim X}[\rho(x)]$.

Next we define the notion of conditional min-entropy. The conditional min-entropy of X given $\rho(X)$ measures the average success probability of predicting x given the encoding $\rho(x)$. Formally,

Definition 6.4. The *conditional min-entropy of X given an encoding ρ* is

$$H_\infty(X;\rho) = -\log \sup_F \mathop{\mathbb{E}}_{x\sim X}[\mathrm{Tr}(F_x\rho(x))],$$

where the supremum ranges over all POVMs $F = \{F_x\}_{x\in\Lambda}$.

We remark that there exists another definition of conditional min-entropy in the quantum setting, which is more algebraic in flavor. However, the two definitions are equivalent, as shown in [87].

Proposition 6.5 ([88, Proposition 2])**.** *If ρ is a b-storage encoding of X then $H_\infty(X;\rho) \geq H_\infty(X) - b$.*

We shall need the following standard lemmas regarding min-entropy that can be found, e.g., in [121]. The first lemma says that cutting ℓ bits from a source cannot reduce the min-entropy by more than ℓ.

Lemma 6.6. *Let $X = X_1 \circ X_2$ be a distribution over bit strings and ρ be an encoding such that $H_\infty(X;\rho) \geq k$, and suppose that X_2 is of length ℓ. Let ρ' be the encoding of X_1 defined by $\rho'(x_1) = \mathbb{E}_{x\sim(X|X_1=x_1)}[\rho(x)]$. Then, $H_\infty(X_1;\rho') \geq k - l$.*

Proof: Given any predictor P' which predicts X_1 from ρ', we can construct a predictor P for X (from ρ) as follows: P simply runs P' to obtain a prediction for the prefix x_1, and then appends it with a randomly chosen string from $\{0,1\}^\ell$. Then,

$$\mathrm{Pr}_{x_1\circ x_2\sim X}[P(\rho(x_1\circ x_2)) = x_1\circ x_2] \quad = \quad \mathrm{Pr}_{x_1\circ x_2\sim X}[P'(\rho(x_1\circ x_2)) = x_1]\cdot 2^{-\ell}$$

$$= \Pr_{x_1 \sim X_1}[P'(\rho'(x_1)) = x_1] \cdot 2^{-\ell}.$$

Thus, if $H_\infty(X_1; \rho') < k - l$ then there would have been a predictor which predicts X with probability greater than 2^{-k} and this cannot be the case since $H_\infty(X; \rho) \geq k$. ∎

The second lemma says that if a source has high min-entropy, then revealing a short prefix (with high probability) does not change much the min-entropy. The lemma is a generalization of a well known classical lemma.

Lemma 6.7. *Let $X = X_1 \circ X_2$ be a distribution and ρ be an encoding such that $H_\infty(X; \rho) \geq k$, and suppose that X_1 is of length ℓ. For a prefix x_1, let ρ_{x_1} be the encoding of X_2 defined by $\rho_{x_1}(x_2) = \rho(x_1 \circ x_2)$. Call a prefix x_1* bad *if $H_\infty(X_2 \mid X_1 = x_1; \rho_{x_1}) \leq r$ and denote by B the set of bad prefixes. Then,*

$$\Pr[X_1 \in B] \leq 2^\ell \cdot 2^r \cdot 2^{-k}.$$

Proof: Let the prefix $x_1' \in B$ be the one with the largest probability mass. Then, $\Pr[X_1 = x_1'] \geq \Pr[X_1 \in B] \cdot 2^{-\ell}$. For any $z \in B$, let A_z denote the optimal predictor that predicts X_2 from ρ_z, conditioned on $X_1 = z$. By the definition of min-entropy, for any $z \in B$,

$$\mathop{\mathbb{E}}_{x_2 \sim (X_2 | X_1 = z)} \Pr[A_z(\rho_z(x_2)) = x_2] \geq 2^{-r}.$$

In particular this holds for $z = x_1'$.

Now, define a predictor P for X from ρ by

$$P(\rho(x)) = x_1' \circ A_{x_1'}(\rho(x)),$$

that is, P simply "guesses" that the prefix is x_1' and then applies the optimal predictor $A_{x_1'}$. The average success probability of P is

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \sim X}\Big[\Pr[P(\rho(x)) = x]\Big] &= \mathop{\mathbb{E}}_{x_1 \sim X_1}\left[\mathop{\mathbb{E}}_{x_2 \sim (X_2 | X_1 = x_1)}\Big[\delta_{x_1, x_1'} \cdot \Pr[A_{x_1'}(\rho_{x_1'}(x_2)) = x_2]\Big]\right] \\
&= \Pr[X_1 = x_1'] \cdot \mathop{\mathbb{E}}_{x_2 \sim (X_2 | X_1 = x_1')}\Big[\Pr[A_{x_1'}(\rho_{x_1'}(x_2)) = x_2]\Big] \\
&\geq \Pr[X_1 \in B] \cdot 2^{-\ell} \cdot 2^{-r}
\end{aligned}$$

On the other hand, since $H_\infty(X; \rho) \geq k$, the average success probability of P is at most 2^{-k}. Altogether, $\Pr[X_1 \in B] \leq 2^\ell \cdot 2^r \cdot 2^{-k}$. ∎

6.1.2 Quantum-proof extractors

We now define the three different classes of extractors against quantum adversaries that we deal with in this chapter. We begin with the most general (and natural) definition:

Definition 6.8. A function $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a *quantum-proof* (n,k,ϵ) *strong extractor* if for every distribution X over $\{0,1\}^n$ and every encoding ρ such that $H_\infty(X;\rho) \geq k$,

$$\|U_t \circ E(X,U_t) \circ \rho(X) - U_{t+m} \times \bar{\rho}_X\|_{\mathrm{tr}} \leq \epsilon.$$

We use $\circ$ to denote correlated values. Thus, $U_t \circ E(X,U_t) \circ \rho(X)$ denotes the mixed state obtained by sampling $x{\sim}X, y{\sim}U_t$ and outputting $|y,E(x,y)\rangle\langle y,E(x,y)| \otimes \rho(x)$. Notice that all 3 registers are correlated. When a register is independent of the others we use $\times$ instead of $\circ$. Thus, $U_{t+m} \times \bar{\rho}_X$ denotes the mixed state obtained by sampling $x{\sim}X, w{\sim}U_{t+m}$ and outputting $|w\rangle\langle w| \otimes \rho(x)$.

Next we define quantum-proof extractors for *flat distributions*:

Definition 6.9. A function $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is a *quantum-proof* (n,f,k,ϵ) *strong extractor for flat distributions* if for every *flat* distribution X over $\{0,1\}^n$ with exactly f min-entropy and every encoding ρ of X with $H_\infty(X;\rho) \geq k$,

$$\|U_t \circ E(X,U_t) \circ \rho(X) - U_{t+m} \times \bar{\rho}_X\|_{\mathrm{tr}} \leq \epsilon.$$

We remark that in the classical setting every extractor for flat distributions is also an extractor for general distributions, since every distribution with min-entropy k can be expressed as a convex combination of flat distributions over 2^k elements.

Finally we define extractors against quantum storage:

Definition 6.10. A function $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ is an (n,k,b,ϵ) *strong extractor against quantum storage* if for every distribution X over $\{0,1\}^n$ with $H_\infty(X) \geq k$ and every b-storage encoding ρ of X,

$$\|U_t \circ E(X,U_t) \circ \rho(X) - U_{t+m} \times \bar{\rho}_X\|_{\mathrm{tr}} \leq \epsilon.$$

The next lemma shows it sufficient to consider only flat distributions when arguing about the correctness of extractors against quantum storage.

Lemma 6.11. *If E is* not *an (n,k,b,ϵ) strong extractor against quantum storage then there exists a set X of cardinality 2^k and a b-storage encoding ρ such that E fails on $(X;\rho)$, that is,*

$$\|U_t \circ E(X,U_t) \circ \rho(X) - U_{t+m} \times \bar{\rho}_X\|_{\mathrm{tr}} > \epsilon.$$

Proof: We prove the contrapositive, i.e., we assume that E works for flat distributions of min-entropy exactly k and prove that it also works for general distributions with at least k min-entropy.

Suppose X is a distribution with $H_\infty(X) \geq k$. Then X can expressed as a convex combination of flat distributions X_i each with $H_\infty(X_i) = k$. If ρ is a b-storage encoding of X then it is also a b-storage encoding of each of these flat distributions X_i. Thus, by assumption,

$$\|U_t \circ E(X_i, U_t) \circ \rho(X_i) - U_{t+m} \times \bar{\rho}_{X_i}\|_{\mathrm{tr}} \leq \epsilon.$$

Now by convexity,

$$\|U_t \circ E(X, U_t) \circ \rho(X) - U_{t+m} \times \bar{\rho}_X\|_{\mathrm{tr}} \leq \epsilon,$$

as desired. ■

Combining this with Proposition 6.5 we get:

Lemma 6.12. *Every quantum-proof (n, f, k, ϵ) strong extractor for flat distributions, is an $(n, f, f-k, \epsilon)$ strong extractor against quantum storage.*

6.1.3 Lossless condensers

Definition 6.13 (strong condenser)**.** A mapping $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{n'}$ is an $(n, k_1) \to_\epsilon (n', k_2)$ *strong condenser* if for every distribution X with k_1 min-entropy, $U_d \circ C(X, U_d)$ is ϵ-close to a distribution with $d + k_2$ min-entropy.

One typically wants to maximize k_2 and bring it close to k_1 while minimizing n' (it can be as small as $k_1 + O(\log \epsilon^{-1})$) and d (it can be as small as $\log((n-k)/(n'-k)) + \log \epsilon^{-1} + O(1)$). For a discussion of the parameters, see [35, Appendix B]. We call the condenser *lossless* if $k_2 = k_1$.

The property of lossless condensers that we shall use is the following.

Fact 6.14 ([131, Lemma 2.2.1])**.** Let $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{n'}$ be an $(n, k) \to_\epsilon (n', k)$ lossless condenser. Consider the mapping

$$C' : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{n'} \times \{0,1\}^d$$

$$C'(x, y) = C(x, y) \circ y.$$

Then, for every set $X \subseteq \{0,1\}^n$ of size $|X| \leq 2^k$, there exists a mapping $C'' : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{n'} \times \{0,1\}^d$ that is injective on $X \times \{0,1\}^d$ and agrees with C' on at least $1 - \epsilon$ fraction of the set $X \times \{0,1\}^d$.

6.2 A reduction to full classical entropy

A popular approach for constructing explicit extractors in the classical setting is as follows:

- Construct an explicit extractor for the *high* min-entropy regime, i.e. for sources X distributed over $\{0,1\}^n$ that have k min-entropy for some large k close to n, and,
- Show a reduction from the general case to the high min-entropy case.

In the classical setting this is often achieved by composing an extractor for the high min-entropy regime with a classical lossless condenser. Specifically, assume:

- $C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^{n'}$ is an $(n,k) \to_{\epsilon_1} (n',k)$ strong lossless condenser, and,
- $E : \{0,1\}^{d+n'} \times \{0,1\}^t \to \{0,1\}^m$ is a $(d+n', d+k, \epsilon_2)$ strong extractor.

Define $EC : \{0,1\}^n \times (\{0,1\}^d \times \{0,1\}^t) \to \{0,1\}^m$ by

$$EC(x,(y_1,y_2)) \quad = \quad E((C(x,y_1),y_1),y_2).$$

In the classical setting, [132, Section 5] prove that EC is a strong $(n,k,\epsilon_1+\epsilon_2)$ extractor. In this section we try to generalize this result to the quantum setting. We prove:

Theorem 6.15. *Let C and EC be as above.*

- *If E is a quantum-proof $(d+n', d+k, k_2, \epsilon_2)$ strong extractor for flat distributions, then EC is a $(n,k,k_2,\epsilon = \epsilon_2 + 2\epsilon_1)$ strong extractor for flat distributions.*
- *If E is a $(d+n', d+k, d+b, \epsilon_2)$ strong extractor against quantum storage, then EC is an $(n,k,b,\epsilon = \epsilon_2 + 2\epsilon_1)$ strong extractor against quantum storage.*

The intuition behind the theorem is the following. When the condenser C is applied on a flat source, it is essentially a one-to-one mapping between the source X and its image $C(X)$. Therefore, roughly speaking, any quantum information about x can be translated to quantum information about $C(x)$ and vice-versa. To make this precise we need to take care of the condenser's seed, and this incurs a small loss in the parameters.

We first prove the second item.

Proof (second item): Assume, by contradiction that EC is not an $(n,k,b,\epsilon = \epsilon_2 + 2\epsilon_1)$ strong extractor against quantum storage. Then, by Lemma 6.11, there exists a subset $X \subseteq \{0,1\}^n$ of

cardinality 2^k and a b-storage encoding ρ of X such that, given this encoding, the output of the extractor EC is not ϵ-close to uniform. That is,

$$\left\| U_{t+d} \circ EC(X, U_{t+d}) \circ \rho(X) - U_{t+d+m} \times \bar{\rho}_X \right\|_{\mathrm{tr}} \quad > \quad \epsilon.$$

In particular, by Fact 6.1, there exists some Boolean measurement that ϵ-distinguishes the two distributions. Since the first two components are classical, we can represent this measurement as follows. For every $y \in \{0,1\}^{t+d}$ and $z \in \{0,1\}^m$ there exists a Boolean measurement $\{F^{y,z}, I - F^{y,z}\}$ on the quantum component such that

$$\left| \mathop{\mathbb{E}}_{x \sim X,\, y \sim U} \left[\mathrm{Tr}\left(F^{y,EC(x,y)} \rho(x) \right) \right] - \mathop{\mathbb{E}}_{y,z \sim U} \left[\mathrm{Tr}\left(F^{y,z} \bar{\rho}_X \right) \right] \right| \quad > \quad \epsilon.$$

We now show how this can be used to break the extractor E. Consider the set $A = X \times \{0,1\}^d$. By Fact 6.14, there exists a mapping D that is injective on A and agrees with the condenser on at least $1 - \epsilon_1$ fraction of A. Denoting $B = D(A)$, it is clear that $H_\infty(B) \geq d + k$.

For $(\tilde{x}, \tilde{y}) \in B$ we define the encoding

$$\rho'(\tilde{x}, \tilde{y}) = |y_1\rangle\langle y_1| \otimes \rho(D^{\leftarrow}(\tilde{x}, \tilde{y})),$$

where $(x, y_1) = D^{-1}(\tilde{x}, \tilde{y}) \in A$ is the unique element such that $D(x, y_1) = (\tilde{x}, \tilde{y})$, and $D^{\leftarrow}(\tilde{x}, \tilde{y}) = x$.

Next, we define a measurement $\left\{ \overline{F}^{y_2,z}, I - \overline{F}^{y_2,z} \right\}$ that given the input $y_2 \in \{0,1\}^t, z \in \{0,1\}^m$ and $\rho'(\tilde{x}, \tilde{y}) = |y_1\rangle\langle y_1| \otimes \rho(x)$, sets $y = (y_1, y_2)$ and applies the measurement $\{F^{y,z}, I - F^{y,z}\}$ on the quantum register $\rho(x)$.

Now,

$$\left| \mathop{\mathbb{E}}_{b \sim B,\, y_2 \sim U_t} \left[\mathrm{Tr}\left(\overline{F}^{y_2, E(b,y_2)} \rho'(b) \right) \right] - \mathop{\mathbb{E}}_{x \sim X,\, y \sim U_{d+t}} \left[\mathrm{Tr}\left(F^{y,EC(x,y)} \rho(x) \right) \right] \right| \quad \leq \quad \epsilon_1,$$

since the flat distribution over B is ϵ_1-close to the distribution obtained by sampling $x \in X, y_1 \in U_d$ and outputting $(C(x, y_1), y_1)$. For the same reason, averaging over B for $\overline{F}$ is almost as averaging over X for F. Namely,

$$\left| \mathop{\mathbb{E}}_{y_2,z \sim U} \left[\mathrm{Tr}\left(\overline{F}^{y_2,z} \bar{\rho'}_B \right) \right] - \mathop{\mathbb{E}}_{y,z \sim U} \left[\mathrm{Tr}\left(F^{y,z} \bar{\rho}_X \right) \right] \right| \quad \leq \quad \epsilon_1.$$

It follows that

$$\left| \mathop{\mathbb{E}}_{b\sim B,\, y_2\sim U} \left[\mathrm{Tr}\left(\overline{F}^{y_2,E(b,y_2)} \rho'(b) \right) \right] - \mathop{\mathbb{E}}_{y_2,z\sim U} \left[\mathrm{Tr}\left(\overline{F}^{y_2,z} \bar{\rho}'_B \right) \right] \right| \;\geq$$

$$\left| \mathop{\mathbb{E}}_{x\sim X,\, y\sim U} \left[\mathrm{Tr}\left(F^{y,EC(x,y)} \rho(x) \right) \right] - \mathop{\mathbb{E}}_{y,z\sim U} \left[\mathrm{Tr}\left(F^{y,z} \bar{\rho}_X \right) \right] \right| - 2\epsilon_1 \;>\; \epsilon - 2\epsilon_1 = \epsilon_2.$$

Clearly ρ' is a $(d+b)$-storage encoding of B. This contradicts the fact that E is a strong extractor against $d+b$ quantum storage. ■

We now prove the first item.

Proof (first item): Assume, for contradiction, that EC is not a quantum-proof (n,k,k_2,ϵ) strong extractor for flat distributions. Then there exists a subset $X \subseteq \{0,1\}^n$ of cardinality exactly 2^k and an encoding ρ of X such that the conditional min-entropy is at least k_2 but given this encoding the output of the extractor EC is not ϵ-close to uniform. The proof proceeds as before, defining the Boolean measurement F, the sets A and B, the encoding ρ' and the measurement $\overline{F}$. If we can show that $H_\infty(B;\rho') \geq k_2$ then we break the extractor E and reach a contradiction. Indeed:

Claim 6.16. $H_\infty(B;\rho') \geq k_2$.

Proof: Assume, for contradiction, that $H_\infty(B;\rho') < k_2$. Then, there exists a predictor W' such that

$$\Pr_{b\sim B}[W'(\rho'(b)) = b] > 2^{-k_2}.$$

Define a new predictor, W, that given $\rho(x)$ works as follows. First W chooses $y\sim U_d$ and runs W' on $|y\rangle\langle y| \otimes \rho(x)$ to get some answer $\widetilde{b}$. It then outputs $D^{\leftarrow}(\widetilde{b})$.

The success probability of the predictor W is

$$\begin{aligned} \Pr_{x\sim X}[W(\rho(x)) = x] &= \Pr_{x\sim X, y\in\{0,1\}^d}[D^{\leftarrow}(W'(|y\rangle\langle y| \otimes \rho(x))) = x] \\ &\geq \Pr_{x\sim X, y\in\{0,1\}^d}[W'(|y\rangle\langle y| \otimes \rho(x)) = D(x,y)] \\ &= \Pr_{b\sim B}[W'(\rho'(b)) = b] > 2^{-k_2}. \end{aligned}$$

This contradicts the fact that $H_\infty(X;\rho) \geq k_2$. ■

■

We remark that we do not know how to extend the proof to work with lossy condensers.

6.3 An explicit quantum-proof extractor for the high-entropy regime

In this section we describe a construction of a short-seed quantum-proof (n, k, ϵ) strong extractor that works whenever $k \gg n/2$. In the classical setting this scenario was studied in [35], developing and improving techniques from [110] and other papers. Here we only need the techniques developed in [110].

Intuitively, the extractor E that we construct works as follows. First, it divides the source to two parts of equal length. Since the min-entropy is larger than $n/2$, for almost any fixing of the first part of the source, the distribution on the second part has $\Omega(n)$ min-entropy. Hence, applying an extractor E_2 on the second part results in output bits that are close to uniform. Since this is true for almost every fixing of the first part, these output bits are essentially independent of the first part of the source. Therefore, these output bits can serve as a seed for another extractor, E_1, that is applied on the first part of the source.

Formally, assume:

- $E_1 : \{0,1\}^{n/2} \times \{0,1\}^{d_1} \to \{0,1\}^{m_1}$ is a quantum-proof $(\frac{n}{2}, \frac{n}{2} - b, \epsilon_1)$ strong extractor, and,
- $E_2 : \{0,1\}^{n/2} \times \{0,1\}^{d_2} \to \{0,1\}^{d_1}$ is a quantum-proof $(\frac{n}{2}, k, \epsilon_2)$ strong extractor.

Define $E : \{0,1\}^n \times \{0,1\}^{d_2} \to \{0,1\}^{m_1}$ by

$$E(x, y) = E_1(x_1, E_2(x_2, y)),$$

where $x = x_1 \circ x_2$ and $x_1, x_2 \in \{0,1\}^{n/2}$.

Theorem 6.17. *Let E_1, E_2 and E be as above with $k = \frac{n}{2} - b - \log \epsilon^{-1}$. Then E is a quantum-proof $(n, n - b, \epsilon + \epsilon_1 + \epsilon_2)$ strong extractor.*

Proof: Let $X = X_1 \circ X_2$ be a distribution on $\{0,1\}^n = \{0,1\}^{n/2} \times \{0,1\}^{n/2}$ and ρ be an encoding such that $H_\infty(X; \rho) \geq n - b$. For a prefix $x_1 \in \{0,1\}^{n/2}$, let ρ_{x_1} be the encoding of X_2 defined by $\rho_{x_1}(x_2) = \rho(x_1 \circ x_2)$. A prefix x_1 is said to be *bad* if $H_\infty(X_2 \mid X_1 = x_1; \rho_{x_1}) \leq k$. By Lemma 6.7, the probability x_1 (sampled from X_1) is bad is at most

$$\frac{2^{n/2} \cdot 2^k}{2^{n-b}} = \frac{2^{n/2} \cdot 2^{n/2-b-\log \epsilon^{-1}}}{2^{n-b}} = \epsilon.$$

Whenever x_1 is not bad, $H_\infty(X_2 \mid X_1 = x_1; \rho_{x_1}) > k$, that is, the extractor E_2 is applied on a distribution with k min-entropy. Therefore, by the assumption on E_2, its output is ϵ_2-close to uniform. That is, for every good x_1,

$$\| U_{d_2} \circ x_1 \circ E_2(X_2, U_{d_2}) \circ \rho_{x_1}(X_2) - U_{d_2} \circ x_1 \circ U_{d_1} \circ \rho_{x_1}(X_2) \|_{\mathrm{tr}} \leq \epsilon_2.$$

Hence, the distribution $U_{d_2} \circ X_1 \circ E_2(X_2, U_{d_2}) \circ \rho(X)$ is $(\epsilon+\epsilon_2)$-close to $U_{d_2} \circ X_1 \circ U_{d_1} \circ \rho(X)$. In particular,

$$\begin{aligned} & \|U_{d_2} \circ E(X, U_{d_2}) \circ \rho(X) - U_{d_2+d_1} \circ \bar{\rho}_X\|_{\mathrm{tr}} \\ = \; & \|U_{d_2} \circ E_1(X_1, E_2(X_2, U_{d_2})) \circ \rho(X) - U_{d_2+d_1} \circ \bar{\rho}_X\|_{\mathrm{tr}} \\ \leq \; & \epsilon + \epsilon_2 + \|U_{d_2} \circ E_1(X_1, U_{d_1}) \circ \rho(X) - U_{d_2+d_1} \circ \bar{\rho}_X\|_{\mathrm{tr}}, \end{aligned}$$

where the last inequality follows from Fact 6.2.

Since, $H_\infty(X;\rho) \geq n - b$, by Lemma 6.6, if we define an encoding ρ' of X_1 by $\rho'(x_1) = \mathbb{E}_{x \sim (X|X_1=x_1)}[\rho(x)]$, then $H_\infty(X_1;\rho') \geq n - b - n/2 = n/2 - b$. Therefore, by the assumption on E_1 we get

$$\|E_1(X_1, U_{d_1}) \circ \rho(X) - U_{m_1} \otimes \bar{\rho}_X\|_{\mathrm{tr}} \leq \epsilon_1,$$

and thus

$$\|U_{d_2} \circ E(X, U_{d_2}) \circ \rho(X) - U_{d_2+d_1} \otimes \bar{\rho}_X\|_{\mathrm{tr}} \quad \leq \quad \epsilon + \epsilon_1 + \epsilon_2.$$

■

6.3.1 Plugging in explicit constructions

We use Trevisan's extractor, which was already shown to be quantum-proof in [41, 40]. Specifically, we use the following two instantiations of this extractor:

Theorem 6.18 ([40]). *For every constant $\delta > 0$, there exists $E_1 : \{0,1\}^{\frac{n}{2}} \times \{0,1\}^{O(\log^2(n/\epsilon_1))} \to \{0,1\}^{(1-\delta)(\frac{n}{2}-b)}$ which is a quantum-proof $(\frac{n}{2}, \frac{n}{2} - b, \epsilon_1)$ strong extractor.*

Theorem 6.19 ([40]). *For every constants $\gamma_1, \gamma_2 > 0$, there exists $E_2 : \{0,1\}^{\frac{n}{2}} \times \{0,1\}^{O(\log(n/\epsilon_2))} \to \{0,1\}^{k^{1-\gamma_1}}$ which is a quantum-proof $(\frac{n}{2}, k, \epsilon_2)$ strong extractor, for $k > n^{\gamma_2}$.*

Plugging these two constructions into Theorem 6.17 gives Theorem 1.3 which we now restate.

Theorem 1.3. For any $\beta < \frac{1}{2}, \gamma > 0$ and $\epsilon \geq 2^{-n^{(1-\gamma)/2}}$, there exists an explicit quantum-proof $(n, (1-\beta)n, \epsilon)$ strong extractor $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log \epsilon^{-1})$ and output length $m = \Omega(n)$.

Proof: We set $\epsilon_1 = \epsilon_2 = \epsilon,\ b = \beta n,\ k = \frac{n}{2} - \beta n - \log \epsilon^{-1},\ \gamma_2 = \delta = \frac{1}{2}$ and $\gamma_1 < \gamma$. In order to apply Theorem 6.17 we need to verify that the output length of E_2 is not shorter than the seed length of E_1. This is indeed the case since

$$k^{1-\gamma_1} \geq (\frac{n}{2} - \beta n - n^{\frac{1-\gamma}{2}})^{1-\gamma_1} \geq n^{1-\gamma} \geq O(\log^2(\frac{n}{\epsilon})).$$

The output length of E is $\frac{1}{2}(\frac{1}{2}-\beta)n = \Omega(n)$. ■

6.4 The final extractor for the bounded storage model

We need the classical lossless condenser of [59].

Theorem 6.20 ([59])**.** *For every $\alpha > 0$ there exists an $(n,k) \to_\epsilon ((1+\alpha)k, k)$ strong lossless condenser C with seed length $O(\log n + \log \epsilon^{-1})$.*

Plugging the condenser C and the extractor E of Theorem 1.3 into Theorem 6.15 gives Theorem 1.2, which we now restate.

Theorem 1.2. For any $\beta < \frac{1}{2}$ and $\epsilon \geq 2^{-k^\beta}$, there exists an explicit $(n, k, \beta k, \epsilon)$ strong extractor against quantum storage, $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log \epsilon^{-1})$ and output length $m = \Omega(k)$.

Proof: Let $\zeta > 0$ be a constant to be fixed later. The extractor E from Theorem 1.3, when the source length is set to be $2(1-\beta)(1-\zeta)k$, is a quantum-proof $\big(2(1-\beta)(1-\zeta)k, (1-\beta)k, \epsilon\big)$ strong extractor. In particular, it is a $\big(2(1-\beta)(1-\zeta)k, k, \beta k, \epsilon\big)$ strong extractor against quantum storage. Its output length is $\Omega(k)$. The theorem follows by applying Theorem 6.15, using the condenser of Theorem 6.20 with $\alpha = 2(1-\beta)(1-\zeta) - 1$. Since $\beta < \frac{1}{2}$ there is a way to fix ζ such that $\alpha > 0$. ■

Since Theorem 6.15 works in the more general model of flat distributions, and since the extractor from Theorem 1.3 already works in the most general setting, we get Theorem 1.1:

Theorem 1.1. For any $\beta < \frac{1}{2}$ and $\epsilon \geq 2^{-k^\beta}$, there exists an explicit quantum-proof $(n, k, (1-\beta)k, \epsilon)$ strong extractor for flat distributions, $E : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$, with seed length $t = O(\log n + \log \epsilon^{-1})$ and output length $m = \Omega(k)$.

Bibliography

[1] S. Aaronson and A. Ambainis. Quantum search of spatial regions. In *Proceedings of 44th IEEE FOCS*, pages 200–209, 2003. quant-ph/0303041. 100, 112

[2] J. Adams. Character tables for $GL(2), SL(2), PGL(2)$ and $PSL(2)$ over a finite field. `http://www.math.umd.edu/~jda/characters/characters.pdf`, 2002. 56

[3] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. 4

[4] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k–wise independent random variables. *Random Structures and Algorithms*, 3(3):289–303, 1992. 76, 83

[5] 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 FOCS*, pages 630–637, 2001. 5

[6] N. Alon and V. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory. Series B*, 38(1):73–88, 1985. 4

[7] N. Alon and Y. Roichman. Random Cayley Graphs and Expanders. *Random Structures and Algorithms*, 5(2):271–285, 1994. 43

[8] A. Ambainis, A. Nayak, A. Ta-Shma, and U. V. Vazirani. Quantum dense coding and quantum finite automata. *Journal of the ACM*, 49:496–511, 2002. Earlier version in 31st ACM STOC, 1999, pp. 376-383. 69, 96

[9] A. Ambainis and A. Smith. Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. In *RANDOM*, pages 249–260, 2004. 6, 43, 45, 46, 65

[10] L. Babai, T. P. Hayes, and P. G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001. Earlier version in STOC'98. 110

[11] K. Ball, E. Carlen, and E. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones Mathematicae*, 115:463–482, 1994. 7, 8, 93, 96, 98, 102

[12] R. Beals. Quantum computation of Fourier transforms over symmetric groups. *STOC*, pages 48–53, 1997. 44

[13] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of set disjointness. *Computational Complexity*, 15(4):391–432, 2006. Earlier version in Complexity'05. 100, 110, 111

[14] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159–182, 1975. 94

[15] A. Ben-Aroya, , O. Schwartz, and A. Ta-Shma. Quantum expanders: motivation and constructions. *Theory of Computing*, 6(3):47–79, 2010. Earlier version in CCC'08. 6

[16] A. Ben-Aroya, K. Efremenko, and A. Ta-Shma. Local list decoding with a constant number of queries. In *Proceedings of 51st IEEE FOCS*, 2010. 1

[17] A. Ben-Aroya, K. Efremenko, and A. Ta-Shma. A note on amplifying the error-tolerance of locally decodable codes. Technical report, http://eccc.hpi-web.de/report/2010/134/, 2010. 1

[18] A. Ben-Aroya, O. Regev, and R. d. Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing. quant-ph/0705.3806, 2007. 97

[19] A. Ben-Aroya, O. Regev, and R. d. Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In *Proceedings of 49th IEEE FOCS*, pages 477–486, 2008. 8, 97

[20] A. Ben-Aroya and A. Ta-Shma. Quantum expanders and the quantum entropy difference problem. Technical report, arXiv:quant-ph/0702129, 2007. 44

[21] A. Ben-Aroya and A. Ta-Shma. Constructing small-bias sets from algebraic-geometric codes. In *Proceedings of 50th IEEE FOCS*, pages 191–197, 2009. 7

[22] A. Ben-Aroya and A. Ta-Shma. Approximate quantum error correction for correlated noise. *IEEE Transactions on Information Theory*, 57(6):3982–3988, 2010. Earlier version in the twelfth workshop on Quantum Information Processing (QIP)'09. 1

[23] A. Ben-Aroya and A. Ta-Shma. On the complexity of approximating the diamond norm. *Quantum Information and Computation*, 10(1 & 2):77–86, 2010. 1

[24] A. Ben-Aroya and A. Ta-Shma. Better short-seed extractors against quantum knowledge. *Theoretical Computer Science*, 2011. To appear. 12

[25] A. Ben-Aroya and A. Ta-Shma. A combinatorial construction of almost-Ramanujan graphs using the zig-zag product. *SIAM Journal on Computing*, 40(2):267–290, 2011. Earlier version in STOC'08. 5

[26] C. Bennett, G. Brassard, C. Crepeau, and U. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6, Part 2):1915–1923, 1995. 9

[27] C. Bennett, G. Brassard, and J. Robert. Privacy amplification by public discussion. *SIAM Journal on Computing*, 17(2):210–229, 1988. 9

[28] R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1997. 113

[29] Y. Bilu and N. Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006. 5

[30] S. G. Bobkov. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Annals of Probability*, 25(1):206–214, 1997. 96

[31] A. Bonami. Etude des coefficients de Fourier des fonctions de $L^p(G)$. *Annales de l'Institut Fourier*, 20(2):335–402, 1970. 94

[32] C. Borell. On the integrability of Banach space valued Walsh polynomials. In *Séminaire de Probabilités, XIII (Univ. Strasbourg, 1977/78)*, volume 721 of *Lecture Notes in Math.*, pages 1–3. Springer, Berlin, 1979. 96

[33] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of 30th ACM STOC*, pages 63–68, 1998. quant-ph/9802040. 100

[34] H. Buhrman and R. d. Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of 16th IEEE Conference on Computational Complexity*, pages 120–130, 2001. cs.CC/9910010. 100

[35] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree expansion beyond the degree / 2 barrier. In *Proceedings of the 34th STOC*, pages 659–668, 2002. 4, 10, 42, 124, 128

[36] E. A. Carlen and E. H. Lieb. Optimal hypercontractivity for Fermi fields and related noncommutative integration inequalities. *Communications in Mathematical Physics*, 155(1):27–46, 1993. 96

[37] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical report, ECCC TR–08–002, 2008. Available at http://www.eccc.uni-trier.de/eccc/. 110

[38] M. Christandl, R. Renner, and A. Ekert. A Generic Security Proof for Quantum Key Distribution, 2004. arXiv:quant-ph/0402131. 9

[39] C. M. Dawson and M. A. Nielsen. The Solovay-Kitaev algorithm. *Quantum Inf. Comput.*, 6(1):81–95, 2006. 59

[40] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan's extractor in the presence of quantum side information, 2009. arXiv:0912.5514. 129

[41] A. De and T. Vidick. Near-optimal extractors against quantum storage. In *Proc. 42nd ACM Symp. on Theory of Computing (STOC)*, 2010. 9, 10, 97, 129

[42] P. Dickinson and A. Nayak. Approximate randomization of quantum states with fewer bits of key. In *AIP Conference Proceedings*, volume 864, pages 18–36, 2006. 45

[43] Y. Dodis and A. Smith. Correcting errors without leaking partial information. In *Proc. 37th ACM Symp. on Theory of Computing (STOC)*, pages 654–663, 2005. 10

[44] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Trans. Amer. Math. Soc.*, 284(2):787–794, 1984. 4

[45] Z. Dvir, S. Kopparty, S. Saraf, and M. Sudan. Extensions to the method of multiplicities, with applications to kakeya sets and mergers. In *Proc. 50th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 181–190. IEEE, 2009. 10

[46] Z. Dvir and A. Wigderson. Kakeya sets, new mergers and old extractors. In *Proc. 49th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 625–633, 2008. 10

[47] K. Efremenko. 3-query locally decodable codes of subexponential length. In *STOC*, pages 39–44, 2009. 100, 101

[48] S. Fehr and C. Schaffner. Randomness extraction via delta-biased masking in the presence of a quantum attacker. In *Proceedings of Theory of Cryptography (TCC)*, pages 465–481, 2008. 10, 95

[49] J. Friedman. A proof of Alon's second eigenvalue conjecture. *Memoirs of the AMS*, to appear. 4, 35

[50] W. Fulton. *Algebraic Curves*. Third edition, 2008. 80

[51] O. Gabber and Z. Galil. Explicit Constructions of Linear-Sized Superconcentrators. *Journal of Computer and System Sciences*, 22(3):407–420, 1981. 4

[52] A. Garcia and E. Stichtenoth, H. *Topics in Geometry, Coding Theory and Cryptography (Algebra and Applications)*. Springer-Verlag, 2006. 82

[53] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. *SIAM Journal on Computing*, 38(5):1695–1708, 2008. 10, 93

[54] O. Goldreich, H. Karloff, L. Schulman, and L. Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *Computational Complexity*, 15(3):263–296, 2006. Earlier version in Complexity'02. Also on ECCC. 100

[55] O. Goldreich and A. Wigderson. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Structures and Algorithms*, 11(4):315–343, 1997. 10, 65

[56] V. Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and computation*, 112(1):51–54, 1994. 110

[57] D. Gross and J. Eisert. Quantum Margulis expanders. *Quantum Inf. Comput.*, 8(8&9):722–733, 2008. 44

[58] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975. 96

[59] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. *Journal of the ACM*, 56(4):1–34, 2009. 10, 130

[60] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. Reprint of the 1952 edition. 103

[61] J. Harris and W. Fulton. *Representation Theory*. Springer, 1991. 48, 56

[62] A. Harrow. Quantum expanders from any classical Cayley graph expander. *Quantum Inf. Comput.*, 8(8&9):715–721, 2008. 44, 45

[63] J. Håstad. Some optimal inapproximability results. In *Proceedings of 29th ACM STOC*, pages 1–10, 1997. 93

[64] M. Hastings and A. Harrow. Classical and quantum tensor product expanders. *Quantum Inf. Comput.*, 9(3&4):336–360, 2009. 45

[65] M. B. Hastings. Entropy and entanglement in quantum ground states. *Phys. Rev. B*, 76(3):035114, 2007. 6, 43, 45

[66] M. B. Hastings. Random unitaries give quantum expanders. *Phys. Rev. A*, 76(3):032315, 2007. 44, 59

[67] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English translation in *Problems of Information Transmission*, 9:177–183, 1973. 96

[68] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the AMS*, 43(4):439–561, 2006. 4, 6

[69] S. Hoory and A. Wigderson. Universal sequences for expander graphs. *Information Processing Letters*, 46:67–69, 2 1993. 23

[70] P. Høyer and R. d. Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *Proceedings of 19th Annual Symposium on Theoretical Aspects of Computer Science (STACS'2002)*, volume 2285 of *Lecture Notes in Computer Science*, pages 299–310. Springer, 2002. quant-ph/0109068. 100

[71] R. Impagliazzo, L. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proc. 21st ACM Symp. on Theory of Computing (STOC)*, pages 12–24, 1989. 10

[72] J. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55(3):414–440, 1997. Earlier version in FOCS'94. 93

[73] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. John Wiley New York, 2000. 33

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

[75] S. Jukna. *Extremal Combinatorics*. EATCS Series. Springer, 2001. 109

[76] N. Kahale. Eigenvalues and expansion of regular graphs. *Journal of the ACM*, 42(5):1091–1106, 1995. 4

[77] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proceedings of 29th IEEE FOCS*, pages 68–80, 1988. 93

[78] G. Kalai and S. Safra. Threshold phenomena and influence: perspectives from mathematics, computer science, and economics. *Computational complexity and statistical physics*, pages 25–60, 2006. 93

[79] M. Kassabov. Symmetric groups and expanders. *Electron. Res. Announc. Amer. Math. Soc.*, 11, 2005. 44

[80] J. Katz and L. Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of 32nd ACM STOC*, pages 80–86, 2000. 101, 115, 116

[81] I. Kerenidis and R. d. Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69(3):395–420, 2004. Special issue on STOC'03. quant-ph/0208062. 8, 100, 101, 113, 114, 117

[82] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of 42nd IEEE FOCS*, pages 288–297, 2001. quant-ph/0106160. 93

[83] H. Klauck, R. Špalek, and R. d. Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of 45th IEEE FOCS*, pages 12–21, 2004. quant-ph/0402123. 100, 110, 111

[84] M. M. Klawe. Limitations on Explicit Constructions of Expanding Graphs. *SIAM J. Comput.*, 13(1):156–166, 1984. 43

[85] R. König, U. Maurer, and R. Renner. On the power of quantum memory. *IEEE Transactions on Information Theory*, 51(7):2391–2401, 2005. 9, 10

[86] R. König and R. Renner. Sampling of min-entropy relative to quantum knowledge, 28 Dec 2007. quant-ph/0712.4291. 9, 98, 99

[87] R. Konig, R. Renner, and C. Schaffner. The operational meaning of min-and max-entropy. *IEEE Transactions on Information theory*, 55(9):4337–4347, 2009. 121

[88] R. König and B. Terhal. The bounded-storage model in the presence of a quantum adversary. *IEEE Transactions on Information Theory*, 54(2):749–762, 2008. 9, 121

[89] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997. 108

[90] J. D. Lafferty and D. Rockmore. Fast fourier analysis for SL_2 over a finite field and related numerical experiments. *Experiment. Math.*, 1(2):115–139, 1992. 44

[91] S. Lang. *Algebra*. Springer, revised third edition, 2002. 79

[92] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geometric and Functional Analysis*, 14(4):745–747, 2004. 96

[93] T. Lee, G. Schechtman, and A. Shraibman. Lower bounds on quantum multiparty communication complexity. In *24th Annual IEEE Conference on Computational Complexity*, pages 254–262. IEEE, 2009. 110

[94] T. Lee and A. Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. In *Proceedings of 23rd IEEE Conference on Computational Complexity*, 2008. arXiv:0712.4279. 110

[95] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, 1993. Earlier version in FOCS'89. 93

[96] A. Lubotzky, R. Philips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988. 4, 43, 51

[97] Y. Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995. Earlier version in COLT'92. 93

[98] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredaci Informacii*, 9(4):71–80, 1973. 4, 44, 45

[99] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988. 4, 43

[100] R. Meshulam and A. Wigderson. Expanders in group algebras. *Combinatorica*, 24(4):659–680, 2004. 5

[101] M. Morgenstern. Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q. *Journal of Combinatorial Theory. Series B*, 62(1):44–62, 1994. 4

[102] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of mathematics*, 171(1):295–341, 2010. 93

[103] E. Mossel, R. O'Donnell, and R. Servedio. Learning functions of k relevant variables. *Journal of Computer and System Sciences*, 69(3):421–434, 2004. Earlier version in STOC'03. 93

[104] J. Naor and M. Naor. Small–bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. 76

[105] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of 40th IEEE FOCS*, pages 369–376, 1999. quant-ph/9904093. 96, 97

[106] A. Nayak and A. Vishwanath. Quantum walk on the line. quant-ph/0010117, Oct 2000. 95

[107] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000. 66, 69

[108] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000. 102

[109] A. Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207–210, 1991. 4, 22

[110] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996. 10, 11, 128

[111] R. O'Donnell. *Computational applications of noise sensitivity*. PhD thesis, MIT, 2003. 93

[112] R. O'Donnell. Lecture notes for a course "Analysis of Boolean functions", 2007. Available at http://www.cs.cmu.edu/~odonnell/boolean-analysis/. 94

[113] R. O'Donnell. Some topics in analysis of Boolean functions. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 569–578. ACM, 2008. 93

[114] M. Pinsker. On the complexity of a concentrator. In *7th Internat. Teletraffic Confer.*, pages 318/1–318/4, 1973. 4

[115] S. Popescu and D. Rohrlich. Thermodynamics and the measure of entanglement. *Physical Review A*, 56(5):3319–3321, 1997. 47

[116] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, 2000. 10

[117] R. Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3/4):205–221, 1995. 93

[118] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, mathematics*, 67(1):159–176, 2003. quant-ph/0204025. 100

[119] O. Reingold. Undirected st-connectivity in log-space. *Journal of the ACM*, 55(4):1–24, 2008. 4, 23

[120] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):157–187, 2002. 4, 5, 13, 14, 20, 22, 36, 42, 44, 57, 60, 61

[121] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, 2005. 121

[122] G. B. Robinson. On the Representations of the Symmetric Group. *American Journal of Mathematics*, 60(3):745–760, 1938. 54

[123] E. Rozenman, A. Shalev, and A. Wigderson. Iterative construction of cayley expander graphs. *Theory of Computing*, 2(5):91–120, 2006. 5

[124] A. Sahai and S. Vadhan. Manipulating statistical difference, 1998. 70, 71

[125] C. Schensted. Longest increasing and decreasing subsequences. *Canad. J. Math*, 13(2), 1961. 54

[126] J. P. Serre. *Linear representations of finite groups*, volume 42 of *Graduate texts in Mathematics*. Springer, 1977. 48, 54

[127] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002. 7

[128] A. Srinivasan and D. Zuckerman. Computing with very weak random sources. *SIAM Journal on Computing*, 28(4):1433–1459, 1999. 10

[129] H. Stichtenoth. *Algebraic function fields and codes*. Springer Verlag, 1993. 76, 81, 82, 84, 86, 87, 88, 89

[130] H. Stichtenoth. Private communication, 2009. 87

[131] A. Ta-Shma, C. Umans, and D. Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *Proc. 33th ACM Symp. on Theory of Computing (STOC)*, 2001. 124

[132] A. Ta-Shma, C. Umans, and D. Zuckerman. Lossless condensers, unbalanced expanders, and extractors. *Combinatorica*, 27(2):213–240, 2007. 11, 125

[133] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner. Leftover hashing against quantum side information, 2010. arXiv:1002.2436. 10

[134] N. Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of trace classes $S_p(1 \leq p < \infty)$. *Studia Mathematica*, 50:163–182, 1974. 96

[135] L. Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001. 10

[136] L. Trevisan. Some applications of coding theory in computational complexity. *Quaderni di Matematica*, 13:347–424, 2004. 100, 101

[137] M. Tsfasman, S. Vladutx, and T. Zink. Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound. *Mathematische Nachrichten*, 109(1), 1982. 75

[138] S. Vadhan. *A Study of Statistical Zero-Knowledge Proofs*. PhD thesis, Massachusetts Institute of Technology, 1999. 66

[139] S. Vadhan. The unified theory of pseudorandomness. *SIGACT News*, 38(3):39–54, 2007. 2

[140] E. Viola and A. Wigderson. One-way multi-party communication lower bound for pointer jumping with applications. In *Proceedings of 48th IEEE FOCS*, pages 427–437, 2007. 98, 110

[141] F. Voloch. Special divisors of large dimension on curves with many points over finite fields. To appear in Portugaliae Mathematica, 2009. 7, 75, 77, 87

[142] J. Watrous. Limits on the power of quantum statistical zero-knowledge. In *FOCS*, pages 459–470, 2002. 46, 68, 69, 71

[143] J. Watrous. Zero-knowledge against quantum attacks. In *STOC*, pages 296–305, 2006. 46

[144] R. d. Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 287(1):337–353, 2002. 108

[145] D. Woodruff. New lower bounds for general locally decodable codes. Technical report, ECCC TR07–006, 2006. 101

[146] S. Yekhanin. Towards 3-query locally decodable codes of subexponential length. In *Proceedings of 39th ACM STOC*, pages 266–274, 2007. 100, 101