Electronic Colloquium on Computational Complexity, Report No. 27 (2022)

New Near-Linear Time Decodable Codes Closer to the GV Bound

Guy Blanc*
Stanford
gblanc@stanford.edu

Dean Doron†
Ben Gurion University
deand@bgu.ac.il

Abstract

We construct a family of binary codes of relative distance $\frac{1}{2}-\varepsilon$ and rate

$$\varepsilon^2 \cdot 2^{-\log^{\alpha}(1/\varepsilon)}$$

for $\alpha \approx \frac{1}{2}$ that are decodable, probabilistically, in near-linear time. This improves upon the rate of the state-of-the-art near-linear time decoding near the GV bound due to Jeronimo, Srivastava, and Tulsiani, who gave a randomized decoding of Ta-Shma codes with $\alpha \approx \frac{5}{6}$ [TS17, JST21]. Each code in our family can be constructed in probabilistic polynomial time, or deterministic polynomial time given sufficiently good explicit 3-uniform hypergraphs.

Our construction is based on a new graph-based bias amplification method. While previous works start with some base code of relative distance $\frac{1}{2}-\varepsilon_0$ for $\varepsilon_0 \gg \varepsilon$ and amplify the distance to $\frac{1}{2}-\varepsilon$ by walking on an expander, or on a carefully tailored product of expanders, we walk over very sparse, highly mixing, hypergraphs. Study of such hypergraphs further offers an avenue toward achieving rate $\widetilde{\Omega}(\varepsilon^2)$. For our unique- and list-decoding algorithms, we employ the framework developed in [JST21].

*Supported by NSF CAREER Award 1942123.
†Part of this work was done while at Stanford, supported by NSF Award CCF-1763311.

ISSN 1433-8092

1 Introduction

The Gilbert–Varshamov (GV) bound, for binary codes, tells us that there exist codes, even linear ones, with relative distance $\frac{1-\varepsilon}{2}$ and rate $\Omega(\varepsilon^2)$ [Gil52, Var57]. Namely, there exist codes $\mathcal{C} \subseteq \mathbb{F}_2^n$ such that for any two distinct codewords $x, y \in \mathcal{C}$ it holds that $\Delta(x, y) \geq \frac{1-\varepsilon}{2}$, for Δ being the normalized Hamming distance, such that $\frac{\log|\mathcal{C}|}{n} = \Omega(\varepsilon^2)$. Finding such small-redundancy codes, hopefully accompanied by an efficient decoding algorithm, has been subject to extensive and fruitful research in the past decades (see, e.g., [NN93, ABN+92, AGHP92, BATS13, GI05, TS17]). In a breakthrough result, Ta-Shma [TS17] constructed explicit linear codes of relative distance $\frac{1-\varepsilon}{2}$ having rate $\varepsilon^{2+o(1)}$. Ta-Shma's codes are also ε-*balanced*, i.e., $\Delta(x, y) \in \left[\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right]$, and thus give rise to explicit ε-biased sample spaces, which are ubiquitous in pseudorandomness and derandomization.

No decoding algorithm was given in [TS17], and this was later ameliorated by Jeronimo, Quintana, Srivastava, and Tulsiani [JQST20, JST21], who showed that a slight variant of Ta-Shma's codes are indeed efficiently decodable, and even in time $\widetilde{O}_\varepsilon(n)$.

Theorem 1 ([TS17, JST21])**.** *There exists an explicit family of ε-balanced binary linear codes $\mathcal{C}_{\mathsf{TS}} \subseteq \mathbb{F}_2^n$ of rate*

$$r_{\mathsf{TSD}} = \varepsilon^2 \cdot 2^{-O\left(\log(1/\varepsilon)^{5/6}\right)},$$

such that:

1. *There exists a randomized algorithm that uniquely decode $\mathcal{C}_{\mathsf{TS}}$ up to half the distance in time $c_1(\varepsilon) \cdot \widetilde{O}(n)$. That is, given a noisy word $\tilde{z} \in \mathbb{F}_2^n$, the algorithm returns, with high probability, the unique $z \in \mathcal{C}$ such that $\Delta(z, \tilde{z}) \leq \frac{1-\varepsilon}{4}$ (if such exists).*

2. *There exists a randomized algorithm that list-decodes $\mathcal{C}_{\mathsf{TS}}$ up to radius*

$$\rho_{\mathsf{TSD}} = \frac{1}{2} - 2^{-O\left((\log(1/\varepsilon))^{1/6}\right)}$$

in time $c_2(\varepsilon) \cdot \widetilde{O}(n)$. That is, given a noisy word $\tilde{z} \in \mathbb{F}_2^n$, the algorithm returns, with high probability, a list $\mathcal{L} = \{z \in \mathcal{C} : \Delta(\tilde{z}, z) \leq \rho\}$ of size $|\mathcal{L}| = O(1/\varepsilon)$.[1]

We note that without any guarantee on the decoding capabilities, the codes in [TS17] achieve a better rate of

$$r_{\mathsf{TS}} = \varepsilon^2 \cdot 2^{-\widetilde{O}\left(\log(1/\varepsilon)^{2/3}\right)}.$$

Randomized constructions of binary codes, namely, randomized algorithms that output a good code with high probability, are also well-studied, where the goal is to achieve enough structure to allow for efficient decoding. If we focus on decoding in time $n^{1+o(1)}$, the current state-of-the-art is due to Hemenway, Wootters, and Ron-Zewi, that reaches the GV bound with a randomized construction.[2]

Theorem 2 ([HRZW19])**.** *There exists a family of ε-balanced binary codes $\mathcal{C}_{\mathsf{HRW}} \subseteq \mathbb{F}_2^n$ of rate $\Omega(\varepsilon^2)$ that can be constructed in probabilistic polynomial time, such that:*

[1]The guarantee on the list size is not a unique property of $\mathcal{C}_{\mathsf{TS}}$, but follows from the Johnson bound (see, e.g., [GRS15, Section 7.3]), observing that $\rho_{\mathsf{TSD}} \leq \frac{1}{2} - \sqrt{\varepsilon}$.

[2]The foregoing theorem appears in the arXiv version, and some of the parameters are only implicit there.

1. *There exists a randomized algorithm that uniquely decodes $\mathcal{C}_{\mathsf{HRW}}$ up to half the distance in time $c_3(\varepsilon) \cdot n^{1+1/t}$, where $t \approx \log\log\log n$.*

2. *There exists a randomized algorithm that list-decodes*[3] *$\mathcal{C}_{\mathsf{HRW}}$ up to radius $\frac{1}{2} - O(\sqrt{\varepsilon})$ in time $c_3(\varepsilon) \cdot n^{1+1/t}$.*

In this work, we continue the study of near-linear time decodable binary codes near the GV bound, and give a randomized construction with improved rate.

Theorem 3 (see also Theorems 7 and 9)**.** *There exists a family of ε-balanced binary codes $\mathcal{C} \subseteq \mathbb{F}_2^n$ that can be constructed in probabilistic polynomial time, of rate*

$$r = \varepsilon^2 \cdot 2^{-\widetilde{O}\left(\sqrt{\log(1/\varepsilon)}\right)},$$

such that:

1. *There exists a randomized algorithm that uniquely decodes $\mathcal{C}$ up to half the distance in time $c_1(\varepsilon) \cdot \widetilde{O}(n)$.*

2. *There exists a randomized algorithm that list-decodes $\mathcal{C}$ up to radius*

$$\rho = \frac{1}{2} - 2^{-O(\sqrt{\log(1/\varepsilon)})}$$

in time $c_2(\varepsilon) \cdot \widetilde{O}(n)$.

Thus, our codes achieve a better rate (and a better list decoding radius) than in [TS17, JST21], while maintaining the $\widetilde{O}(n)$ runtime. Compared to state-of-the-art randomized constructions, we do not reach the GV bound, nor do we reach the Johnson radius for list decoding, but our decoding is faster, and as we shall soon see, our code is more structured. (The [HRZW19] result concatenate an outer code over a large alphabet with uniformly and independently chosen inner binary codes.)

In terms of the dependence on ε, for Theorems 1 and 3, $c_1(\varepsilon)$ is doubly-exponential in $\log^{\alpha}(1/\varepsilon)$ for some $\alpha < 1$ (that is slightly better in Theorem 3), and $c_2(\varepsilon)$ is triply-exponential in $\log^{\alpha}(1/\varepsilon)$. In Theorem 2, $c_3(\varepsilon)$ is triply-exponential in $\mathrm{poly}(1/\varepsilon)$.[4]

Our construction, which we shall soon describe, is arguably *simpler* than the constructions of Theorems 1 and 2.[5] Moreover, it gives an avenue toward achieving an even better rate of $\widetilde{\Omega}(\varepsilon^2)$ if we assume the existence of better primitives. In slightly more details, our construction utilizes hypergraphs with a strong mixing property, dubbed λ-*mixing*, and we show that a random 3-regular hypergraph achieves a good enough λ. A better dependence between λ and the regularity of the hypergraph readily gives better rate (for the details, see Section 6). We thereby put forward a challenge that warrants revisiting mixing properties of 3-uniform hypergraphs, which is interesting in itself.

[3] The randomized list decoding algorithm of [HRZW19] was later derandomized in [KRRZ+20].

[4] More accurately, it is also doubly-exponential in $\log(1/\varepsilon) \cdot t$. The original analysis of [HRZW19] implies a quadruple-exponential dependence on $\mathrm{poly}(1/\varepsilon)$, but a better bound on the output list size of random list decodable codes, given in [LW18], can be used to reduce it to triply-exponential. Using the concatenation scheme of [HRZW19] with a different outer code given in [KRZSW18] may be used to reduce the dependence on ε but at a cost of making the dependence on n worse. Finally, we note that the failure probability in Theorem 2 is sub-exponentially small, whereas the failure probability in Theorems 1 and 3 is exponentially small.

[5] By this we refer to our probabilistic construction, in which we draw a favorable hypergraph H at random. Admittedly, making our construction deterministic by constructing an explicit family of good H-s is likely to make it less simple.

1.1 Our construction

Our construction goes via distance amplification. We start with some base code $\mathcal{C}_0 \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ and construct our code $\mathcal{C} \colon \mathbb{F}_2^k \to \mathbb{F}_2^{\bar{n}}$ so that for every coordinate $i \in [\bar{n}]$ and $x \in \mathbb{F}_2^k$, $\mathcal{C}(x)_i$ is a function of the bits $\mathcal{C}_0(x)_{\Gamma(i)}$, where $\Gamma(i) \subseteq [n]$ is a small, carefully chosen, subset of the coordinates of $\mathcal{C}_0$. When we take the aforementioned function to be the parity function, i.e.,

$$\mathcal{C}(x)_i = \bigoplus_{j \in \Gamma(i)} \mathcal{C}_0(x)_j,$$

the code $\mathcal{C}$ is called the direct sum lift of $\mathcal{C}_0$ w.r.t. Γ. The goal is thus to start with $\mathcal{C}_0$ that is ε_0-balanced and argue that the lifted code $\mathcal{C}$ is ε-balanced, for $\varepsilon \ll \varepsilon_0$. A good Γ, that would fulfil this goal, is dubbed a *parity sampler*. See Section 2.1 for a slightly more general definition. Also, see [NN93, ABN+92, Bog12, TS17] for previous works that utilize direct sum lifting for distance amplification.

Our Γ, roughly speaking, consists of short walks over a hypergraph over n vertices. Toward giving a more detailed overview, let us define the desired hypergraphs more formally.

Mixing hypergraphs. Let $H = (V, E)$ be a d-regular 3-uniform hypergraph over $V = [n]$. That is, E contains "hyperedges" of the form (w_1, w_2, w_3), and for each $v \in V$ and $j \in [3]$ we have that $v = w_j$ for exactly d hyperedges. We say H is λ-*mixing* if for any $S_1, S_2, S_3 \subseteq V$, it holds that

$$\left| \frac{E(S_1, S_2, S_3)}{d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2} \right| \le \lambda \cdot \sqrt{|S_1| \cdot |S_3|},$$

where $E(S_1, S_2, S_3)$ is the number of hyperedges $(w_1, w_2, w_3) \in E$ where $w_j \in S_j$ for $j = 1, 2, 3$. This notion and some variants of it were studied before, and we refer to Section 3 for the relevant discussion. In this work we show that a random hypergraphs is $\lambda = O(1/\sqrt{d})$-mixing (see Corollary 3.12), but unfortunately we are not aware of any explicit construction that achieves such a good dependence on d.

Walks on hypergraphs. Set $t = t(\varepsilon)$ be a desired walk length. Starting from a random $\boldsymbol{v}_0 \sim V$, we walk on H according to uniformly random $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_t \sim [d]$ as follows. For each $j \in [t]$,

1. Let $\boldsymbol{e}_j$ be the $\boldsymbol{i}_j$-th hyperedge that touches $\boldsymbol{v}_{j-1}$ according to some fixed ordering. In particular, we require that $\boldsymbol{v}_{j-1} = (\boldsymbol{e}_j)_1$.
2. Denote $\boldsymbol{v}_j = (\boldsymbol{e}_t)_3$.
3. Denote $\boldsymbol{w}_j = (\boldsymbol{e}_t)_2$.

Γ comprises all the walks

$$\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_t).$$

Note that we choose to query $\boldsymbol{w}_j$, but use $\boldsymbol{v}_j$ to determine the next step of our walk. Our lifted $\mathcal{C} \subseteq \mathbb{F}_2^{\bar{n}}$ will therefore have blocklength $\bar{n} = n \cdot d^t$.[6] Choosing parameters appropriately and using a λ-mixing H that satisfies $\lambda = O(1/\sqrt{d})$, we achieve the rate $\frac{k}{\bar{n}}$ that is given in Theorem 3.

In Section 1.2 below we briefly discuss how we analyze these walks, and how we are able to improve upon previous constructions that are also based on random walk over expanders.

[6] We defer subtleties regarding sampling a single walk multiple times to the technical sections.

Non-backtracking walks on λ-spectral hypergraphs. It turns out that we can get an even better rate, of $\widetilde{\Omega}(\varepsilon^2)$, by walking over hypergraphs with an even better dependence on d. Toward this end, we need a strengthening of our λ-mixing property, which we call λ-spectral. We say that H is λ-spectral if for any $x, y, z \in \mathbb{R}^n$, it holds that

$$\left| \frac{1}{d} \cdot \sum_{(i,j,k)\in E} x_i y_j z_k - \frac{1}{n^2} \cdot \sum_{i\in V} x_i \cdot \sum_{i\in V} y_i \cdot \sum_{i\in V} z_j \right| \le \lambda \cdot \|x\|_2 \cdot \|y\|_\infty \cdot \|z\|_2.$$

In Section 3, we show that a λ-spectral hypergraph is readily a λ-mixing one, and that a λ-mixing hypergraph is λ'-spectral for $\lambda' = O(\lambda \log(1/\lambda))$.

Conjecturing the existence of λ-spectral hypergraphs with λ approaching $2/\sqrt{d}$ (see Open Problem 1), we can slightly modify the above construction to yield a rate of $\widetilde{\Omega}(\varepsilon^2)$, bringing us astonishingly close to the GV bound.

Theorem 4 (informal; see Corollary 6.2)**.** *Assuming the existence of explicit λ-spectral hypergraphs with λ approaching $\frac{2}{\sqrt{d}}$, there exists an explicit family of ε-balanced codes $\mathcal{C} \subseteq \mathbb{F}_2^n$ of rate*

$$\varepsilon^2 \cdot \frac{1}{\mathrm{poly}(\log(1/\varepsilon))}$$

that are list- and uniquely-decodable in (probabilistic) near-linear time. The list decoding radius is $\frac{1}{2} - \frac{1}{\mathrm{poly}(\log(1/\varepsilon))}$.

For our modified construction, we replace the above random walks over H with *non-backtracking* walks, thus not "wasting" any randomness on returning steps. Analyzing the refined construction naturally requires working with non-symmetric operators, and in Section 6 we extend upon spectral decomposition results of Lubetzky and Peres [LP16]. We remark that we are not aware of many cases in which *directed* spectral graph theory is used in TCS, and our work demonstrates such an application.

Explicitness. An ε-biased sample space over $\{0,1\}^k$ is a set $S \subseteq \{0,1\}^k$ such that for any nonzero test $\alpha \subseteq [k]$, it holds that

$$\left| \Pr_{\boldsymbol{s}\sim S}\left[\bigoplus_{i\in\alpha} \boldsymbol{s}_i = 0\right] - \Pr_{\boldsymbol{s}\sim S}\left[\bigoplus_{i\in\alpha} \boldsymbol{s}_i = 1\right] \right| \le \varepsilon.$$

It is well-known that linear ε-balanced codes are equivalent to ε-biased sample space, by letting the elements of S correspond to rows in the generator matrix of a binary code $\mathcal{C}$. Thus, an *explicit* ε-balanced code $\mathcal{C} \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ gives rise to an explicit ε-biased sample space $S \subseteq \{0,1\}^k$ of cardinality n. In our construction, the only non-explicit ingredient is the λ-mixing, or λ-spectral, hypergraph. Thus, coming up with such explicit hypergraphs would readily yield explicit (or even fully-explicit) small-biased spaces with better dependence on ε than the ones implied by Theorem 1.[7]

[7] For ε-biased sample spaces we don't need to take a base code $\mathcal{C}_0$ that is efficiently encodable. Thus, given explicit good hypergraph and a suitable $\mathcal{C}_0$ (say, from [NN93]), we would be able to construct our ε-biased sample spaces in time polynomial in n and $1/\varepsilon$ for any $\varepsilon > 0$.

1.2 On (re)breaking the rate-$\Omega(\varepsilon^4)$ barrier of random walks

Recall that our construction uses a parity sampler Γ to amplify the distance of a base code that is ε_0-balanced to a one that is ε-balanced. Toward that goal, we will require that for a worst-case $S \subseteq [n]$ satisfying $|\mathbb{E}_{\boldsymbol{i}\in[n]}[(-1)^{\mathbb{1}[\boldsymbol{i}\in S]}| \le \varepsilon_0$, that

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{w}\in\Gamma} \left[\prod_{j=1}^{|\boldsymbol{w}|} (-1)^{\mathbb{1}[\boldsymbol{w}_j \in S]} \right] \right| \le \varepsilon. \tag{1}$$

There is a simple, albeit inefficient, way to construct such a parity sampler: Take Γ to include all elements of $[n]^t$. This ensures $\varepsilon = \varepsilon_0^t$, however, then $|\Gamma| \triangleq \bar{n} = n^t$, which is obviously too large and would lead to a code with a vanishing rate. Thus, we seek a sparsification of the trivial parity sampler.

Random walks on graphs and the ε^4-barrier. Building off of ideas of Rozenman and Wigderson, Ta-Shma [TS17] suggested replacing the above fully independent construction with a random walk of length t on an n-vertex expander, associating each vertex of the expander with an element of $[n]$. If the graph used, G, is d-regular, then this construction leads to $\bar{n} = nd^{t-1}$, a substantial improvement from $|\Gamma| = n^t$.

We overload G to represent the graph's normalized adjacency matrix and let Π be the diagonal matrix in which $\Pi_{i,i} = (-1)^{\mathbb{1}[i\in S]}$. One can verify that if Γ consists of all length-t walks on G, Equation (1) is satisfied for

$$\varepsilon = \left| \frac{1}{n} \mathbf{1}^\dagger \left(\Pi G\right)^t \mathbf{1} \right| \le \left\| (\Pi G)^t \right\|_{\mathrm{op}},$$

where $\mathbf{1}$ is the all-ones vector and $\|\cdot\|_{\mathrm{op}}$ is the operator norm $\|A\|_{\mathrm{op}} = \max_{x\neq 0} \|Ax\|_2/\|x\|_2$. As a first attempt, we could try to bound $\left\|(\Pi G)^t\right\|_{\mathrm{op}} \le \|\Pi G\|_{\mathrm{op}}^t$. When a vector v is perpendicular to $\mathbf{1}$, we have that $\|\Pi G v\|_2 \le \|Gv\|_2 \le \lambda\|v\|_2$. Unfortunately, when a vector v is parallel to $\mathbf{1}$, we have that $\|\Pi G v\|_2 = \|Gv\|_2 = \|v\|_2$ because $G\mathbf{1} = \mathbf{1}$, meaning that $\|\Pi G\|_{\mathrm{op}} = 1$.

Ta-Shma observed that in the latter case, the *second* step works in our favor. This is because $\Pi\mathbf{1}$ is "mostly" (depending on how small ε_0 is) perpendicular to $\mathbf{1}$. In particular, he showed that $\|\Pi G \Pi G \mathbf{1}\|_2 \le (\lambda + \varepsilon_0)\|\mathbf{1}\|_2$. Intuitively, at least one out of every two steps "works",[8] which is sufficient to guarantee a rate of $\approx \varepsilon^4$ by taking a good enough G. That is still far from the GV bound of $\approx \varepsilon^2$.

Breaking the ε^4-barrier. To break the barrier, Ta-Shma uses an intricately-designed random walk on a graph product called the *s-wide replacement product* (introduced in [BATS11]), to guarantee that $s - O(1)$ out of every s steps work, for some $s < t$. Here, we diverge from Ta-Shma's approach. We will only aim for one out of every two steps to work, but will share randomness between the two steps in order to make them as cheap as a single step.

Specifically, let G_1 and G_2 be two degree-d expanders on the same n vertices. In order to take two coupled steps from a vertex v_1, we draw a random $\boldsymbol{j} \in [d]$ and move to v_2, the $\boldsymbol{j}^{\text{th}}$ neighbor of v_1 in G_1 (according to some fixed ordering). Then, we move to v_3, the $\boldsymbol{j}^{\text{th}}$ neighbor of v_2 in G_2. As we use the same label $\boldsymbol{j}$ for both steps, this walk can take ℓ "double steps" with a support size

[8] That phenomenon, of losing one λ factor in every two steps, is not a mere artifact of the proof, at least if one makes not further assumptions on the construction's primitives. See [BATS11, TS17] for relevant discussions.

of only nd^{ℓ}. In contrast, if the steps were chosen independently, the support size would be $nd^{2\ell}$. If we could guarantee that a double step is as productive as two independent steps, the rate of the resulting code would be $\varepsilon^{2+o(1)}$.

For the double step to work, clearly there must be some relation between the two expanders. Otherwise, G_2 could always reverse the step taken by G_1. Hence, we would like to think of G_1 and G_2 together as a single primitive: For each vertex v_1, there are d choices for the pair (v_2, v_3). As a result, we can think of G_1 and G_2 together as a single d-regular 3-uniform hypergraph, and consider walks on that hypergraph, $H = (V, E_H)$, motivating the construction in Section 1.1.

To analyze this walk, we introduce an operator, $A^{(S)} \in \mathbb{R}^{V \times V}$. For each $i, k \in [n]$, we set

$$A^{(S)}_{i,k} \triangleq \frac{1}{d} \cdot \sum_{j:(i,j,k)\in E_H} (-1)^{\mathbb{1}[j\in S]}.$$

Then, for Γ corresponding to the length-t "double-step" construction, we show Equation (1) holds for

$$\varepsilon = \left| \frac{1}{n} \mathbf{1}^{\dagger} \left(\Pi A^{(S)} \right)^{t} \mathbf{1} \right| \leq \left\| \left(\Pi A^{(S)} \right)^{t} \right\|_{\mathrm{op}}.$$

If H is λ-spectral, it is simple to bound $\|\Pi A^{(S)}\|_{\mathrm{op}} = \|A^{(S)}\|_{\mathrm{op}} \leq \lambda + \varepsilon_0$ (see Proposition 4.5), which gives a bound of $\varepsilon = (\lambda + \varepsilon_0)^t$ and is sufficient for rate $\approx \varepsilon^2$.

Lastly, we note that because Π is unitary, the entire analysis goes through if instead of bounding $\|(\Pi A^{(S)})^t\|_{\mathrm{op}}$, we instead bound $\|(A^{(S)})^t\|_{\mathrm{op}}$. This corresponds to the a double-step construction where we only record every other vertex visited (starting with the second), which is exactly our construction in Section 1.1.

Bounding $A^{(S)}$, given the right notion of hypergraph expansion, is easier than analyzing Ta-Shma's s-wide replacement product, so we think of our construction as conceptually simpler. For our approach, the challenge is to construct sufficiently good hypergraphs. We are not aware of any explicit constructions, but are able to show that a random hypergraph suffices for decoding our code and obtaining Theorem 3.

1.3 Decoding our codes

Our decoding result in Theorem 3 follows the framework of Jeronimo et al. [JST21]. They used a novel algorithmic weak regularity lemma to show that direct sum lifts are decodable, roughly speaking, given that the parity sampler Γ used for the lifting satisfies the *splittability* condition (we refer the reader to [JST21] for the precise definition). While we suspect that our Γ is *not* splittable, we distill a weaker property that suffices for the [JST21] framework to work.

This property, which we call τ-sampling, tells us that we can use $\Gamma \subseteq [n]^t$ to sample any set $S \subseteq [n]$, starting from any prefix. Namely, for every $i \in [t]$ and $X \subseteq [n]^{i-1}$, we require that

$$\left| \Pr_{\boldsymbol{w}\in\Gamma} \left[\boldsymbol{w}_i \in S \mid (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{i-1}) \in X \right] - \rho(S) \right| \leq \frac{\tau}{\rho(X)},$$

where $\rho(A)$, for some subset $A \subseteq [m]$, is its density $\frac{|A|}{m}$. For the more general definition, and further discussion, see Section 5.2. We believe that this strong mixing property, which still falls short of full-fledged splittability, is an interesting notion in itself. In Section 5.2, we show that our Γ is indeed τ-sampling, thereby allowing us to unique- and list-decode our code $\mathcal{C}$ in $\widetilde{O}_{\varepsilon}(n)$ time.

2 Preliminaries

For integers a, b, we use $[a, b]$ to denote the set $\{a, \ldots, b\}$ and $[n]$ as a shorthand for $[1, n]$. Given a set $S \subseteq [n]$, when the ground set $[n]$ is clear from context, we denote $\rho(S) = \frac{|S|}{n}$, and its ± 1 variant as $\text{bias}(S) = 1 - 2\rho(S)$. For $z \in \mathbb{F}_2^n$, we similarly denote $\text{bias}(z)$ as the bias of its characteristic set, i.e., $\mathbb{E}_{i \in [n]}[(-1)^{z_i}]$. We use boldface letters to denote random variables, except for $\mathbf{1} \in \mathbb{R}^n$, which we use for the all-ones vector. Also, when bounding running time, by writing $g(n) = \exp(f(n))$ we mean that $g(n) \leq 2^{c \cdot f(n)}$ for some universal constant $c > 0$.

Definition 2.1 (discretizable distribution)**.** *For $M \in \mathbb{N}$, We say that a distribution $\mathcal{W}$ is M-discretizable if it satisfies either of the following two equivalent properties.*

1. *For any x in the support of $\mathcal{W}$, $\Pr_{\boldsymbol{x} \sim \mathcal{W}}[\boldsymbol{x} = x] = i/M$ for some $i \in \mathbb{N}$.*
2. *Let $\mathcal{U}_M$ be the uniform distribution over $[M]$. Then, there is some function f mapping $[M]$ to the support of $\mathcal{W}$ for which $f(\boldsymbol{i})$, where $\boldsymbol{i} \sim \mathcal{U}_M$, has the same distribution as a sample from $\mathcal{W}$.*

We say that $\mathcal{W}$ is computable in (deterministic or probabilistic) time T if f above is computable in time T. In particular, $\mathcal{W}$ is explicit if it is computable in deterministic time $\text{poly}(M)$*, and fully explicit if it is computable in deterministic time* $\text{poly}(\log M)$*.*

Definition 2.2 (homogeneous distribution)**.** *For $n, t \in \mathbb{N}$, we say that a distribution $\mathcal{W}$ over $[n]^t$ is* homogeneous *if its restriction to any coordinate is uniform over $[n]$, i.e., if for any $i \in [t]$ and $a \in [n]$ it holds that $\Pr_{\boldsymbol{w} \sim \mathcal{W}}[\boldsymbol{w}_i = a] = \frac{1}{n}$.*

For any domain $\mathcal{X}$ and two distributions $\mathcal{D}, \mathcal{D}'$ over $\mathcal{X}$ we define the *total variation distance* of $\mathcal{D}$ and $\mathcal{D}'$ in terms of the optimal test distinguishing the distributions, i.e.,

$$d_{\text{TV}}(\mathcal{D}, \mathcal{D}') \triangleq \sup_{T \colon \mathcal{X} \to [0,1]} \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{D}}[T(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{D}'}[T(\boldsymbol{x}')] \right\}.$$

For a matrix $A \in \mathbb{R}^{n \times n}$, we denote by $\|A\|_{\text{op}}$ its operator norm $\|A\|_{\text{op}} = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$, which is also the maximum of $x^\dagger A y$ over all norm-1 vectors $x, y \in \mathbb{R}^n$.

Error correcting codes. A binary error correcting code of message length k and blocklength n is a mapping $\mathcal{C} \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$, which we will often identify with its image $\text{Im}(\mathcal{C}) \subseteq \mathbb{F}_2^n$. The *rate* of $\mathcal{C}$ is $\frac{k}{n}$, and its *relative distance* is $\Delta(\mathcal{C}) = \frac{1}{n} \min_{z \neq z'} \Delta(z, z')$ for $z, z' \in \mathcal{C}$, and $\Delta(z, z') = |\{i \in [n] : z_i \neq z'_i\}|$ being the Hamming distance. The Hamming ball of (relative) radius β centered at z is the set $B(z, \beta) = \{z' \in \mathbb{F}_2^n : \Delta(z, z') \leq \beta\}$.

We denote $\text{bias}(\mathcal{C})$ as the maximal bias, in absolute value, of every nonzero $z \in \mathcal{C}$. Thus, $\text{bias}(\mathcal{C}) \leq \varepsilon$ if the Hamming weight of any nonzero codeword is in $\left[\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right]$.

Definition 2.3 (balanced codes)**.** *A linear binary error correcting code is ε-balanced if* $\text{bias}(\mathcal{C}) \leq \varepsilon$*.*

In particular, an ε-balanced code $\mathcal{C}$ has distance at least $\frac{1-\varepsilon}{2}$.

Unique and list decoding. We say that $\mathcal{C}$ is (combinatorially) (β, L) list decodable if for every $z \in \mathbb{F}_2^n$, $|\mathcal{C} \cap B(z, \beta)| \leq L$. The Johnson bound tells us that any ε-balanced code is $(1/2 - \sqrt{\varepsilon}, L)$ list decodable for $L = O(1/\varepsilon)$. The *algorithmic* list decoding problem aims at finding the list $\mathcal{L}_{\mathcal{C},\beta}(z) \triangleq \mathcal{C} \cap B(z, \beta)$. When $\beta \leq \frac{1-\varepsilon}{4}$ and $\mathcal{C}$ is ε-balanced, we know that $\mathcal{L}_{\mathcal{C},\beta}$ always contains at most one codeword, which corresponds to the unique decoding problem.

2.1 Parity samplers and direct sum codes

We will be interested in constructing a binary linear code $\mathcal{C} \subseteq \mathbb{F}_2^{\bar{n}}$ with small bias by amplifying the (moderate) bias of some base code $\mathcal{C}_0 \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$. One natural way to do so is by XORing t-tuples of $\mathcal{C}_0$ according to some distribution $\mathcal{W} \sim [n]^t$.

Definition 2.4 (direct sum codes). *For $t, n, \bar{n} \in \mathbb{N}$, let $\mathcal{W} \sim [n]^t$ be an $\bar{n}$-discretizable distribution equipped with a corresponding mapping function $f_{\mathcal{W}} \colon [\bar{n}] \to [n]^t$. For $z \in \mathbb{F}_2^n$, we let* $\mathrm{dsum}_{\mathcal{W}}(z) \in \mathbb{F}_2^{\bar{n}}$ *be such that*

$$\mathrm{dsum}_{\mathcal{W}}(z)[\ell] = \sum_{i=1}^{t} z\left[f_{\mathcal{W}}(\ell)_i\right]$$

where the addition is taken over $\mathbb{F}_2$. Given a code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$, the direct sum lift of $\mathcal{C}_0$ according to $\mathcal{W}$ is the code

$$\mathrm{dsum}_{\mathcal{W}}(\mathcal{C}_0) = \{\mathrm{dsum}_{\mathcal{W}}(z) : z \in \mathcal{C}_0\} \subseteq \mathbb{F}_2^{\bar{n}}.$$

Definition 2.5 (parity sampler). *For $t, n \in \mathbb{N}$, and $0 \leq \varepsilon < \varepsilon_0 \leq 1$, we say that $\mathcal{W} \sim [n]^t$ is an $(\varepsilon_0, \varepsilon)$ parity sampler if for every $z \in \mathbb{F}_2^n$ with $|\mathrm{bias}(z)| \leq \varepsilon_0$ it holds that $|\mathrm{bias}(\mathrm{dsum}_{\mathcal{W}}(z))| \leq \varepsilon$.*

Clearly, if $\mathcal{C}_0 \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ is ε_0-balanced and $\mathcal{W} \sim [n]^t$ is an $\bar{n}$-discretizable $(\varepsilon_0, \varepsilon)$ parity sampler, the lifted code $\mathcal{C} = \mathrm{dsum}_{\mathcal{W}}(\mathcal{C}_0)$ is ε-balanced with rate $\frac{k}{\bar{n}}$.

3 Expanding 3-Uniform Hypergraphs

Our construction uses a family of expanding d-regular 3-uniform hypergraphs.

Definition 3.1 (d-regular 3-uniform hypergraph). *A 3-uniform hypergraph consists of a set of vertices, V, and hyperedges $E \subseteq V^3$. The hypergraph $H = (V, E)$ is d-regular if, for each $v \in V$ and $j \in [3]$, the number of hyperedges $(w_1, w_2, w_3) \in E$ for which $v = w_j$ is d.*

We will set d carefully in our construction, but for now, it can be thought of as an arbitrary constant. For the remainder of this section, we will use “hypergraph” as shorthand for d-regular 3-uniform hypergraph.

There are various notions of expansion for hypergraphs and they are not all equivalent. In this work, we consider two such notions.

Definition 3.2 (λ-mixing hypergraph). *A d-regular hypergraph $H = (V, E)$ on n vertices is λ-mixing if, for any $S_1, S_2, S_3 \subseteq V$,*

$$\left| \frac{E(S_1, S_2, S_3)}{d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2} \right| \leq \lambda \cdot \sqrt{|S_1| \cdot |S_3|}, \tag{2}$$

where $E(S_1, S_2, S_3)$ is the number of hyperedges $(w_1, w_2, w_3) \in E$ where $w_j \in S_j$ for $j = 1, 2, 3$.

Note that the right-hand side of the above definition does not depend on the size of S_2.[9] That parallels the definition below, in which we use $\|y\|_\infty$ instead of $\|y\|_2$.

Definition 3.3 (λ-spectral hypergraph)**.** *A d-regular hypergraph $H = (V, E)$ on n vertices is λ-spectral if, for any $x, y, z \in \mathbb{R}^n$,*

$$\left| \frac{1}{d} \cdot \sum_{(i,j,k)\in E} x_i y_j z_k - \frac{1}{n^2} \cdot \sum_{i\in V} x_i \cdot \sum_{i\in V} y_i \cdot \sum_{i\in V} z_j \right| \le \lambda \cdot \|x\|_2 \cdot \|y\|_\infty \cdot \|z\|_2. \tag{3}$$

We will only care about cases where $y \in \{\pm 1\}^n$.[10] In those cases, we have $\|y\|_\infty = 1$ and $\|y\|_2 = \sqrt{n}$. It is thus tempting to replace the $\|y\|_\infty$ in the right-hand side of Equation (3) with $\frac{\|y\|_2}{\sqrt{n}}$, as ℓ_2 norms are often easier to work with than ℓ_∞ norms. Unfortunately, no "good" λ-spectral hypergraphs would exist with that modification: Whenever $n \gg d^2$, it is straightforward to bound $\lambda = \Omega(\sqrt{n}/d)$. For our definition, as we shall soon see, it is possible to achieve $\lambda \approx 1/\sqrt{d}$.

A variant of the spectral definition, where indeed one takes $\|y\|_2$ instead of $\|y\|_\infty$, was first studied by Friedman and Wigderson [FW95], who also showed that the spectral definition implies combinatorial mixing. They considered much larger edge densities than us (corresponding to $d > n$ for our definition). Hypergraphs with similar combinatorial mixing properties were previously constructed from random walks on expanders [BH04], from Ramanujan complexes and other spectral properties of simplicial complexes (e.g., [LM15, PRT16, CMRT16, Par17, GP19]), and from Cayley graphs [CTZ20]. However, to the best of our knowledge, no explicit construction achieves λ smaller than $\approx \frac{1}{d^{1/3}}$. A simple hypergraph construction would be to take all length-2 walks on a Ramanujan expander as the hyperedges. This was considered by [BH04], who showed it can achieve $\lambda \approx \frac{1}{d^{1/4}}$.[11]

Similar to standard graphs, spectral expansion implies mixing.

Proposition 3.4 (spectral $\implies$ mixing)**.** *For any $\lambda > 0$, if H is a λ-spectral hypergraph, it is also a λ-mixing hypergraph.*

Proof. For any $S_1, S_2, S_3 \subseteq V$, let $x_i = \mathbb{1}[i \in S_1]$, $y_i = \mathbb{1}[i \in S_2]$, and $z_i = \mathbb{1}[i \in S_3]$. Then,

$$E(S_1, S_2, S_3) = \sum_{(i,j,k)\in E} x_i y_j z_k.$$

Equation (2) follows directly from Equation (3). $\square$

By applying the converse to the expander mixing lemma for ordinary graphs [BL06], we can show that for symmetric hypergraphs, mixing implies spectral expansion with only a minor quantitative gap.

Definition 3.5 (symmetric 3-uniform hypergraph)**.** *We say a 3-uniform hypergraph $H = (V, E)$ is* symmetric *if, for any edge $e = (v_1, v_2, v_3) \in E$, the edge (v_3, v_2, v_1) is also in E.*

[9] Some works consider the stronger requirement of $\sqrt{|S_{\sigma(1)}| \cdot |S_{\sigma(2)}|}$ instead of $\sqrt{|S_1| \cdot |S_3|}$, for $S_{\sigma(1)}$ and $S_{\sigma(2)}$ being the two smallest sets.

[10] In fact, for any fixed choice of x and z, the left-hand side of Equation (3) is linear in y. Therefore, it is maximized for some $y \in \{\pm 1\}^n$ and so in general it is sufficient to consider only such y.

[11] In [BH04], Bilu and Hoory used hypergraphs for the construction of asymptotically good codes, generalizing Tanner's expander codes [Tan81] and their decoding [SS96, Zém01].

Proposition 3.6 (mixing $\implies$ spectral)**.** *There exists a universal constant $c_{\mathsf{spec}} > 0$ for which the following holds. Let $H = (V = [n], E)$ be any d-regular hypergraph, and $\lambda \leq \frac{1}{2}$. If H is λ-mixing and symmetric, then H is a λ'-spectral for $\lambda' = c_{\mathsf{spec}} \cdot \lambda \log(1/\lambda)$.*

The proof of Proposition 3.6 is a simple application of a similar result for graphs.

Lemma 3.7 (Lemma 3.3 of [BL06])**.** *There exists a universal constant c_{BL} for which the following hods. For any $n \times n$ real symmetric matrix A and $\lambda \leq \frac{1}{2}$, suppose each row has an ℓ_1 norm of at most 1 and for any two vectors $u, v \in \{0,1\}^n$,*[12]

$$|u^\dagger A v| \leq \lambda \cdot \|u\| \cdot \|v\|. \tag{4}$$

Then, the spectral radius of A is at most $c_{\mathsf{BL}} \cdot \lambda \log(1/\lambda)$.

If G is the (normalized) transition matrix of a d-regular graph, and $A = G - \frac{1}{n}J$ for J being the all-ones matrix, then Lemma 3.7 shows a converse to the expander mixing lemma: Any graph with good mixing is also a good spectral expander. We show that their result can be lifted to 3-uniform hypergraphs.

Proof of Proposition 3.6. Fix any $y \in \mathbb{R}^n$ and let $A^{(y)} \in \mathbb{R}^{n \times n}$ be defined as

$$A^{(y)}_{i,k} \triangleq \frac{1}{2d} \cdot \left(\sum_{j \in [n]} \mathbb{1}[(i,j,k) \in E] \cdot y_j \right) - \frac{1}{2n^2} \cdot \sum_{j \in [n]} y_j.$$

Then, for any $x, z \in \mathbb{R}^n$,

$$2 \cdot (z^\dagger A^{(y)} x) = \frac{1}{d} \cdot \sum_{(i,j,k) \in E} x_i y_j z_k - \frac{1}{n^2} \cdot \sum_{i \in V} x_i \cdot \sum_{i \in V} y_i \cdot \sum_{i \in V} z_j.$$

Therefore, in order to prove that H is an λ'-spectral expander, it is sufficient to show that for all $y \in \mathbb{R}^n$, the operator norm of $A^{(y)}$ is at most $\|y\|_\infty \cdot \frac{\lambda'}{2}$. We observe that:

1. For any fixed $x, z \in \mathbb{R}^n$, the quantity $z^\dagger A^{(y)} x$ is a linear function of y. Thus, we can consider y-s with $\|y\|_\infty = 1$ without loss of generality. Furthermore, the y maximizing $z^\dagger A^{(y)} x$ among those with $\|y\|_\infty = 1$ will be in $\{\pm 1\}^n$. Therefore we assume $y \in \{\pm 1\}^n$ also without loss of generality.

2. Since H is symmetric, the operator norm and spectral radius of $A^{(y)}$ are equal, so we instead bound the spectral radius.

We will apply Lemma 3.7 to each $A^{(y)}$. Note that:

- $A^{(y)}$ is symmetric, which follows immediately from the fact that H is symmetric.
- The ℓ_1 norm of each row of $A^{(y)}$ is bounded by 1:

$$\sum_{k \in [n]} \left| \frac{1}{2d} \cdot \left(\sum_{j \in [n]} \mathbb{1}[(i,j,k) \in E] \cdot y_j \right) + \frac{1}{2n^2} \cdot \sum_{j \in [n]} y_j \right| \leq \frac{1}{2d} \cdot d + \frac{1}{2n^2} \cdot n \leq 1.$$

[12] Note that Bilu and Linial's Lemma statement only has the weaker requirement that this hold for orthogonal u and v. As a result, they have an additional condition that the diagonal entries of A not be too large, which is not needed for our version of the Lemma.

Finally, fix some $u, v \in \{0,1\}^n$, and define the sets

$$\begin{aligned} S_1 &= \{i \in V : u_i = 1\} \\ S_3 &= \{i \in V : v_i = 1\} \\ S_2^+ &= \{i \in V : y_i = 1\} \\ S_2^- &= \{i \in V : y_i = -1\}. \end{aligned}$$

Then,

$$\begin{aligned} \left|u^\dagger A^{(y)} v\right| &= \frac{1}{2} \cdot \left|\frac{1}{d} \cdot \left(E(S_1, S_2^+, S_3) - E(S_1, S_2^-, S_3)\right) - \frac{|S_1| \cdot |S_3| \cdot (|S_2^+| - |S_2^-|)}{n^2}\right| \\ &\le \frac{1}{2} \cdot \left|\frac{1}{d} \cdot E(S_1, S_2^+, S_3) - \frac{|S_1| \cdot |S_3| \cdot |S_2^+|}{n^2}\right| + \frac{1}{2} \cdot \left|\frac{1}{d} \cdot E(S_1, S_2^-, S_3) - \frac{|S_1| \cdot |S_3| \cdot |S_2^-|}{n^2}\right| \\ &\le \lambda \sqrt{|S_1| \cdot |S_3|}, \end{aligned}$$

where the last inequality follows from fact that H is λ-mixing (applied to both expressions). Thus, $|u^\dagger A^{(y)} v| \le \lambda \cdot \|u\| \cdot \|v\|$ and we can apply Lemma 3.7 to obtain $\|A^{(y)}\|_{\mathrm{op}} = c_{\mathsf{BL}} \cdot \lambda \log(1/\lambda)$, implying that H is $(\lambda' = 2c_{\mathsf{BL}} \cdot \lambda \log(1/\lambda))$-spectral. □

3.1 Random hypergraphs mix well

In this section, we'll show that given an expanding graph G, its *random hypergraph completion* is, with high probability, a good expander. Throughout, we say that an undirected regular graph G is a λ-expander if the second largest eigenvalue of its normalized adjacency matrix, in magnitude, is at most λ.

Definition 3.8 (random hypergraph completion of a graph.)**.** *Let $G = (V = [n], E_G)$ be a d-regular graph. To sample a random hypergraph completion, $\boldsymbol{H}$ of G, we choose a uniformly random ordering of G's edges, $\{(\boldsymbol{u}_1, \boldsymbol{v}_1), \dots, (\boldsymbol{u}_{nd}, \boldsymbol{v}_{nd})\} = E_G$. Then, we set $\boldsymbol{H} = (V, \boldsymbol{E_H})$, where*

$$\boldsymbol{E_H} \triangleq \{(\boldsymbol{u}_i, \lceil i/d \rceil, \boldsymbol{v}_i) \mid i \in [nd]\}.$$

Equivalently, for each $(u, v) \in E_G$, we choose an independent and uniform $\boldsymbol{w} \in V$ and add the hyperedge $(u, \boldsymbol{w}, v)$ to E_H, conditioned on the resulting hypergraph $\boldsymbol{H}$ being d-regular.

Next, we prove that the random hypergraph completion mixes well with high probability.

Lemma 3.9. *Let $G = (V = [n], E)$ be a d-regular λ-expander and $\boldsymbol{H}$ be a random hypergraph completion of G. With probability at least $1 - 2^{-n}$, $\boldsymbol{H}$ is a λ'-mixing hypergraph for $\lambda' = 2\lambda + \frac{2}{\sqrt{d}}$.*

In order to prove Lemma 3.9, we'll use Hoeffding's for sampling *without* replacements.

Fact 3.10 (Hoeffding's inequality, [Hoe63])**.** *For any integers $a, k \le m$, suppose there are m items, of which k of them are marked. Let $\boldsymbol{x}$ be a random variable indicating the number of marked items when a of the m items are sampled uniformly and independently* without replacement*. Then, for any $t \ge 0$,*

$$\Pr\left[\left|\boldsymbol{x} - \frac{k}{m} \cdot a\right| \ge t\right] \le 2 \exp\left(\frac{-2t^2}{a}\right).$$

Proof of Lemma 3.9. Fix arbitrary $S_1, S_2, S_3 \subseteq V$. We will show that with probability at least $1 - 2^{-4n}$ it holds that

$$\left|\frac{E_{\boldsymbol{H}}(S_1, S_2, S_3)}{d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2}\right| \le \left(\frac{2}{\sqrt{d}} + 2\lambda\right) \cdot \sqrt{|S_1| \cdot |S_3|} \tag{5}$$

where $E_{\boldsymbol{H}}(S_1, S_2, S_3)$ is the number of edges (v_1, v_2, v_3) in $\boldsymbol{H}$ where $v_j \in S_j$ for each $j \in [3]$. The desired result then follows from a union bound over the $(2^n)^3 = 2^{3n}$ choices for S_1, S_2, S_3. Let $E_G(S_1, S_3)$ be the number of edges, (u, v), of G, such that $u \in S_1$ and $v \in S_3$. By the expander mixing lemma applied to G, we have that

$$\left|\frac{E_G(S_1, S_3)}{d} - \frac{|S_1||S_3|}{n}\right| \le \lambda\sqrt{|S_1||S_3|}.$$

Let us define $\mu \triangleq \frac{|S_1||S_3|}{n}$ and $\Delta \triangleq \lambda\sqrt{|S_1||S_3|}$. We consider two cases.

1. In the first case, $\mu \le \Delta$. Here, we use the simple bound

   $$0 \le \frac{E_{\boldsymbol{H}}(S_1, S_2, S_3)}{d} \le \frac{E_G(S_1, S_3)}{d} \le 2\Delta$$

   that holds with probability 1. This, along with the fact $\frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2} \le \mu \le \Delta$ implies that Equation (5) always holds.

2. In the second case, $\mu > \Delta$. Here, we will apply Fact 3.10. G has a total of nd edges, of which $d \cdot |S_2|$ are matched to a vertex in S_2. $E_{\boldsymbol{H}}(S_1, S_2, S_3)$ samples $E_G(S_1, S_3)$ of those nd edges (without replacement) and counts how many were among the $d \cdot |S_2|$ assigned to S_2. Hence, by Hoeffding's inequality, we have for any $t \ge 0$,

   $$\Pr\left[\left|E_{\boldsymbol{H}}(S_1, S_2, S_3) - \frac{d \cdot |S_2|}{nd} \cdot E_G(S_1, S_3)\right| \ge t\right] \le 2\exp\left(\frac{-2t^2}{E_G(S_1, S_3)}\right).$$

   Then, setting $t = 2\sqrt{d|S_1| \cdot |S_3|}$ and using the fact that $E_G(S_1, S_3) \le d(\mu + \Delta) \le 2d\mu$,

   $$\begin{aligned}\Pr\left[\left|E_{\boldsymbol{H}}(S_1, S_2, S_3) - \frac{|S_2|}{n} \cdot E_G(S_1, S_3)\right| \ge 2\sqrt{d|S_1| \cdot |S_3|}\right] &\le 2\exp\left(\frac{-8d|S_1| \cdot |S_3|}{2d \cdot \frac{|S_1||S_3|}{n}}\right). \\ &= 2\exp(-4.5n) \le 2^{-4n}.\end{aligned}$$

   As the above shows, $E_{\boldsymbol{H}}(S_1, S_2, S_3)$ is within $\pm t$ of its expectation with probability at least 2^{-4n}. When that occurs, we have that

   $$\begin{aligned}\left|\frac{E_{\boldsymbol{H}}(S_1, S_2, S_3)}{d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2}\right| &= \frac{1}{d}\left|E_{\boldsymbol{H}}(S_1, S_2, S_3) - \frac{|S_2|d\mu}{n}\right| \\ &\le \frac{t}{d} + \frac{|S_2|}{n} \cdot \left|\frac{E_G(S_1, S_3)}{d} - \mu\right| \\ &\le \frac{t}{d} + \frac{|S_2|}{n} \cdot \Delta \qquad \text{(expander mixing lemma)} \\ &\le 2\sqrt{\frac{|S_1| \cdot |S_3|}{d}} + \lambda\sqrt{|S_1| \cdot |S_3|}. \qquad (|S_2|/n \le 1)\end{aligned}$$

Hence, Equation (5) holds with probability at least $1 - 2^{-4n}$ in both cases, so we can union bound over the 2^{3n} choices for S_1, S_2, S_3.

□

For our purposes, we will want the hypergraph to be symmetric. It is quite easy to "symmetrize" any hypergraph at only a modest cost to the degree.

Proposition 3.11. *For any $n, d \in \mathbb{N}$, there is an algorithm running in time $O(nd)$ that takes as input any d-regular λ-mixing hypergraph $H = (V = [n], E_H)$ and outputs a $2d$-regular λ-mixing hypergraph $H' = (V = [n], E_{H'})$ over the same vertices.*

Proof. For every edge $(u, v, w) \in E_H$ we include both (u, v, w) and the reverse edge (w, v, u) in $E_{H'}$.[13] Clearly, this results in the degree of H' being $2d$. Then, for any $S_1, S_2, S_3 \subseteq V$,

$$\begin{aligned}
&\left|\frac{E_{H'}(S_1, S_2, S_3)}{2d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2}\right| \\
&\quad = \left|\frac{E_H(S_1, S_2, S_3) + E_H(S_3, S_2, S_1)}{2d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2}\right| \\
&\quad \le \frac{1}{2}\left|\frac{E_H(S_1, S_2, S_3)}{d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2}\right| + \frac{1}{2}\left|\frac{E_H(S_3, S_2, S_1)}{d} - \frac{|S_1| \cdot |S_2| \cdot |S_3|}{n^2}\right| \\
&\le \lambda\sqrt{|S_1| \cdot |S_3|}.
\end{aligned}$$

Therefore, H' is λ-mixing, as desired. □

We have explicit (and even fully explicit) constructions of Ramanujan graphs, i.e., λ-expanders for $\lambda \le \frac{2\sqrt{d-1}}{d}$, albeit with some restrictions on d [LPS88, Mar88]. By manipulating Ramanujan graphs, Alon gave a construction of d-regular λ-expanders over n vertices for any d and n while suffering only a tiny loss in λ (see [Alo21], and also [MRSV21, Gol19] for weaker constructions). In particular, there exist explicit expanders with $\lambda = O(1/\sqrt{d})$ for all n-s. We thus get the following corollary.

Corollary 3.12. *There exists a probabilistic algorithm such that for any integer n and even integer $6 \le d \le n$, runs in time $\mathrm{poly}(n)$ and with probability at least $1 - 2^{-n}$ outputs a 3-uniform symmetric d-regular hypergraph that is $\lambda = \frac{c_{\mathsf{rand}}}{\sqrt{d}}$-mixing, where $c_{\mathsf{rand}} \ge 2$ is some universal constant.*

We will refer to the above probabilistic construction as our *preprocessing step.*

Unfortunately, we do not know how to construct *explicit* mixing, or spectral, hypergraphs with $\lambda \approx 2/\sqrt{d}$. We put forward a concrete goal of constructing "nearly Ramanujan" spectral hypergraphs.

Open Problem 1. *Construct a sufficiently dense infinite family of explicit 3-uniform d-regular hypergraphs which are λ-spectral for $\lambda \le \frac{2}{\sqrt{d}} \cdot (1 + d^c)$, where $c < 0$ is any absolute constant.*

Getting such hypergraphs is an interesting goal on its own right, and it seems that Ramanujan *complexes* are in some sense too strong to yield a hypergraph construction with such a small λ. As we will later see, fulfilling Open Problem 1 would readily give *explicit* ε-balanced codes with rate

[13]Note that, we do this even if it results in duplicated hyperedges: If (u, v, w) and (w, v, u) are both in E_H, then there will be two copies of (u, v, w) and two copies of (w, v, u) in $E_{H'}$.

$\widetilde{\Omega}(\varepsilon^2)$ and efficient decoding, and moreover, by the known connection to small-biased distributions, also an explicit ε-biased distributions over $\mathbb{F}_2^n$ with support size $n \cdot \widetilde{O}(\varepsilon^{-2})$. See Section 6 for the details.

4 From Hypergraphs to Parity Samplers

Fix some $n, t, d \in \mathbb{N}$ and a d-regular 3-uniform hypergraph $H = (V, E_H)$ over n vertices. We will construct a parity sampler $\mathcal{W}_{H,t} \sim [n]^t$ from H and show that when H is a good spectral expander, $\mathcal{W}$ is a good parity sampler.

The construction. For each $v \in V$, there are d edges $(v_1, v_2, v_3) \in E_H$ with $v_1 = v$. For any $i \in [d]$, let $e_H(v, i)$ be the i^{th} such edge (i.e., $e_H(v, i)_1 = v$). Without loss of generality, we assume the vertices are each labeled with a unique integer between 1 and n (i.e $V = [n]$).

To sample $\boldsymbol{w} \sim \mathcal{W}_{H,t}$, independently sample a starting vertex $\boldsymbol{v}_0 \in [n]$ and edge labels $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_t \sim [d]$ uniformly. $\boldsymbol{w}$ will be a deterministic function of $\boldsymbol{v}_0$ and $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_t$ computed as follows. For each $j = 1, \ldots, t$,

1. Let $\boldsymbol{e}_j = e_H(\boldsymbol{v}_{j-1}, \boldsymbol{i}_j)$.
2. Let $\boldsymbol{v}_j = (\boldsymbol{e}_j)_3$.
3. Let $\boldsymbol{w}_j = (\boldsymbol{e}_j)_2$.

The sample is then $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_t)$.

We present two simple claims about $\mathcal{W}_{H,t}$.

Claim 4.1. *For any $t \in \mathbb{N}$ and a d-regular hypergraph H over n vertices, $\mathcal{W}_{H,t}$ is (nd^t)-discretizable.*

Proof. The sample $\boldsymbol{w} \sim \mathcal{W}_{H,t}$ is a deterministic function of $\boldsymbol{v}_0$ and $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_t$, and those variables are set to a uniform choice out of nd^t possibilities. The claim then follows from Definition 2.1. □

Claim 4.2. *For any $t \in \mathbb{N}$, d-regular hypergraph H, and $j \in [t]$, $\boldsymbol{e}_j$ is uniform over all nd edges of H. As a result, $\mathcal{W}_{H,t}$ is homogeneous.*

Proof. We will first prove that each $\boldsymbol{e}_j$ is uniform over E_H by induction on j. $\boldsymbol{e}_1$ is uniform over the nd edges in E_H as it is sampled by independently choosing a starting vertex $\boldsymbol{v}_0 \sim V$ uniformly and then uniformly choosing one of its d-neighbors. Furthermore, for any $j \in [t-1]$, if $\boldsymbol{e}_j$ is uniform, then $\boldsymbol{v}_j$ is also uniform. Hence, $\boldsymbol{e}_{j+1}$ is sampled by selecting a uniform vertex and then (independently) one of its d-neighbors, so $\boldsymbol{e}_{j+1}$ is also uniform over E_H.

Next, we show $\mathcal{W}_{H,t}$ is homogeneous. Fix any $a \in [n]$ and $j \in [t]$. As $\boldsymbol{w}_j = a$ if and only if $\boldsymbol{e}_j$ is one of the d-edges in E_H whose second vertex is a and $\boldsymbol{e}_j$ is uniform over the nd edges in E_H,

$$\Pr[\boldsymbol{w}_j = a] = \frac{d}{nd} = \frac{1}{n}.$$ □

Finally, we show that whenever H is a good expander, $\mathcal{W}_{H,t}$ is a good parity sampler.

Theorem 5. *For any $n, d, t \in \mathbb{N}$, $\lambda, \varepsilon_0 > 0$, and H a λ-spectral d-regular 3-uniform hypergraph on n vertices, $\mathcal{W}_{H,t}$ is an $(\varepsilon_0, \varepsilon \triangleq (\varepsilon_0 + \lambda)^t)$-parity sampler.*

As an immediate corollary of Theorem 5 and Proposition 3.6:

Corollary 4.3. *There exists an absolute constant $c_{\mathsf{spec}} > 0$ for which the following holds. For any $n, d, t \in \mathbb{N}$, $\lambda, \varepsilon_0 > 0$, and H a λ-mixing symmetric d-regular 3-uniform hypergraph on n vertices, $\mathcal{W}_{H,t}$ is an $(\varepsilon_0, \varepsilon \triangleq (\varepsilon_0 + c_{\mathsf{spec}} \cdot \lambda \log(1/\lambda))^t)$-parity sampler.*

Throughout the remainder of this section, we will use the shorthand $\mathcal{W} \triangleq \mathcal{W}_{H,t}$. For proving Theorem 5, it will be convenient to consider $\sigma \in \{\pm 1\}^n$ rather than $z \in \mathbb{F}_2^n$ (as in Definition 2.5) using the mapping $\sigma_i = (-1)^{z_i}$. Equivalent to Definition 2.5, $\mathcal{W}$ is an $(\varepsilon_0, \varepsilon)$ parity sampler if $|\mathrm{bias}_{\mathcal{W}}(\sigma)| \le \varepsilon$ for all $\sigma \in \{\pm 1\}^n$ satisfying $|\mathrm{bias}(\sigma)| = \left|\mathbb{E}_{\boldsymbol{i} \sim [n]}[\sigma_{\boldsymbol{i}}]\right| \le \varepsilon_0$, where

$$\mathrm{bias}_{\mathcal{W}}(\sigma) \triangleq \mathop{\mathbb{E}}_{\boldsymbol{w} \sim \mathcal{W}} \left[\prod_{j=1}^{t} \sigma_{\boldsymbol{w}_j} \right].$$

Similarly to [TS17], we express that bias algebraically.

Lemma 4.4. *Let $A^{(\sigma)} \in \mathbb{R}^{n \times n}$ be defined as*

$$A^{(\sigma)}_{i,k} \triangleq \frac{1}{d} \cdot \sum_{(i',j',k') \in E_H} \sigma_{j'} \cdot \mathbb{1}[i' = i \wedge k' = k].$$

Then,

$$\mathrm{bias}_{\mathcal{W}}(\sigma) = \frac{1}{n} \cdot \mathbf{1}^\dagger \left(A^{(\sigma)}\right)^t \mathbf{1}.$$

Proof. Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_t$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_t$ be the random variables defined in the construction of $\mathcal{W}$. We claim that for each $j \in \{0, 1, \ldots, t\}$ and $v \in [n]$, that

$$\mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_j = v] \prod_{k=1}^{j} \sigma_{\boldsymbol{w}_k}\right] = \frac{1}{n} \cdot \left(\mathbf{1}^\dagger \left(A^{(\sigma)}\right)^j\right)_v \tag{6}$$

By induction on j. Clearly, for $j = 0$, Equation (6) holds as both sides are equal to $\frac{1}{n}$ for any $v \in [n]$. For $j \ge 1$,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_j = v] \prod_{k=1}^{j} \sigma_{\boldsymbol{w}_k}\right] &= \sum_{v' \in [n]} \mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_j = v \wedge \boldsymbol{v}_{j-1} = v'] \prod_{k=1}^{j} \sigma_{\boldsymbol{w}_k}\right] \\
&= \sum_{v' \in [n]} \mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_{j-1} = v'] \prod_{k=1}^{j-1} \sigma_{\boldsymbol{w}_k}\right] \cdot \mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_j = v] \cdot \sigma_{\boldsymbol{w}_j} \,\middle|\, \boldsymbol{v}_{j-1} = v'\right] \\
&= \sum_{v' \in [n]} \frac{1}{n} \cdot \left(\mathbf{1}^\dagger \left(A^{(\sigma)}\right)^{j-1}\right)_{v'} \cdot \mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_j = v] \cdot \sigma_{\boldsymbol{w}_j} \,\middle|\, \boldsymbol{v}_{j-1} = v'\right] \\
&= \frac{1}{n} \sum_{(a,b,c) \in E_H} \left(\mathbf{1}^\dagger \left(A^{(\sigma)}\right)^{j-1}\right)_a \cdot \frac{1}{d} \cdot \sigma_b \cdot \mathbb{1}[v = c] \\
&= \frac{1}{n} \cdot \left(\mathbf{1}^\dagger \left(A^{(\sigma)}\right)^j\right)_v,
\end{aligned}$$

where the third equality is the inductive hypothesis. Then,

$$\mathrm{bias}_{\mathcal{W}}(\sigma) = \mathop{\mathbb{E}}_{\boldsymbol{w}\sim\mathcal{W}}\left[\prod_{j=1}^{t}\sigma_j\right] = \sum_{v\in[n]}\mathbb{E}\left[\mathbb{1}[\boldsymbol{v}_j = v]\prod_{k=1}^{j}\sigma_{\boldsymbol{w}_k}\right] = \frac{1}{n}\cdot\mathbf{1}^{\dagger}\left(A^{(\sigma)}\right)^{t}\mathbf{1}.$$

□

Next, we use the fact that H is a λ-spectral expander to reason about $A^{(\sigma)}$.

Proposition 4.5. *Let $J_n \in \mathbb{R}^{n\times n}$ be the matrix in which every element is $1/n$. For any $\sigma \in \{\pm 1\}^n$,*

$$\left\|A^{(\sigma)} - \mathrm{bias}(\sigma)J_n\right\|_{\mathrm{op}} \le \lambda.$$

Proof. Fix any $x, z \in \mathbb{R}^n$. Then,

$$\begin{aligned}
z^{\dagger}\left(A^{(\sigma)} - \mathrm{bias}(\sigma)J_n\right)x &= \frac{1}{d}\cdot\sum_{(i,j,k)\in E_H} x_k\sigma_j z_i - \frac{\mathrm{bias}(\sigma)}{n}\cdot\sum_{i\in[n]}x_i\sum_{j\in[n]}z_j \\
&= \frac{1}{d}\cdot\sum_{(i,j,k)\in E_H} x_i\sigma_j z_k - \frac{1}{n^2}\cdot\sum_{k\in[n]}\sigma_k\sum_{i\in[n]}x_i\sum_{j\in[n]}z_j \\
&\le \lambda\|x\|_2\|z\|_2. && \text{(Definition 3.3)}
\end{aligned}$$

□

As an immediate consequence, we have:

Corollary 4.6. *For any $\sigma \in \{\pm 1\}^n$,*

$$\left\|A^{(\sigma)}\right\|_{\mathrm{op}} \le |\mathrm{bias}(\sigma)| + \lambda.$$

Proof. As $\|J_n\|_{\mathrm{op}} = 1$, the desired result follows from the reversed triangle inequality applied to the operator norm. □

Finally, we prove Theorem 5.

Proof of Theorem 5. For any $\sigma \in \{\pm 1\}^n$ satisfying $|\mathrm{bias}(\sigma)| \le \varepsilon_0$, we have:

$$\begin{aligned}
|\mathrm{bias}_{\mathcal{W}}(\sigma)| &= \frac{\mathbf{1}^{\dagger}}{\sqrt{n}}\left(A^{(\sigma)}\right)^{t}\frac{\mathbf{1}}{\sqrt{n}} && \text{(Lemma 4.4)} \\
&\le \left\|\left(A^{(\sigma)}\right)^{t}\right\|_{\mathrm{op}} \\
&\le \left\|A^{(\sigma)}\right\|_{\mathrm{op}}^{t} \le (|\mathrm{bias}(\sigma)| + \lambda)^t \le (\varepsilon_0 + \lambda)^t. && \text{(Corollary 4.6)}
\end{aligned}$$

□

5 Codes Closer to the GV Bound

5.1 The construction

We use our parity sampler to amplify the distance of a given base code via direct sum lifting. More formally, given $k \in \mathbb{N}$ and $\varepsilon > 0$, let $\varepsilon_0 = \varepsilon_0(\varepsilon)$ soon to be determined, and let

- $\mathcal{C}_0 \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ be an ε_0-balanced code, and,
- $\mathcal{W} \sim [n]^t$ be an $(\varepsilon_0, \varepsilon)$-parity sampler which is M-discretizable.

Denote $\bar{n} = |\mathrm{Supp}(\mathcal{W})| \le M$. Thus, the lifted code $\mathcal{C} = \mathrm{dsum}_{\mathcal{W}}(\mathcal{C}_0) \subseteq \mathbb{F}_2^{\bar{n}}$ is clearly ε-balanced. For the base code, we use the codes by Guruswami and Indyk, that admit $O_{\varepsilon_0}(n)$-time encoding and decoding.

Theorem 6 ([GI05])**.** *For every integer n and any $\varepsilon_0 > 0$ there exists an ε_0-balanced code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ of rate $\Omega(\varepsilon_0^3)$. Furthermore, $\mathcal{C}_0$ is encodable in time $\exp(\mathrm{poly}(1/\varepsilon_0))n$ and decodable from $\frac{1}{4} - \varepsilon_0$ fraction of errors in time $\exp(\mathrm{poly}(1/\varepsilon_0))n$.*[14]

Setting parameters Given $\varepsilon > 0$, we henceforth set:

1. The initial bias of $\mathcal{C}_0$ to
$$\varepsilon_0 = \frac{1}{2} \cdot 2^{-\sqrt{\log(1/\varepsilon)}}.$$
2. The number of steps over H to
$$t = \left\lceil \sqrt{\log(1/\varepsilon)} \right\rceil .$$
3. The degree d of H to be the smallest even integer for which
$$c_{\mathsf{spec}} \cdot c_{\mathsf{rand}} \cdot \frac{1}{\sqrt{d}} \log\left(\frac{\sqrt{d}}{c_{\mathsf{rand}}}\right) \le \varepsilon_0.$$
This is chosen so that Corollary 3.12 gives, with high probability, a symmetric hypergraph H such that, by Corollary 4.3, $\mathcal{W}_{H,t}$ is a $(\varepsilon_0, (2\varepsilon_0)^t)$-parity sampler.

Using the above parameters in Theorem 5, together with the random hypergraph of Corollary 3.12. we get the following (randomized) error correcting code $\mathcal{C}$.

Theorem 7. *There exists an efficient randomized algorithm such that for every k and any $\varepsilon > 0$, outputs with probability $1 - 2^{-\Omega(k)}$ an ε-balanced linear code $\mathcal{C} \colon \mathbb{F}_2^k \to \mathbb{F}_2^{\bar{n}}$ of rate*
$$\frac{k}{\bar{n}} = 2^{-c_{\mathsf{r}} \log\log(1/\varepsilon)\sqrt{\log(1/\varepsilon)}} \cdot \varepsilon^{-2}$$
for some universal constant c_{r}, that is encodable in deterministic time $\exp(\exp(\sqrt{\log(1/\varepsilon)})) \cdot k$.[15]

[14] One can get a better *randomized* encoding time of $\mathrm{poly}(1/\varepsilon_0)$.

[15] Following the previous footnote, one can get a randomized encoding in time $\mathrm{poly}(1/\varepsilon) \cdot k$ by using a randomized encoding of the based code.

More precisely, there exists a randomized preprocessing step that runs in time $\exp(\sqrt{\log(1/\varepsilon)}) \cdot k$ *and succeeds with probability* $1 - 2^{-\Omega(k)}$*. Once it succeeds, it fixes a* deterministic *mapping* $\mathcal{C}$ *such that for every* $x \in \mathbb{F}_2^k$*,* $\mathcal{C}(x)$ *can be computed in deterministic in the above* $O_\varepsilon(\bar{n})$ *time.*

If, moreover, for a large enough constant d *we are given an explicit family of* λ*-spectral* d*-regular* 3*-uniform hypergraphs satisfying* $\lambda = \frac{\mathrm{poly}(\log d)}{\sqrt{d}}$*, there is no need for a preprocessing step, and* $\mathcal{C}$ *is explicit.*

Proof. Given $k \in \mathbb{N}$ and $\varepsilon > 0$, we set ε_0, t, and d as above. Theorem 6 gives us a code $\mathcal{C}_0 \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ which is ε_0-balanced, for $n = O(k/\varepsilon_0^3)$. By Corollary 3.12 we can output a $\lambda = \frac{c_{\mathsf{rand}}}{\sqrt{d}}$-mixing H with probability at least $1 - 2^{-n}$ in polynomial time. This is our preprocessing step. By Corollary 4.3 and our choice of parameters, indeed

$$\left(\varepsilon_0 + c_{\mathsf{spec}} \cdot \lambda \log \frac{1}{\lambda}\right)^t \leq (2\varepsilon_0)^t \leq \varepsilon,$$

so $\mathcal{W} = \mathcal{W}_{H,t}$ appropriately amplifies the distance. Moreover, $\mathcal{W}$ is M-discretizable for

$$M = nd^t = O\left(\frac{k}{\varepsilon_0^3}\right) \cdot d^{\left\lceil \sqrt{\log(1/\varepsilon)} \right\rceil} = k \cdot 2^{(3+\log d)\sqrt{\log(1/\varepsilon)} + \log d + O(1)}.$$

We proceed to bounding $\log d$ in terms of ε. Recalling how we set d, we see that $\frac{c}{\sqrt{d}} \log d = 2^{-\sqrt{\log(1/\varepsilon)}}$, where c is some universal positive constants. From this we can infer that

$$d \leq c^2 2^{2\sqrt{\log(1/\varepsilon)}} \cdot \log^3\left(c \cdot 2^{\sqrt{\log(1/\varepsilon)}}\right), \tag{7}$$

so $\log d = 2\sqrt{\log(1/\varepsilon)} + O(\log\log(1/\varepsilon))$. (We are assuming that ε is smaller than some small constant, which we can without loss of generality.) Hence,

$$M = \frac{k}{\varepsilon^2} \cdot 2^{O\left(\sqrt{\log(1/\varepsilon)} \cdot \log\log(1/\varepsilon)\right)}.$$

Since $\mathcal{W}$ is M-discretizable, $\mathcal{C} = \mathrm{dsum}_{\mathcal{W}}(\mathcal{C}_0)$ has block length $\bar{n} \leq M$, as required.

Drawing H at random, by Corollary 3.12 takes

$$O(dn) = \widetilde{O}\left(\frac{1}{\varepsilon_0^3}\right) \cdot k = \exp(\sqrt{\log(1/\varepsilon)}) \cdot k$$

time. Given the hypergraph H, the encoding amounts to computing $\mathcal{C}_0(x)$ and taking parities of its coordinates according to $\mathcal{W}$. This takes

$$\exp(\mathrm{poly}(1/\varepsilon_0)) \cdot k + M \cdot t$$

time, following Theorem 6.

For the "moreoever" part, one can verify that we can tolerate a $\mathrm{polylog}(d)$ multiplicative factor in λ with no substantial loss in parameters. ☐

Note that an explicit construction of H would also give an explicit ε-biased sample space over $\mathbb{F}_2^k$ with support size $\bar{n}$, improving upon the state-of-the-art

$$\frac{k}{\bar{n}} = 2^{-\widetilde{O}(\log(1/\varepsilon))^{2/3}} \cdot \varepsilon^{-2}$$

by Ta-Shma [TS17].

5.2 τ-sampling

Toward establishing efficient decoding, we will need the notion of τ-sampling.

Definition 5.1 (τ-sampling)**.** *We say that a distribution $\mathcal{W}$ over $[n]^t$ is* τ-sampling *if for any $i \in [t-1]$, $S \subseteq [n]$, and $X \subseteq [n]^i$,*

$$\operatorname*{Cov}_{\boldsymbol{w}\sim\mathcal{W}}\Big[\mathbb{1}[\boldsymbol{w}_{i+1} \in S], \mathbb{1}[(\boldsymbol{w}_1, \ldots \boldsymbol{w}_i) \in X]\Big] \leq \tau.$$

To highlight the fact that it is indeed a (strong) sampling property, assume for simplicity that $\mathcal{W}$ is homogeneous. Fix any $i \in [t-1]$, $S \subseteq [n]$, and $X \subseteq [n]^i$. The property of τ-sampling thus tells us we can use $\mathcal{W}$ to sample S starting from any prefix. Namely, that

$$\left|\Pr_{\boldsymbol{w}\in\mathcal{W}}[\boldsymbol{w}_{i+1} \in S \mid (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_i) \in X] - \rho(S)\right| \leq \frac{\tau}{\rho(X)}.$$

Note that the $t = 2$ case corresponds to the setting of the expander mixing lemma.

A τ-sampling distribution $\mathcal{W}$ satisfies the property of the "Splittable Mixing Lemma" of Jeronimo, Srivastava, and Tulsiani [JST21, Lemma 4.6] for the family of "± 1 cut functions". This fact is crucial us, and for completeness we establish this in Appendix A.

Given a τ-sampling homogeneous $\mathcal{W} \sim [n]^t$, a standard hybrid argument shows the following.

Lemma 5.2. *Let $\mathcal{W} \sim [n]^t$ be homogeneous and τ-sampling, and let $z \in \mathbb{F}_2^n$ be arbitrary. Then,*

$$\left|\mathop{\mathbb{E}}_{\boldsymbol{w}\in\mathcal{W}}\left[(-1)^{z_{\boldsymbol{w}_1} + \ldots z_{\boldsymbol{w}_t}}\right] - \operatorname{bias}(z)^t\right| \leq 2(t-1)\tau.$$

Thus, high-order mixing in particular implies that $\mathcal{W}$ is an $(\varepsilon_0, \varepsilon = \varepsilon_0^t + (t-1)\tau)$ parity sampler for all ε_0. However, for us, and also in [JST21], $t \cdot \tau$ is too large so we prove the parity sampling property separately.

We conclude our discussion about τ-sampling by showing that our $\mathcal{W}$ is indeed τ-sampling.

Lemma 5.3. *For any $n, d, t \in \mathbb{N}$ and $\lambda > 0$, if $H = (V = [n], E)$ is a λ-spectral d-regular 3-uniform hypergraph, then $\mathcal{W}_{H,t}$ is $\tau = \frac{\lambda}{4}$-sampling.*

Proof. As in Definition 5.1, fix any $i \in [t-1]$, $S \subseteq [n]$ and $X \subseteq [n]^i$. Let $S_i(\boldsymbol{w})$ be the indicator that $\boldsymbol{w}_{i+1} \in S$, and $X_i(w)$ the indicator that $(w_1, \ldots, w_t) \in X$. Using the identity $\operatorname{Cov}[1 - 2\boldsymbol{a}, 1 - 2\boldsymbol{b}] = 4\operatorname{Cov}[\boldsymbol{a}, \boldsymbol{b}]$, it is sufficient to prove that

$$\operatorname*{Cov}_{\boldsymbol{w}\sim\mathcal{W}_{H,t}}\left[(-1)^{S_i(\boldsymbol{w})}, (-1)^{X_i(\boldsymbol{w})}\right] \leq \lambda.$$

Let $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_t$, $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_t$, and $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_t$ be the random variables defined in the construction of $\mathcal{W}_{H,t}$. Furthermore, let $x, y, z \in \mathbb{R}^n$ be the vectors defined, for each $j \in [n]$, by

$$\begin{aligned} x_j &\triangleq \mathbb{E}\left[(-1)^{X_i(\boldsymbol{w})} \,\middle|\, \boldsymbol{v}_i = j\right], \\ y_j &\triangleq (-1)^{\mathbb{1}[j\in S]}, \\ z_j &\triangleq \frac{1}{n}. \end{aligned}$$

Note that $\|z\|_2 = 1/\sqrt{n}$, $\|y\|_\infty = 1$, and $\|x\|_2 \leq \sqrt{n}$ (which follows from $\|x\|_\infty \leq 1$). Our goal in this proof will be to show that the following equation holds:

$$\operatorname*{Cov}_{\boldsymbol{w}}\left[(-1)^{S_i(\boldsymbol{w})}, (-1)^{X_i(\boldsymbol{w})}\right] = \frac{1}{d} \cdot \sum_{(a,b,c)\in E} x_a y_b z_c - \frac{1}{n^2} \cdot \sum_{a\in[n]} x_a \cdot \sum_{a\in[n]} y_a \cdot \sum_{a\in[n]} z_a \tag{8}$$

Once we do, the desired result follows from Definition 3.3. In order to compute the covariance, we first expand:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{w}}\left[(-1)^{S_i(\boldsymbol{w})} \cdot (-1)^{X_i(\boldsymbol{w})}\right] &= \frac{1}{nd} \cdot \sum_{e\in E} \mathop{\mathbb{E}}_{\boldsymbol{w}}\left[(-1)^{S_i(\boldsymbol{w})} \cdot (-1)^{X_i(\boldsymbol{w})} \mid \boldsymbol{e}_{i+1} = e\right] && \text{(Claim 4.2)} \\
&= \frac{1}{nd} \cdot \sum_{(a,b,c)\in E} (-1)^{\mathbb{1}[b\in S]} \cdot \mathop{\mathbb{E}}_{\boldsymbol{w}}\left[(-1)^{X_i(\boldsymbol{w})} \,\middle|\, \boldsymbol{v}_i = a\right] \\
&= \frac{1}{d} \cdot \sum_{(a,b,c)\in E} x_a y_b z_c. && (9)
\end{aligned}$$

Next, directly from the definition of y and Claim 4.2,

$$\mathop{\mathbb{E}}_{\boldsymbol{w}}\left[(-1)^{S_i(\boldsymbol{w})}\right] = \frac{1}{n} \sum_{a\in[n]} y_a. \tag{10}$$

Similarly,

$$\mathop{\mathbb{E}}_{\boldsymbol{w}}\left[(-1)^{X_i(\boldsymbol{w})}\right] = \frac{1}{n} \sum_{a\in[n]} x_a. \tag{11}$$

Equation (8) follows from Equations (9) to (11) and the fact that $\sum_{a\in[n]} z_a = 1$. The desired result then follows Definition 3.3.

□

5.3 Decoding $\mathcal{C}$

We follow the work of Jernoimo et al. who gave a near-linear time list- and unique-decoding algorithm for Ta-Shma's code via an efficient weak regularity lemma, and prove that their algorithm also applies to our code as well. In the language of τ-sampling distributions, they prove:[16]

Theorem 8 ([JST21]). *There exists a constant c_{JST} such that the following holds for any integers d, t, k, n and any $\tau, \varepsilon_0, \varepsilon > 0$. Let $\mathcal{C}_0 \colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ be a code with bias at most ε_0 which is uniquely decodable to within distance $\frac{1-\varepsilon_0}{4}$ in time $T_0 = T_0(n, \varepsilon_0)$. Let $\mathcal{W} \sim [n]^t$ be a homogeneous τ-sampling distribution, let $\mathcal{C} = \mathrm{dsum}_{\mathcal{W}}(\mathcal{C}_0)$ be the corresponding direct sum lifting, and assume that the bias of $\mathcal{C}$ is at most ε. Let β be such that*

$$\beta \geq \max\left\{\sqrt{\varepsilon}, \sqrt{c_{\mathsf{JST}} \cdot t^3 \tau}, 2 \cdot \left(\frac{1}{2} + 2\varepsilon_0\right)^t\right\}.$$

[16] Jernoimo et al. prove their result under stronger requirements, however one can verify that the mixing requirement suffices.

Then, there exists a randomized algorithm, which given $\tilde{y} \in \mathbb{F}_2^{|\mathcal{W}|}$, recovers the list $\mathcal{C} \cap B(\tilde{y}, \frac{1}{2} - \beta)$ with probability at least $1 - \frac{1}{\varepsilon} \cdot 2^{-\Omega(\varepsilon_0^2 n)}$ in time $\widetilde{O}(c_{\beta,t,\varepsilon_0} \cdot (|\mathcal{W}| + T_0))$, for $c_{\beta,t,\varepsilon_0} = (6/\varepsilon_0)^{2^{O(t^3/\beta^2)}}$. Moreover, if we are able to set $\beta = \frac{1+\varepsilon}{4}$, we can uniquely decode $\mathcal{C}$ to within distance $\frac{1-\varepsilon}{4}$ with probability at least $1 - 2^{-\Omega(\varepsilon_0^2 n)}$.

Plugging-in our code $\mathcal{C}$ (using $\mathcal{W}_{H,t}$ and $\mathcal{C}_0$ as described in Section 5.1), we get our main result.

Theorem 9. *Given $k \in \mathbb{N}$ and $\varepsilon > 0$, let $\mathcal{C} \colon \mathbb{F}_2^k \to \mathbb{F}_2^{\bar{n}}$ be the ε-balanced, linear-time encodable code guaranteed to us by Theorem 7 with high probability. Then, $\mathcal{C}$ also admits the following decoding capabilities.*

1. *$\mathcal{C}$ is list decodable up to radius $\frac{1}{2} - \beta$ for $\beta = 2^{-\frac{1}{2}\sqrt{\log(1/\varepsilon)}}$ by a randomized algorithm that runs in time $c_1(\varepsilon) \cdot k$ and succeeds with probability $1 - 2^{-\Omega(k)}$.*
2. *$\mathcal{C}$ is uniquely decodable to within distance $\frac{1-\varepsilon}{4}$ by a randomized algorithm that runs in time $c_2(\varepsilon) \cdot k$ and succeeds with probability $1 - 2^{-\Omega(k)}$.*

Above, $c_1(\varepsilon) = \exp(\exp(\exp(\sqrt{\log(1/\varepsilon)})))$ and $c_2(\varepsilon) = \exp(\exp(\sqrt{\log(1/\varepsilon)}))$.

Proof. By Lemma 5.3, our parity sampler $\mathcal{W}$ which we use for $\mathcal{C}$ is τ-sampling for

$$\tau = \frac{\lambda}{4} = O\left(\frac{\log d}{\sqrt{d}}\right) = \widetilde{O}\left(2^{-2\sqrt{\log(1/\varepsilon)}}\right),$$

where we used the fact that our randomized construction of H gives us $\lambda = O\left(\frac{1}{\sqrt{d}} \log d\right)$ and the bound on d from Equation (7). All that is left is to show how the two items follow from Theorem 8. For the list decoding result, we take β to be as small as possible. For us,

$$\left(\frac{1}{2} + 2\varepsilon_0\right)^t \leq 2^{-\frac{1}{2}\sqrt{\log(1/\varepsilon)}},$$

and

$$\sqrt{c_{\mathsf{JST}} \cdot t^3 \tau} = \widetilde{O}\left(2^{-\sqrt{\log(1/\varepsilon)}}\right).$$

We can thus conclude that $\beta \leq 2^{-\frac{1}{2}\sqrt{\log(1/\varepsilon)}}$. (Again, we are assuming ε is smaller than some small constant.) For the running time, note that

$$T_0 = \exp(\mathrm{poly}(1/\varepsilon_0))n = \exp\left(2^{O\left(\sqrt{\log(1/\varepsilon)}\right)}\right) \cdot k,$$

and $c_{\beta,t,\varepsilon_0}$ is thus triply-exponential in $\sqrt{\log(1/\varepsilon)}$, and overall

$$\widetilde{O}\left(c_{\beta,t,\varepsilon_0} \cdot (|\mathcal{W}| + T_0)\right) = \widetilde{O}\left(c_{\beta,t,\varepsilon_0}\right) \cdot k.$$

For the unique decoding result, we take $\beta = \frac{1+\varepsilon}{4}$, and the running time becomes doubly-exponential in $\sqrt{\log(1/\varepsilon)}$. $\square$

6 Assuming a Ramanujan Hypergraph

In Section 4 we showed how to construct a parity sampler given a mixing, or spectral, hypergraph. Applying that construction with a $\lambda = \frac{\mathrm{polylog}(d)}{\sqrt{d}}$-spectral hypergraph allows us to prove Theorem 3, giving a code that approaches the GV bound.

One can also hope for better spectral hypergraphs. For (non-hyper) graphs, the best λ possible is roughly $\frac{2\sqrt{d-1}}{d}$, and graphs with such good expansion are the celebrated *Ramanujan graphs*. In this section, we show how hypergraphs with similar expansion proprieties would give codes even closer to the GV bound.

Definition 6.1 (almost Ramanujan hypergraph)**.** *For any $\delta \geq 0$, we say that a d-regular 3-uniform hypergraph is δ-*almost Ramanujan *if it is λ-spectral for*

$$\lambda = \frac{2(1+\delta)\sqrt{d-1}}{d}.$$

We will be interested in δ-almost Ramanujan hypergraphs with $\delta \leq d^c$ for any constant $c < 0$. The goal of this section is to prove the following theorem.

Theorem 10. *For any absolute constants $c_1, c_2 > 0$, there is a deterministic algorithm that given any $n \in \mathbb{N}$ and $\varepsilon, \tau > 0$, and a d-regular 3-uniform $(\delta = d^{-c_1})$-almost Ramanujan hypergraph on n vertices for any d in the range*

$$d_{\min} \leq d \leq d_{\min}^{c_2} \quad \text{where} \quad d_{\min} = \mathrm{poly}\left(\log\frac{1}{\varepsilon}, \frac{1}{\tau}\right),$$

constructs an $(\varepsilon_0, \varepsilon)$ parity sampler $\mathcal{W} \sim [n]^t$ that is homogeneous, M-discretizable, and τ-sampling for

$$\begin{aligned}
\varepsilon_0 &= \frac{1}{\mathrm{poly}(\log(1/\varepsilon), 1/\tau)}, \\
t &= O\left(\log(1/\varepsilon)\right), \\
M &= \frac{n}{\varepsilon^2} \cdot \mathrm{poly}(\log(1/\varepsilon), 1/\tau).
\end{aligned}$$

Moreover, the algorithm runs in time $O(Mt)$.

We prove Theorem 10 in Section 6.5. As a corollary of the above theorem, assuming explicit almost Ramanujan hypergraphs, we get explicit codes with rate $\widetilde{\Omega}(\varepsilon^2)$ which are (probabilistically) decodable.

Corollary 6.2. *Given an explicit family of almost Ramanujan expanders, as described in Theorem 10, for any $k \in \mathbb{N}$ and $\varepsilon > 0$ there exists an explicit*[17] *ε-balanced code $\mathcal{C}\colon \mathbb{F}_2^k \to \mathbb{F}_2^{\bar{n}}$ of rate*

$$\frac{k}{\bar{n}} = \varepsilon^{-2} \cdot \frac{1}{\mathrm{poly}(\log(1/\varepsilon))}$$

with the following decoding capabilities.

[17] Here we assume the explicitness of the base code $\mathcal{C}_0$. See Theorem 6 for the exact dependence of the encoding time on ε_0.

1. $\mathcal{C}$ is list decodable up to radius $\frac{1}{2} - \beta$ for $\beta = \frac{1}{\operatorname{poly}(\log(1/\varepsilon))}$ by a randomized algorithm that runs in time $c(\varepsilon) \cdot k$ and succeeds with probability $1 - 2^{-\Omega(k)}$.

2. $\mathcal{C}$ is uniquely decodable to within distance $\frac{1-\varepsilon}{4}$ by a randomized algorithm that runs in time $c(\varepsilon) \cdot k$ and succeeds with probability $1 - 2^{-\Omega(k)}$.

Above, $c(\varepsilon) = \exp(\exp(\operatorname{poly}(\log(1/\varepsilon))))$.

By choosing an appropriate $\tau = \frac{1}{\operatorname{poly}(\log(1/\varepsilon))}$ in Theorem 10, the proof of Corollary 6.2 is essentially identical, up to parameters, to the proof of Theorem 9 so we omit it.

6.1 The non-backtracking parity sampler

In order to take advantage of an almost Ramanujan hypergraph $H = (V, E_H)$, we need a more efficient parity sampler than the one presented in Section 4. First, we recall that construction: For any edge $e \in E_H$, let the set of neighboring edges, $N_H(e)$, be all edges $e' \in E_H$ satisfying $e_3 = e'_1$. To sample from our original parity sampler $\boldsymbol{w} \sim \mathcal{W}_{H,t}$, we first sampled a starting hyperedge $\boldsymbol{e}_1$. Then, for each $j \in [2, n]$, we sampled $\boldsymbol{e}_j$ uniformly from the d edges in $N(\boldsymbol{e}_{j-1})$. The final sample $\boldsymbol{w}$ comprises $\boldsymbol{w}_j = (\boldsymbol{e}_j)_2$ for each $j \in [t]$.

When H is symmetric (as in Definition 3.5), the previous construction is wasteful, as there is a $\frac{1}{d}$ chance that $\boldsymbol{e}_{j+1}$ will just be $\boldsymbol{e}_j$ in reverse. To remedy that inefficiency, we define a non-backtracking parity sampler that avoids taking any reverse steps. Let $H = (V, E_H)$ be a symmetric d-regular hypergraph. For any edge $e \in E_H$, let $N_H^{(\mathrm{nb})}(e)$ be the $(d-1)$-sized set consisting of all neighboring edges except for the reversed edge. Namely,

$$N_H^{(\mathrm{nb})}(e) = N_H(e) \setminus \{(e_3, e_2, e_1)\}.$$

The parity sampler $\mathcal{W}_{H,t}^{(\mathrm{nb})}$. The non-backtracking parity sampler is identical to the original parity sampler, except $N_H^{(\mathrm{nb})}$ is used instead of N_H to sample $\boldsymbol{e}_{j+1}$ given $\boldsymbol{e}_j$. In more detail, to sample from it, $\boldsymbol{w} \sim \mathcal{W}_{H,t}^{(\mathrm{nb})}$, we first sample a starting edge $\boldsymbol{e}_1$ uniformly. Then, for each $j \in [2, n]$, we sample $\boldsymbol{e}_j$ uniformly and independently from $N_H^{(\mathrm{nb})}(\boldsymbol{e}_{j-1})$. The final sample $\boldsymbol{w}$ is once again the set $\boldsymbol{w}_j = (\boldsymbol{e}_j)_2$ for each $j \in [t]$.

Remark 6.3 (unique edges)**.** For convenience, we will assume that in the hypergraph $H = (V, E_H)$, for any $v_1, v_3 \in V$, there is at most one edge $e \in E_H$ satisfying $e_1 = v_1$ and $e_3 = v_3$. This assumption is not essential but simplifies notation. If H has multiple identical edges, then $N_H(e)$ is a multiset instead of a set, and in order to ensure $N_H^{(\mathrm{nb})}(e)$ has size exactly $d-1$, we only want to remove one copy of the reverse edge (e_3, e_2, e_1) from $N_H(e)$.

Much of the analysis of $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is similar to that of $\mathcal{W}_{H,t}$. We defer the more repetitive proofs to the appendix and will instead focus this section on novel machinery. For example, the proof of Claim 6.4 is given in Appendix B.

Claim 6.4. *For any $t \in \mathbb{N}$ and any d-regular symmetric hypergraph H over n vertices, the following holds.*

Proposition B.1: $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is homogeneous, and,

Proposition B.2: $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is $nd(d-1)^{t-1}$-discretizable.

6.2 Expressing the bias algebraically

Fix some $\varepsilon_0 > 0$. In order to prove that $\mathcal{W}_{H,t}$ is an $(\varepsilon_0, \varepsilon \triangleq (\varepsilon_0 + \lambda)^t)$-parity sampler (Theorem 5), we considered any $\sigma \in \{\pm 1\}^n$ satisfying $\left|\mathbb{E}_{\boldsymbol{i} \sim [n]}[\sigma_{\boldsymbol{i}}]\right| \leq \varepsilon_0$ and proved that $\left|\mathrm{bias}_{\mathcal{W}_{H,t}}(\sigma)\right| \leq \varepsilon$, where

$$\mathrm{bias}_{\mathcal{W}_{H,t}}(\sigma) = \mathop{\mathbb{E}}_{\boldsymbol{w} \sim \mathcal{W}_{H,t}}\left[\prod_{j=1}^{t} \sigma_{\boldsymbol{w}_j}\right].$$

To do so, we expressed $\mathrm{bias}_{\mathcal{W}_{H,t}}(\sigma)$ algebraically in terms of the matrix[18] $A^{(\sigma)} \in \mathbb{R}^{V \times V}$, where

$$A^{(\sigma)}_{i,k} \triangleq \sum_{(i',j',k') \in E_H} \sigma_{j'} \cdot \mathbb{1}[i' = i, k' = k]. \tag{12}$$

Then, we showed that

$$\mathrm{bias}_{\mathcal{W}_{H,t}}(\sigma) = \frac{\mathbf{1}^\dagger \left(A^{(\sigma)}\right)^t \mathbf{1}}{nd^t} \leq \frac{\left\|(A^{(\sigma)})^t\right\|_{\mathrm{op}}}{d^t}.$$

Our approach to proving that $\mathcal{W}^{(\mathrm{nb})}_{H,t}$ is a parity sampler will be similar. Rather than analyzing $A^{(\sigma)}$, we will analyze the non-backtracking operator $B^{(\sigma)} \in \mathbb{R}^{E_H \times E_H}$, where

$$B^{(\sigma)}_{e',e} \triangleq \mathbb{1}\left[e \in N^{(\mathrm{nb})}_H(e')\right] \cdot \sigma_{e_2}. \tag{13}$$

This is a slight extension, to hypergraphs, of the classical non-backtracking operator commonly used to analyze non-backtracking walks on graphs (a walk of vertices $v_0, v_1, \ldots, v_t$ is non-backtracking if $v_i \neq v_{i+2}$ for each $i \in [t-2]$). Just as in the proof of Theorem 5, we'll be able to bound $\mathrm{bias}_{\mathcal{W}^{(\mathrm{nb})}_{H,t}}(\sigma)$ in terms of an appropriate operator norm.

Lemma 6.5. *For any d-regular symmetric hypergraph $H = (V, E_H)$, $\sigma \in \{\pm 1\}^n$, letting $B^{(\sigma)}$ be the non-backtracking operator defined in Equation (13), we have that*

$$\left|\mathrm{bias}_{\mathcal{W}^{(\mathrm{nb})}_{H,t}}(\sigma)\right| = \left|\frac{\mathbf{1}^\dagger (B^{(\sigma)})^t \mathbf{1}}{nd(d-1)^t}\right| \leq \frac{\left\|(B^{(\sigma)})^t\right\|_{\mathrm{op}}}{(d-1)^t}.$$

As the proof of Lemma 6.5 is similar to that of Theorem 5, we defer it to Appendix B.

6.3 Bounding $\|B^t\|_{\mathrm{op}}$

Throughout this subsection, $\sigma \in \{\pm 1\}^n$ is fixed so we will use A and B as a shorthand for $A^{(\sigma)}$ and $B^{(\sigma)}$ respectively. In proving Theorem 5, we bounded the operator norm of A^t using $\|A^t\|_{\mathrm{op}} \leq \|A\|^t_{\mathrm{op}}$. As A is real and symmetric, that inequality is tight. The corresponding inequality for B, $\|B^t\|_{\mathrm{op}} \leq \|B\|^t_{\mathrm{op}}$, is *not* tight. We will later see that $\|B\|_{\mathrm{op}} = d-1$, and bounding $\|B^t\|_{\mathrm{op}} \leq (d-1)^t$ would be useless for Lemma 6.5 as it would only upper bound the bias at the trivial bound of 1.

In a different context, Lubetzky and Peres were able to analyze non-backtracking walks on Ramanujan *graphs* [LP16]. While we cannot use their results verbatim, by reasoning about A and

[18] In this section ,we scale up $A^{(\sigma)}$ by a factor of d relative to in Section 4 so that all of its elements are integers. This simplifies notation.

B as operators on a graph (rather than the hypergraph H), we will be able to apply their ideas and techniques.

Let $G = (V, E_G)$ be the graph on the same vertices as H with the edge set

$$E_G \triangleq \{(e_1, e_3) \mid e \in E_H\}.$$

Hence, each edge $e \in E_H$ corresponds to an edge $(e_1, e_3) \in E_G$. With that edge, we associate the sign σ_{e_2}. The matrix A is then the *signed* adjacency matrix of G. If instead of viewing B as a matrix in $\mathbb{R}^{E_H \times E_H}$ (as in Equation (13)), we consider the equivalent matrix $B \in \mathbb{R}^{E_G \times E_G}$, then

$$B_{ab,cd} \triangleq \begin{cases} A_{cd} & \text{if } b = c \text{ and } a \neq d \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

The goal of this subsection is to prove the following bound on $\|B^t\|_{\mathrm{op}}$.

Lemma 6.6. *For any $d \geq 2$ and a d-regular edge-signed graph $G = (V, E)$, let A be its (signed) adjacency matrix and B be defined as in Equation (14). For*

$$\theta_{\max} \triangleq \begin{cases} \frac{\|A\|_{\mathrm{op}}}{2} + \sqrt{\frac{\|A\|_{\mathrm{op}}^2}{4} - (d-1)} & \text{if } \|A\|_{\mathrm{op}} \geq 2\sqrt{d-1} \\ \sqrt{d-1} & \text{otherwise} \end{cases}$$

and any $t \geq 1$,

$$\left\|B^t\right\|_{\mathrm{op}} \leq 2(d-1)t(\theta_{\max})^{t-1}. \tag{15}$$

To prove Lemma 6.6, we will show that B is *almost*-diagonalizable. In particular, that it is unitarily equivalent to a block-diagonal matrix in which each block has size at most 2×2. Lubetzky and Peres proved the below lemma in the case where G is not edge-signed (or equivalently, all edges have the sign $+1$) [LP16, Proposition 3.1]. Our proof follows theirs in spirit, but we aimed to provide additional details for some parts of the argument (see Remark 6.15 for a more detailed comparison).

Lemma 6.7. *Let $G = (V, E)$ be a d-regular edge-signed graph on n vertices and $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of its (signed) adjacency matrix. For each $i \in [n]$, let*

$$R_i \triangleq \begin{cases} [d-1] & \text{if } \lambda_i = d \\ [-(d-1)] & \text{if } \lambda_i = -d \\ \begin{bmatrix} \theta_i & \alpha_i \\ 0 & \theta_i' \end{bmatrix} & \text{otherwise,} \end{cases}$$

for some $\alpha_i \in \mathbb{C}$ satisfying $|\alpha_i| \leq d-1$ and $\theta_i, \theta_i' \in \mathbb{C}$ being the two solutions of

$$\theta^2 - \lambda_i\theta + (d-1) = 0.$$

Then, for $k = nd - \sum_{i \in [n]} \dim(R_i)$ and some $b_i \in \{\pm 1\}$ for each $i \in [k]$, the operator B from Equation (14) is unitarily equivalent to

$$\Lambda \triangleq \mathrm{diag}(R_1, \ldots, R_n, b_1, \ldots, b_k).$$

First, we show how Lemma 6.6 follows from Lemma 6.7.

Proof of Lemma 6.6 assuming Lemma 6.7. By Lemma 6.7, we know that B is unitarily equivalent to

$$\Lambda \triangleq \mathrm{diag}(R_1, \ldots, R_n, b_1, \ldots, b_k),$$

where $R_1, \ldots, R_n$ are as defined in Lemma 6.7 and $b_i \in \{\pm 1\}$ for all $i \in [k]$. Therefore, B^t is unitarily equivalent to

$$\Lambda^t = \mathrm{diag}(R_1^t, \ldots, R_n^t, b_1^t, \ldots, b_k^t).$$

As a result,

$$\left\|B^t\right\|_{\mathrm{op}} = \left\|\Lambda^t\right\|_{\mathrm{op}} = \max\left\{1, \max_{i \in [n]} \|R_i^t\|_{\mathrm{op}}\right\}.$$

Our goal is to show that the above is bounded by the expression in Equation (15). Since that bound is larger than 1, it is enough to show that $\|R_i^t\|_{\mathrm{op}} \le 2(d-1)t(\theta_{\max})^{t-1}$ for each $i \in [n]$.

First, consider the case where $|\lambda_i| = d$. By Lemma 6.7, we have $R_i = [\pm(d-1)]$, and so $\|R_i^t\|_{\mathrm{op}} = (d-1)^t$. In this case, we must have $\|A\|_{\mathrm{op}} = d$ implying that $\theta_{\max} = d-1$. Hence, the right hand side of Equation (15) is at least $2t(d-1)^t$ which is at least as large as $\|R_i^t\|_{\mathrm{op}}$, as desired.

In the other case, $|\lambda_i| < d$. By Lemma 6.7,

$$R_i = \begin{bmatrix} \theta_i & \alpha_i \\ 0 & \theta_i' \end{bmatrix}$$

for some $\alpha_i \in \mathbb{C}$ satisfying $|\alpha_i| \le d-1$ and $\theta_i, \theta_i' \in \mathbb{C}$ the two solutions of

$$\theta^2 - \lambda_i \theta + (d-1) = 0.$$

As $|\lambda_i| \le \|A\|_{\mathrm{op}}$, the two solutions of the above equation have their magnitudes bounded by $\theta_{\max}$. We'll use that to bound $\|R_i^t\|_{\mathrm{op}}$. By an easy inductive argument,

$$R_i^t = \begin{bmatrix} (\theta_i)^t & \alpha_i \sum_{j=0}^{t-1} (\theta_i)^j (\theta_i')^{t-1-j} \\ 0 & (\theta_i')^t \end{bmatrix}.$$

Therefore,

$$\begin{aligned} \|R_i^t\|_{\mathrm{op}} &\le \left\| \begin{bmatrix} (\theta_i)^t & 0 \\ 0 & (\theta_i')^t \end{bmatrix} \right\|_{\mathrm{op}} + \left\| \begin{bmatrix} 0 & \alpha_i \sum_{j=0}^{t-1} (\theta_i)^j (\theta_i')^{t-1-j} \\ 0 & 0 \end{bmatrix} \right\|_{\mathrm{op}} \\ &\le \max(|\theta_i|, |\theta_i'|)^t + t \max(|\theta_i|, |\theta_i'|)^{t-1} |\alpha_i| \\ &\le (\theta_{\max})^t + t(\theta_{\max})^{t-1}(d-1) \\ &\le 2t(\theta_{\max})^{t-1}(d-1). \qquad (\theta_{\max} \le (d-1) \text{ and } t \ge 1) \end{aligned}$$

We conclude that $\|R_i^t\|_{\mathrm{op}}$ is at most the bound of Equation (15) for every $i \in [t]$, as desired. □

In order to prove Lemma 6.7, we decompose $\mathbb{C}^E$ into subspaces on which B acts independently, namely they are orthogonal and B-invariant.

Fact 6.8. *Let $B : \Omega \to \Omega$ a linear operator, and let $S_1, \ldots, S_m$ be disjoint subspaces for which $S_1 \oplus \cdots \oplus S_m = \Omega$. Assume that the following holds:*

1. *For any $i \ne j \in [m]$, $S_i \perp S_j$, and,*

2. *For any $i \in [m]$, S_i is an invariant subspace of B, meaning $BS_i \subseteq S_i$.*

For each $i \in [m]$, let $R_i \in \mathbb{C}^{(\dim S_i)\times(\dim S_i)}$ be a matrix computing the restriction $B|_{S_i} : S_i \to S_i$ with respect to some orthonormal basis of S_i. Then, B is unitarily equivalent to

$$\Lambda \triangleq \mathrm{diag}(R_1, \ldots, R_m).$$

Proof. For each $i \in [m]$, there is some orthonormal basis $v_{i,1}, \ldots, v_{i,k_i}$ in which the operator $B|_{S_i}$ corresponds to the matrix R_i. As $S_1 \oplus \cdots \oplus S_m = B$ and S_i are orthogonal subspaces, $v_{1,1}, \ldots, v_{1,k_1}, \ldots, v_{m,1}, \ldots, v_{m,k_m}$ is an orthonormal basis for all of Ω, and Λ computes B with respect to that basis. □

We break $\mathbb{C}^E$ into subspaces satisfying the requirements of Fact 6.8. For any $w \in \mathbb{C}^V$, we define $w^{(\mathrm{in})}, w^{(\mathrm{out})} \in \mathbb{C}^E$.[19]

$$\begin{aligned} w^{(\mathrm{in})}_{xy} &\triangleq w_y \\ w^{(\mathrm{out})}_{xy} &\triangleq A_{xy} w_x. \end{aligned} \tag{16}$$

For $A \in \mathbb{R}^{V\times V}$ being the adjacency matrix defined in Lemma 6.7, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A with corresponding eigenvectors $w_1, \ldots, w_n$. For each $i \in [n]$, we define

$$S_i \triangleq \mathrm{Span}\left\{ w_i^{(\mathrm{in})}, w_i^{(\mathrm{out})} \right\}, \tag{17}$$

and define S_{rem} to be the orthogonal compliment of $S_1 \oplus \cdots \oplus S_n$ within $\mathbb{C}^E$. Our first goal is to show that $S_1, \ldots, S_n, S_{\mathrm{rem}}$ meet the requirements of Fact 6.8.

A large portion of this proof will require straightforward calculations. We collect all of these calculations into the following proposition, proved in Appendix B.

Proposition 6.9.

1. *For any $w, w' \in \mathbb{C}^V$,*

$$\begin{aligned} \langle w^{(\mathrm{in})}, (w')^{(\mathrm{in})} \rangle &= d \cdot \langle w, w' \rangle \\ \langle w^{(\mathrm{out})}, (w')^{(\mathrm{out})} \rangle &= d \cdot \langle w, w' \rangle \\ \langle w^{(\mathrm{in})}, (w')^{(\mathrm{out})} \rangle &= \langle Aw, w' \rangle \end{aligned} \tag{18}$$

2. *For any $w \in \mathbb{C}^V$,*

$$\begin{aligned} Bw^{(\mathrm{in})} &= (Aw)^{(\mathrm{in})} - w^{(\mathrm{out})} \\ Bw^{(\mathrm{out})} &= (d-1) w^{(\mathrm{in})}. \end{aligned} \tag{19}$$

3. *For any $v \in \mathbb{C}^E$ satisfying $v \perp w^{(\mathrm{out})}$ for all $w \in \mathbb{C}^V$,*

$$v_{xy} = -A_{yx} v_{yx} \qquad \text{for every edge } xy. \tag{20}$$

[19] We use the name $w^{(\mathrm{in})}$ as all edges going into a vertex y have weight w_y. Similarly, for $w^{(\mathrm{out})}$, all edges going out of the vertex x have weight w_x multiplied by that edge's sign.

4. The operator norm of B is

$$\|B\|_{\mathrm{op}} = d - 1. \tag{21}$$

We use Proposition 6.9 to prove Lemma 6.7 by showing that $S_1, \ldots, S_n, S_{\mathrm{rem}}$ meet the requirements of Fact 6.8.

Proposition 6.10. *For B defined in Lemma 6.7 and $S_1, \ldots, S_n$ defined in Equation (17), for any $i \neq j \in [n]$, $S_i \perp S_j$.*

Proof. It is sufficient to prove that for any $i \neq j \in [n]$, $w_i^{(\mathrm{in})} \perp w_j^{(\mathrm{in})}$, $w_i^{(\mathrm{out})} \perp w_j^{(\mathrm{out})}$, and $w_i^{(\mathrm{in})} \perp w_j^{(\mathrm{out})}$. As A is real and symmetric, its eigenvectors are orthogonal, and so $w_i \perp w_j$ and $Aw_i \perp w_j$. The desired result follows from Equation (18).

□

In order to apply Fact 6.8, we also need to prove that each S_i and S_{rem} are all invariant under B.

Proposition 6.11. *For B defined in Lemma 6.7 and $S_1, \ldots, S_n$ defined in Equation (17), $BS_i \subseteq S_i$ for each $i \in [n]$. Furthermore, for S_{rem}, the orthogonal compliment of $S_1 \oplus \cdots \oplus S_n$ within $\mathbb{C}^E$, we also have that $BS_{\mathrm{rem}} \subseteq S_{\mathrm{rem}}$.*

Proof. First we show that $BS_i \subseteq S_i$ for any $i \in [n]$. Recall that S_i is the span of $w_i^{(\mathrm{in})}$ and $w_i^{(\mathrm{out})}$, so it suffices to show that $Bw_i^{(\mathrm{in})} \in S_i$ and $Bw_i^{(\mathrm{out})} \in S_i$. This follows from Equation (19) and the fact that w_i is an eigenvector of A.

Secondly, we show that for any $v \in S_{\mathrm{rem}}$, $Bv \in S_{\mathrm{rem}}$. As $v \in S_{\mathrm{rem}}$, we know that $v \perp w_i^{(\mathrm{out})}$ and $v \perp w_i^{(\mathrm{in})}$ for each eigenvector w_i. As the eigenvectors span all of $\mathbb{C}^V$, we further have that $v \perp w^{(\mathrm{in})}$ and $v \perp w^{(\mathrm{out})}$ for any $w \in \mathbb{C}^V$. In order to prove that $Bv \in S_{\mathrm{rem}}$, we will need to prove that $Bv \perp w^{(\mathrm{in})}$ and $Bv \perp w^{(\mathrm{out})}$ for any $w \in \mathbb{C}^V$.

1. We show that $Bv \perp w^{(\mathrm{in})}$:

$$\begin{aligned}
\langle Bv, w^{(\mathrm{in})} \rangle &= \sum_{xy \in E} (Bv)_{xy} w_y \\
&= \sum_{xy \in E} -A_{yx} v_{yx} w_y && \text{(Equation (20))} \\
&= -\sum_{xy \in E} (A_{xy} w_x) v_{xy} && \text{(switching names of } x \text{ and } y\text{)} \\
&= -\langle w^{(\mathrm{out})}, v \rangle = 0. && (v \perp w^{(\mathrm{out})})
\end{aligned}$$

2. Similarly, we show that $Bv \perp w^{(\text{out})}$:

$$
\begin{aligned}
\langle Bv, w^{(\text{out})}\rangle &= \sum_{xy\in E} (Bv)_{xy} A_{xy} w_x \\
&= \sum_{xy\in E} -A_{yx} v_{yx} A_{xy} w_x && \text{(Equation (20))} \\
&= -\sum_{xy\in E} v_{yx} w_x && (A_{xy} = A_{yx} \in \{\pm 1\}) \\
&= -\sum_{xy\in E} v_{xy} w_y && \text{(switching names of } x \text{ and } y) \\
&= -\langle v, w^{(\text{in})}\rangle = 0. && (v \perp w^{(\text{in})})
\end{aligned}
$$

Therefore, $Bv \in S_{\text{rem}}$, as desired.

□

Using Propositions 6.10 and 6.11 we can apply Fact 6.8 with the decomposition $\mathbb{C}^E = S_1 \oplus \cdots \oplus S_n \oplus S_{\text{rem}}$. In order for Fact 6.8 to be useful, we need to understand the restricted operator $B|_{S_i}$. Toward this end, we recall the Schur decomposition.

Fact 6.12 (Schur decomposition)**.** *For any linear operator $T : \Omega \to \Omega$, there is an orthonormal basis in which T can be written as an upper triangular matrix $U \in \mathbb{C}^{(\dim\Omega)\times(\dim\Omega)}$. In particular, the diagonal entries of U are the eigenvalues of T.*

Proposition 6.13. *Let $A, B, \lambda_1, \ldots, \lambda_n$ be as defined in Lemma 6.7 and $S_1, \ldots, S_n$ as in Equation (17). For any $i \in [n]$, there is an orthonormal basis of S_i under which $B|_{S_i} : S_i \to S_i$ is computed by the following matrix.*

$$
R_i \triangleq \begin{cases} [d-1] & \text{if } \lambda_i = d \\ [-(d-1)] & \text{if } \lambda_i = -d \\ \begin{bmatrix} \theta_i & \alpha_i \\ 0 & \theta_i' \end{bmatrix} & \text{otherwise,} \end{cases}
$$

for some $\alpha_i \in \mathbb{C}$ satisfying $|\alpha_i| \le d-1$, and $\theta_i, \theta_i' \in \mathbb{C}$ are the two solutions of

$$
\theta^2 - \lambda_i\theta + (d-1) = 0.
$$

Proof. We first compute the dimension of $S_i = \operatorname{Span}\left\{w_i^{(\text{in})}, w_i^{(\text{out})}\right\}$. This dimension is 1 if $w_i^{(\text{in})}$ and $w_i^{(\text{out})}$ are parallel, and 2 otherwise. They are parallel if and only if

$$
\left|\langle w_i^{(\text{in})}, w_i^{(\text{out})}\rangle\right| = \left\|w_i^{(\text{in})}\right\|_2 \left\|w_i^{(\text{out})}\right\|_2.
$$

Assuming that w_i is normalized to satisfy $\|w_i\|_2 = 1$, we have from Equation (18) that $\|w_i^{(\text{in})}\|_2 = \|w_i^{(\text{out})}\|_2 = \sqrt{d}$ and that $\langle w_i^{(\text{in})}, w_i^{(\text{out})}\rangle = \lambda_i$. There are three cases depending on the value of λ_i:

1. If $\lambda_i = d$, then $\langle w_i^{(\text{in})}, w_i^{(\text{out})} \rangle = \|w_i^{(\text{in})}\|_2 \|w_i^{(\text{out})}\|_2$, implying that $w_i^{(\text{in})} = w_i^{(\text{out})}$. In this case, the dimension of S_i is 1, and we can just set R_i to

$$[\|Bv\|_2 / \|v\|_2]$$

for a single nonzero $v \in S_i$. Recall from Equation (19) that $Bw_i^{(\text{out})} = (d-1)w_i^{(\text{in})} = (d-1)w_i^{(\text{out})}$. Therefore, we set $R_i = [d-1]$.

2. If $\lambda_i = -d$, then $\langle w_i^{(\text{in})}, w_i^{(\text{out})} \rangle = -\|w_i^{(\text{in})}\|_2 \|w_i^{(\text{out})}\|_2$, implying that $w_i^{(\text{in})} = -w_i^{(\text{out})}$. Once again, the dimension of S_i is 1, and we set R_i to

$$[\|Bv\|_2 / \|v\|_2]$$

for a single nonzero $v \in S_i$. Also from Equation (19), $Bw_i^{(\text{out})} = (d-1)w_i^{(\text{in})} = -(d-1)w_i^{(\text{out})}$. Therefore, we set $R_i = [-(d-1)]$.

3. Otherwise, $\dim(S_i) = 2$. Recall from Equation (19) that

$$\begin{aligned} Bw_i^{(\text{in})} &= \lambda_i w_i^{(\text{in})} - w_i^{(\text{out})}, \\ Bw_i^{(\text{out})} &= (d-1)w_i^{(\text{in})}. \end{aligned}$$

Therefore, the characteristic polynomial of $B|_{S_i}$ is $p(\theta) = \theta^2 - \theta\lambda_i + (d-1)$. Applying Fact 6.12, for θ_i, θ_i' being the two roots of $p(\theta)$ and some $\alpha_i \in \mathbb{C}$, we can set

$$R_i = \begin{bmatrix} \theta_i & \alpha_i \\ 0 & \theta_i' \end{bmatrix}.$$

Finally, we bound $|\alpha_i|$:

$$|\alpha_i| \leq \|R_i\|_{\text{op}} \leq \|B\|_{\text{op}} \leq d-1,$$

where $\|B\|_{\text{op}} \leq d-1$ is Equation (21).

□

Similar to Proposition 6.13, we characterize $B|_{S_{\text{rem}}}$.

Proposition 6.14. *Let A, B be as defined in Lemma 6.7, $S_1, \ldots, S_n$ as in Equation (17), and S_{rem} be the orthogonal compliment of $S_1 \oplus \ldots \oplus S_n$ within $\mathbb{C}^E$. Then, there is an orthonormal basis for S_{rem} in which the restriction $B|_{S_{\text{rem}}}$ is expressed by $\operatorname{diag}(b_1, \ldots, b_{\dim(S_{\text{rem}})})$ where $b_i \in \{\pm 1\}$ for each $i \in [\dim(S_{\text{rem}})]$.*

Proof. For any $v \in S_{\text{rem}}$ and any edge $xy \in E$, by Equation (20),

$$(Bv)_{xy} = -A_{yx} v_{yx}.$$

Hence, the operator $B|_{S_{\text{rem}}}$ preserves the ℓ_2 norm. By Fact 6.12, there is some orthonormal basis in which the matrix representing $B|_{S_{\text{rem}}}$ is an upper triangular matrix. The only upper triangular matrices that are ℓ_2 norm-preserving are diagonal matrices in which all diagonal entries are ± 1. □

Combining the above propositions completes the proof of Lemma 6.7.

Proof of Lemma 6.7. We apply Fact 6.8 with the subspaces $S_1, \dots, S_n, S_{\mathrm{rem}}$. By Propositions 6.10 and 6.11, those subspaces meet the requirement of Fact 6.8. Applying Propositions 6.13 and 6.14 gives the desired form for Λ. □

Remark 6.15 (comparison with [LP16])**.** Our proof of Lemma 6.7 follows that of [LP16, Proposition 3.1] with a few (minor) technical differences. In order to handle edge signs, we must adjust the definitions in Equation (16) and carry that change throughout the computation. Additionally, to prove (the analogous statements of) Propositions 6.10 and 6.11, Lubetzky and Peres reason about the operator $C \triangleq (d-1)B^{-1} + B$, whereas we never need consider the inverse of B.

6.4 τ-sampling

In this section, we prove that $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is τ-sampling for an appropriately chosen τ. Recall that this is needed for the decoding result.

Lemma 6.16. *For any $n, d, t \in \mathbb{N}$ and $\lambda > 0$, if $H = (V = [n], E)$ is a λ-spectral d-regular 3-uniform hypergraph, then $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is τ-sampling for $\tau = \frac{\lambda}{4} + \frac{2}{d}$.*

In order to prove Lemma 6.16, we will construct a distribution $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ that is close, in total variation distance, to $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ and show that $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ is $\tau = \frac{\lambda}{4}$-sampling. Once we do this, we will use the closeness of $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ and $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ to transfer that result to a bound on the τ-sampling of $\mathcal{W}_{H,t}^{(\mathrm{nb})}$. As the definition of τ-sampling involves computing a covariance, the following proposition will useful.

Proposition 6.17. *Let $\mathcal{D}, \widehat{\mathcal{D}}$ be distributions each over some domain Ω. For any bounded functions $f, g \colon \Omega \to [0, 1]$,*

$$\left| \operatorname*{Cov}_{\boldsymbol{x} \sim \mathcal{D}}[f(\boldsymbol{x}), g(\boldsymbol{x})] - \operatorname*{Cov}_{\widehat{\boldsymbol{x}} \sim \widehat{\mathcal{D}}}[f(\widehat{\boldsymbol{x}}), g(\widehat{\boldsymbol{x}})] \right| \leq 2 d_{\mathrm{TV}}(\mathcal{D}, \widehat{\mathcal{D}}).$$

Proof. We use the following covariance identity:

$$\operatorname*{Cov}_{\boldsymbol{x} \sim \mathcal{D}}[f(\boldsymbol{x}), g(\boldsymbol{x})] = \frac{1}{2} \cdot \mathop{\mathbb{E}}_{\boldsymbol{x}_1, \boldsymbol{x}_2 \sim \mathcal{D}^2}[(f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2))(g(\boldsymbol{x}_1) - g(\boldsymbol{x}_2))].$$

Defining $h(x) \triangleq (f(x_1) - f(x_2))(g(x_1) - g(x_2))$, we have

$$\begin{aligned}
\left| \operatorname*{Cov}_{\boldsymbol{x} \sim \mathcal{D}}[f(\boldsymbol{x}), g(\boldsymbol{x})] - \operatorname*{Cov}_{\widehat{\boldsymbol{x}} \sim \widehat{\mathcal{D}}}[f(\widehat{\boldsymbol{x}}), g(\widehat{\boldsymbol{x}})] \right| &\leq \frac{1}{2} \left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{D}^2}[h(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\widehat{\boldsymbol{x}} \sim \widehat{\mathcal{D}}^2}[h(\widehat{\boldsymbol{x}})] \right| \\
&\leq \frac{1}{2} d_{\mathrm{TV}}(\mathcal{D}^2, \widehat{\mathcal{D}}^2) \cdot \left(\sup_{x \in \Omega^2} h(x) - \inf_{x \in \Omega^2} h(x) \right) \\
&\leq \frac{1}{2} (2 d_{\mathrm{TV}}(\mathcal{D}, \widehat{\mathcal{D}})) \cdot (1 - (-1)) = 2 d_{\mathrm{TV}}(\mathcal{D}, \widehat{\mathcal{D}}).
\end{aligned}$$

□

In order to prove that $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ is close, in TV distance, to $\mathcal{W}_{H,t}^{(\mathrm{nb})}$, we'll use the following easy fact.

Fact 6.18. *For any distributions $\mathcal{D}$ and $\mathcal{D}'$ over the same domain, and any coupling of $\boldsymbol{x} \sim \mathcal{D}$ and $\boldsymbol{x}' \sim \mathcal{D}'$,*

$$d_{\mathrm{TV}}(\mathcal{D}, \mathcal{D}') \leq \Pr[\boldsymbol{x} \neq \boldsymbol{x}'].$$

The proof of Fact 6.18 follows readily from the definition of TV distance. Indeed, it is well known that the TV distance is equal to the infimum of $\Pr[\boldsymbol{x} \neq \boldsymbol{x}']$ over all couplings $\boldsymbol{x} \sim \mathcal{D}, \boldsymbol{x}' \sim \mathcal{D}'$.

Using Fact 6.18, it's not hard to show that the TV distance of $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ and $\mathcal{W}_{H,t}$ is at most $\frac{t}{d}$. Therefore, Lemma 5.3 and Proposition 6.17 are sufficient to show that $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is $\frac{\lambda}{4} + \frac{2t}{d}$-sampling. In the following proof, we'll construct a distribution $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ that is closer to $\mathcal{W}_{H,t}^{(\mathrm{nb})}$, giving a better bound on τ.

Proof of Lemma 6.16. As in Definition 5.1, fix any $i \in [t]$, $S \subseteq [n]$ and $X \subseteq [n]^i$. Let $S_i(\boldsymbol{w})$ be the indicator that $\boldsymbol{w}_{i+1} \in S$, and $X_i(w)$ the indicator that $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_t) \in X$. Our goal is to prove that

$$\operatorname*{Cov}_{\boldsymbol{w} \sim \widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}} \Big[S_i(\boldsymbol{w}), X_i(\boldsymbol{w}) \Big] \leq \frac{\lambda}{4} + \frac{2}{d}.$$

To do so, we will define a distribution $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ satisfying $d_{\mathrm{TV}}(\mathcal{W}_{H,t}^{(\mathrm{nb})}, \widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}) \leq \frac{1}{d}$ and prove that

$$\operatorname*{Cov}_{\widehat{\boldsymbol{w}} \sim \widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}} \Big[S_i(\widehat{\boldsymbol{w}}), X_i(\widehat{\boldsymbol{w}}) \Big] \leq \frac{\lambda}{4}. \tag{22}$$

At a high level, $\widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$ will perform "non-backtracking" steps during the first i time steps and a normal ("backtracks" with probability $\frac{1}{d}$) step at time step $j = i+1$. In contrast, $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ performs "non-backtracking" steps at every time step. Formally, to sample $\widehat{\boldsymbol{w}} \sim \widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$, we first sample $\boldsymbol{w} \sim \mathcal{W}_{H,t}^{(\mathrm{nb})}$. Then, with probability $1 - \frac{1}{d}$, $\widehat{\boldsymbol{w}}$ is set to $\boldsymbol{w}$. Otherwise, $\widehat{\boldsymbol{w}}_j = \boldsymbol{w}_j$ for each $j \neq i+1$, and $\widehat{\boldsymbol{w}}_{i+1} = \boldsymbol{w}_i$. Through the remainder of this proof, we assume that $\boldsymbol{w}$ and $\widehat{\boldsymbol{w}}$ are coupled according to this generation process.

Clearly, $d_{\mathrm{TV}}(\mathcal{W}_{H,t}^{(\mathrm{nb})}, \widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}) \leq 1/d$. Hence, by Proposition 6.17 it suffices to prove Equation (22). The remainder of this proof is similar to that of Lemma 5.3. Instead of showing Equation (22), we prove the following (equivalent equation) holds:

$$\operatorname{Cov}\Big[(-1)^{S_i(\widehat{\boldsymbol{w}})}, (-1)^{X_i(\widehat{\boldsymbol{w}})}\Big] \leq \lambda.$$

Let $\boldsymbol{e}_1, \dots, \boldsymbol{e}_t$ be the random variables defined in the construction of $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ (which are coupled to the sample $\boldsymbol{w} \sim \mathcal{W}_{H,t}^{(\mathrm{nb})}$ and hence $\widehat{\boldsymbol{w}} \sim \widehat{\mathcal{W}}_{H,t}^{(\mathrm{nb})}$). If the probability-$\big(1 - \frac{1}{d}\big)$ event occurred where we set $\widehat{\boldsymbol{w}} = \boldsymbol{w}$, then we define $\widehat{\boldsymbol{e}}_{i+1} \triangleq \boldsymbol{e}_{i+1}$.

Recall that $\boldsymbol{e}_{i+1}$ is sampled uniformly from the $(d-1)$ hyperedges satisfying $(\boldsymbol{e}_i)_3 = (\boldsymbol{e}_{i+1})_1$ and $\boldsymbol{e}_{i+1} \neq ((\boldsymbol{e}_i)_3, (\boldsymbol{e}_i)_2, (\boldsymbol{e}_i)_1)$. Therefore, the generation process for $\widehat{\boldsymbol{e}}_{i+1}$ is equivalent to sampling uniformly from the d hyperedges satisfying $(\boldsymbol{e}_i)_3 = (\widehat{\boldsymbol{e}}_{i+1})_1$.

Let $x, y, z \in \mathbb{R}^n$ be the vectors defined, for each $j \in [n]$,

$$\begin{aligned}
x_j &\triangleq \mathbb{E}\left[(-1)^{X_i(\widehat{\boldsymbol{w}})} \,\middle|\, (\widehat{\boldsymbol{e}}_{i+1})_1 = j\right] \\
y_j &\triangleq (-1)^{\mathbb{1}[j \in S]} \\
z_j &\triangleq \frac{1}{n}.
\end{aligned}$$

Note that $\|z\|_2 = 1/\sqrt{n}$, $\|y\|_\infty = 1$, and $\|x\|_2 \le \sqrt{n}$ (which follows from $\|x\|_\infty \le 1$). Our goal will be to prove that the following equation holds.

$$\operatorname*{Cov}_{\widehat{\boldsymbol{w}}}\left[(-1)^{S_i(\widehat{\boldsymbol{w}})}, (-1)^{X_i(\widehat{\boldsymbol{w}})}\right] = \frac{1}{d} \cdot \sum_{(a,b,c)\in E} x_a y_b z_c - \frac{1}{n^2} \cdot \sum_{a\in[n]} x_a \cdot \sum_{a\in[n]} y_a \cdot \sum_{a\in[n]} z_a. \tag{23}$$

Once we prove the above equation holds, the desired result follows from Definition 3.3. As we proved in Claim 6.4, the distribution of $\boldsymbol{e}_i$ is uniform over all nd hyperedges, and hence that of $\widehat{\boldsymbol{e}}_{i+1}$ is also uniform over all hyperedges. In order to compute the covariance, we first expand:

$$\begin{aligned}
\operatorname*{\mathbb{E}}_{\widehat{\boldsymbol{w}}}\left[(-1)^{S_i(\widehat{\boldsymbol{w}})} \cdot (-1)^{X_i(\widehat{\boldsymbol{w}})}\right] &= \frac{1}{nd} \cdot \sum_{e\in E} \operatorname*{\mathbb{E}}_{\widehat{\boldsymbol{w}}}\left[(-1)^{S_i(\widehat{\boldsymbol{w}})} \cdot (-1)^{X_i(\widehat{\boldsymbol{w}})} \mid \widehat{\boldsymbol{e}}_{i+1} = e\right] \\
&= \frac{1}{nd} \cdot \sum_{(a,b,c)\in E} (-1)^{\mathbb{1}[b\in S]} \cdot \mathbb{E}\left[(-1)^{X_i(\widehat{\boldsymbol{w}})} \,\middle|\, (\widehat{\boldsymbol{e}}_{i+1})_1 = a\right] \\
&= \frac{1}{d} \cdot \sum_{(a,b,c)\in E} x_a y_b z_c. \qquad (24)
\end{aligned}$$

Next, directly from the definition of y and Claim 6.4,

$$\operatorname*{\mathbb{E}}_{\widehat{\boldsymbol{w}}}\left[(-1)^{S_i(\widehat{\boldsymbol{w}})}\right] = \frac{1}{n} \sum_{a\in[n]} y_a. \tag{25}$$

Similarly,

$$\operatorname*{\mathbb{E}}_{\widehat{\boldsymbol{w}}}\left[(-1)^{X_i(\widehat{\boldsymbol{w}})}\right] = \frac{1}{n} \sum_{a\in[n]} x_a. \tag{26}$$

Equation (23) follows from Equations (24) to (26) and the fact that $\sum_{a\in[n]} z_a = 1$. Lemma 6.16 follows from Definition 3.3, and the desired result from Proposition 6.17. $\square$

6.5 Setting the parameters

In this subsection, we set the parameters needed to prove the following theorem, restated for convenience.

Theorem 10. *For any absolute constants $c_1, c_2 > 0$, there is a deterministic algorithm that given any $n \in \mathbb{N}$ and $\varepsilon, \tau > 0$, and a d-regular 3-uniform ($\delta = d^{-c_1}$)-almost Ramanujan hypergraph on n vertices for any d in the range*

$$d_{\min} \le d \le d_{\min}^{c_2} \quad \text{where} \quad d_{\min} = \operatorname{poly}\left(\log\frac{1}{\varepsilon}, \frac{1}{\tau}\right),$$

constructs an $(\varepsilon_0, \varepsilon)$ parity sampler $\mathcal{W} \sim [n]^t$ that is homogeneous, M-discretizable, and τ-sampling for

$$\begin{aligned}
\varepsilon_0 &= \frac{1}{\operatorname{poly}(\log(1/\varepsilon), 1/\tau)}, \\
t &= O\left(\log(1/\varepsilon)\right), \\
M &= \frac{n}{\varepsilon^2} \cdot \operatorname{poly}(\log(1/\varepsilon), 1/\tau).
\end{aligned}$$

Moreover, the algorithm runs in time $O(Mt)$.

Proof. We set

$$d_{\min} \triangleq \max\left(\frac{9}{\tau^2}, \log\left(\frac{1}{\varepsilon}\right)^{2/c_1}, 145\right).$$

Let H be the d-regular 3-uniform ($\delta \triangleq d^{-c_1}$)-almost Ramanujan hypergraph on n vertices for some $d_{\min} \le d \le (d_{\min})^{c_2}$ given to the algorithm. It will return $\mathcal{W} \triangleq \mathcal{W}_{H,t}^{(\mathrm{nb})}$ for some t to be later specified.

Regardless of what that t is, by Lemma 6.16, we have that $\mathcal{W}$ is τ'-sampling for

$$\tau' \le \frac{4\sqrt{d-1}}{4d} + \frac{2}{d} \le \frac{3}{\sqrt{d}} \le \tau.$$

To complete this proof, we need to set t large enough so that $\mathcal{W}$ is an $(\varepsilon_0, \varepsilon)$-parity sampler and small enough so that it is M-discretizable. Set

$$\varepsilon_0 = \frac{2\delta\sqrt{d-1}}{d}.$$

In order to set t, we first define,

$$\varepsilon(t') \triangleq \frac{2t'(1+5d^{-c_1/2})^{t'-1}}{(d-1)^{(t'-3)/2}}$$

and then set t to the minimum integer so that $\varepsilon(t) \le \varepsilon$. As $d \ge d_{\min} \ge 145$,

$$\varepsilon(t') \ge \frac{2t' \cdot 6^{t'-1}}{12^{t'-3}} = c \cdot t' \cdot 2^{-t'}$$

for an absolute constant c. Therefore, the minimum t for which $\varepsilon(t) \le \varepsilon$ is $O(\log(1/\varepsilon))$, as desired.

We proceed to bound M. For any $t' \in \mathbb{N}$, $\frac{\varepsilon(t')}{\varepsilon(t'+1)} \le \sqrt{d-1}$, so $\varepsilon(t) \ge \frac{\varepsilon}{\sqrt{d-1}}$. Therefore,

$$(d-1)^{t-1} = \frac{4t^2(d-1)^2(1+5d^{-c_1/2})^{2t-2}}{\varepsilon(t)^2} \le \frac{4t^2(d-1)^3(1+5d^{-c_1/2})^{2t-2}}{\varepsilon^2}.$$

By Claim 6.4, $\mathcal{W}$ is M-discretizable for

$$M \le \frac{n^2}{\varepsilon} \cdot 4dt^2(d-1)^3(1+5d^{-c_1/2})^{2t-2}$$

Using the bounds $(1+a)^b \le \exp(ab)$, $d \ge d_{\min} \ge \log(1/\varepsilon)^{2/c_1}$, and $t = O(\log(1/\varepsilon)$, we have $(1+5d^{-c_1/2})^{2t-2} = O(1)$. We've also already bounded d and t, each at most $\mathrm{poly}(\log(1/\varepsilon), 1/\tau)$, so we can give the desired bound on M,

$$M \le \frac{n^2}{\varepsilon} \cdot \mathrm{poly}(\log(1/\varepsilon), 1/\tau).$$

Lastly, we need to show that $\mathcal{W}$ is an $(\varepsilon_0, \varepsilon)$-parity sampler. Choose any $\sigma \in \{\pm 1\}^n$ satisfying $|\mathbb{E}_{\boldsymbol{i}\sim[n]}[\sigma_{\boldsymbol{i}}]| \le \varepsilon_0$. By Lemma 6.5, it is sufficient to show, for $B^{(\sigma)}$ defined in Equation (13), that

$$\frac{\left\|\left(B^{(\sigma)}\right)^t\right\|_{\mathrm{op}}}{(d-1)^t} \le \varepsilon.$$

Lemma 6.6 allows us to bound $\|(B^{(\sigma)})^t\|_{\text{op}}$ in terms of $\|(A^{(\sigma)})^t\|_{\text{op}}$ for $A^{(\sigma)}$ defined in Equation (12). Applying Corollary 4.6 (and noting that $A^{(\sigma)}$ of this section is scaled up by a factor of d relative to that in Corollary 4.6),

$$\begin{aligned}\left\|A^{(\sigma)}\right\|_{\text{op}} &\leq d\left(|\text{bias}(\sigma)| + \lambda\right)\\ &\leq d\varepsilon_0 + 2(1+\delta)\sqrt{d-1}\\ &= 2\delta\sqrt{d-1} + 2(1+\delta)\sqrt{d-1}\\ &= 2(1+2\delta)\sqrt{d-1}.\end{aligned}$$

The quantity $\theta_{\max}$ defined in Lemma 6.6 satisfies

$$\begin{aligned}\theta_{\max} &= \frac{\|A\|_{\text{op}}}{2} + \sqrt{\frac{\|A\|_{\text{op}}^2}{4} - (d-1)}\\ &\leq (1+2\delta)\sqrt{d-1} + \sqrt{d-1}\sqrt{(1+2\delta)^2 - 1}\\ &= \sqrt{d-1}\cdot\left(1 + 2\delta + \sqrt{4\delta^2 + 4\delta}\right)\\ &\leq \sqrt{d-1}\cdot\left(1 + 2\delta + \sqrt{8\delta}\right) && (\delta \leq 1)\\ &\leq \sqrt{d-1}\cdot\left(1 + 2\delta + 3\sqrt{\delta}\right)\\ &\leq \sqrt{d-1}\cdot\left(1 + 5\sqrt{\delta}\right)\\ &\leq \sqrt{d-1}\cdot\left(1 + 5d^{-c_1/2}\right).\end{aligned}$$

By Lemma 6.6, we then have that

$$\begin{aligned}\frac{\left\|\left(B^{(\sigma)}\right)^t\right\|_{\text{op}}}{(d-1)^t} &\leq \frac{2(d-1)t(\theta_{\max})^{t-1}}{(d-1)^{t-1}}\\ &\leq \frac{2(d-1)t(1+5d^{-c_1/2})^{t-1}}{(d-1)^{(t-1)/2}}\\ &= \varepsilon(t) \leq \varepsilon.\end{aligned}$$

Therefore, $\mathcal{W}$ is an $(\varepsilon_0, \varepsilon)$ parity sampler. □

7 Acknowledgements

We are grateful to Victor Lecomte and Omer Reingold for stimulating discussions and collaboration in the early stages of the project. We also thank Ori Parzanchevski, Madhur Tulsiani, and Mary Wootters and for interesting and useful discussions.

References

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *Information Theory, IEEE Transactions on*, 38(2):509–516, 1992.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[Alo21] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, pages 1–17, 2021.

[BATS11] Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-Ramanujan graphs using the zig-zag product. *SIAM Journal on Computing*, 40(2):267–290, 2011.

[BATS13] Avraham Ben-Aroya and Amnon Ta-Shma. Constructing small-bias sets from algebraic-geometric codes. *Theory of Computing*, 9(5):253–272, 2013.

[BH04] Yonatan Bilu and Shlomo Hoory. On codes from hypergraphs. *European Journal of Combinatorics*, 25(3):339–354, 2004.

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006.

[Bog12] Andrej Bogdanov. A different way to improve the bias via expanders. *Topics in (and out) the theory of computing, Lecture*, 2012.

[CMRT16] Emma Cohen, Dhruv Mubayi, Peter Ralli, and Prasad Tetali. Inverse expander mixing for hypergraphs. *The Electronic Journal of Combinatorics*, 23(2):P2–20, 2016.

[CTZ20] David Conlon, Jonathan Tidor, and Yufei Zhao. Hypergraph expanders of all uniformities from Cayley graphs. *Proceedings of the London Mathematical Society*, 121(5):1311–1336, 2020.

[FW95] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995.

[GI05] Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005.

[Gil52] Edgar N. Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3):504–522, 1952.

[Gol19] Oded Goldreich. On constructing expanders for any number of vertices. October 2019. Available at https://www.wisdom.weizmann.ac.il/~oded/R1/ex4all.pdf.

[GP19] Konstantin Golubev and Ori Parzanchevski. Spectrum and combinatorics of two-dimensional Ramanujan complexes. *Israel Journal of Mathematics*, 230(2):583–612, 2019.

[GRS15] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory.* 2015.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[HRZW19] Brett Hemenway, Noga Ron-Zewi, and Mary Wootters. Local list recovery of high-rate tensor codes and applications. *SIAM Journal on Computing*, pages FOCS17–157, 2019.

[JQST20] Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. Unique decoding of explicit ε-balanced codes near the Gilbert–Varshamov bound. In *Proceedings of the 61st Annual Symposium on Foundations of Computer Science (FOCS 2020)*, pages 434–445. IEEE, 2020.

[JST21] Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. Near-linear time decoding of Ta-Shma's codes via splittable regularity. In *Proceedings of the 53rdth Annual Symposium on Theory of Computing (STOC 2021)*, pages 1527–1536. ACM, 2021.

[KRRZ$^+$20] Swastik Kopparty, Nicolas Resch, Noga Ron-Zewi, Shubhangi Saraf, and Shashwat Silas. On list recovery of high-rate tensor codes. *IEEE Transactions on Information Theory*, 67(1):296–316, 2020.

[KRZSW18] Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. Improved decoding of folded Reed–Solomon and multiplicity codes. In *Proceedings of the 59th Annual Symposium on Foundations of Computer Science (FOCS 2018)*, pages 212–223. IEEE, 2018.

[LM15] John Lenz and Dhruv Mubayi. Eigenvalues and linear quasirandom hypergraphs. In *Forum of Mathematics, Sigma*, volume 3. Cambridge University Press, 2015.

[LP16] Eyal Lubetzky and Yuval Peres. Cutoff on all Ramanujan graphs. *Geometric and Functional Analysis*, 26(4):1190–1216, 2016.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LW18] Ray Li and Mary Wootters. Improved list-decodability of random linear binary codes. In *APPROX–RANDOM*, volume 116 of *LIPIcs*, pages 50:1–50:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[Mar88] Grigorii Aleksandrovich Margulis. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problemy peredachi informatsii*, 24(1):51–60, 1988.

[MRSV21] Jack Murtagh, Omer Reingold, Aaron Sidford, and Salil Vadhan. Deterministic approximation of random walks in small space. *Theory of Computing*, 17(1):1–35, 2021.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993.

[Par17] Ori Parzanchevski. Mixing in high-dimensional expanders. *Combinatorics, Probability and Computing*, 26(5):746–761, 2017.

[PRT16] Ori Parzanchevski, Ron Rosenthal, and Ran J. Tessler. Isoperimetric inequalities in simplicial complexes. *Combinatorica*, 36(2):195–227, 2016.

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

[Tan81] R. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on information theory*, 27(5):533–547, 1981.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, ε-balanced codes. In *Proceedings of the 49th Annual Symposium on Theory of Computing (STOC 2017)*, pages 238–251. ACM, 2017.

[Var57] Rom Rubenovich Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

[Zém01] Gillés Zémor. On expander codes. *IEEE Transactions on Information Theory*, 47(2):835–837, 2001.

A Splittability and τ-Sampling

Our decoding result follows from the work of Jernoimo, Srivastava, and Tulsiani [JST21]. In [JST21], for the result of Theorem 8 to hold, they require $\mathcal{W}$ to be τ-*splittable*. τ-splittability is stronger than, and in fact implies, τ-sampling. We will not define splittability here, since for their result to hold, only a "Splittable Mixing Lemma" is required. The mixing property used in [JST21] goes as follows.

Definition A.1 (splittable mixing for ± 1 cut functions)**.** *Given positive integers n, t, and $s < t-1$, we denote by $\mathcal{F}_s$ the set of all functions $f\colon [n]^t \to \{1,-1\}$ of the form*

$$f(x_1,\dots,x_t) = b \cdot \chi_{A_1}(x_1) \cdot \dots \cdot \chi_{A_s}(x_s) \cdot \chi_B(x_{s+1},\dots,x_t), \tag{27}$$

where $b \in \{1,-1\}$, $A_1,\dots,A_s \subseteq [n]$, $B \subseteq [n]^{t-s}$, and where $\chi_S(x) = -1$ if $x \in S$ and 1 otherwise.

For a distribution $\mathcal{W} \sim [n]^t$, we denote by $\mathcal{W}_{[i,j]}$ the distribution over $[n]^{j-i+1}$ that is obtained by sampling $\boldsymbol{x} \sim \mathcal{W}$ and outputting $\boldsymbol{x}_{[i,j]}$. For simplicity, we abbreviate $\mathcal{W}_i = \mathcal{W}_{[i,i]}$. Given $s \in [t-1]$, let ν_s be the probability measure over $[n]^t$ for which $\nu_s(x) = \prod_{i\in[s]} \Pr[\mathcal{W}_i = x_i]\cdot\Pr[\mathcal{W}_{[s+1,t]} = x_{[s+1,t]}]$. We say $\mathcal{W} \sim [n]^t$ satisfies splittable mixing with error τ, *if for every $s \in [t-1]$ and every $f, f' \in \mathcal{F}_s$ it holds that*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\nu_s}\left[f(\boldsymbol{x})f'(\boldsymbol{x})\right] - \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\nu_{s-1}}\left[f(\boldsymbol{x})f'(\boldsymbol{x})\right] \right| \le \tau. \tag{28}$$

In [JST21], they show that the above property follows from τ-splittablility. Here we show that the above property also follows from our weaker notion of τ-sampling. For the sake of being compatible with [JST21], we will use a slightly different definition of τ-sampling than what was given in Definition 5.1.

Definition A.2. *We say that a distribution $\mathcal{W}$ over $[n]^t$ is* τ-sampling *if for any $i \in [t-1]$, $S \subseteq [n]$, and $X \subseteq [n]^{t-i}$,*

$$\mathop{\mathrm{Cov}}_{\boldsymbol{w}\sim\mathcal{W}}\Big[\mathbb{1}[\boldsymbol{w}_i \in S], \mathbb{1}[(\boldsymbol{w}_{i+1},\dots \boldsymbol{w}_t) \in X]\Big] \le \tau.$$

For decoding lifted sum codes, both definitions are essentially the same, since we can always work with $\mathcal{W}_{\mathrm{rev}}$, wherein we sample $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_t) \sim \mathcal{W}$ and output $(\boldsymbol{w}_t, \ldots, \boldsymbol{w}_1)$. Moreover, our $\mathcal{W}$ satisfies $\mathcal{W} = \mathcal{W}_{\mathrm{rev}}$.

Claim A.3. *Let $\mathcal{W} \subseteq [n]^t$ be homogeneous and τ-sampling. Then, $\mathcal{W}$ satisfies splittable mixing with error 4τ.*

Proof. Let $s \in [t-1]$, $b, b' \in \{1, -1\}$ and sets $A_1, \ldots, A_s, A'_1, \ldots, A'_s \subseteq [n]$ and $B, B' \subseteq [n]^{t-s}$ that correspond to the functions f and f'. For $i \in [s]$ denote $C_i = A_i \triangle A'_i$ and $D = B \triangle B'$. We then have

$$f(x)f'(x) = \sigma \cdot \chi_{C_1}(x_1) \cdot \ldots \cdot \chi_{C_s}(x_s) \cdot \chi_D(x_{s+1}, \ldots, x_t),$$

for $\sigma = bb' \in \{1, -1\}$. The subtraction on the left hand side of Equation (28) now reads

$$\sum_{x \in [n]^{s-1}} \prod_{i \in [s-1]} \Pr[\mathcal{W}_i = x_i] \cdot \sigma \cdot \chi_{C_1}(x_1) \cdot \ldots \cdot \chi_{C_{s-1}}(x_{s-1}) \cdot \\ \sum_{y \in [n], z \in [n]^{t-s}} \chi_{C_s}(y) \cdot \chi_D(z) \cdot \big(\Pr[\mathcal{W}_s = y] \Pr[\mathcal{W}_{[s+1,t]} = z] - \Pr[\mathcal{W}_{[s,t]} = y \circ z]\big). \tag{29}$$

The second line of the above expression can be written as

$$\mathop{\mathbb{E}}_{\boldsymbol{y} \sim \mathcal{W}_s}\left[(-1)^{\mathbb{1}[\boldsymbol{y} \in C_s]}\right] \cdot \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathcal{W}_{[s+1,t]}}\left[(-1)^{\mathbb{1}[\boldsymbol{z} \in D]}\right] - \mathop{\mathbb{E}}_{(\boldsymbol{y},\boldsymbol{z}) \sim (\mathcal{W}_s, \mathcal{W}_{[s+1,t]})}\left[(-1)^{\mathbb{1}[\boldsymbol{y} \in C_s]}(-1)^{\mathbb{1}[\boldsymbol{z} \in D]}\right], \tag{30}$$

which is simply the covariance between the random variables $(-1)^{\mathbb{1}[\boldsymbol{y} \in C_s]}$ and $(-1)^{\mathbb{1}[\boldsymbol{z} \in D]}$ where $\boldsymbol{y}$ and $\boldsymbol{z}$ are drawn appropriately. As $\mathcal{W}$ is τ-sampling, we know that

$$|\mathrm{Cov}\left[\mathbb{1}[\boldsymbol{y} \in C_s], \mathbb{1}[\boldsymbol{z} \in D]\right]| \leq \tau,$$

and so Equation (30), in absolute value, amounts to

$$\left|\mathrm{Cov}\left[(-1)^{\mathbb{1}[\boldsymbol{y} \in C_s]}, (-1)^{\mathbb{1}[\boldsymbol{z} \in D]}\right]\right| \leq 4\tau,$$

which readily follows from the fact that $(-1)^b = 1 - 2b$ for $b \in \{0, 1\}$, as we argued in Lemma 5.3. Taking absolute value, by the triangle inequality, Equation (29) can be bounded by

$$4\tau \cdot \sum_{x \in [n]^{s-1}} \prod_{i \in [s-1]} \Pr[\mathcal{W}_i = x_i].$$

As $\mathcal{W}$ is homogeneous, $\Pr[\mathcal{W}_i = x_i] = \frac{1}{n}$, and we are done. □

B Properties of the Non-Backtracking Parity Sampler

First, we prove the two propositions of Claim 6.4.

Proposition B.1. *For any $t \in \mathbb{N}$ and d-regular symmetric hypergraph H over n vertices, $\mathcal{W}^{(\mathrm{nb})}_{H,t}$ is homogeneous.*

Proof. $\boldsymbol{e}_1$ is uniform over E_H. Then, for every $j \in [t-1]$, if $\boldsymbol{e}_j$ is uniform, the $\boldsymbol{e}_{j+1}$ is uniform. By induction, $\boldsymbol{e}_j$ is uniform for every $j \in [t]$.

Next, fix any $a \in [n]$ and $j \in [t]$. As $\boldsymbol{w}_j = a$ if and only if $\boldsymbol{e}_j$ is one of the d-edges in E_H who's second vertex is a and $\boldsymbol{e}_j$ is uniform over the nd edges in E_H,

$$\Pr[\boldsymbol{w}_j = a] = \frac{d}{nd} = \frac{1}{n}.$$

□

Proposition B.2. *For any $t \in \mathbb{N}$ and d-regular symmetric hypergraph H over n vertices, $\mathcal{W}_{H,t}^{(\mathrm{nb})}$ is $(nd(d-1)^{t-1})$-discretizable.*

Proof. $\boldsymbol{e}_1$ is picked uniformly from nd options. Then, for each $j \in [2,t]$, $\boldsymbol{e}_{j+1}$ is picked uniformly (and independently from prior choices) from $N_H^{(\mathrm{nb})}(e)$, which has $(d-1)$ elements. Therefore, $\boldsymbol{w}$ is picked uniformly from $nd(d-1)^t$ possible items, potentially with duplicates. □

Next, we prove the following Lemma, restated for convenience.

Lemma 6.5. *For any d-regular symmetric hypergraph $H = (V, E_H)$, $\sigma \in \{\pm 1\}^n$, letting $B^{(\sigma)}$ be the non-backtracking operator defined in Equation (13), we have that*

$$\left|\mathrm{bias}_{\mathcal{W}_{H,t}^{(\mathrm{nb})}}(\sigma)\right| = \left|\frac{\mathbf{1}^\dagger (B^{(\sigma)})^t \mathbf{1}}{nd(d-1)^t}\right| \leq \frac{\left\|(B^{(\sigma)})^t\right\|_{\mathrm{op}}}{(d-1)^t}.$$

Proof. Our goal is to prove that

$$\mathrm{bias}_{\mathcal{W}_{H,t}^{(\mathrm{nb})}}(\sigma) = \mathop{\mathbb{E}}_{\boldsymbol{w} \sim \mathcal{W}_{H,t}}\left[\prod_{j=1}^{t} \sigma_{\boldsymbol{w}_j}\right] = \frac{\mathbf{1}^\dagger (B^{(\sigma)})^t \mathbf{1}}{nd(d-1)^t}.$$

Draw some $\boldsymbol{w} \sim \mathcal{W}_{H,t}^{(\mathrm{nb})}$, and let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_t$ be the random variables (coupled to $\boldsymbol{w}$) defined in the construction of $\mathcal{W}_{H,t}^{(\mathrm{nb})}$. We claim that for each $j \in [t]$ and $e \in E_H$,

$$\mathbb{E}\left[\mathbb{1}[\boldsymbol{e}_j = e] \prod_{k=1}^{j} \sigma_{\boldsymbol{w}_k}\right] = \frac{(\mathbf{1}^\dagger (B^{(\sigma)})^j)_e}{nd(d-1)^j}. \tag{31}$$

For $j = 1$, $\boldsymbol{e}_1$ is initialized uniformly among the nd edges, so the above expression should be equal to $\frac{\sigma_{e_2}}{nd}$. Indeed,

$$\begin{aligned}
\frac{(\mathbf{1}^\dagger B^{(\sigma)})_e}{nd(d-1)} &= \frac{1}{nd(d-1)} \sum_{e' \in E_H} B_{e',e} \\
&= \frac{1}{nd(d-1)} \sum_{e' \in E_H} \mathbb{1}\left[e \in N_H^{(\mathrm{nb})}(e')\right] \cdot \sigma_{e_2} \\
&= \frac{\sigma_{e_2}}{nd}.
\end{aligned}$$

So Equation (31) holds for $j = 1$. For $j \geq 2$, we proceed by induction.

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}[\boldsymbol{e}_j = e] \prod_{k=1}^{j} \sigma_{\boldsymbol{w}_k}\right] &= \sum_{e' \in E_H} \mathbb{E}\left[\mathbb{1}[\boldsymbol{e}_j = e, \boldsymbol{e}_{j-1} = e'] \prod_{k=1}^{j} \sigma_{\boldsymbol{w}_k}\right] \\
&= \sum_{e' \in E_H} \mathbb{E}\left[\mathbb{1}[\boldsymbol{e}_{j-1} = e'] \prod_{k=1}^{j-1} \sigma_{\boldsymbol{w}_k}\right] \cdot \mathbb{E}\left[\mathbb{1}[\boldsymbol{e}_j = e] \cdot \sigma_{\boldsymbol{w}_j} \mid \boldsymbol{e}_{j-1} = e'\right] \\
&= \sum_{e' \in E_H} \frac{(\mathbf{1}^\dagger (B^{(\sigma)})^{j-1})_{e'}}{nd(d-1)^{j-1}} \cdot \mathbb{E}\left[\mathbb{1}[\boldsymbol{e}_j = e] \cdot \sigma_{\boldsymbol{w}_j} \mid \boldsymbol{e}_{j-1} = e'\right] \\
&= \sum_{e' \in E_H} \frac{(\mathbf{1}^\dagger (B^{(\sigma)})^{j-1})_{e'}}{nd(d-1)^{j-1}} \cdot \frac{\mathbb{1}\left[e \in N_H^{(\mathrm{nb})}(e')\right] \cdot \sigma_{e_2}}{d-1} \\
&= \frac{1}{nd(d-1)^j} \cdot \sum_{e' \in E_H} (\mathbf{1}^\dagger (B^{(\sigma)})^{j-1})_{e'} \cdot B^{(\sigma)}_{e',e} \\
&= \frac{(\mathbf{1}^\dagger (B^{(\sigma)})^{j})_{e}}{nd(d-1)^j}
\end{aligned}$$

Hence, Equation (31) holds by induction. The desired result follows by summing Equation (31) over all edges. □

Next, we do all the calculations necessary for Proposition 6.9, restated below for convenience.

Proposition B.3.

1. *For any $w, w' \in \mathbb{C}^V$,*

$$\begin{aligned}
\langle w^{(\mathrm{in})}, (w')^{(\mathrm{in})} \rangle &= d \cdot \langle w, w' \rangle \\
\langle w^{(\mathrm{out})}, (w')^{(\mathrm{out})} \rangle &= d \cdot \langle w, w' \rangle \\
\langle w^{(\mathrm{in})}, (w')^{(\mathrm{out})} \rangle &= \langle Aw, w' \rangle
\end{aligned} \tag{18}$$

2. *For any $w \in \mathbb{C}^V$,*

$$\begin{aligned}
Bw^{(\mathrm{in})} &= (Aw)^{(\mathrm{in})} - w^{(\mathrm{out})} \\
Bw^{(\mathrm{out})} &= (d-1)w^{(\mathrm{in})}.
\end{aligned} \tag{19}$$

3. *For any $v \in \mathbb{C}^E$ satisfying $v \perp w^{(\mathrm{out})}$ for all $w \in \mathbb{C}^V$,*

$$v_{xy} = -A_{yx} v_{yx} \qquad \textit{for every edge } xy. \tag{20}$$

4. *The operator norm of B is*

$$\|B\|_{\mathrm{op}} = d - 1. \tag{21}$$

Proof. First, for Equation (18):

1. We compute $\langle w^{(\mathrm{in})}, (w')^{(\mathrm{in})} \rangle$:

$$\begin{aligned} \langle w^{(\mathrm{in})}, (w')^{(\mathrm{in})} \rangle &= \sum_{xy \in E} w_y w'_y \\ &= d \cdot \sum_{y \in V} w_y w'_y \\ &= d \cdot \langle w, w' \rangle \end{aligned}$$

2. We compute $\langle w^{(\mathrm{out})}, (w')^{(\mathrm{out})} \rangle$:

$$\begin{aligned} \langle w^{(\mathrm{out})}, (w')^{(\mathrm{out})} \rangle &= \sum_{xy \in E} (A_{xy})^2 w_x w'_x \\ &= d \sum_{x \in V} w_x w'_x && (A_{xy} \in \{\pm 1\} \text{ for any edge } xy) \\ &= d \cdot \langle w, w' \rangle \end{aligned}$$

3. We compute $\langle w^{(\mathrm{in})}, (w')^{(\mathrm{out})} \rangle$:

$$\begin{aligned} \langle w^{(\mathrm{in})}, (w')^{(\mathrm{out})} \rangle &= \sum_{xy \in E} A_{xy} w_y w'_x \\ &= \sum_{x \in V} w'_x \cdot \sum_{y : xy \in E} A_{xy} w_y \\ &= \sum_{x \in V} w'_x (Aw)_x \\ &= \langle Aw, w' \rangle \end{aligned}$$

Next, we verify Equation (19). Consider any $w \in \mathbb{C}^V$.

1. We analyze $Bw^{(\mathrm{in})}$. For any $xy \in E$,

$$\begin{aligned} (Bw^{(\mathrm{in})})_{xy} &= \sum_{z : yz \in E, z \neq x} A_{yz} \cdot (w^{(\mathrm{in})})_{yz} \\ &= \sum_{z : yz \in E, z \neq x} A_{yz} w_z \\ &= \sum_{z : yz \in E} A_{yz} w_z - A_{yx} w_x \\ &= (Aw)_y - A_{xy} \cdot w_x && (A_{yx} = A_{xy}) \end{aligned}$$

Therefore, $Bw^{(\mathrm{in})} = (Aw)^{(\mathrm{in})} - w^{(\mathrm{out})}$.

2. We analyze $Bw^{(\text{out})}$. For any $xy \in E$,

$$\begin{aligned}(Bw^{(\text{out})})_{xy} &= \sum_{z:yz\in E, z\neq x} A_{yz} \cdot (w^{(\text{out})})_{yz} \\ &= \sum_{z:yz\in E, z\neq x} (A_{yz})^2 (w)_y \\ &= (d-1)\cdot w_y && ((A_{yz})^2 = 1 \text{ for any } yz \in E.)\end{aligned}$$

Therefore, $Bw^{(\text{out})} = (d-1)w^{(\text{in})}$.

Next, we verify Equation (20). Choose any $v \in \mathbb{C}^E$ satisfying $v \perp w^{(\text{out})}$ for all $w \in \mathbb{C}^V$. In particular, v is orthogonal to $(e_x)^{(\text{out})}$ for every $x \in V$, where e_x is the vector that has weight 1 on vertex x and weight 0 everywhere else. This implies that, for all $x \in V$,

$$\sum_{y:xy\in E} v_{xy} A_{xy} = \langle v, (e_x)^{(\text{out})} \rangle = 0.$$

We compute $(Bv)_{xy}$ for any edge xy.

$$\begin{aligned}(Bv)_{xy} &= \sum_{z:yz\in E, z\neq x} A_{yz} v_{yz} \\ &= \sum_{z:yz\in E} A_{yz} v_{yz} - A_{yx} v_{yx} \\ &= -A_{yx} v_{yx} && (v \perp (e_y)^{(\text{out})})\end{aligned}$$

Finally, we prove Equation (21). In the proof of Proposition 6.11, we showed that if $v' \in \mathbb{C}^E$ is orthogonal to $w^{(\text{out})}$ for all $w \in \mathbb{C}^V$, then Bv' is orthogonal to $w^{(\text{in})}$ for all $w \in \mathbb{C}^v$. Furthermore, in the proof of Proposition 6.14, we showed that under the same condition, $\|Bv'\|_2 = \|v'\|_2$.

Now, consider any $v \in \mathbb{C}^E$. We can decompose it into

$$v = w^{(\text{out})} + v'$$

for some $w \in \mathbb{C}^V$ and v' that is orthogonal to $(w')^{(\text{out})}$ for all $w' \in \mathbb{C}^V$. Then, we have that

$$\begin{aligned}\|Bv\|_2 &= \|(d-1)w^{(\text{in})} + Bv'\|_2 \\ &= \sqrt{(d-1)^2 \|w^{(\text{in})}\|_2^2 + \|Bv'\|_2^2} && (Bv' \text{ is orthogonal to } w^{(\text{in})}) \\ &= \sqrt{(d-1)^2 \|w^{(\text{out})}\|_2^2 + \|v'\|_2^2} \\ &= \sqrt{(d-1)^2 \|v\|_2^2 - ((d-1)^2 - 1)\|v'\|_2^2} && (\|v\|_2^2 = \|w^{(\text{out})}\|_2^2 + \|v'\|_2^2) \\ &\leq \sqrt{(d-1)^2 \|v\|_2^2} \\ &= d-1.\end{aligned}$$

Hence, $\|Bv\|_2 \leq (d-1)\|v\|_2$, and with equality whenever $v = w^{(\text{out})}$ for some $w \in \mathbb{C}^V$, proving that $\|B\|_{\text{op}} = d-1$. $\square$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il